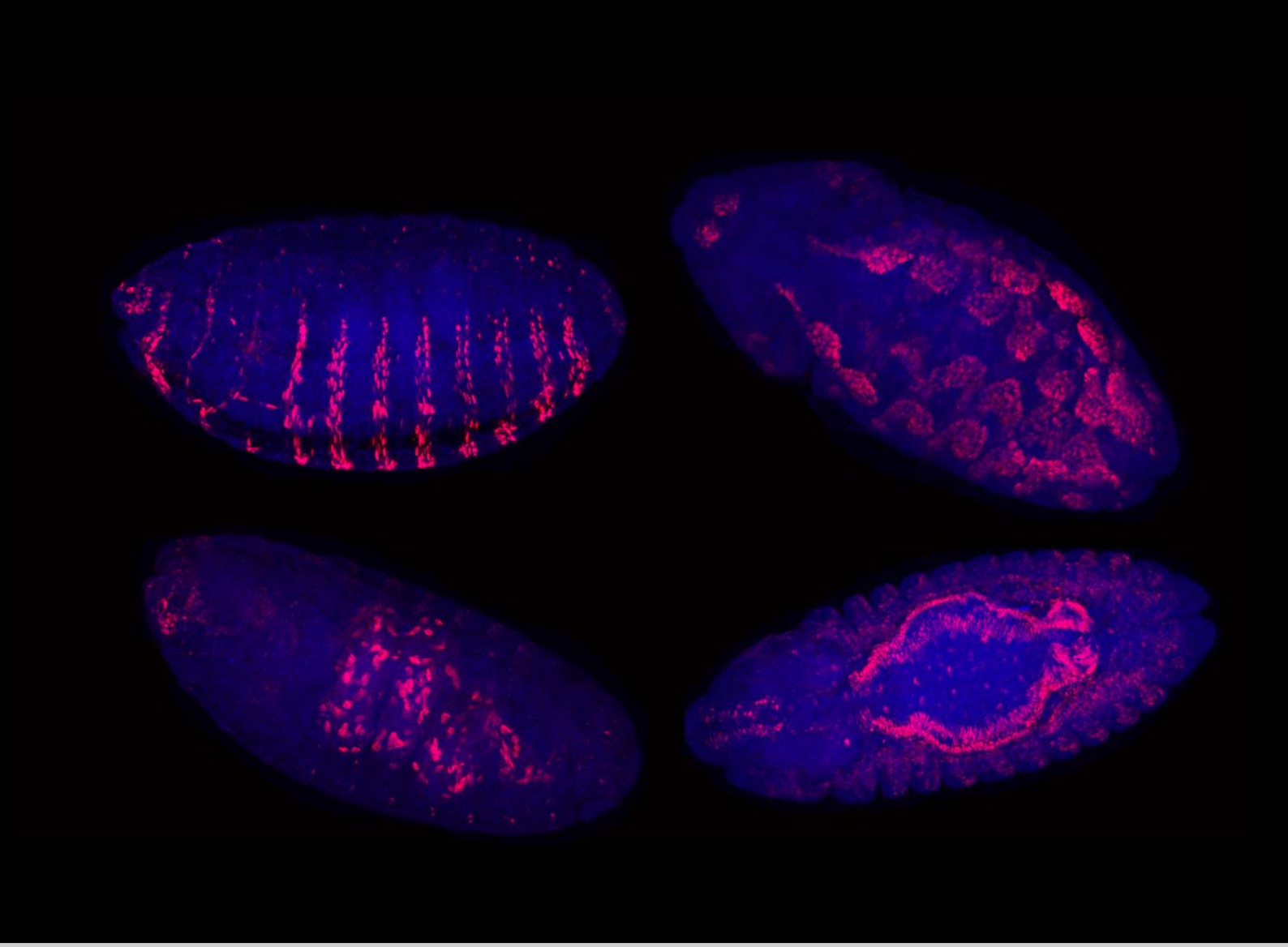


Interrogating the complex role of Ubx & multi-enhancer transcriptional hubs throughout development

Mariana Rama Pedro Alves

2021

Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Mariana Rama Pedro Alves

born in: Coimbra, Portugal

Oral examination: December 14th, 2021

Interrogating the complex role of Ubx and multi-enhancer transcriptional hubs throughout development

Referees: Prof. Dr. Ingrid Lohmann
Dr. Eileen Furlong

Cover picture: *Drosophila* embryos, with nuclear staining in blue and ßGal staining in magenta.

Summary

During development, complex gene expression patterns are formed, relying on the spatiotemporal coordination of proteins and genomic regulatory regions, among many other players and layers of processes. Low-affinity binding sites efficiently contribute to transcription through brief interactions with transcription factors. Concentrating transcription factors in localized environments, driven by clusters of enhancers, have been suggested as a mechanism by which low-affinity sites are efficiently used. Such localized transcriptional environments have been observed in *Drosophila melanogaster* embryos for the transcription factor Ultrabithorax (Ubx) (Tsai *et al*., 2017) at the locus of its target *shavenbaby* (*svb*). In this thesis, I interrogate the complex role of Ubx and multi-enhancer transcriptional hubs throughout development.

Firstly, I showed that active *svb* enhancers on different chromosomes tend to co-localize. I then tested the hypothesis that multi-enhancer interactions contribute to low-affinity transcriptional microenvironments and show that defects from a deletion of a redundant enhancer from the *svb* locus at elevated temperatures can be rescued by introducing *svb*'s *cis*-regulatory sequence driving a reporter in a different chromosome, which suggests that multiple enhancers can reinforce local transcriptional hubs to buffer against environmental stresses.

Secondly, I explored whether microenvironments are specific to low-affinity enhancers or if microenvironments can form around Ubx high-affinity enhancers. I found that Ubx local enrichment is overall maintained when low-affinity sites in a *svb* enhancer are substituted with high-affinity sites. Then, I screened a library of short genomic fragments containing either (endogenous) Ubx high-affinity sites or mutations of these to assay their expression patterns and transcriptional microenvironment features. I show that sequences containing high-affinity sites can function as transcriptional enhancers across development and can exhibit features of multi-enhancer transcriptional microenvironments such as Ubx local enrichment and transcript co-localization with Ubx target *svb*.

Finally, it remained to be tested what is the extent of the contribution of Ubx to these phenomena, so I developed a Ubx recruitment system to explore the role of this transcription factor in enhancer clustering using high-resolution microscopy.

Overall, this project sheds new light on transcriptional microenvironments, providing a more in-depth understanding of their components, exploring them in the context of gene regulation during animal development.

Zusammenfassung

Während der Entwicklung, bilden sich komplexe Genexpressionsmuster heraus, die neben vielen anderen Akteuren und Prozessschichten auf die räumlich-zeitliche Koordination von Proteinen und genomischen Regulationsregionen, angewiesen sind. Bindungsstellen mit geringer Affinität tragen durch kurze Interaktionen mit Transkriptionsfaktoren effizient zur Transkription bei. Die Konzentration von Transkriptionsfaktoren in lokalisierten Umgebungen, die durch Cluster (Gruppierungen) von Enhancer (DNA-Sequenzen, die die Gentranskription regeln, wenn sie durch Transkriptionsfaktoren verbunden sind) gesteuert werden, wurde als Mechanismus zur effizienten Nutzung von Bindestellen mit geringer Affinität vorgeschlagen. Solche lokalisierten Transkriptionsumgebungen wurden in Embryonen von *Drosophila melanogaster* für den Transkriptionsfaktor Ultrabithorax (Ubx) (Tsai *et al*., 2017) am Genlocus *shavenbaby* (*svb*), beobachtet. In dieser Doktorarbeit, untersuche ich die komplexe Rolle von Ubx und von Multi-Enhancer Transkriptionskompartimenten während der Entwicklung.

Zunächst habe ich gezeigt, dass aktive *svb* Enhancer auf verschiedenen Chromosomen zur Kolokalisierung neigen. Anschließend testete ich die Hypothese, dass Interaktionen zwischen mehreren Enhancern zu transkriptionellen Mikroumgebungen mit geringer Affinität beitragen. Ebenfalls zeigte ich, dass Defekte die durch die Deletion eines redundanten Enhancers am *svb* Locus bei erhöhten Temperaturen entstanden sind, durch die Einführung einer cisregulatorische Sequenz von *svb*, was zur Expression eines Reporters auf einem anderen Chromosom führt, behoben werden können. Dies lässt darauf schließen, dass Multienhancer örtliche transkriptionelle Kompartimente verstärken können, um Umweltbelastungen abzufedern.

Zweitens, untersuchte ich die Möglichkeit, dass Mikroumgebungen spezifisch für Enhancer mit geringer Affinität sind oder ob sie sich um hochaffine Ubx-Enhancer bilden können. Ich fand heraus, dass die lokale Anreicherung von Ubx insgesamt erhalten bleibt, wenn die niederaffinen Bindenstellen in einem *svb* Enhancer durch hochaffine Bindenstellen ausgetauscht werden.

Anschließend, analysierte ich eine Bibliothek mit kurzen Genomfragmenten, die entweder (endogene) hochaffine Bindestelle für Ubx oder dere Mutationen an dieser Stelle enthalten, um die resultierenden Expressionsmuster und die Eigenschaften der transkriptionale Mikroumgebungen zu untersuchen. Ich konnte zeigen, dass die Sequenzen, die hochaffine Bindestellen enthalten, während der gesamten Entwicklung

als transkriptionelle Enhancer fungieren und dass sie Merkmale von transkriptionaler Mikroumgebungen mit Multienhancern aufweisen können, wie, zum Beispiel, eine lokale Anreicherung von Ubx und Kolokalisierung von Transkripten am Zielgen *svb*.
Schließlich blieb noch zu prüfen, inwieweit Ubx zu diesen Phänomenen beiträgt. Deswegen, entwickelte ich ein Ubx-Rekrutierungssystem, um die Rolle dieses Transkriptionsfaktors bei der Clusterbildung von Enhancern mit Hilfe von hochauflösender Mikroskopie zu erforschen.
Insgesamt wirft dieses Projekt ein neues Licht auf transkriptionelle Mikroumgebungen und ermöglicht dadurch ein tieferes Verständnis der Komponenten, besonders im Zusammenhang mit der Genregulation während der Entwicklung.

Resumo

Durante o desenvolvimento animal são formados padrões de expressão genética (ou gênica) complexos que se baseiam na coordenação espácio-temporal entre proteínas e regiões de regulação genómica, além de muitos outros fatores e processos. Locais/sítios de ligação de baixa afinidade no ADN/DNA contribuem para a transcrição de forma eficiente, através de interações breves com fatores de transcrição. A concentração de fatores de transcrição em microambientes localizados no núcleo celular, impulsionada por 'clusters' (agrupamentos) de '*enhancers*' (sequências de ADN/DNA que regulam a transcrição de genes quando ligadas por fatores de transcrição), foi sugerida como mecanismo pelo qual locais/sítios de ligação de baixa afinidade são utilizados eficientemente. Estes ambientes de transcrição localizados foram observados em embriões de *Drosophila melanogaster* para o fator de transcrição Ultrabithorax (Ubx) (Tsai *et al.*, 2017) no *locus* genético do seu alvo, *shavenbaby* (*svb*). Nesta tese, procuro esclarecer o papel complexo do Ubx e de compartimentos transcricionais com múltiplos *enhancers* ao longo do desenvolvimento.

Em primeiro lugar, mostrei que *enhancers* de *svb* ativos em diferentes cromossomas tendem a se co-localizar. Em seguida, testou-se a hipótese de que as interações de múltiplos *enhancers* contribuem para microambientes transcricionais de baixa afinidade. Demonstrei que anomalias resultantes da eliminação de um *enhancer* redundante do *locus* genético de *svb* a temperaturas elevadas podem ser revertidas ao introduzir, num cromossoma diferente, a sequência *cis*-regulatória de *svb*, que, neste caso, controla a expressão de um gene repórter. Essa observação sugere que múltiplos *enhancers* podem reforçar compartimentos transcricionais locais/sítios para proteger contra pressões/estressores ambientais.

Em segundo lugar, explorei a possibilidade de os microambientes serem específicos dos *enhancers* de baixa afinidade ou de se conseguirem formar em torno de *enhancers* de alta afinidade para Ubx. Concluí que o enriquecimento local em Ubx é mantido, globalmente, quando os locais/sítios de ligação de baixa afinidade num *enhancer* de *svb* são substituídos por locais/sítios de ligação de alta afinidade.

Em seguida, analisei um(a) banco/biblioteca de pequenos fragmentos genómicos que contêm ou locais/sítios (endógenos) de ligação de alta afinidade para o Ubx, ou mutações dos mesmos, para avaliar os respectivos padrões de expressão a que conduzem/que dirigem e as características dos microambientes transcricionais. Concluí que as sequências que contêm locais/sítios de ligação de elevada afinidade podem funcionar como

enhancers transcricionais ao longo do desenvolvimento e podem exibir características de microambientes transcricionais de múltiplos *enhancers*, como um enriquecimento local de Ubx e co-localização transcricional com um dos seus alvos, o *svb*.

Por fim, restava testar a contribuição do Ubx para estes fenómenos. Desenvolvi então um sistema de recrutamento de Ubx para explorar o papel deste fator de transcrição no agrupamento de *enhancers*, utilizando microscopia de alta resolução.

Em conjunto, os resultados deste projeto revelam novas informações sobre os microambientes transcricionais, ao apresentar uma análise mais aprofundada dos respetivos componentes, explorando-os no contexto da regulação genética durante o desenvolvimento animal.

To every person who wished to be a scientist but couldn't because of the social injustice(s) that structure(s) society and academia.

In the loving memory of Avó Nana, Avô Manel and Nandita.

“Se não está bem assim

não duvides de mim

ensina-me lá

Eu quero ser melhor e maior

eu aguento

ensina-me já

E tudo o que eu não sei

o que eu não faço bem

ensinaste-me mal

Tem paciência para mim

dá-me amor e carinho

quando eu me sinta pior

E se não resultar o melhor é mudar de método”

B Fachada in *Rapazes e raposas* (2020)

Acknowledgments

It is hard to cover all corners of gratitude during a span of four years, especially when memory abilities can deteriorate especially fast over a stressful time. If you read this and I forgot to mention you, please let me know, and I apologize in advance.

I would like to start by thanking Justin. When we first met, there was just you, your exciting ideas and an empty lab space. 4 years later, the Crocker lab is full of people, flies and Genscript vials. I am very grateful that you trusted me the opportunity of joining you and your lab. I do not take for granted the opportunities you provided me, starting right from the start and continuing along the way. Thank you for letting me ask you for help when I needed. I am grateful for the flexibility and freedom you gave me throughout. Thank you for being open, for your good humour and, mostly, for being human.

It was a formative experience to see a lab being built from the very start, and I am thankful to every lab member (I have met all so far!) for their help, support and/or collaboration: Lautaro, Mindy, Aref, Luis, Harshit, Noa, Blanca, Marlize, Kian, Tim, Esther, and the ones referred to in the following sentences. Thank you, Tin, for helping me sediment some perspective. Thank you, Kerstin, for introducing me to mindfulness and for the German-related support. Thank you Natalia for being so refreshingly nice. Thank you, Xueying, for your sorority, kindness and thoughtful feedback. Thank you, Gilberto, for your generosity both socially and professionally; this work would not be the same without your contribution. Thank you also for your generosity and patience in a particular challenging time. Thank you, Albert, for your collaboration and especially for your infinite patience and for your honesty: teaching me how to do *in-situs*, helping me in the microscope, with analysis, sharing your scripts, your settings, your feedback, explaining the same things several times when needed… your honesty helped me gain perspective and coordinates to navigate several aspects of this adventure, thank you also for valuing me for certain sides of myself which I sometimes underestimate(d).

I would like to express my gratitude for Tim Pollex, Julie Carnesecchi and Catarina Ramos Carmo, for guidance and advice. Tim, you were an absolute patient star, guiding me and listening to me, being a true guidepost especially crucial until Rafael arrived. Julie, I was always so in awe by your brilliance and thank you for making me feel welcomed in a community I did not belong to and for sharing interesting insights, ideas and enthusiasm. Catarina, it is impossible to fully capture the multidimensions in which you have been and are an influence and an inspiration. You taught me *Drosophila* manners, you welcomed and supported me, always leading by example as an outstanding scientist. Most importantly, you are a good friend.

Thank you to my TAC members Ingrid Lohmann, Eileen Furlong and Alexander Aulehla for your insightful feedback and important peer-review (in the broad sense). Thank you to Alexis Maizel for joining my examination committee.

Thank you to the Developmental Biology Unit and the Fly Labs (de Renzis, Ephrussi, Furlong and Leptin labs). Thank you to Maria for being so approachable. Thank you to Ale and to Anna not only for your work but also your constant patience and kindness. There are corridor smiles or small acts of kindness that can fill your day. Thank you Paul for being a good support since before the start (starting with good advice in interview week at the ISG bus) until the very end. You are such a star. Thank you for showing up so many times to say, in other words, that you saw me. Thank you Carina for leaving me the legacy of and supporting a push for a mentally safer environment for the Unit's PhD students. Thank you for giving me the right business card at the right time. Thank you Cristina! for your constant support, and also Jonas, Daniel, Emiliano, Joaquim, Juan, Fabi, Kevin, Raquel, Mainak, Emilia, Phil, Anniek, Nezha, James, Naomi, Vlad, Stefano and so many others!

Thank you to everyone that collaborated with me and helped me in experiments and projects that eventually failed and did not make it to this thesis: the entire GeneCore (particularly Vladimir, Dinko, Bianka, Anja, Ferris), Charles Girardot, Rebecca Viales, Gabi, David Stern, Kim Remans, Claire Deo, Michael Stadler, Luca Giorgetti, Sagar Pratapsi, Rebecca Delker, and possibly others. Thank you also to the ALMF, PepCore and Sven from Zeiss Support. Thank you to the people from Genetivision and Genscript that worked in my orders. Thank you to everyone at IT and electronic and mechanic workshops who helped me so many times! There would be no thesis without a functional laptop. Thank you to Massimo for your patience, flexibility and encouragement when printing this thesis.

I am grateful that the EIPP promotes the PhD Symposium, as it allowed me to meet my dear friend and talented scientist Amos Abolaji. Thank you to Monika but especially to Carolina and Matija for the administrative support when I needed. Thank you to Patricia for your incredible work at the Fellows' Careers Service.

EMBL is full of people who deserve acknowledgment in this thesis. There can be no science without all of the people involved in all of the jobs that sometimes are forgotten by the "scientists" and that make EMBL work. I acknowledge that my privilege means that I do not know the names of a lot of these people, and I do not even think about who is doing X or Y function, as some things appear to arrive or exist "already made". Thank you to the cleaning services, to security for making me feel safe working at times when EMBL was deserted, to transport services, to the EMBL Gardeners for providing beautiful flowers and tree landscapes to look at which can be especially reinvigorating when returning home from work. Thank you to

Raffaele for your warm greetings. Thank you to all the cantine and cafeteria staff. Thank you especially to Michael for your complicity to my sweet tooth and Fernando, Maxi and Stefan for always friendly interactions. Thank you to the Kitchen Ladies for spoiling us and saving us from preparing reagents, food, washing material.

Thank you to Roshni Mooneeram for being an inspiration. I wish you the best for your work as Head of Equality, Diversity & Inclusion. You can do it superbly, if given the right chance by management. Thank you to Francesca and Luisa for teaching me so much about EDI too.

Thank you to everyone in the wider scientific community. I would like to firstly thank Alexandra Elbakyan. Then, I would like to thank those that I had the pleasure to interact around the world during these years, in particular the legendary MBL GRN Class of 2018 and the incredible people I met in Nigeria in 2020. Thank you to Amos, ore mi, for your kind invitation for this benchmark experience of my PhD journey, and to your family as well. Thank you to the absolute stars in the Abolaji lab, to my travel buddy Alex Whitworth, to inspiration Andreas Prokop and to the gracious Helen Ochuko Kwanashie. Thank you to the DrosAfrica crew, but especially Isa, for welcoming me as a volunteer and for your essential support. Thank you to Doros, Jacqueline, Verena, Yvonne (also for the soap!), Edith, Catherine and everyone at EMBL that gave a bit of their time and generosity to contribute in different ways to different dimensions of this trip. I would also like to thank Holger Breithaupt for his incredible revision and editing of my EMBO Reports piece and to Prof. Ana Luísa Carvalho for always remembering about me and supporting me from afar.

For various interactions at EMBL throughout these years, involving one or more bench-unrelated free-time passion projects, thank you to Photolab, especially Jan, Stefan, Massimo and Christopher. Thank you to Verena, Anniek, Phil, Cristina, Madalina, Anne-Flore, Tabea, Lucia, Anne-Marie, Patrick, Laura, Haipeng, Adam, Mathew, Luca, Berta, Fabian, Joshua, Sonia, Mustafa, Lothar, Shweta, Ed, Eva. Thank you to Alexander for listening and trusting. A very special thank you to Iris and to Jodie.

Thank you, Martina, for being a patient listener, a fierce supporter, and for bringing so much light and humour to my life. Thank you for the projects we built together, the challenges you supported me in navigating through. Everyone needs a Martina in their life.

Thank you Sonja, you took care of my mind and my feet and you are such a warm human being.

Thank you, Claire Standley, for your kindness and supportive generosity.

Thank you to everyone who gave a bit of their time and accepted the invitation for informal chats about career and life, also called “informational interviews”.

Merci beaucoup, Catherine Koleda.

Thank you to Veronica for making home a peaceful place and being an adult and humane housemate. Thank you to Lupe for your magic fairy dust. Thank you Luna for your sweet and pure love. Thank you Mr Fadani and neighbours for making it so eventless in foreign land.

When I think of Heidelberg, DTH automatically pops into my mind. Thank you to everyone at DTH for their warm welcome and making me feel a part of a (non-scientific) community in Heidelberg, which basically translates to feeling less like a visitor and more like a normal citizen. Thank you to past and present ensemble, to Svea, Jenny, Leonardo. Obrigada querido Renan, Heidelberg fica sempre mais bonito (no *lato sensu*) com você, e minha vida ficou também! Thank you so much, Iván, for being an inspiration and for your generosity and kindness in so many different dimensions at so many different times. How lucky am I to be alive at the same time as you (and Beyoncé!). Even just as an audience member, without meeting you (the whole company) personally, the light you (the whole company) bring to someone’s life is immense!

Gostaria também de agradecer à Elisa, Helena, Mamadu, Mariana, Moato, Ana, Abdulai, Gislaine, Margarida, Cecilia, Eliane, Lucas, Maria João, Patrícia e Romana pela paciência enquanto estive a escrever.

Thank you very much Hannah Neubauer, my psychotherapist. I don’t think I would graduate if it was not for the work we did during our appointments. Evans *et al.* showed in 2018 in Nature Biotechnology that graduate students are 6 times as likely to experience depression and anxiety as compared to the general population. I would also like to thank Rita Ribeiro and João Gomes for helping me take care of my physical wellbeing.

We are made of the world around us. I wish I could list the names of all the albums, movies, dance pieces, concerts and theatre shows that accompanied and inspired me through these years. Gloria and DTH in Heidelberg and the Wuppertal Opernhaus were sacred places to be in. Carolina, André, Karen, Kristina, Kai, Sara and Ju, thank you for hosting me and showing me around Germany.

Thank you to all my friends, close and far, recent and old, the ones who visited, the ones who didn’t, who write and call and stay and show there is no mountain high enough, thank you for your patience and your love. Thank you, Christophe, your friendship does not fit in a few sentences, thank you for truly committing to making the distance not a barrier or an excuse. Sofias, obrigada por acreditarem sempre em mim. Elsa, Gonçalo, Tomás,

Joana, finalmente cheguei aos vossos calcanhares. Tadeja, Mara, Michele, Nika, Jemal and the Coimbra group, Veronica, Stine, JP, Flávia, Eva, Professora Catroga, Professor Francisco, Beza, Ledia, Celine, Renata, João André, thank you for support and love. Obrigada Bois por estarem sempre lá. Particularly, thank you to Joana Cunha for your help translating my summary to German and Portuguese and Sara, Gabi and Carol for the respective revision. Obrigada Paula Roberto, pelas visitas. Obrigada Catarina Almeida, lindas memórias, minha amiga-furacão. Keel e Bela, obrigada por se manterem cá. Thank you Denis for sharing your wisdom, your wit, your compassion, your patience, your incredible humour – I feel very privileged and don't take it for granted.

My incredible friend Sandra Raquel, you welcomed me unbelievably well. Thank you for your hospitality, for going above and beyond. Although it is pretty epic that you stored my suitcases for 4 weeks beneath the benches of your previous lab while I transited from Cambridge to Heidelberg and found me my apartment, that does not come close to the importance of your friendship, your generosity, the wisdom you share. Also, together with Filipinha and Rafael, we were quite the bombastic crew (pitty we did not form a band!). Thank you also to AJ, Shyamal, Catarina Pechincha, Sofia e Filipe.

Catarina e Luís, obrigada por estarem sempre cá e lá (literalmente), pela inspiração, pelo apoio, pela generosidade, pelas nossas chamadas e risotas de sábado a noite. Não consigo imaginar melhores amigos senior que uma jovem poderia desejar. Fora de brincadeiras, aprendo e aprendi muito convosco.

Obrigada, minha amiga Carol, por ser uma luz de maturidade, por ser paciente, por partilhar sua sabedoria comigo, por me aceitar por quem sou.

Thank you Sarita, for the friend you are. I am grateful for the memories we have together, for all your support, all our laughs. You are the only good thing coming out of the predoc course. I cherish the uniqueness of our friendship. Thank you for going along with it, even though you might not understand me at times.

Thank you Gabi for being such an amazing colleague, a PhD-"brother from another mother", I am very grateful for your help and patience throughout these years, including through the finish line. Mas isso não se compara com sua amizade. Mozão, fica muito difícil resumir, vou ter muita saudade sua. Você é literalmente um amigo para os bons e maus momentos. O seu abraço e solidariedade são igualmente incríveis como a sua boa disposição, o seu humor. Você é incansável, foi tão generoso em tantas ocasiões, poxa...obrigada por ter normalizado procurar ajuda em 2018...obrigada por tudo o que você aturou. Obrigada pelo maracujá, o grelo de diamante, e todos os outros bops.

Muchas muchas muchas gracias guapísima Inés. Tan dificil es encontrar palabras... In chasing dreams, I was not expecting to gain a true friend. I am so glad I did. You are so unbelievably patient, intelligent, generous, funny, wise, and incredible. I admire you so much. I am so thankful for your friendship and for being so comfortable to being unapologetically myself around you. Thank you for being like a sister, for your unbelievable support in so many dimensions. You have taught me so much… Heidelberg would not be the same without you.

I would like to thank all of my family for their support. Tios, tias, primas, primos e Flávio. Muita ajuda, generosidade, boleias, merendas, e mais do que tudo, apoio incondicional. Sabe muito bem.

Obrigada minha mana manita por todo o teu apoio, por seres sempre honesta e por me pores em cheque, por partilhares outros mundos comigo e por seres confidente. Obrigada pelos teus Headspace nudges. Estou com muita vontade de ir viver para mais perto de ti.

Para os meus pais, que como já não tenho avós, são agora as primeiras pessoas do mundo a ligar a este tipo de coisas como uma dissertação, ou se saio no jornal. Obrigada por me terem dado um crescimento altamente privilegiado e por todos os sacrifícios que fizeram para que eu pudesse chegar até aqui. Amo-vos.

I would like to thank João for being my partner through thick and thin. For I've got a secret that no one else can know, that keeps my temperament even during times of snow...Between ~2500km, you always reminded me to keep the dream light and fire burning. Obrigada por seres compreensivo, e excelente de conversa (o cabelo mais bonito, coroa da mais linda testa). Tão pra inventar um mar grande o bastante que me assuste e que eu desista de você. In the cracks of light, you were there.

This thesis is written in the loving memory of sweet Nandita, my grandfather Avô Manel and my grandmother Avó Nana - who I miss deeply and with whom a lot of time I sacrificed because of this PhD. I started writing this section months before my submission date, when Avô Manel was still alive and eager that his granddaughter finished her degree and came back home. It was quite sad to have to add him to the dedication, but even sadder that nor him, Avó Nana, Avô Nelson or Avó Odete could see me graduate, as they would probably be one of the few people to care to read this dissertation and keep it with pride.

Lastly, I would like to thank Rafael Galupa, friend, mentor, co-founder of dreams, colleague, human. It was the absolute lottery winning to have crossed paths with you. It started with the reassurance I needed to be able to stay afloat on one path and it ended up with you shaping so much of who I became after these 4 years. It is important to highlight all the guidance and supervision you tirelessly did for this work, including designing together

a project that I was so excited about and all the “babysitting” to make sure it was completed. Through the finish line, you made sure to break down for me the steps to some light at the end of the tunnel, and held my hand throughout. You taught me that one could design ways to answer questions and do exciting science while having and maintaining structure. You showed me how one can be a critical and rigorous scientist, in the narrow sense, but also a complete, compassionate and multidimensional scientist, in the broad sense. Please do not forget about me if you become a famous youtuber. You truly listened and you saw me. You always pointed me to the essential of matters (big or small), to what ‘is invisible to the eye’, to the ‘things made of truth and joy’. You taught me that there is no such thing as far away. You reminded me to say ‘Silenzio Bruno!’. You taught me a different way of thinking about the world. And you accepted me 100% the way I am. Thank you for cheering for me when not even I managed to. I am honoured of what we built together and will always cherish our multiple adventures and our endless loladas. Para sempre: coragem hoje, abraços hoje.

Table of Contents

Table of Figures

1 Introduction

"It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner, have all been produced by laws acting around us ... Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows. There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved."

Charles Darwin, On the Origin of Species (1859)

1.1 Learning from Life

Life is as beautiful as it is complex, and so are living organisms, which have "endless forms" (as Darwin coined in the quote above). A contributor to these endless forms is the diversification of multicellular organisms. Within all multicellular organisms, humans are part of metazoans, a group that includes not only mammals but also fish, amphibians, reptiles, insects, and other animals (UCMP Virtual Museum of Paleontology). My thesis will touch on the theme of how such complex and different forms are born from "so simple a beginning" (see quote above), using the model organism *Drosophila melanogaster* (reviewed in Markow, 2015), which can be informally called "fruit fly". In this dissertation, I will focus on evidence and mechanisms studied using this organism, depicted in Figure 1.1.

A

B

Figure 1.1: **Representations of *Drosophila*.**
(A) Drawing of *Drosophila* by John Curtis (Curtis, 1833).
Reproduced with permission License Number: 5140241302053.
(B) Fabric Collage of *Drosophila* gifted to me by the *Drosophila* Research and Training Centre, Nigeria (https://*Drosophila*rtc.org/)

1.2 Learning with and from the fruit fly, *Drosophila*

In the quest to discover more about the mysteries of life (or "Learning from Life/Von Leben Lernen", as one can see written around EMBL's campus), scientists use and have used several model organisms (Alfred and Baldwin, 2015). Model organisms are organisms from different species that can be maintained in the laboratory and from which scientists can often infer or speculate about other species, and are well described. Well-described means that there is a lot of evidence being built over time about them, which makes it easier to go further in the questions researchers ask and try to answer. Leonelli and Ankeny describe it as "non-human species that are extensively studied to understand a range of biological phenomena, with the hope that data, models, and theories generated will apply to other organisms, particularly those that are in some way more complex than the original" (Leonelli and Ankeny, 2013).

Often overlooked by modern pressures for the anthropocentric utility of research and a focus on studying the mechanisms of biology using human cells, or mammal model organisms (Gregory Petsko, 2011), fruit flies are "unsung heroes" (Brookes, 2001) of biological scientific discovery (Prokop, 2018). Since the beginning of the 1900s, they have contributed to many

discoveries (reviewed in Roberts, 2006; Prokop, 2016), including implications for human health, such as the development of a chemotherapy drug (Briscoe and Thérond, 2005). Around 77% of disease-associated human genes have an orthologous gene (in other words, have a gene that matches them) in *Drosophila melanogaster* (Reiter *et al*., 2001). Among a wide variety of topics, fruit flies can even be used to study Parkinson's disease (Biosa *et al*., 2018) or neurotoxicity (Oyetayo *et al*., 2020; Abolaji *et al*., 2020).

Drosophila-based science includes discoveries that were highlighted by receiving the Nobel prize (Manchester Fly Facility, 2017) - meaning that a (regarded as) selected group chose them as some of the most important of all. These discoveries include “the role played by the chromosome in heredity” (Nobel Foundation, 1933) and the “molecular mechanisms controlling the circadian rhythm” (Nobel Foundation, 2017). Most relevant to this thesis, discoveries in the fruit fly “concerning the genetic control of early embryonic development" were performed at EMBL by Eric F. Wieschaus and Christiane Nüsslein-Volhard, which in conjunction with Edward B. Lewis, led to great insight into animal development (Nobel Foundation, 1995). I refer to Lewis’ studies later in this chapter.

The fruit fly and its related research can also be a great tool to strengthen science where cost-effectiveness is a bottleneck to do it (Palacios *et al*., 2020, Martín-Bermudo *et al*., 2017, Marta Vicente-Crespo, 2015). Furthermore, through it, researchers can share and involve society with science, with *Drosophila* being a great model for public engagement communication or science education (as reviewed by Patel and Prokop, 2017 with the case study of the Manchester Fly Facility; see also Patel *et al*., 2017).

1.3 From one egg to arms and wings

While most humans have arms and legs, flies have wings, legs, and halteres. However, both share a similar body organization, divided by different segments such as head, thorax, and abdomen, as represented in Figure 1.2.

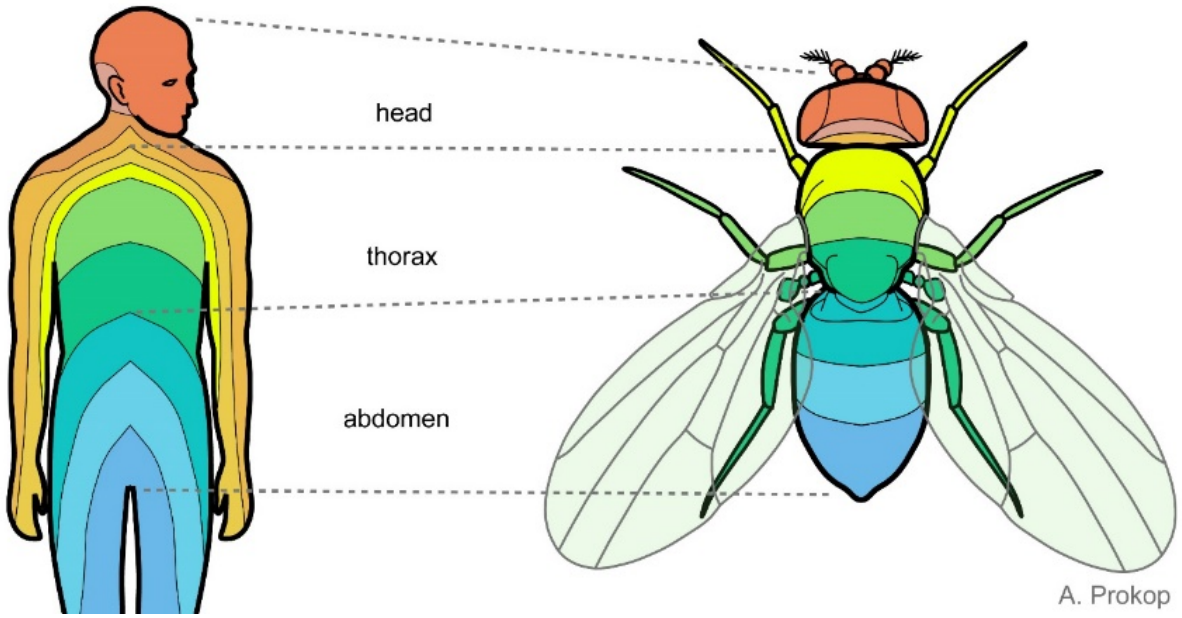


Figure 1.2: **Human and Fly share similar body plans.**
Scheme portraying the different body segments and sub-segments (each represented by a different colour and delineated by a grey line) in humans (left) and flies (right), and highlighting the similarities (dotted lines) through similar division between head, thorax, and abdomen. Between the fly's two pairs of legs closer to the head and the pair of legs closer to the "tail", you can see drawn two small organs that are the halteres, which help the fly to balance.
Reproduced from Manchester Fly Facility (2015). droso4schools: Online resources for school lessons using the fruit fly *Drosophila*. https://droso4schools.wordpress.com/ under Attribution 4.0 International (CC BY 4.0), license at https://creativecommons.org/licenses/by/4.0/legalcode.

This "body plan" consists of three "body axes", anterior-posterior (A-P), dorsal-ventral (D-V), and left-right (L-R). After egg fertilization, a single cell divides and multiplies. Progressively, different cells acquire distinct features (called cell differentiation) and start arranging in different groups of similar types, forming different tissues and segments. Over time, these groups of cells become more and more specified (also referred to as fate specification), giving rise to different organs and finally an adult form (Gilbert, 2000b).

The process by which cells in the embryo start arranging in different groups can be called "body patterning" (Takahashi *et al*., 2000). This process could be compared to building a new house, as Takahashi and colleagues did: "This mass of cells becomes subdivided into distinct groups (rooms) that eventually will exhibit functional specializations (furniture) later in development" (Takahashi *et al*., 2000). Nevertheless, there is a lot yet to be understood about these fascinating processes at a mechanistic level.

1.4 Hox and their paradox

The anterior-posterior ('head' to 'tail') axis segmentation is controlled by Hox genes, which, in the fly, are expressed and translated into Hox proteins in different embryo parts. These Hox proteins define segments that will form the different body parts in the adult (reviewed in Pearson *et al*., 2005) as depicted by Figure 1.3.

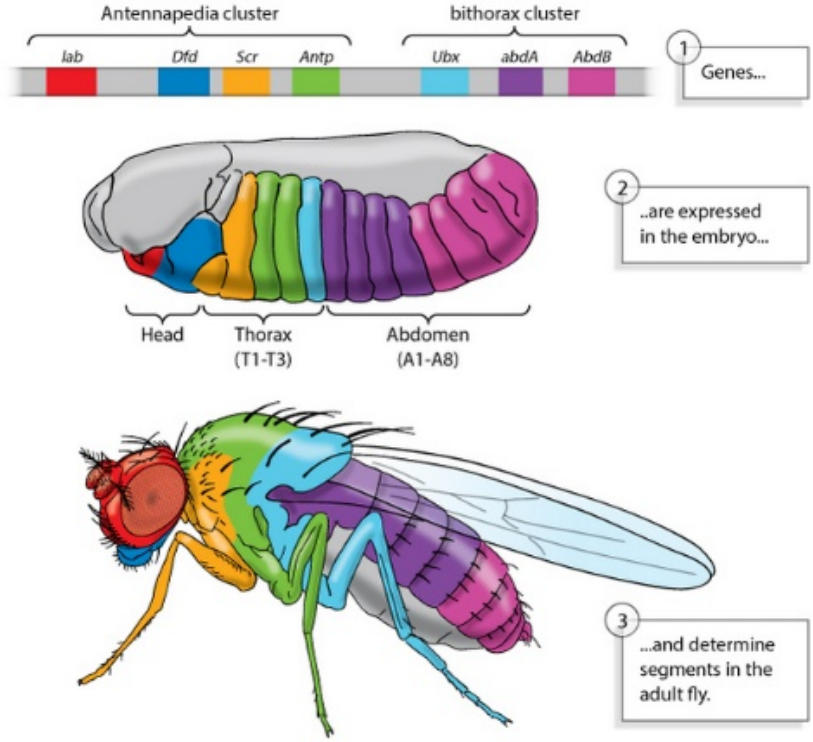


Figure 1.3: **Hox genes determine body segmentation.**
Genes (represented on the top, divided into two clusters and colour coded to correspond to middle and bottom parts), are expressed in different segments in the embryo (represented in the middle) and control the formation of different adult body parts (represented in the bottom). This is a general scheme; it is missing the *proboscipedia* gene and is not depicting the overlapping regions between Antp, Ubx, and AbdA in the embryo. Scheme reproduced with permission from Nature Education © 2013.

Hox proteins are transcription factors, which can be defined as "a protein containing at least one DNA binding domain along with domains that mediate interactions with cofactors and transcriptional machinery" (Kribelbauer *et al*., 2019). Hox genes share 180 base pairs (bp) in their sequence, which is called the 'homeobox' (McGinnis *et al*., 1984) and encodes for 60 amino acids which make a DNA-binding 'homeodomain' (reviewed in Scott *et al*., 1989). This system is conserved in many other animals, from humans to lampreys (reviewed in (Garcia-Fernàndez, 2005; Pascual-Anaya *et al*., 2013).

Drosophila melanogaster has eight Hox genes organized in two chromosomal clusters that exhibit a feature called “spatial collinearity”: their order in the genome matches their expression distribution and function in the anterior-posterior axis (McGinnis and Krumlauf, 1992). This order is depicted in Figure 1.3. In this genomic representation, the top, left side corresponds to the 3’-end and the right side to 5’-end, and the Hox *proboscipedia* (*pb*) gene, which has no loss of function phenotype in embryos, is missing between *labial* (*lab*) and *Deformed* (*Dfd*) (reviewed in McGinnis and Krumlauf, 1992). The embryonic expression of *Antennapedia* (*Antp)*, *Ultrabithorax* (*Ubx),* and *abdominal-A (abdA)* is partially overlapping in specific segments, which the coloured embryo in the middle is not representing but for which an illustrative depiction can be found in Pearson *et al*., 2005.

As a testament to the power of investigating with the fly, the understanding of what I have described in this section has origins in early experiments where one can observe the adult phenotypes that are a consequence of interfering with Hox genes, in particular their DNA-binding-domain (Lewis, 1978; Kaufman *et al*., 1990). On a personal level, I do not know anyone who is not in awe when I show them a picture of Ed Lewis’ four-winged fly. In fact, I have kept a similar picture on the wall of my lab desk throughout this Ph.D. thesis research. The four-winged fly results from mutations to the Hox gene *Ubx* that lead to the transformation of the third thoracic segment into the likeness of the second thoracic segment, which gives rise to the pair of wings (Bender *et al*., 1983). This is called a ‘homeotic transformation’ which can be generally described as the “transformation of one body region into the likeness of another” (Pearson *et al*., 2005) (Figure 1.4 A and B). Among more extreme examples of transformations, while Ed Lewis tried to make a ten-legged fly without success (Crow and Bender, 2004), he managed to make an eight-legged fly (Lewis, 1963).

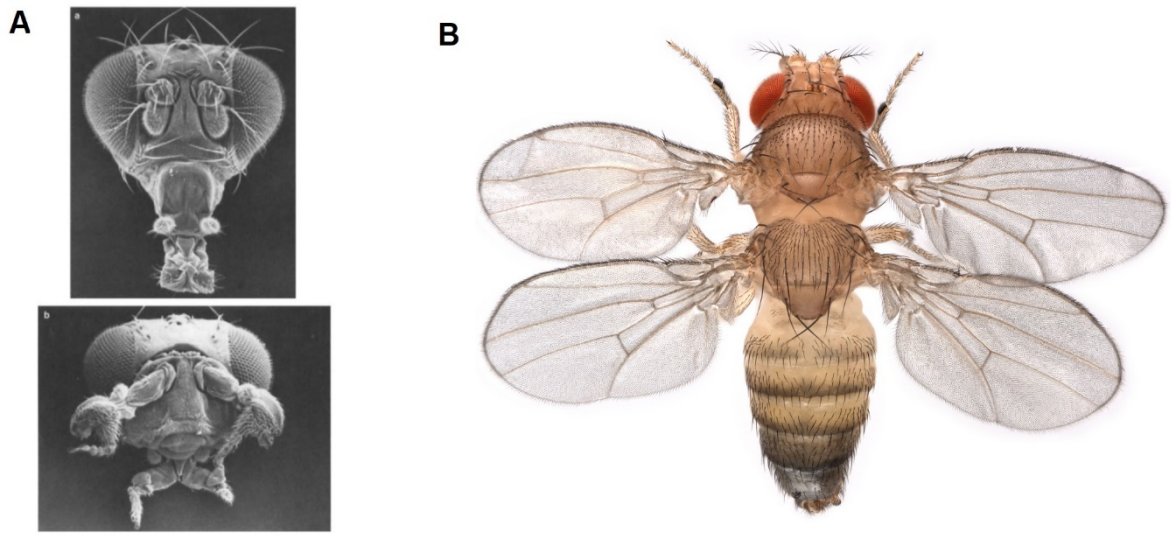


Figure 1.4: **Transforming head structures into legs or doubling the number of wings.**
(A)Through the double mutation of the *Antp* and *pb* genes, the antennae and labial palps - seen in the non-mutant fly in (a) - transform into (first and second thoracic) legs – seen in mutant-fly in (b). Reprinted from Advances in Genetics, Vol 27, Issue C, Molecular and genetic organization of the antennapedia gene complex of *Drosophila* melanogaster, Pages 309-362, 1990, with permission from Elsevier License Number: 5140831234675. (Kaufman *et al*., 1990)
(B) *Drosophila melanogaster* Ubx mutant (abx bx[3] pbx/Df(3R)P2) with four wings, instead of the usual two wings. Reproduced from Nicolas Gompel's Lab (2021). https://gompel.org/ under Attribution-NonCommercial-ShareAlike 4.0 International (CC -NC-SA BY 4.0), license at https://creativecommons.org/licenses/by-nc-sa/4.0/

These are striking observations. In humans, the phenotypes of mutations to the HOX genes can also be observed, in cases of, for example, limb alterations such as polydactyly (Goodman, 2002). Hox proteins, expressed throughout the embryo, contribute to the formation of different body segments by their action as transcription factors (but probably not exclusively, as I explain later in this chapter): they bind to genomic regions called 'enhancers' promoting or repressing gene expression (reviewed by Spitz and Furlong, 2012). This way, they, directly and indirectly, regulate several downstream target genes that have various biological functions, from cell death to cell differentiation regulators, among several others (reviewed in (Hueber and Lohmann, 2008). Hox target enhancers have been described as sharing common properties such as tissue specificity (that can be due to regulation by other molecules) and requiring the binding of multiple Hox-monomers and exhibiting the sites for that (Pearson *et al*., 2005). The different Hox Transcription Factors share a preference for binding a similar sequence *in vitro* – more frequently TAATTA, with slight variations (Berger *et al*., 2008; Noyes *et al*., 2008).

Despite binding to similar sequences with high affinities, Hox proteins have specific functions, as they control different genes that contribute to the formation of different anatomic features. Understanding how this class of proteins regulate their target genes *in vivo* will be critical to unravel the mechanisms that explain this paradox. A lot has been done (that I describe in the continuation of this chapter) but there is also a lot to still be understood (that I refer to in the Conclusions and Perspective chapter). It is, therefore, important to highlight that *in vit*ro-discovered or theoretically predicted binding sites are not necessarily equivalent to *in vivo* binding sites, as conditions differ among cell types including chromatin states, developmental stages, other molecular components.

This paradox could be explained by having the expression of each Hox protein being differential and cell-type specific, but this is not always the case. For example, they can partially overlap (Gould *et al*., 1997), as is the case in *Drosophila* of Antennapedia (Antp), Ultrabithorax (Ubx) and abdominal-A (abdA) (McGinnis and Krumlauf, 1992). Another clue could lie in the protein domains other than the DNA-binding domain, such as the ones that drive interactions with other molecules that also bind DNA. This is a strategy used by other transcription factors, such as bZIP proteins - transcription factors that exhibit a Basic Leucine Zipper DNA-binding domain – which combine very diversely with various proteins (Rodríguez-Martínez *et al*., 2017). For Hox Transcription Factors in *Drosophila*, there are at least two known co-factors called Extradenticle (Exd) and Homothorax (Hth) (reviewed in Mann *et al*., 2009). Complexes between Hox proteins and these cofactors exhibit what has been called “latent specificity”, where the binding to the cofactor alters the structure of the Hox protein, and “reveals” an emergent biding preference (Slattery *et al*., 2011).

The questions about this paradox do not extinguish themselves in DNA-binding. Hox Transcription Factors have been described to affect gene regulation beyond direct enhancer-binding, through interacting with a multitude of molecules and affecting a variety of processes, in what has been defined as “multi-step” or “multi-level regulatory functions” (Carnesecchi *et al*., 2018). These include binding to promoters and other genomic regions, interacting with transcriptional machinery (such as the Polymerase Initiation Complex), acting on chromatin state - by regulating histone modifications -, affecting chromatin conformation - by regulating loop formation -, and acting on mRNA processing (reviewed in Carnesecchi *et al*., 2018).

Finally, locus specificity – “the ability of a specific genomic site to preferentially bind a transcription factor as opposed to other potential sites in the genome” (Kribelbauer *et al*., 2019), remains an outstanding question. In some cases, different Hox proteins can also target the same genes. If these different Hox proteins are co-expressed, then there are other layers of specificity here at play (reviewed in Mann *et al*., 2009), such as paralog specificity - “the ability of a DNA ligand to recruit a particular paralog among multiple available transcription factors from the same family” (as defined by Kribelbauer *et al*., 2019).

1.5 Low-affinity Binding Sites

One mechanism to confer transcription factor specificity during animal development is by using low-affinity binding sites (Crocker *et al*., 2015, 2016; Farley *et al*., 2015, 2016; Ramos and Barolo, 2013). A low-affinity binding site (reviewed in Kribelbauer *et al*., 2019; Crocker *et al*., 2016) can be defined as a “DNA site bound up to 1,000-fold more weakly than the optimal DNA sequence, but still more strongly than the immediately surrounding sequence” (Kribelbauer *et al*., 2019). These have also been described as “suboptimal” (Farley *et al*., 2015) or “submaximal” (Bhimsaria *et al*., 2018). Although common in eukaryotes, these sites are not easy to predict or distinguish from non-specific binding by most models used to determine the binding preferences of transcription factors (reviewed in Kribelbauer *et al*., 2019). Nevertheless, recent algorithms have been developed to identify them (reviewed in Kribelbauer *et al*., 2019; Rastogi *et al*., 2018).

Concerning embryonic development, low-affinity binding sites have been described to be relevant to regulate varied features of gene expression, not only specificity but also timing, location, and level, as depicted in Figure 1.5 (Crocker *et al*., 2016).

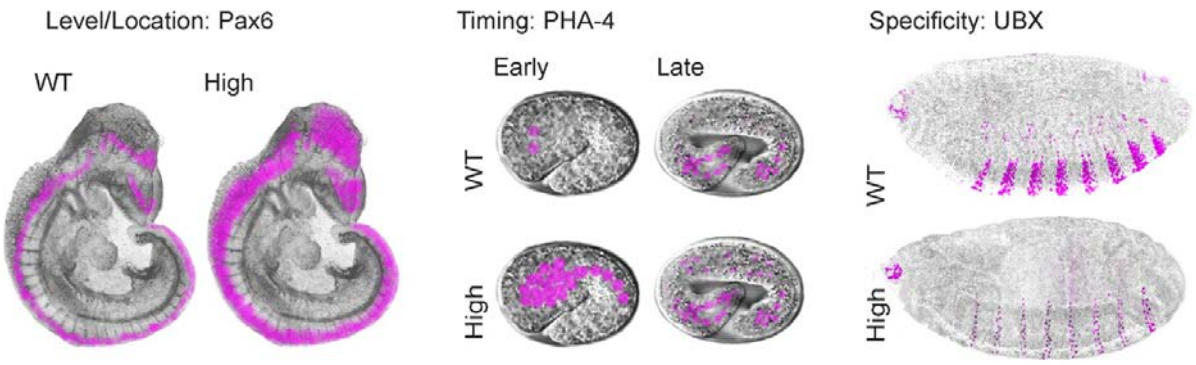


Figure 1.5: **Examples of low-affinity binding site regulation of embryonic gene expression.**
Evidence from consequences of substituting low-affinity for high-affinity sites. The first example from the left concerns PAX6 sites from Scardigli *et al*., 2003. The example in the middle concerns PHA-4 sites from Gaudet and Mango, 2002. The example on the right concerns Ubx sites from Crocker *et al*.,2015.
Reprinted from Current Topics in Developmental Biology, Vol 117, Chapter Twenty-Seven - The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution, Pages 455-469, 2016, with permission from Elsevier License Number: 5143290650580. (Crocker *et al*., 2016)

In the case of the Hox Transcription Factor Ubx, low-affinity sites have also been reported to mediate their own negative autoregulation (Delker *et al*., 2019). For this protein, replacing low-affinity sites with high-affinity sites - which can be occupied simultaneously by several Hox proteins (Berger *et al*., 2008) – resulted in a reduction of specificity (Crocker *et al*., 2015). As shown in Figure 1.6 (and included in the scheme of Figure 1.7B), mutating low-affinity sites in an enhancer sequence regulated by Ubx in a series of predicted increasing binding affinities, resulted in ectopic expression and modified expression levels in places of usual expression (Crocker *et al*., 2015).

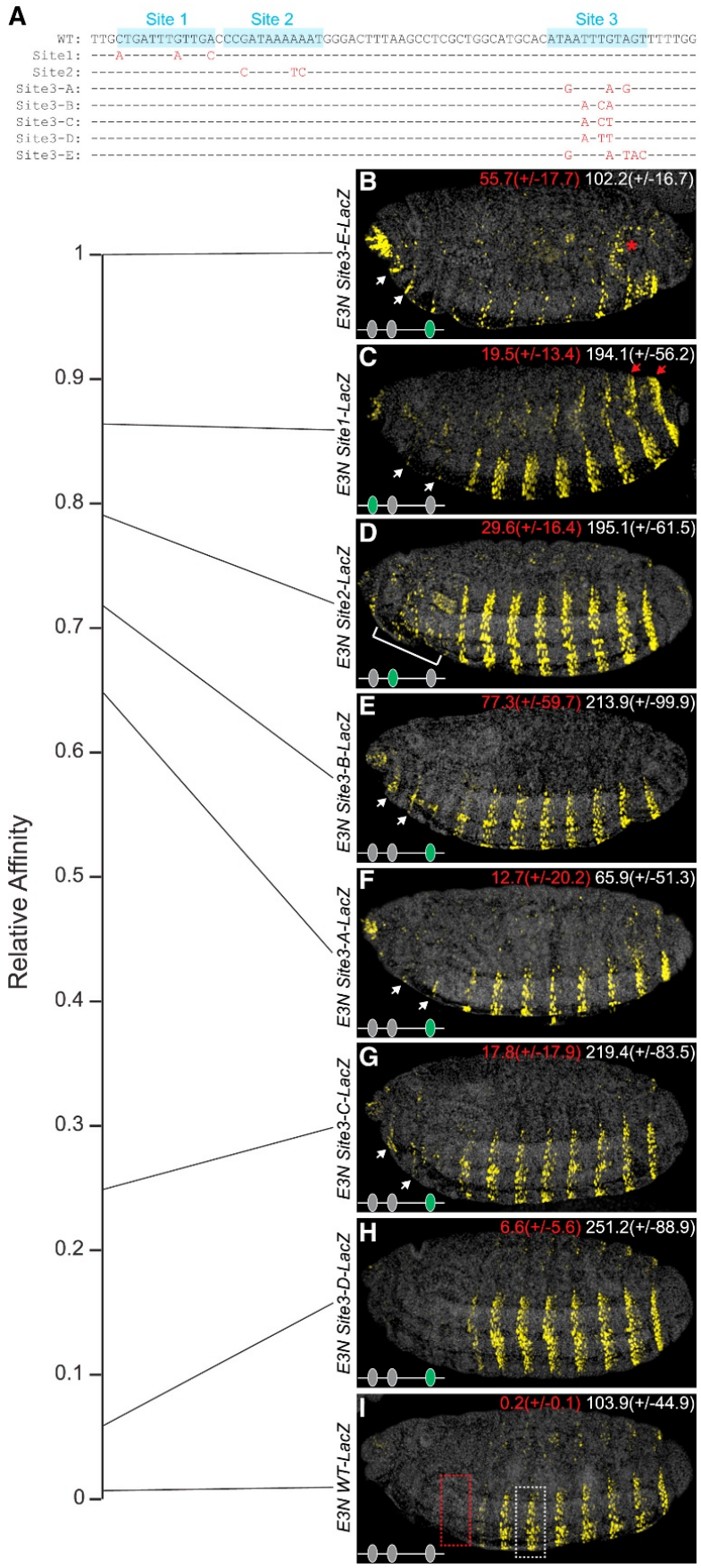


Figure 1.6: **Substituting low-affinity Ubx sites for higher affinity sites leads to gene expression changes.**

(A) Series of sequences aligned: starting from *E3N* wildtype and then the mutated sequences, where red letters correspond to modified sequence (unaltered sequence in dashes).
(B-I) Embryos ordered by SELEX-seq predicted relative affinities of respective sequences (from (A)) they carry, stained for βGal protein (sequences are driving *lacZ* reporter). Numbers outside of parentheses in each panel correspond to average levels (in fluorescence intensity arbitrary units) of expression of regions outlined in (I) (n=10 for each line). ±1 SD is represented by numbers inside parentheses.
(B-G) White brackets and arrows mark expression anterior to A1.
(C) Red asterisk (intestine) or arrows (dorsal and lateral) mark ectopic staining and expression.
Reprinted from Cell, Vol 160, Issues 1-2, Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness, Pages 191-203, 2015, with permission from Elsevier License Number: 5143051160626. (Crocker *et al*., 2015).

The same study shows that there appears to be a trade-off between binding affinity and specificity for Hox Transcription Factors (Crocker *et al*., 2015, Figure 1.7). *In vitro*, and in complex with Exd, there is an inverse correlation between relative affinity of 12mer sequences for exclusively Ubx/AbdA-Exd and the proportion of these sequences that are bound by several Hox-Exd complexes (Crocker *et al*., 2015), as shown in Figure 1.7A.

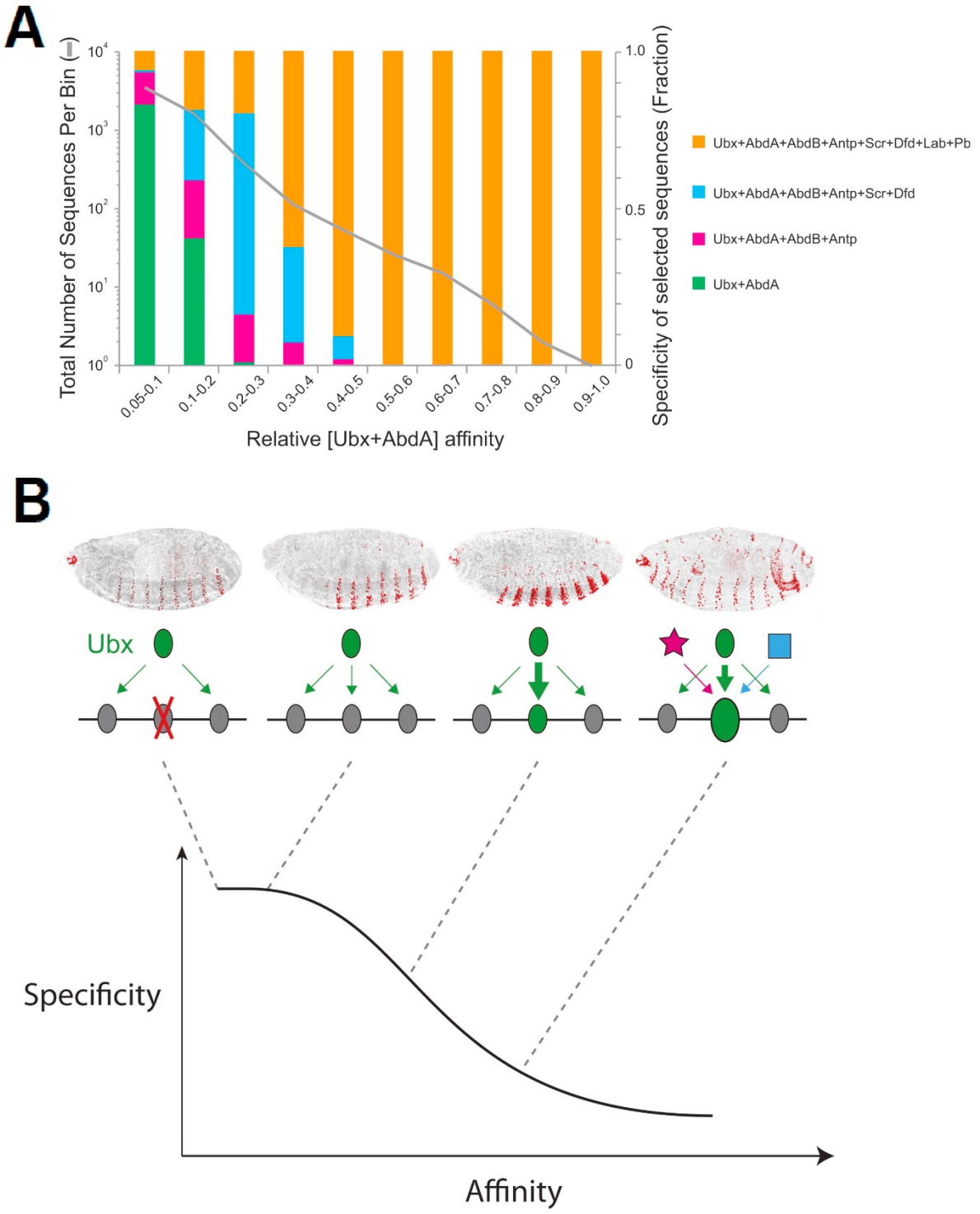


Figure 1.7: **Sequence Affinity and Specificity Trade-off.**
(A) Colored bars show (specificity groups) Ubx/AbdA-Exd relative affinity of 12mer sequences against the proportion of 12mers bound by various Hox-Exd complexes. The grey line corresponds to the number of 12mer sequences in each affinity bin. Coloured in green, sequences specific for Ubx/AbdA-Exd binding are more prevalent in lower affinity bins.
(B) Scheme portrays the sequence-affinity trade-off trend from (A), adding the figurative representation for the context of Ubx, with also representative examples from stainings of enhancer expression. Low-affinity binding sites in grey and high-

affinity in green. The second from the left represents the wild-type situation. The first from the left represents a mutation of a low-affinity binding site. The star and the box represent other Hox factors. The growing oval shape represents higher affinity.
Reprinted from Cell, Vol 160, Issues 1-2, Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness, Pages 191-203, 2015, with permission from Elsevier License Number: 5143051160626. (Crocker *et al*., 2015).

It is still not fully understood how high-affinity Hox transcription factor binding sites, known to have low specificity *in vitro* (Mann *et al.,* 2009), regulate precise patterns of gene expression during development. Nevertheless, the correct tuning of gene regulation can probably be aided by a spectrum of different binding site affinities for different transcription factors. For example, the use of high-affinity sites is essential to pattern lineage specification programmes, such as cardiogenesis in *Drosophila*. In these tissues, the repression of an enhancer requires high-affinity sites of transcription factor Yan. At the same time, its paralog Pnt mediates activation through low-affinity sites (Boisclair Lachance *et al*., 2018). In the case of Yan and Pnt, the preference of different binding affinities facilitates the transcription factor that uses high-affinity sites can more easily (Yan, with lower concentration) to outcompete its paralog that uses low-affinity sites (Boisclair Lachance *et al*., 2018). I discuss the role of transcription factor concentration in section 1.7 of this chapter, with a focus on Hox factor Ubx.

1.6 The case of Ubx and *shavenbaby*

The Hox factor Ubx, mentioned in earlier sections of this chapter, specifies the third thoracic (T3) and abdominal segments of *Drosophila* (Bender *et al*., 1983). This Hox Transcription Factor interacts with many proteins, featuring tissue-specific interactomes. Ubx can specify different tissues and regulate genes as an activator and a repressor (Hersh *et al*., 2005; Domsch *et al*., 2019; Loker *et al*., 2021). Ubx can also regulate splicing by interacting with transcriptional machinery (Carnesecchi *et al*., 2021). It has also been shown recently that Ubx can affect chromatin conformation and accessibility (Domsch *et al*., 2019; Loker *et al*., 2021), and interact with chromatin remodellers (Domsch *et al*., 2019). Ubx interacts with other transcription factors using different protein domains (Carnesecchi *et al*., 2020, Hsiao *et al*., 2014). Besides its DNA-binding Homeodomain, other Ubx protein domains have been described, such as a transcription activation domain (Liu *et al*., 2008), a YPWM and a UbdA motif for interaction with Exd (Merabet *et al*., 2007; Passner *et al*., 1999), a linker region that separates the former from the DNA-binding Homeodomain (Saadaoui *et al*., 2011), and a partial repression domain (Ronshaugen *et al*., 2002).

Throughout the entire Ubx protein structure, including in several domains outside of the DNA-Binding Homeodomain, there have been predicted several protein-protein interaction motifs, several of which have intrinsically disordered regions, which may contribute to *in vivo* binding specificity (Hsiao *et al.*, 2014; Liu *et al.*, 2008; Brodsky *et al.*, 2020).

One of the target genes that Ubx regulates is *shavenbaby (svb)* (Crocker *et al*., 2015). The *shavenbaby* gene specifies the formation of trichomes (Figure 1.8A-B), hair-like projections in the epidermis of *Drosophila* larvae, by activating over 150 genes (reviewed in Kittelmann *et al*., 2021). Ubx controls multiple *svb* enhancers (Crocker *et al*., 2015, Tsai and Alves *et al*., 2019). Ubx is necessary and sufficient for the expression of *svb* mRNA in the embryo and consequential trichome formation in the larvae, mostly in the A1 segment (Crocker *et al*., 2015, Figure 1.8 A-F). The *cis*-regulatory region of *svb* includes several enhancers, of which 3 enhancers, *DG3*, *E3,* and *7* drive overlapping stripe-like patterns of expression along the ventral abdominal segments of the *Drosophila* embryo (Figure 1.8 G-H).

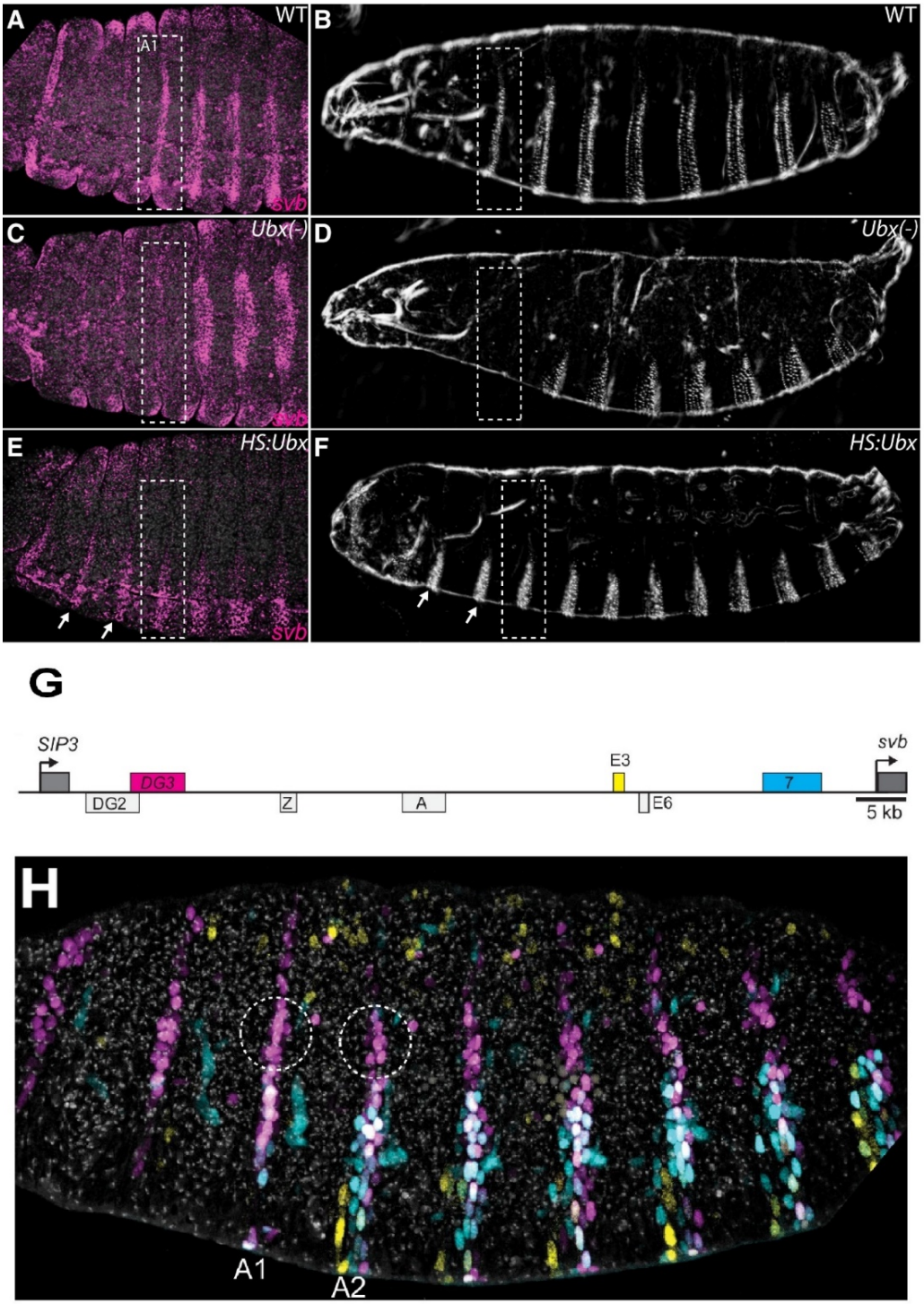


Figure 1.8: **Ubx regulates *svb, whose* ventral enhancers drive overlapping patterns.**

(A-F)On the left, fluorescent stained embryos for *svb* mRNA. On the right, larval cuticle preps. Genotypes are indicated in panels.

(C-D)Segment A1 is transformed into a thoracic segment due to Ubx loss of function. Boxes highlight that this segment lacks *svb* expression and larval trichomes.

(E-F)Arrows point to thoracic segments that transformed into the likeness of A1 due to the ubiquitous expression of Ubx.

(G)Upstream *svb* *cis*-regulatory region schematized. Boxes correspond to embryonic enhancers. Ventral enhancers are coloured *DG3* (pink), *E3* (yellow), and *7* (blue).
(H)Expression patterns driven by ventral enhancers (staining of reporter driven by these sequences) are coloured *DG3* (pink), *E3* (yellow), and *7* (blue). Circled regions highlight A1 and A2 segment areas of exclusive *DG3* coverage. These will be later referenced in the results chapter relating to Figure 2.6.

Associated with Hox co-factors Exd and Hth, Ubx binds to clusters of low-affinity binding sites in these enhancer regions of the *svb* locus, allowing its expression (Crocker *et al*., 2015). The example showing earlier in Figure 1.6 that substituting low-affinity Ubx sites for higher affinity sites lead to gene expression changes, was from a minimally-functional sequence of Ubx-responsive enhancer *E3*, called *E3N* (Crocker *et al*., 2015).

1.7 Transcriptional microenvironments

A possible mechanism for how low-affinity interactions generate a robust transcriptional output is through “transcriptional microenvironments” (reviewed in Tsai *et al*., 2020). Transcriptional microenvironments are “regions within the nucleus that are locally enriched for specific transcription factors and co-factors” (as defined in Tsai *et al*., 2020). With this definition, they differ from ‘transcriptional hubs’ in the sense that microenvironments are locally transcription-factor enriched but can be repressing or activating, whereas transcriptional hubs are “localized nuclear compartments that can sustain high levels of transcriptional output through high local concentrations of polymerases and transcription factors” (as defined in Tsai *et al*., 2020) or “prelooped topologies [that] serve as hubs or traps for accumulation of Pol II and other complexes required for gene expression” (as defined in Furlong and Levine, 2018).

These local regions of high transcription factor concentrations have been found around sites of active *svb* transcription for the transcription factor Ubx (Tsai *et al*., 2017, Figure 1.9 A-J).

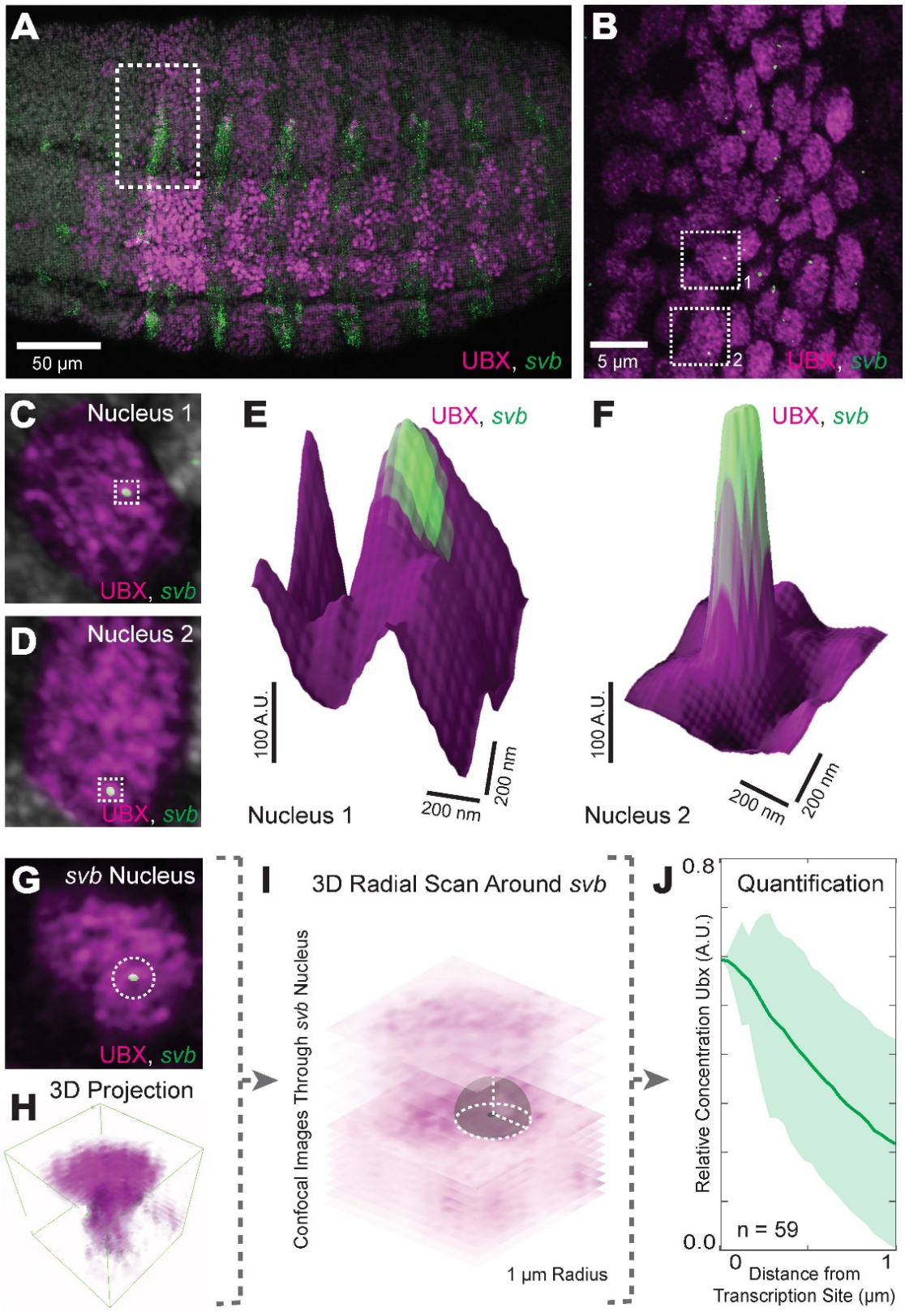


Figure 1.9: **Ubx-*svb* transcriptional microenvironments.**

(A) Staining of embryo shows in magenta Ubx protein and in green *svb* intronic mRNA. Actively transcribed loci are marked by bright spots. White represents the sum of green and magenta, where both protein and mRNA have high levels.
(B) The same embryo as in A, but the picture was taken with Airyscan for higher magnification of the region marked in A. Green corresponds to transcription sites.
(C-D) Boxes in B mark nuclei that are here represented at higher magnification.
(E-F) 3D surface plots of the pictures from C and D. Height corresponds to Ubx intensity. The centre (in green) is the *svb* transcription site.
(G) Representative nucleus for Ubx distribution quantification around transcription site, and respective confocal stack 3D view in (H).
(I) Portrayal of the quantification method. Calculation of 3D radial distribution of average Ubx intensity on the surface of a sphere (example in gray, white dashes represent 1 µm radius (r)) that has the *svb* transcription site as the centre.
(J) Average relative concentration of Ubx (in Arbitrary Units of fluorescence intensity) vs distance from the transcription site. Variance is represented by the shaded region.

These regions are also found around sites of active *svb* ventral enhancer-driven reporter transcription while not being found around transcription sites of a reporter driven by a synthetic network unrelated to Ubx (Tsai *et al*., 2017, Figure 1.10 A-H, M-N). This suggests a relationship between these local high-Ubx regions and Ubx's regulation of *svb*'s transcription. Ubx local enrichment profile changed around reporter transcripts when the transgenic *svb* ventral enhancers driving their expression are modified (Tsai *et al*., 2017, Figure 1.10 G-R).

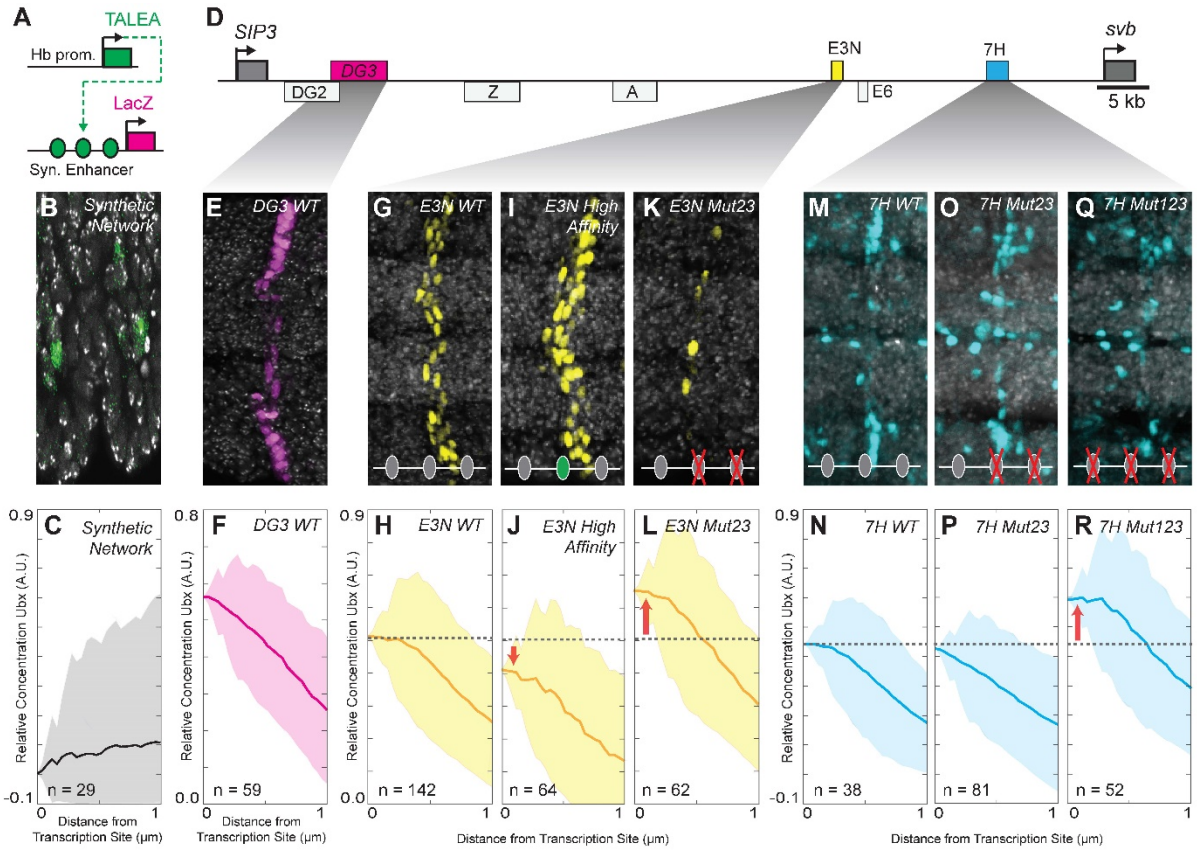


Figure 1.10: **Testing Ubx microenvironments in embryos carrying different sequences.**

(A)Scheme: Hunchback (Hb) promoter drives synthetic TALEA transcription network. Green circles are TALEA-binding sites.

(B)ß-Galactosidase stained early stage 15 embryos that have the TALEA synthetic network.

(C)Relative concentration of Ubx vs the distance from transcription sites (from B).

(D)Scheme: *svb* locus, with coloured embryonic enhancers where shadows connect to respective data.

(E, F) ß-Galactosidase staining and respective Ubx concentration distribution over distance for early stage 15 embryos with *DG3-LacZ* reporter constructs.

(G-L)ß-Galactosidase staining and respective Ubx concentration distribution quantification for early stage 15 embryos with *E3N-lacZ* reporter constructs. Modifications to Ubx-Exd sites are depicted.

(M-R) ß-Galactosidase staining and respective Ubx concentration distribution quantification for early stage 15 embryos with *7H-LacZ* reporter constructs. Modifications to Ubx-Exd sites are depicted.

(C, F, H, J, L, N, P, R) A.U. means Arbitrary Units of fluorescence intensity. Variance in shaded regions.

Furthermore, there was co-enrichment of Hth around transcription sites (Tsai *et al.*, 2017). The idea that Ubx drives transcription of *svb* through fast and short-lived interactions is furthermore supported by the observation that Ubx binds transiently to DNA using live imaging of

expanded early (stage 5) embryos (Tsai *et al*., 2017), which is in accordance with other studies (Morisaki *et al*., 2014). Since these experiments rely on the detection of active transcription sites, they bring transcriptional microenvironments together with probable transcriptional hubs. This finding of Ubx-*svb* microenvironments coincided with a similar finding for transcription factor Bicoid, which were then called “dense Bicoid hubs” (Mir *et al*., 2017).

The formation of subnuclear ‘transcriptional hub’ compartments or ‘condensates’ has been proposed to be a result of cellular biochemical reactions, namely liquid-liquid “phase separation of multi-molecular assemblies” (Hnisz *et al*., 2017). This model integrates different phenomena such as “super-enhancer” clusters (Hnisz *et al*., 2013) and transcriptional bursting (Fukaya *et al*., 2016). RNA Pol II clusters have been observed to form in response to external signals (Cisse *et al*., 2013;Cho *et al*., 2018) and can be predictive of mRNA output (Cho *et al*., 2016). Accordingly, Sox2 enhancers were also observed to cluster and segregate from heterochromatin but overlapping with Pol II clusters (Liu *et al*., 2014). This model is also in line with enhancer-promoter communication through close proximity but not direct interaction (Lim *et al*., 2018a).

Topologically associating domains, or genomic loops, and the molecules that regulate them (such as insulators) contribute to the formation of subcompartments and the heterogeneous nucleus composition (reviewed in Furlong and Levine, 2018). The layer of chromosomal arrangements adds to the dynamic properties of regulatory activity in enhancer regions (Rodríguez-Carballo *et al*., 2020). Thus, it is important not to forget the relevance that chromosome topology, not only in transcriptional microenvironments but also in regards to studying gene regulation in general (reviewed in Cavalheiro *et al*., 2021) and developmental enhancers in particular.

In sum, during development, complex gene expression patterns are formed, relying on the spatiotemporal coordination of proteins and genomic regulatory regions, among many other players and layers of processes. Open questions remain. For example, it is still not known whether high-affinity Hox transcription factor binding sites might use microenvironments or if this is specific to low-affinity enhancers. Furthermore, there is more to learn about how Hox factors achieve specificity, how different binding affinity sites are orchestrated, or more details about transcriptional microenvironments, including their functional importance, their molecular

components, and the involvement of enhancer clustering and/or transcription factors in their formation and function.

1.8 Purpose of the study

The purpose of my Ph.D. thesis, “Interrogating the complex role of Ubx and multi-enhancer transcriptional hubs throughout development”, is to elucidate the mechanisms that surround the activity of Ubx-related gene regulation in *Drosophila* throughout development. Towards this aim, I have attempted to answer the following questions:

1. Do multi-enhancer interactions contribute to low-affinity *svb* transcriptional microenvironments?

2. Are microenvironment features specific to low-affinity enhancers or can Ubx high-affinity-site-containing ‘enhancers’ exhibit similar properties?
    2.1. When and where are Ubx high-affinity-site-containing sequences capable to drive expression?
    2.2. Are Ubx-enriched distribution profiles exclusive to low-affinity enhancers or can they be observed around transcriptionally-active loci from a Ubx high-affinity-site-containing ‘enhancer’?
    2.3 Can transcriptionally-active loci from a Ubx high-affinity-site-containing ‘enhancer’ exhibit *trans*-chromosomal co-localization with Ubx target *svb*, suggestive of multi-enhancer interactions?

3. Is recruitment of Ubx sufficient to drive multi-enhancer clustering?

In the following three chapters of this thesis, I describe and discuss what I have found, including findings in collaboration with colleagues, who I listed in the “Contribution” sections at the end of each chapter.

2 Multi-enhancer interactions contribute to low-affinity *svb* transcriptional microenvironments

As I mentioned in the previous chapter, a lot is still to be understood and discovered regarding the components and mechanisms surrounding low-affinity Ubx-*svb* transcriptional microenvironments, as well as their relevance in the developmental context. What is the contribution of individual enhancers to this microenvironment? Do they interact? If so, is their interaction relevant? Is it a feature of these microenvironments? In Figure 2.1, I depicted two minimalistic ways of imagining a microenvironment, considering (Figure 2.1B) or not (Figure 2.1A) individual enhancers within this context.

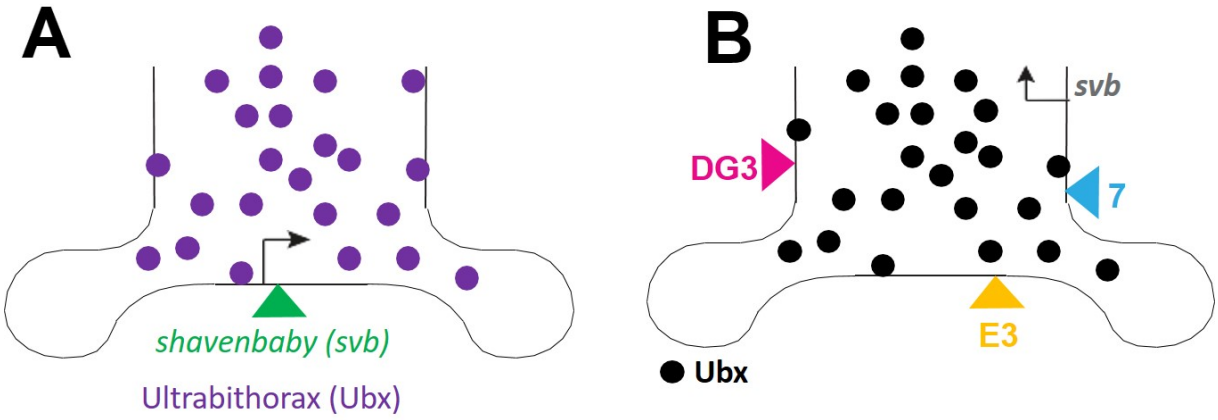


Figure 2.1: **Scheme of Ubx-*svb* microenvironment.**
(A) The thin line represents the genome. The *svb* locus is depicted in green at the bottom and high local concentrations of Ubx in purple dots.
(B) The thin line represents the genome. High local concentrations of Ubx are depicted in black dots and *svb*'s locus is represented by its three ventral enhancers, *DG3* (left, in pink), *E3* (bottom, in yellow), *7* (right, in blue).
For the shape of the genomic line, I took inspiration from a figure Justin once illustrated for a report for a fellowship I applied to.

2.1 Transcription sites from related *svb* enhancers in different chromosomes can co-localize

Working together with Justin and Albert, I found that related enhancers can co-localize to the same microenvironments independently of their chromosomal location. This finding comes from the observation of immunofluorescence stained and *in situ* hybridized fixed embryo samples (see Materials and Methods, 6.5) where transcripts from active *svb* enhancers (*E3N* and *7H*) inserted into a different chromosome co-localize (with distances around 250 nm) with endogenous *svb* transcripts (Figure

2.2 B-C). *E3N* and *7H* are minimal enhancers derived from *svb* (Crocker *et al*., 2015). Co-localization (with distances around 200 nm) of endogenous *svb* transcripts (Figure 2.2 A, 2.2 B first picture from the left, 2.2 C first box from the left, Tsai *et al*., 2017) had been observed, suggesting that homologous *svb* loci might share a Ubx microenvironment. When inserted individually on chromosome 3 to drive *lacZ* expression, both *E3N* and *7H* enhancers possibly share (a) Ubx microenvironment(s) with endogenous *svb,* which is located on chromosome X (Figure 2.2B-D). It does not seem to be the case that endogenous *svb* tends to co-localize with any transcript from chromosome 3, as there is no co-localization between endogenous *fkh,* located on chromosome 3, and endogenous *svb* (Figure 2.2 B last picture from the left, 2.2 C last box from the left). The distances between these transcripts average 1 µm (1000 nm) (Figure 2.2 C last box from the left). The distances observed between transcripts that appear to co-localize are close to the optical resolution of Zeiss LSM 880 Airyscan-images (~140 nm).

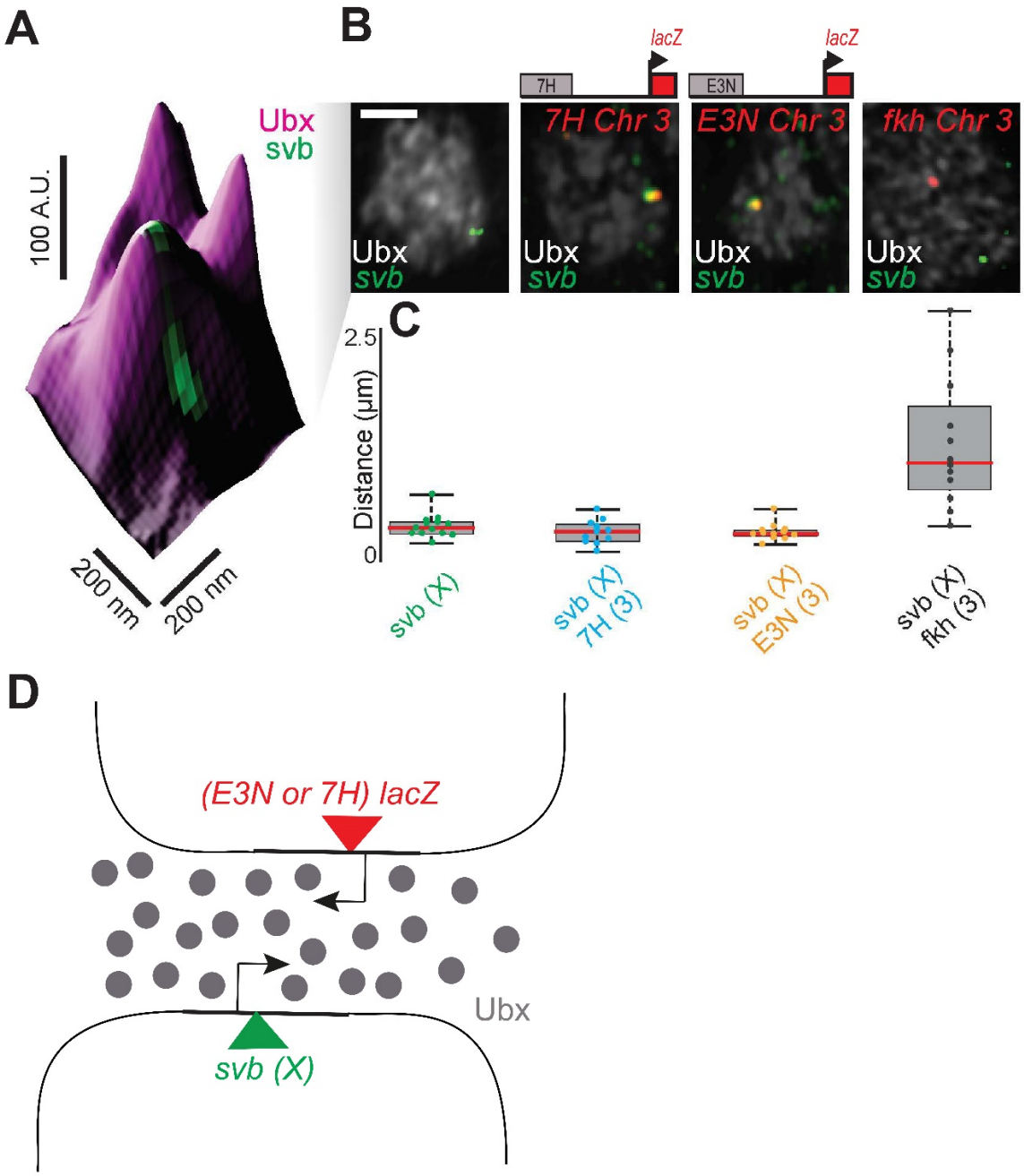


Figure 2.2: **Transcription sites from related *svb* enhancers in different chromosomes can co-localize.**
(A) 3D surface plot - from the leftmost nucleus staining shown in (B) - showing in close proximity two *svb* transcription sites (introns of nascent *svb* mRNA) in green in what might be sharing a Ubx microenvironment. Ubx protein in purple, where the height of the plot indicates Ubx fluorescence intensity in Arbitrary Units (A.U.). (B) Series of nuclei stainings from stage 15 embryos (segment A1). On all of them, Ubx protein is in white, introns of nascent *svb* mRNA in green. In red is mRNA from *lacZ* reporter driven by *7H* (2nd-panel counting from the left), mRNA from *lacZ* reporter driven by *E3N* (3rd panel), or introns of nascent *fkh* mRNA (4th panel), with all of these sites being located on chromosome 3 (Chr 3). Scale bar represents 2 µm. On top of the 2nd and 3rd panels, there is a schematic depiction of the transgenic constructs inserted on chromosome 3, with the respective enhancer represented in a grey box and the *lacZ* reporter in a red box. 3D surface plot of the leftmost nucleus is shown in (A).

(C) Plot of distances between transcription site pairs, for n = 13 (*svb* only), n = 11 (*svb* & *7H*), n = 12 (*svb* & *E3N*), and n = 12 (*svb* & *fkh*) transcription sites. Minima and maxima of the distribution is represented by black bars, mean by red line and standard error by grey box.
(D) Scheme representing potential sharing of Ubx microenvironment by co-localizing transcription sites in different chromosomes. The thin line represents the genome. The *svb* locus is depicted in green at the bottom, the minimal *svb* enhancer-reporter construct is depicted in red on top, and high local concentrations of Ubx in grey dots.
Panel A reproduced from and panels B and C adapted from Tsai *et al.*, 2017 under Attribution 4.0 International (CC BY 4.0), license at https://creativecommons.org/licenses/by/4.0/legalcode.

Overall, these findings suggest that transcriptional microenvironments might involve multiple, higher-order chromosomal interactions. It was still not clear if these co-localizations were essential features of microenvironments nor how low-affinity interactions generate robust transcriptional outputs nor what the developmental relevance of these might be.

2.2 Multi-enhancer interactions compose transcriptional microenvironments and confer robustness against environmental stress

To better understand what is the contribution of these different microenvironment components, it is pertinent to ask whether microenvironments are disrupted or impaired by the loss of a regulatory sequence. The contribution to microenvironments was tested using a mutant with the deletion of a partially redundant *svb* enhancer, *DG3*. For ease of reading, I will call this line 'DG3-deletion line', although the deletion includes other enhancers as well in the vicinity (schematically depicted in Figure 2.5A) (Tsai and Alves *et al.*, 2019), and is officially called and referred to in figures as "Df(X)svb^{108}". To focus on DG3-deletion related effects, the following experiments that I describe are focused on the embryonic A1 segment. Justin had found that this enhancer is more responsive to Ubx on the A1 segment: without Ubx, DG3 reporter gene expression is lost mostly on the ventral side of that segment (Tsai and Alves *et al.*, 2019). Justin had also observed that DG3-deletion mutant exhibits reduced ventral trichome numbers (especially when raised at elevated temperatures (heat stress) (Tsai and Alves *et al.*, 2019) (see Materials and Methods, 6.3, also Figure 2.6). When I refer to samples from this line, it is important to note that solely homozygous embryos or larvae were considered (see Materials and Methods, 6.6 and 6.4 respectively).

To study the features or presence of microenvironments in this mutant line that lacks a regulatory sequence, both wild-type and DG3-deletion embryos were collected, fixed, immunofluorescence stained for Ubx, and *in situ* hybridized for *svb*. At standard temperatures conditions, this enhancer mutant displays a weaker Ubx microenvironment (schematically depicted in Figure 2.5B), with lower Ubx concentration (see Materials and Methods, 6.7) around *svb* transcription sites and lower *svb* transcriptional output (Figure 2.3 A-C). This effect is further exacerbated (Figure 2.3 A-C) when embryos are subjected to elevated temperatures (32°C, see Materials and Methods, 6.3).

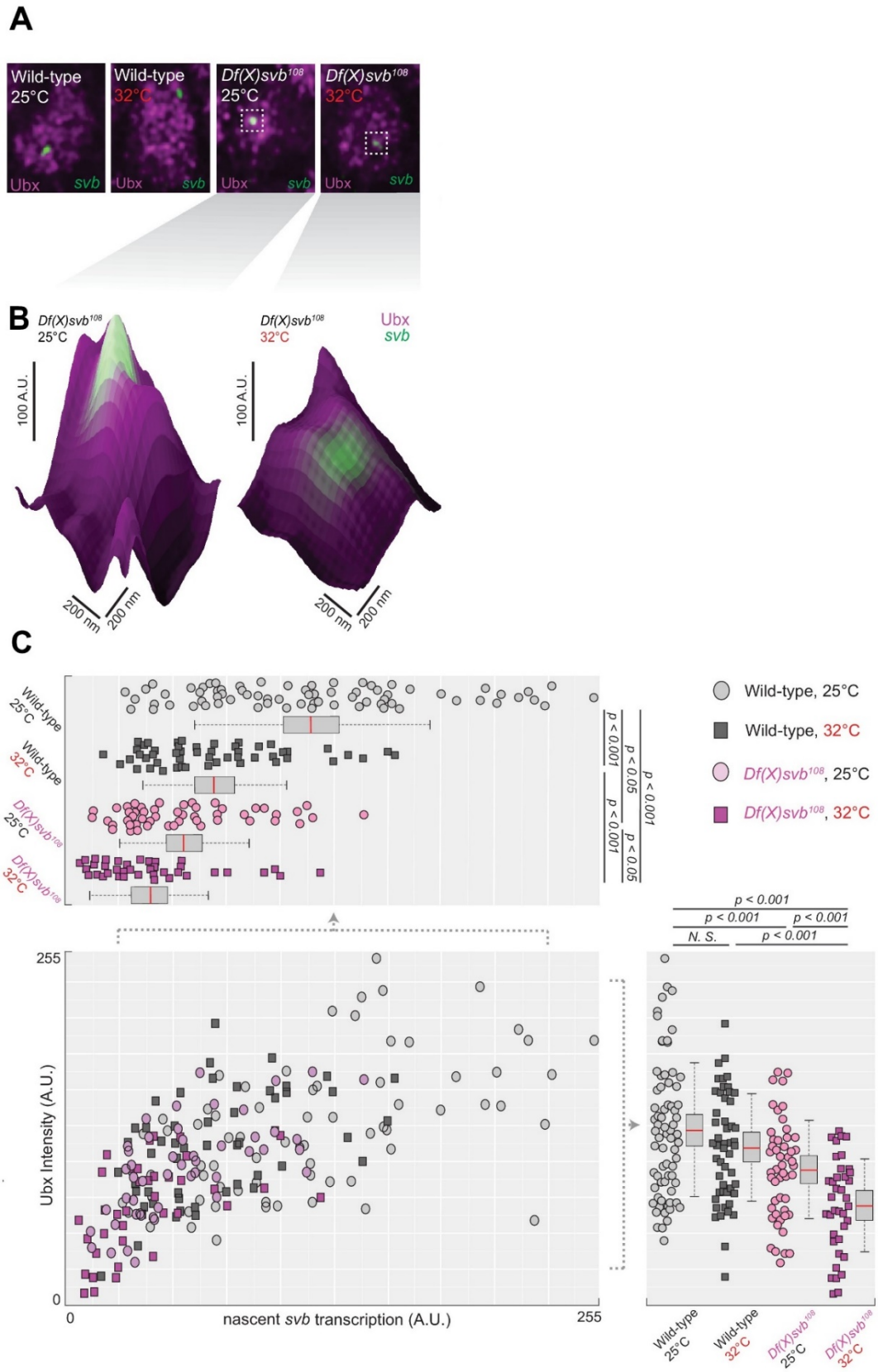


Figure 2.3: **Deletion of *svb* regulatory sequence results in microenvironment impairment, which is exacerbated under heat stress.**

(A) Each panel is a nucleus immunofluorescence stained for Ubx (magenta) and *in situ* hybridized for *svb* (green) and imaged with confocal microscopy. Labels

indicate the temperature (25°C or elevated to 32°C) and the embryo genotype wild-type (w1118) or Df(X)svb^{108} deletion.
(B)Surface plots from panels indicated. Height corresponds to Ubx intensity. Centre is transcription site.
(C)Intensity of Ubx and *svb* transcription at the different conditions and the correlation between them. Of 4 embryos from each genotype/temperature, 71 transcription sites counted for wild-type at 25°C, 50 for Df(X)svb^{108},51 for wild-type at 32°C, 38 for Df(X)svb^{108} at 32°C. Comparisons were made with two-tailed t-tests. Box plots show mean as the center line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers.

These observations suggest that the loss of part of a regulatory sequence can impair microenvironments and that stress conditions amplify this effect. This can indicate that individual enhancers can have a role in microenvironment formation.

Given the hints from Figure 2.1 regarding multi-enhancer interactions, even across chromosomes, it is natural to ask whether this impairment could be rescued by the interaction with the regulatory sequences. And the answer is yes. Strikingly, the effects of this mutated enhancer line on microenvironments were rescued by the insertion of the *cis*-regulatory region of *svb* on another chromosome (svbBAC-*dsRed*), which, as can be seen in Figure 2.4, co-localizes with the *svb* locus (schematically depicted in Figure 2.5D).

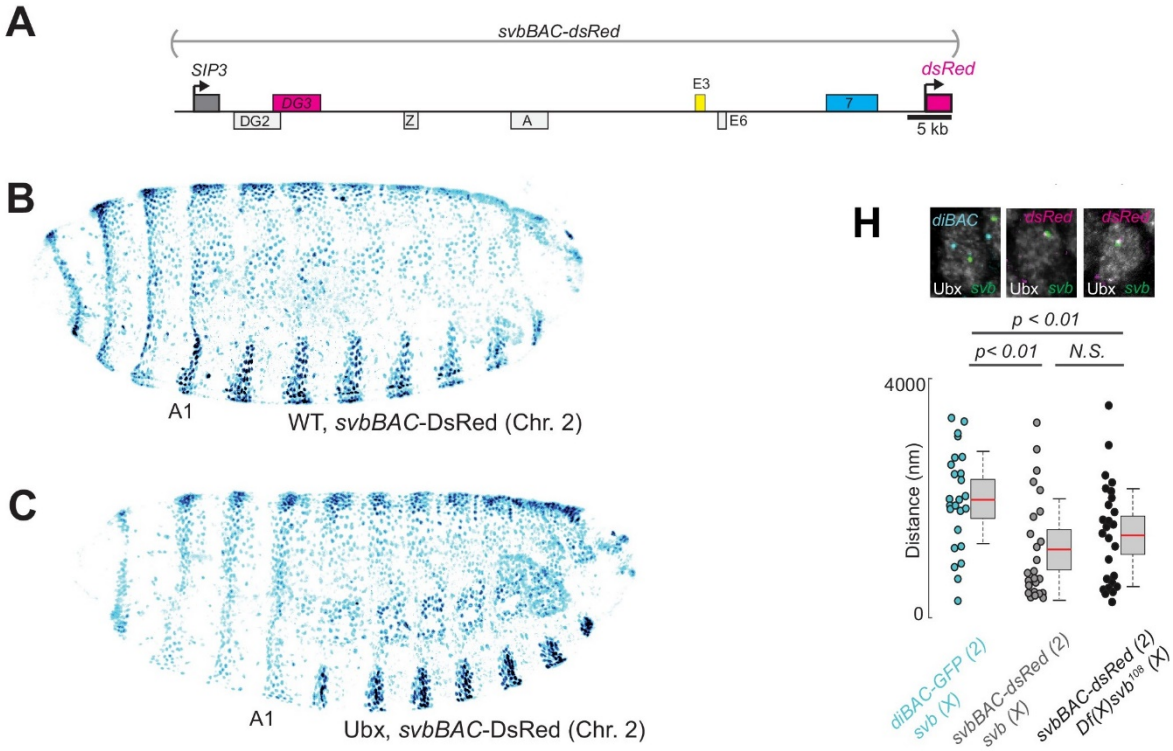
To be able to test the effects of multi-enhancer interaction alone, this rescue had to be done in a way that its effects would not be confounded with an increase of transcriptional *svb* output. Therefore, the *cis*-regulatory region inserted in chromosome 2 is driving the expression of *dsRed*, which acts as a reporter. This insert comprises the entirety of *svb*'s *cis*-regulatory region (svbBAC-*dsRed*, Ella Preger-Ben Noon *et al.*, 2018, Figure 2.4A, Figure 2.5A), and its expression of DsRed protein mimics the embryonic patterns of *svb* expression (Figure 2.4 B) and responsiveness to Ubx (note decrease in expression in segment A1 in Figure 2.4 C).

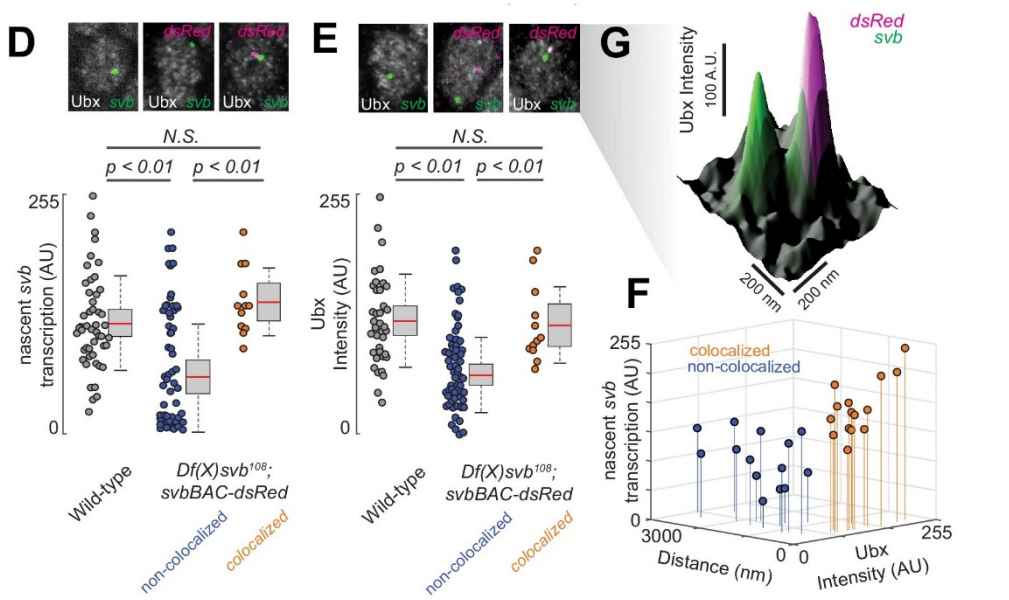
The rescue was tested at 32°C, as it is in these conditions that a bigger effect was seen, as explained in the previous figure. This DNA sequence insertion proved capable of rescuing the defects in Ubx – *svb* microenvironments: when svbBAC-*dsRed* transcripts co-localized with endogenous *svb* transcripts, there was an increase in local Ubx concentration and *svb* transcriptional output, very close to wild-type levels

(Figure 2.4 D-F). Co-localization was considered to be at or below 360nm between transcription sites (see Materials and Methods, 6.8).

The interchromosomal co-localization that was observed in Figure 2.2 also occurred now, even in stress conditions: short distances between endogenous *svb* and *svbBAC-dsRed* transcripts were observed in nuclei that expressed both (Figure 2.4 H). This was seen not only in embryos from the rescue cross between svbBAC-*dsRed* flies and the DG3-deletion mutant flies but also from crosses between svbBAC-*dsRed* and wild-type flies, in both cases at elevated temperatures (Figure 2.4 H, schematically depicted in Figure 2.5C). In embryos from svbBAC-*dsRed* x wild-type crosses, when the *svbBAC-dsRed* sequence is introduced to wild-type embryos, the local Ubx concentration does not suffer any change (Figure 2.4 I). There was also no change in trichome numbers in this latter case (Figure 2.4 J-K).

To control for insertion-related effects being the driver of these co-localizations, the double *in situ* hybridizations were also done (at standard temperatures) for a fly line, which was kindly gifted by Karen Lynn Schulze and Hugo J. Bellen (diBAC-gfp), with a *diachete* regulatory region driving *gfp* expression in the same insertion site as svbBAC-*dsRed*, in chromosome 2. There was no co-localization observed between *gfp* and *svb* transcripts (Figure 2.4 H).


A
svbBAC-dsRed
SIP3
DG3
DG2
Z
A
E3
E6
7
dsRed
5 kb
B
A1
WT, svbBAC-DsRed (Chr. 2)
C
A1
Ubx, svbBAC-DsRed (Chr. 2)
H
diBAC
dsRed
dsRed
Ubx svb
Ubx svb
Ubx svb
p < 0.01
p< 0.01
N.S.
4000
Distance (nm)
0
diBAC-GFP (2)
svb (X)
svbBAC-dsRed (2)
svb (X)
svbBAC-dsRed (2)
Df(X)svb108 (X)


D
dsRed
dsRed
Ubx svb
Ubx svb
Ubx svb
N.S.
p < 0.01
p < 0.01
255
nascent svb transcription (AU)
0
Wild-type
Df(X)svb108; svbBAC-dsRed
non-colocalized
colocalized
E
dsRed
dsRed
Ubx svb
Ubx svb
Ubx svb
N.S.
p < 0.01
p < 0.01
255
Ubx Intensity (AU)
0
Wild-type
Df(X)svb108; svbBAC-dsRed
non-colocalized
colocalized
G
dsRed
svb
Ubx Intensity
100 A.U.
200 nm
200 nm
F
255
colocalized
non-colocalized
nascent svb transcription (AU)
0
3000
Distance (nm)
0
0
Ubx Intensity (AU)
255

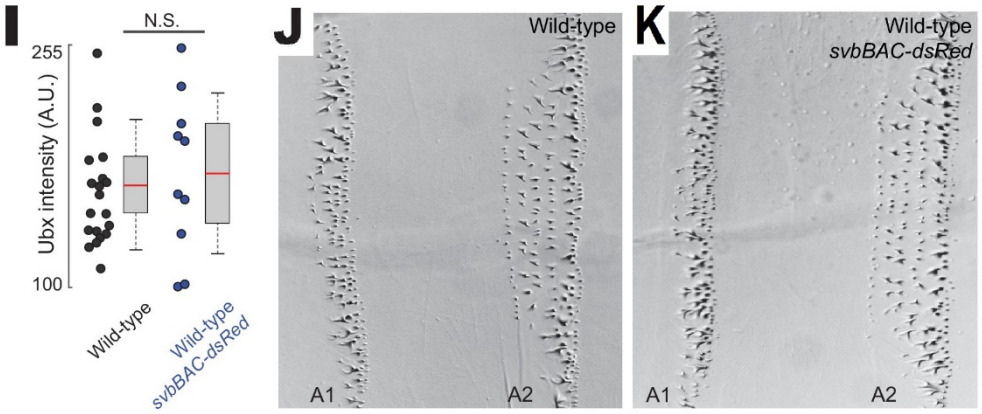

I
N.S.
255
Ubx intensity (A.U.)
100
Wild-type
Wild-type svbBAC-dsRed
J
Wild-type
A1
A2
K
Wild-type
svbBAC-dsRed
A1
A2

Figure 2.4: **DG3-deletion effects in microenvironments are rescued by introducing *svb*'s *cis*-regulatory sequence in a different chromosome.**
(A) Scheme of svbBAC construct: the complete *svb cis*-regulatory region drives expression of *dsRed*.
(B-C) Embryos with insertion of the construct from (A) in Chromosome 2 display an expression pattern to the like of the wild-type *svb*, including a similar response to Ubx.
(D) Intensity of *svb* transcription by *in situ* hybridization represents transcriptional output at the different genotypes at 32°C. Non co-localized correspond to transcription sites not near *dsRed* transcription sites or in nuclei that there was no *dsRed* expression, such as the example of the panel in the center. Transcription site counts: 49 (wild-type), 53 Df(X)svb^{108} non-co-localized, 12 Df(X)svb^{108} co-localized. Comparisons were made with two-tailed t-tests. Box plots show mean as the center line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers.
(E) Intensity of Ubx by staining represents Ubx concentration at the different genotypes at 32°C. Non co-localized correspond to transcription sites not near *dsRed* transcription sites or in nuclei that there was no *dsRed* expression, such as the example of the panel in the center. Transcription site counts: 38 (wild-type), 60 Df(X)svb^{108} non-co-localized, 12 Df(X)svb^{108} co-localized. Comparisons were made with two-tailed t-tests. Box plots show mean as the center line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers.
(F) Intensity of Ubx and *svb* transcription against distances between transcription sites, for nuclei that co-expressed *dsRed* and *svb* at 32°C. The distance of 360nm makes up the threshold between two clusters. Transcription site counts: 14 Df(X)svb^{108} non-co-localized, 15 Df(X)svb^{108} co-localized.
(G) Surface plot from nucleus with two *svb* transcription sites, one of which overlaps with a *dsRed* transcription site. Height corresponds to Ubx intensity.
(H) For nuclei that co-expressed *dsRed* and *svb* at 32°C, distances between them are plotted. Distances between *svb* and *diachete-gfp* sites at 25°C are also plotted. Transcription site counts: 25 *svb-diachete-gfp* from 4 embryos diBAC-gfp x w^{1118} , 25 *svb-svbBAC-dsRed* from 5 embryos svbBAC-*dsRed* x w^{1118} and 26 *Df(X)svb^{108}svb-svbBAC-dsRed* from 4 embryos svbBAC-*dsRed* x Df(X)svb^{108}.
(I) Intensity of Ubx by staining represents Ubx concentration at the different genotypes at 32°C around *svb* transcription sites. Transcription site counts: 20 (3 wild-type embryos), 10 (3 svbBAC-*dsRed* x wild-type embryos). Comparisons were made with two-tailed t-tests. Box plots show mean as the center line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers.
(J-K) Unchanged trichome phenotype in A1 and A2 when svbBAC-*dsRed* is introduced in wild-type (w^{1118}) flies.

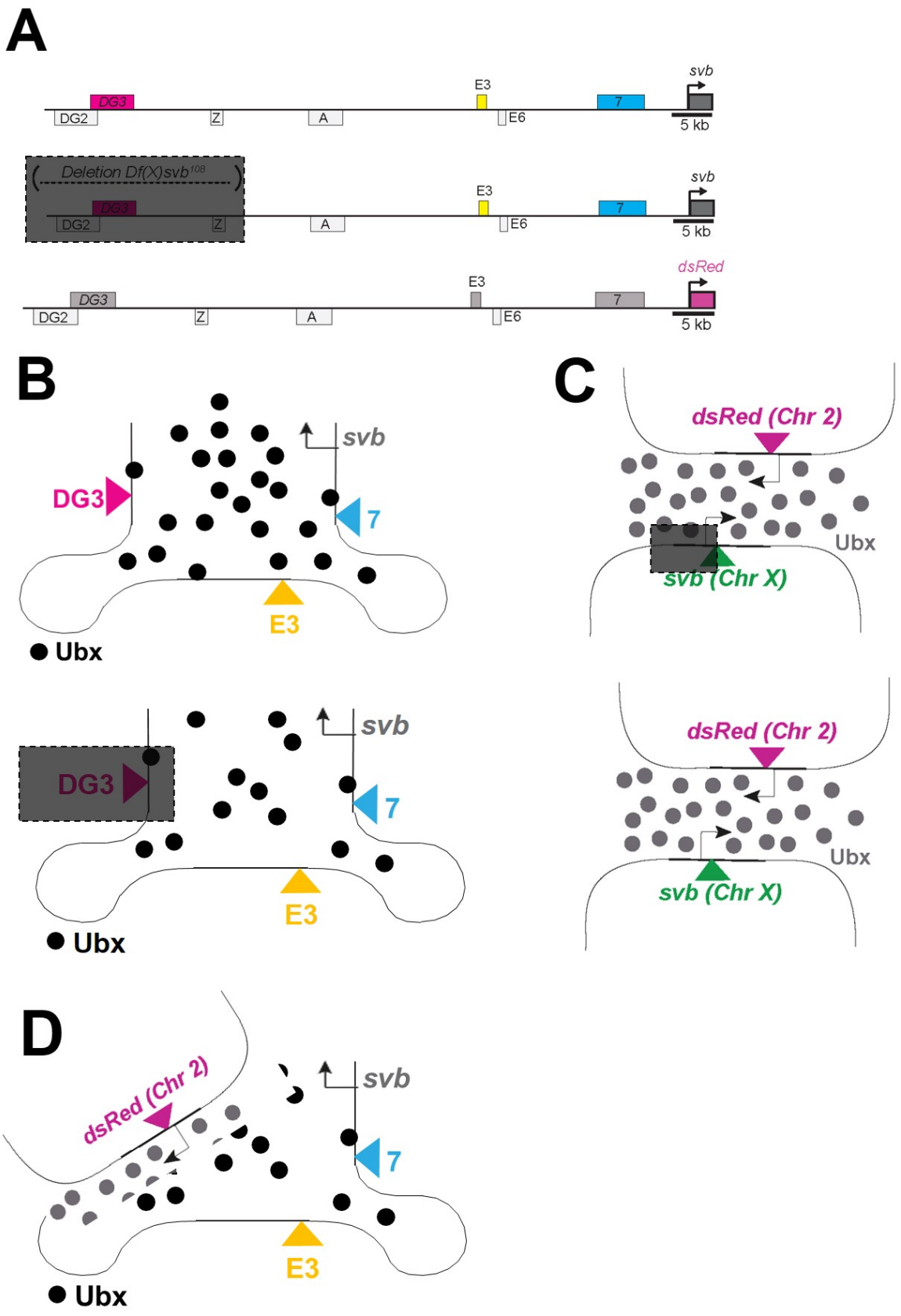


Figure 2.5: **Schematics for multi-enhancer interactions within the Ubx-*svb* microenvironment.**

(A)Upstream *svb cis*-regulatory region schematized. Small boxes correspond to embryonic enhancers. Ventral enhancers are coloured *DG3* (pink), *E3* (yellow), and *7* (blue). DG3-deletion is marked with a grey box in a dashed outline. The last one, the one most in the bottom, indicated the svbBAC construct with the complete *svb cis*-regulatory region driving expression of *dsRed*.
(B) The thin line represents the genome. On the top, high local concentrations of Ubx are depicted in black dots and *svb*'s locus is represented by its three ventral enhancers, *DG3* (left, in pink), *E3* (bottom, in yellow), *7* (right, in blue). On the bottom, lower concentrations of Ubx are depicted in black dots and *svb*'s locus is represented by its three ventral enhancers, *DG3* (left, in pink), *E3* (bottom, in yellow), *7* (right, in blue). DG3-deletion is marked with a grey box in a dashed outline.
(C) Scheme representing potential sharing of Ubx microenvironment by co-localizing transcription sites in different chromosomes. The thin line represents the genome. The *svb* locus is depicted in green at the bottom (with or without the DG3 deletion), the svbBAC-*dsRed* construct is depicted in pink on top, and high local concentrations of Ubx in grey dots.
(D)Schematic representation of svbBAC rescue. The thin line represents the genome. On the bottom, Ubx is depicted in black and grey dots and *svb*'s locus is represented by its two ventral enhancers, *E3* (bottom, in yellow) and *7* (right, in blue). DG3-deletion is covered by the svbBAC-*dsRed* construct, depicted in pink on the top left.
The element representing the linear depiction of *svb*'s full *cis*-regulatory region (first image from panel A) was illustrated by Justin Crocker. For the shape of the genomic line in (B), I took inspiration from a figure Justin once illustrated for a report for a fellowship I applied to.

These observations strengthen the suggestion that related *svb* sequences, including in different chromosomes, can, at least when driving active transcription, interact in Ubx microenvironments (Figure 2.5D). They also suggest that the presence of a *svb* regulatory sequence (only by itself, as in this test it was driving the unrelated reporter *dsRed*) contributes to microenvironment formation and transcriptional output, being able to rescue molecular defects caused by stress conditions. But are these interactions developmentally relevant? Do the observations regarding the rescue of Ubx concentration and *svb* transcriptional output translate into the rescue of the mutant's embryonic phenotype (ventral trichome loss on the A1 segment)?

The answer is: yes, but only partially. At elevated temperatures, when crossing the svbBAC-*dsRed* with the DG3-deletion line, there is a partial trichome number rescue (Figure 2.6 A-D). The introduction of svbBAC-*dsRed* is not able to rescue the loss of trichomes in the area marked by black brackets in Figure 2.6 A-C. This is an area where there is no enhancer redundancy and only *DG3* is expressed (see Figure 1.8H). When testing, at 32°C, the introduction of not the full *svb cis*-regulatory region from svbBAC-*dsRed* but specifically only the DG3 enhancer sequence, was not able to rescue the trichome loss phenotype (Figure 2.6 A-D).

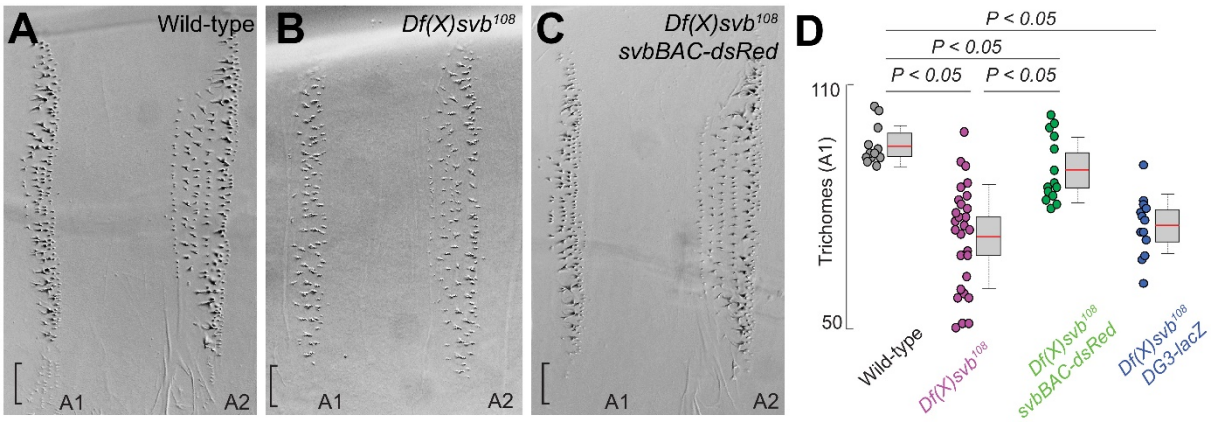


Figure 2.6: **The introduction of *svb*'s *cis*-regulatory sequence in a different chromosome can partially rescue the trichome loss characteristic of the DG3-deletion phenotype.**
(A-C) Trichome phenotype in A1 and A2. Genotypes are indicated. The bracket marks a region where there is no enhancer overlap and only *DG3* is responsible for trichome formation.
(D) Quantification of trichomes in the different genotypes. 12 larvae were counted for wild-type, 28 for Df(X)svb^{108}, 14 for Df(X)svb^{108} x svbBAC-*dsRed* and 13 for Df(X)svb^{108} x DG3-lacZ. Comparisons were made with two-tailed t-tests. Box plots show mean as the center line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers.

2.3 Discussion

These data suggest that multi-enhancer clustering can contribute to developmental robustness against environmental stresses and genetic perturbations.

As detailed in the Introduction, it is important to also bear in mind that many more molecules might be essential microenvironment components or regulators across multiple layers of influence. While here the view is reduced and simplified to focus on players such as Ubx and *svb* ventral enhancers, these components integrate and are influenced by other inputs since they exist in a living and dynamic nuclear, cellular and developmental context.

It has to be stressed that these experiments are based on transcriptionally active loci and markers. Marking and following the DNA sites will be crucial to understand the scale and the timeline of these interactions, and even to confirm that the regulatory sequences are indeed interacting. Here, transcription site location is used as a proxy to their respective

transcriptional microenvironments and the location of their respective regulatory DNA sequences. Marking and following inactive loci will provide more clues to understand several of the missing details surrounding these phenomena. While having that in mind, throughout this thesis (including the next 3 chapters), when I refer to multi-enhancer interactions, these are using the proxy of transcriptionally active loci co-localization.

It should also be noted that there is quite a spread of distances between active transcription sites in the same nuclei, including above 360nm which was the cut-off used here for considering co-localization (Figure 2.4H). The fact that these experiments were made using fixed embryo samples to observe what are dynamic processes is worth having into account. By relying on active markers, what is captured is most likely nuclei at different transcription stages. This adds to these experiments the challenge of finding nuclei where the transcripts of the sequences of interest are co-expressed.

These interchromosomal interactions or “kissing chromosomes” (reviewed in (Maass *et al*., 2019) are in line with other reports from the literature (Gemkow *et al*., 1998; Lim *et al*., 2018). These include reports of “transcription-dependent interchromosomal associations” (Branco and Pombo, 2006) and reports of such interchromosomal interactions being associated with and/or regulating cell fate specification (Monahan *et al*., 2019; Johnston and Desplan, 2014; Maass *et al*., 2018; Ghavi-Helm *et al*., 2014). Co-localized transcripts exhibit higher Ubx concentration (Figure 2.4E-F). To dissect the mechanism of these interchromosomal interactions, it would be important to understand whether DNA sequence features alone drive this clustering or if Ubx has a direct role. DNA sequence features could include TADs and the molecules that structure them, in line with observations that insulators can facilitate long-distance interactions (Postika *et al*., 2018) and that cell-type-specific long-range interactions require chromosome pairing elements between DNA “button” regions that can contain TADs (Viets *et al*., 2019). It is probably a combination of several steps and layers of regulation, nevertheless, it would be interesting to test. The rescue data suggest that trichome recovery comes from microenvironment recovery, but that could be tested more thoroughly too. It would be important to understand whether these multi-enhancer interactions are by-products of other regulatory mechanisms or if their formation constitutes a mechanism on its own in relation to microenvironments. Are these transcriptional sites being driven together to strengthen this microenvironment or do they just happen to be together? Once again, it would be necessary to combine live imaging for time resolution and marking of active and inactive sites for detail resolution. It

is, nevertheless, striking to observe that the introduction of a *cis*-regulatory region of *svb* by itself, without driving the transcription of *svb*, can result in mutant-defect rescues (both molecular and developmental), possibly solely by its interaction in *trans* with the native *svb* locus.

It is interesting to observe that the molecular features related to microenvironments (such as Ubx concentration and transcriptional output) are more responsive than the trichome phenotype to the introduction of the extra *cis*-regulatory sequence in another chromosome. There is, in fact, a specific area that is not responsive to the rescue, and this coincides where (from the redundant ventral enhancers) only *DG3* is expressed (Figures 2.6 A-C, Figure 1.9). Although the svbBAC-*dsRed* sequence contains *DG3*, there seems to still be something missing to be able to drive trichome formation under stress conditions. It could be, for example, another copy of *DG3*. Or it could be that it needs to be localized in its native locus. So, it seems that for the formation of anatomical structures, such as trichomes, in zones where no other enhancers might provide functional redundancy, dosage and/or location of specific regulatory regions might be crucial. In contrast, for rescuing the microscopic microenvironment features, the interaction among multi-enhancer clusters might be sufficient, with individual regions playing a less prominent role, buffered by all the properties emerging from these ensembles. Indeed protection against environmental stresses has been shown to have a contribution from multi-enhancer overlap (Frankel *et al*., 2010; Perry *et al*., 2010) and make sure that the molecules and cells can still execute their function, resulting in maintaining consistent gene expression - with its fine spatiotemporal specificities - for robust embryonic development.

It is also of note that the deletion mutant comprises a region larger than just *DG3*, including other enhancers, and that the lack of this remaining sequence might have effects in the nuclear, cellular, and/or developmental environment that are not yet understood or predictable. When adding exclusively the *DG3* enhancer sequence to the DG3-deletion mutant, the trichome phenotype does not change (Figure 2.6 A-D), suggesting that there is more genomic information needed to see an effect.

In sum, I have shown in this chapter that multi-enhancer interactions contribute to low-affinity *svb* transcriptional microenvironments. It remains to be understood whether transcriptional microenvironments and related multi-enhancer interactions are specific features that only low-affinity enhancer contexts can exhibit.

2.4 Contributions

Figure 2.1
I illustrated the figure, taking aesthetic inspiration and using backbone shapes/elements from a figure that Justin Crocker illustrated.

Figure 2.2
Justin Crocker and Albert Tsai planned this experiment before I arrived at the lab. Justin Crocker oversaw this experiment. Together, Justin and I prepped embryos with immune-fluorescence staining and *in situ* hybridization, which I Airyscan-imaged. I provided input to Justin after he analysed the data and illustrated the original figure (Panels A-C). I illustrated the adaptation to the figure, as detailed in the legend. Justin Crocker acquired funding.

Figures 2.3, 2.4 and 2.6
Justin Crocker proposed these experiments to Albert Tsai and me and oversaw them. Justin Crocker and I maintained different fly lines. When contacted by me, Karen Lynn Schulze and Hugo J. Bellen gifted the diBAC-gfp (CH322-35A16 EGFP) line. Justin Crocker and I did fly crosses. I did the heat-shock experiments and collected embryos and larvae. I designed and prepared the *svb, gfp,* and *lacZ in situ* probes. Justin Crocker and I did stainings, *in situs* and cuticle preps. Albert Tsai did the confocal/Airyscan image acquisition. Justin Crocker and Albert Tsai analysed the data. Justin Crocker illustrated the figures. I contributed to data curation, including providing input to data analysis and figure illustration. Rafael Galupa provided very valuable input, feedback, and discussions to Justin, Albert, and me regarding these experiments. Justin Crocker acquired funding. The Crocker Lab, David Arnosti, Hernan Garcia, and Angela DePace provided insightful comments and discussions.

Figure 2.5
I illustrated the figure, which includes one asset that Justin Crocker illustrated - the linear depiction of *svb*'s full *cis*-regulatory region -, as detailed in the legend.

3 Ubx high-affinity-site-containing genomic regions function as transcriptional enhancers across development and can also exhibit features of multi-enhancer transcriptional microenvironments

As referred to at the end of the previous chapter, it remains to be understood whether transcriptional microenvironments and related multi-enhancer interactions are specific features that only low-affinity enhancer contexts can exhibit. Moreover, as referred to in the Introduction, there is still more to know about how high-affinity Hox transcription factor binding sites can regulate gene expression patterns. In this chapter, I will explore these themes, by assaying Ubx microenvironments in a series of low-to-high Ubx affinity site mutations in transgenic *svb* enhancer reporter lines and by studying a series of endogenous Ubx high-affinity-site-containing genomic regions.

3.1 Ubx local enrichment is overall maintained when low-affinity sites in *E3N* are substituted with high-affinity sites

As mentioned in the Introduction, replacing Ubx low-affinity sites with high-affinity sites resulted in a reduction of specificity (Figure 1.6, Crocker *et al.*, 2015). But what happens to the low-affinity Ubx-*svb* microenvironments? I used the series of fly lines from Figure 1.6, Crocker *et al.*, 2015 to probe the local enrichment of Ubx around transcription sites of the reporter from each transgenic line. This series comprises fly lines that have different mutations of minimal *svb* enhancer *E3N* sites driving a *lacZ* reporter (Figure 3.1 A). These mutations are in different Ubx binding sites, as labelled in Figure 3.1 A, and increase their predicted affinity. These mutations resulted in expression changes - ectopic expression or modified expression levels in places of usual expression (Figure 3.1 B-C).

When looking at the radial Ubx distribution around reporter transcription sites, there is overall maintenance of Ubx enrichment across an increasing relative affinity for Ubx (Figure 3.1 C-D). If Ubx microenvironments were an exclusive feature of its low-affinity binding, then it would be expected to see a disruption of enrichment as the relative affinity is increased. But that does not seem to be the case (Figure 3.1 C-D). Most lines display a similar relative concentration of Ubx at the centre of the transcription sites, close to that of E3N WT, indicated with the dashed line (Figure 3.1 C-D), and slightly lower for lines E3N 3-C and E3N Site 2. Moreover, most lines in this

series display a descending relative Ubx concentration as the distance from transcription site increases, characteristic from transcriptional microenvironments (Tsai *et al*., 2017) indicating that most display relative local Ubx enrichment (Figure 3.1 C-D). Interestingly, for all lines except E3N 3-A, the peak of relative Ubx concentration seems to be proximal to the transcription site centre, with the relative concentration value increasing before entering a descending trajectory (Figure 3.1 C-D). It is also interesting to note that variability changes across different lines (Figure 3.1 C-D). Overall, there does not seem to exist a clear trend correlating with the increase of affinity, but rather a maintenance of the Ubx enrichment profile (Figure 3.1 C-D).

Regarding relative affinity levels, panel B shows the order obtained by SELEX-seq predicted relative affinities of respective sequences from Crocker *et al*., 2015 (Figure 1.6). My colleague Gilberto Alvarez Canales ran an algorithm that predicted the relative affinities for each Hox gene (NRLB, Rastogi *et al*., 2018). When the NRLB predictions are made for these sequences, this prediction suggests a different order of affinities, in which E3N 3-A is the highest affinity line. This could suggest a trend in the variability of Ubx enrichment that I discuss in the discussion section at the end of this chapter. I should note that, while the *n* is relatively lower for the E3N WT control presented, I have repeated that analysis with a higher *n* (118 transcription sites), and the trend was similar (data not shown).

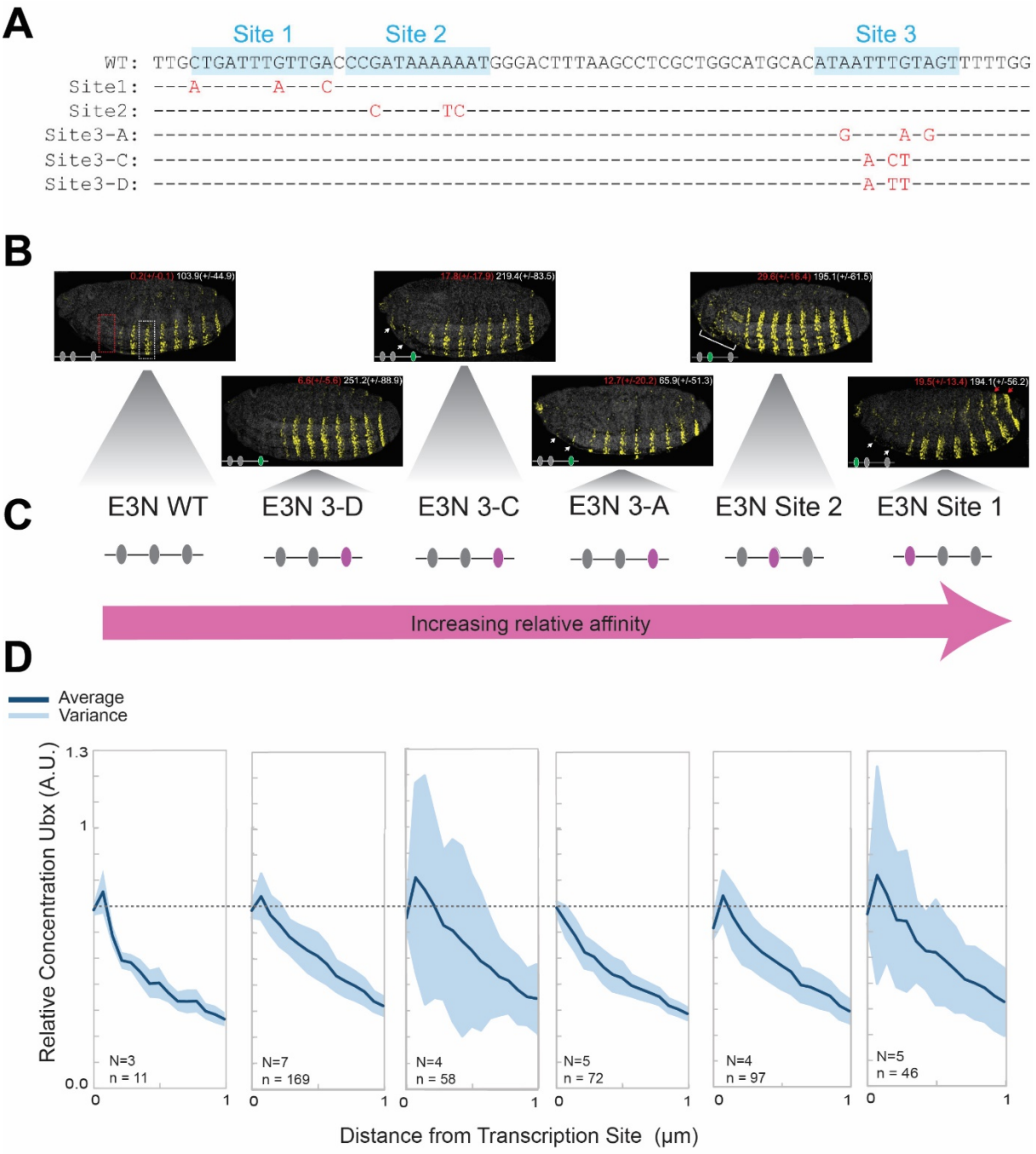


Figure 3.1: **Maintenance of Ubx local enrichment when substituting low-affinity Ubx sites for higher affinity sites.**

(A) Series of sequences aligned: starting from *E3N* wildtype (*E3N* WT) and then the mutated sequences, where red letters correspond to modified sequence (unaltered sequence in dashes).

(B) Embryos stained for βGal protein (sequences schematized in panels A and C are driving *lacZ* reporter). Numbers outside of parentheses in each panel correspond to average levels (in fluorescence intensity arbitrary units) of expression of regions outlined in the first embryo to the left (E3NWT) (n=10 for each line). ±1 SD is represented by numbers inside parentheses. White brackets and arrows mark expression anterior to A1. Red asterisk (intestine) or arrows (dorsal and lateral) mark ectopic staining and expression.

(C) The scheme labels (according to the sequences they carry) the fly lines corresponding with (B) and (D) and depicts Ubx binding sites and highlights where they were modified, in line with (A). These are ordered in increasing SELEX-seq predicted relative affinity for Ubx as indicated by the arrow.
(D) Average relative concentration of Ubx (in Arbitrary Units of fluorescence intensity) vs distance from the transcription site of reporter driven by sequences represented in (A). Variance is represented by shaded region. Dashed line represents starting Ubx level in the transcription site centre for the E3N WT condition. Number of embryos (N) and number of transcription sites (n) is indicated in each plot.
Panel A and B are adapted from Cell, Vol 160, Issues 1-2, Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness, Pages 191-203, 2015, with permission from Elsevier License Number: 5143051160626. (Crocker *et al*., 2015).

I have shown that Ubx local enrichment appears to be overall maintained when low-affinity sites in *E3N* are substituted with high-affinity sites. Would this be a result of the contribution from the sequences in the vicinity, making it robust to these directed changes at only one binding site? What about endogenous Ubx high-affinity binding site sequences?

High-affinity sites are known to have low specificity *in vitro* (Mann *et al.,* 2009), but can regulate precise patterns of gene expression during development, contributing to the activation of certain lineage gene expression programmes (Boisclair Lachance *et al*., 2018), as mentioned in the Introduction.

To understand if Ubx high-affinity binding site sequences can display transcriptional microenvironment features, first I have conducted the screening of a collection of genomic regions bound by Ubx and containing Ubx high-affinity (HA) binding sites (the pipeline of which is schematized in Figure 3.2). This collection, designed and created by Justin Crocker, includes 52 *Drosophila* reporter lines, corresponding to 26 pairs of Ubx high-affinity-site-containing genomic region sequences and their respective mutated sequence, with a mutation in the high-affinity motif (Figure 3.2, Materials & Methods, Supplementary Table 1). These 26 sequences coincide with Ubx binding peaks in Chromosome 3 from Choo and colleagues (Choo *et al*., 2011), and vary in size (Supplementary Table 1).

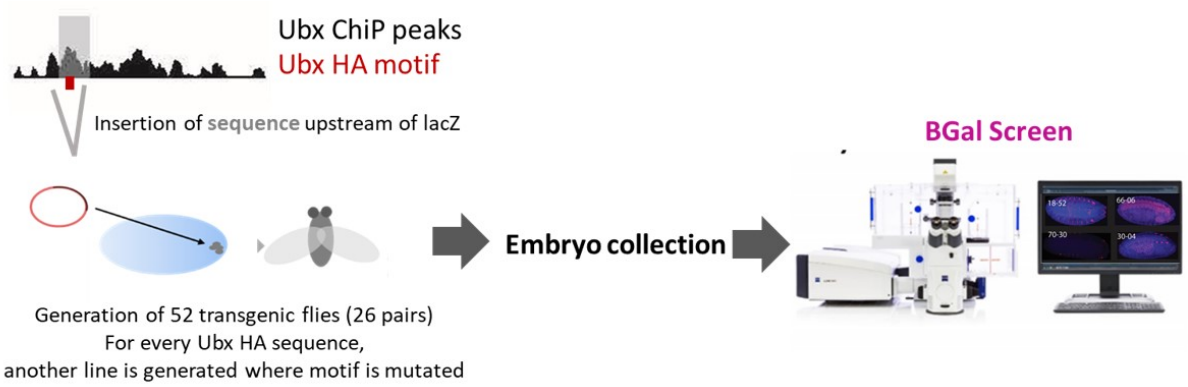


Figure 3.2: **Pipeline of Ubx High-Affinity screen.**
The scheme represents the steps towards the Ubx High-Affinity Screen: starting from identification of 52 sequences to its insertion upstream of *lacZ* reporter, generation of transgenic flies, embryo collection, and finally, confocal imaging of embryos stained for ßGal.

3.2 A characterization of a Ubx high-affinity-site-containing genomic region-screen

Ubx high-affinity-site-containing genomic regions drive embryonic reporter expression broadly both spatially (Figure 3.3 A) and across several stages (Figure 3.3 B). This suggests that these genomic sequences can function as transcriptional enhancers that can be broadly used across development.

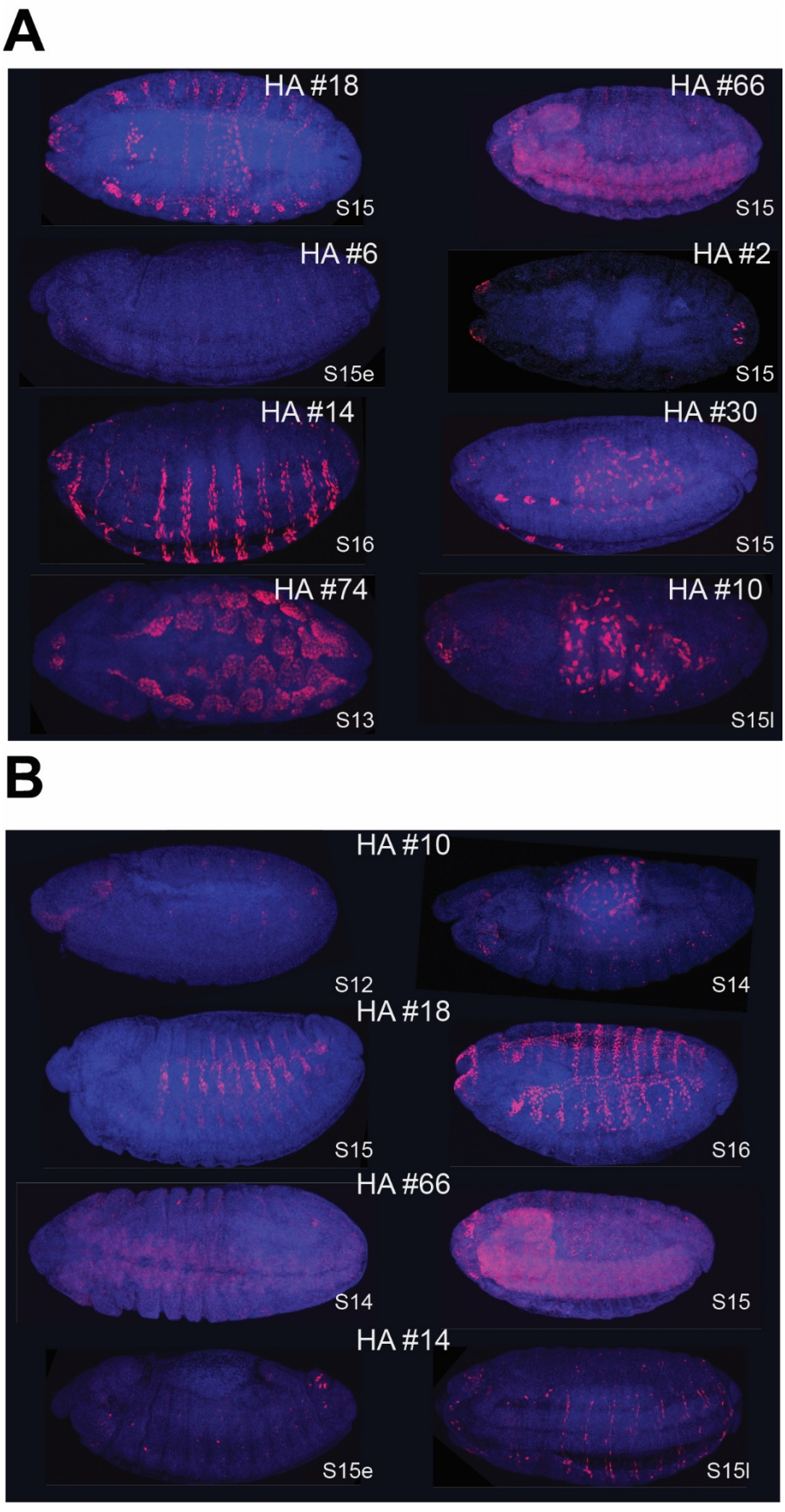


Figure 3.3: **Ubx high-affinity regulatory sequences drive gene expression broadly in the embryo.**

(A) Ubx regulatory sequences drive gene expression across several tissues. Representative embryo stainings, max projections of 20X Zstacks acquired with Zeiss LSM 880, with nuclear staining in blue and ßGal staining (product from enhancer driving *lacZ*) in magenta. Embryo genotypes are WT Ubx High-Affinity lines as indicated. Corresponding stages are indicated ('e' stands for early, 'l' stands for late). For corresponding sequences of genotypes, see Supplementary Table 1.
(B) Ubx regulatory sequences drive gene expression across several stages. Representative embryo stainings, where each line shows two embryos from the same genotype but in different stages of embryonic development. These are max projections of 20X Zstacks acquired with Zeiss LSM 880, with nuclear staining in blue and ßGal staining (product from enhancer driving *lacZ*) in magenta. Embryo genotypes are WT Ubx High-Affinity lines as indicated. Corresponding stages are indicated ('e' stands for early, 'l' stands for late). For corresponding sequences of genotypes, see Supplementary Table 1.

It is important to note that the exact endogenous targets of these sequences are not known. Some have genes in the vicinity and some are in the vicinity of genes with unknown functions (for the list of sequences of this library, see Supplementary Table 1). Furthermore, they could be regulating targets that are not in the vicinity of their sequence. However, the focus of this work is to observe their functionality in the context of reporter gene expression to help understand how sequences that include high-affinity Ubx motifs can be used.

When predicting binding affinity from different Hox Transcription Factors alone and in complex with co-factor Exd, the mutation of the high-affinity sites seems to affect overall mostly the predicted affinity of Hox-Exd complexes (Figure 3.4). This can be seen when looking at the distribution of affinities in all the wild-type sequences vs in all the mutated sequences (Figure 3.4 A). For Ubx and other Hox Transcription Factors, distributions overlap when looking at their predicted affinities (Figure 3.4 A). In contrast, when predicting the affinities of Ubx and other Hox proteins in complex with Exd there is a shift observed, with mutated sequences displaying lower affinities (Figure 3.4 A). This can also be seen when predicting the difference (or delta) of affinities of wild-type minus mutant, for each library pair, for various Hox Transcription Factors, in or not in complex (Figure 3.4 B). In this case, one can see higher difference between the predicted affinities of Hox-Exd complexes than Hox proteins alone (Figure 3.4 B).

One can also see in the distribution profiles for UbxIVa that there is a considerable "bump" in the distribution on the bottom right corner, which is higher affinity in the axis. This prediction further confirms that this collection features a distribution of sequences that contain high-affinity Ubx sites which are being specifically mutated (Figure 3.4 A). Interestingly, this is

also seen for Dfd (Figure 3.4 A). More about this is discussed in the discussion section at the end of this chapter. Overall, there is a varied distribution of predicted absolute affinities for binding of Hox Transcription Factors, in or not in complex with Exd, to the different wild-type sequences (Figure 3.4 C).

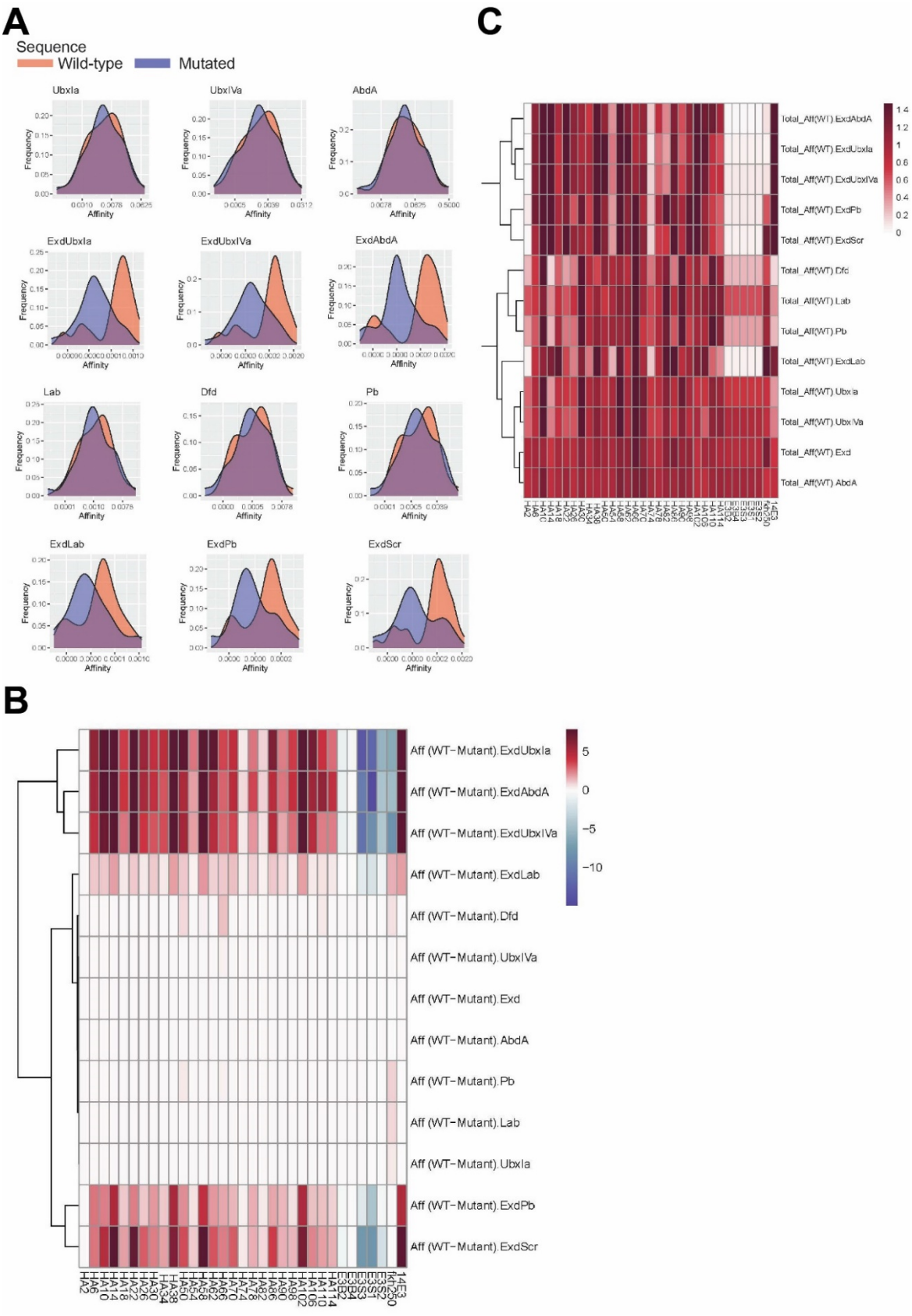


Figure 3.4: **Mutation of Ubx high-affinity binding sites affects predominantly predicted Hox Transcription Factor binding affinity in complex with Exd.**

(A) Density distributions of affinities of putative binding sites for each factor in all the sequences. Y-axis is increasing frequency and X-axis is increasing affinity. Light orange is the wild-type distribution profile and blue is the mutated distribution profile, as stated in the in-figure legend.
(B) Relative change of total affinities due to the mutations, for several of the Hox factors and each of the sequence pairs. To make visually comparable estimations, the values were normalized with the median affinity values from all the sequences for each factor and also normalized by sequence length.
(C) Total affinities for the WT sequences for several of the Hox factors. To make visually comparable estimations, the values were normalized with the median affinity values from all the sequences for each factor and also normalized by sequence length.
The sequences of the Ubx High-Affinity library can be found in Supplementary Table 1. For the B and C, you can also see included 7 other sequences:
1) The E3N series from Figure 3.1, Figure 1.6, Crocker *et al.*, 2015: where E3B4 corresponds to E3N 3-D, E3B2 corresponds to E3N 3-C, E3S3 corresponds to E3N 3-A, E3S2 corresponds to E3N Site 2 and E3S1 corresponds to E3N Site 1. Their wild-type is E3N wild-type.
2) fkh250 (wild-type) is a previously described (Ryoo and Mann, 1999) enhancer element that contains a low-affinity Exd-Hox binding site (AGATTAATCG) preferred by Exd-Scr and fkh250con (mutant), 2 bp mutation which creates an Exd-Hox consensus site that is also bound by Exd-Antp, Exd-Ubx1a, and Exd-AbdA heterodimers (Rastogi *et al*., 2018). These were included as they may be interesting for future experiments, as discussed in the discussion section.
3) 14E3 is a sequence I designed that is a mutant version of HA #14, where the high-affinity motif is substituted for one of the *E3N* low-affinity motifs. These were included as they may be interesting for future experiments.
All panels were generated by Gilberto Alvarez Canales.

Upon mutation of the sequence, I observed different effects on reporter expression (Figure 3.5A). These are overall effects taking into account the several stages of embryonic development -as mentioned before, the Ubx High-Affinity regulatory sequences drive gene expression broadly in the embryo. For a small percentage, 4 in 26, there was an increase in reporter expression upon mutation, either by increase of expression intensity or by new ectopic expression patterns (Figure 3.5 A, B). For half of the library, a decrease in expression upon mutation was observed (Figure 3.5 A, C), either by loss of expression or decrease in intensity. For almost as many lines, 10 in 26, there was no appreciable change between wild-type and mutated Ubx high-affinity sequence (Figure 3.5 A, D).

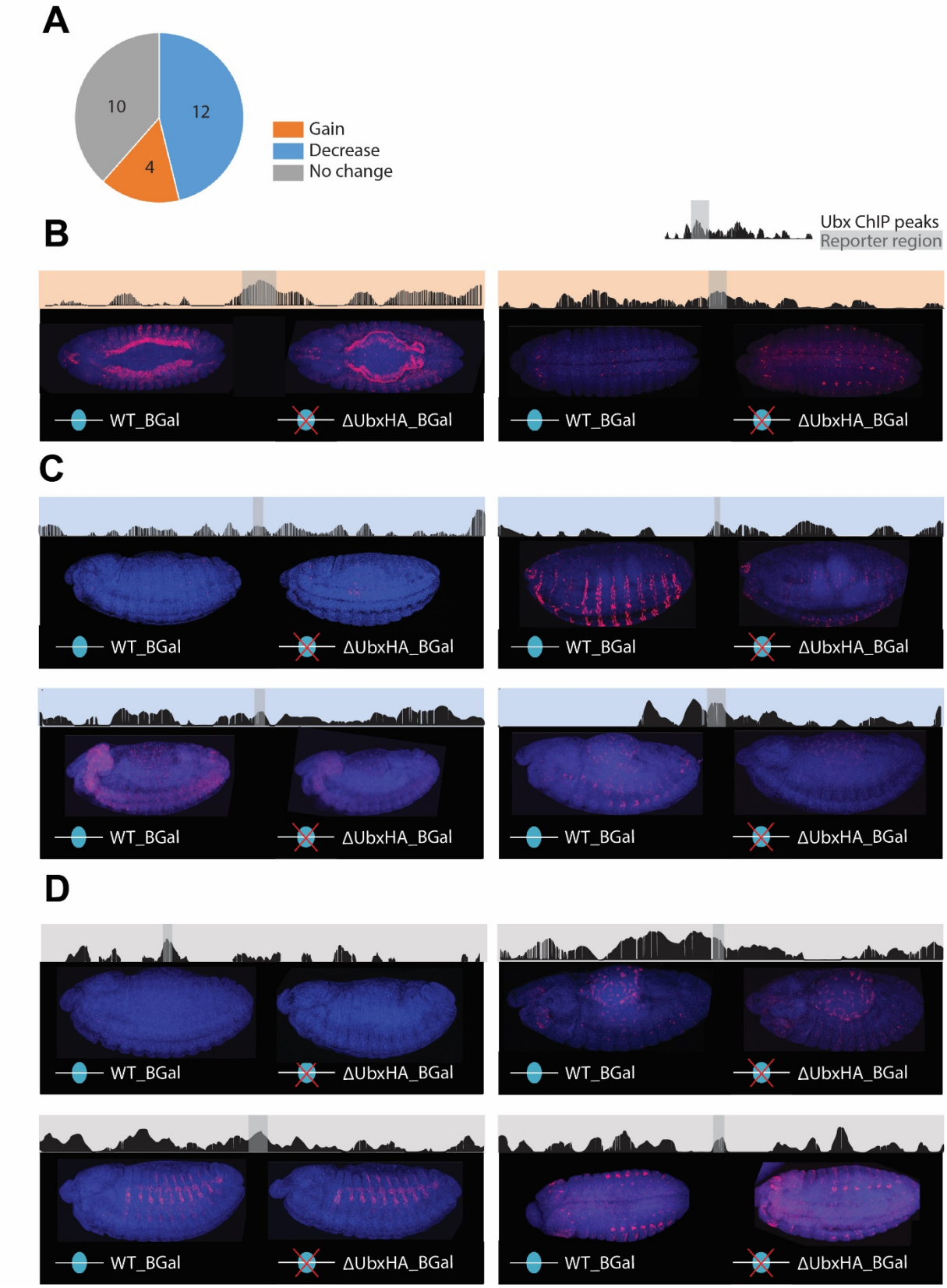


Figure 3.5: **Mutations of HA sites have different effects on reporter expression.**
(A) Pie Chart summarizing overall effects on reporter expression upon HA site mutation across several stages. Gain includes new ectopic expression or increase

in intensity. Conversely, decrease includes loss of expression or decrease in intensity.
(B-D) Embryo stainings showing the comparison of reporter expression in respective Gain of expression (B), Decrease of expression (C), and No change (D) examples. Embryo pictures are max projections of 20X Zstacks acquired with Zeiss LSM 880, with nuclear staining in blue and ßGal staining (product from enhancer driving *lacZ*) in magenta. Above each pair of embryo stainings, one can see a representation of the Ubx peak from (Choo *et al*., 2011), with the reporter enhancer sequence highlighted, related to the respective wild-type of each pair. Embryo genotypes are pairs of Ubx High Affinity library: 74-76 (B'left), 110-112 (B'right), 6-8 (C'top left), 14-16 (C'top right), 66-68 (C'bottom left),34-36 (C'bottom right), 2-4 (D'top left), 10-12 (D'top right), 18-20 (D'bottom left), and 70-72 (D'bottom right). For corresponding sequences see Supplementary Table 1.

The effect on reporter expression upon mutation of the Ubx high-affinity motif does not seem to correlate with the degree of conservation of the high-affinity site (Figure 3.6 A). When looking manually at the sequences from 124 insect species, there is not a statistically significant difference (see Figure 3.6 A legend) in the number of motif-sequence matches between "no change"-effect and "decrease"-effect lines. The number of lines that display an "increase"-effect is too little to make such comparisons.

In regards to the conservation of the wild-type sequences (Ubx high-affinity-site-containing genomic regions), when looking at the PhyloP scores from 27 species in comparison to the *Drosophila melanogaster* genome, the full genomic regions have lower conservation than the Ubx high-affinity-sites and lower conservation than of random sampling of sequences in the 3R and 3L chromosomes (Figure 3.6 B). In contrast, the full genomic regions of this library have significantly higher conservation than 100 *cis*-regulatory modules downloaded from the RedFly DataBase. When compared with the conservation of the full *svb* sequence, the full genomic regions of this library were not significantly different. More details are found in the Figure 3.6B, its legend, and Materials and Methods. As stated in the contribution section, my colleague Gilberto Alvarez Canales did this analysis.

Furthermore, the effect on reporter expression upon mutation of the Ubx high-affinity motif also does not seem to correlate with the size of the full genomic sequences used on the screen (Figure 3.6 C).

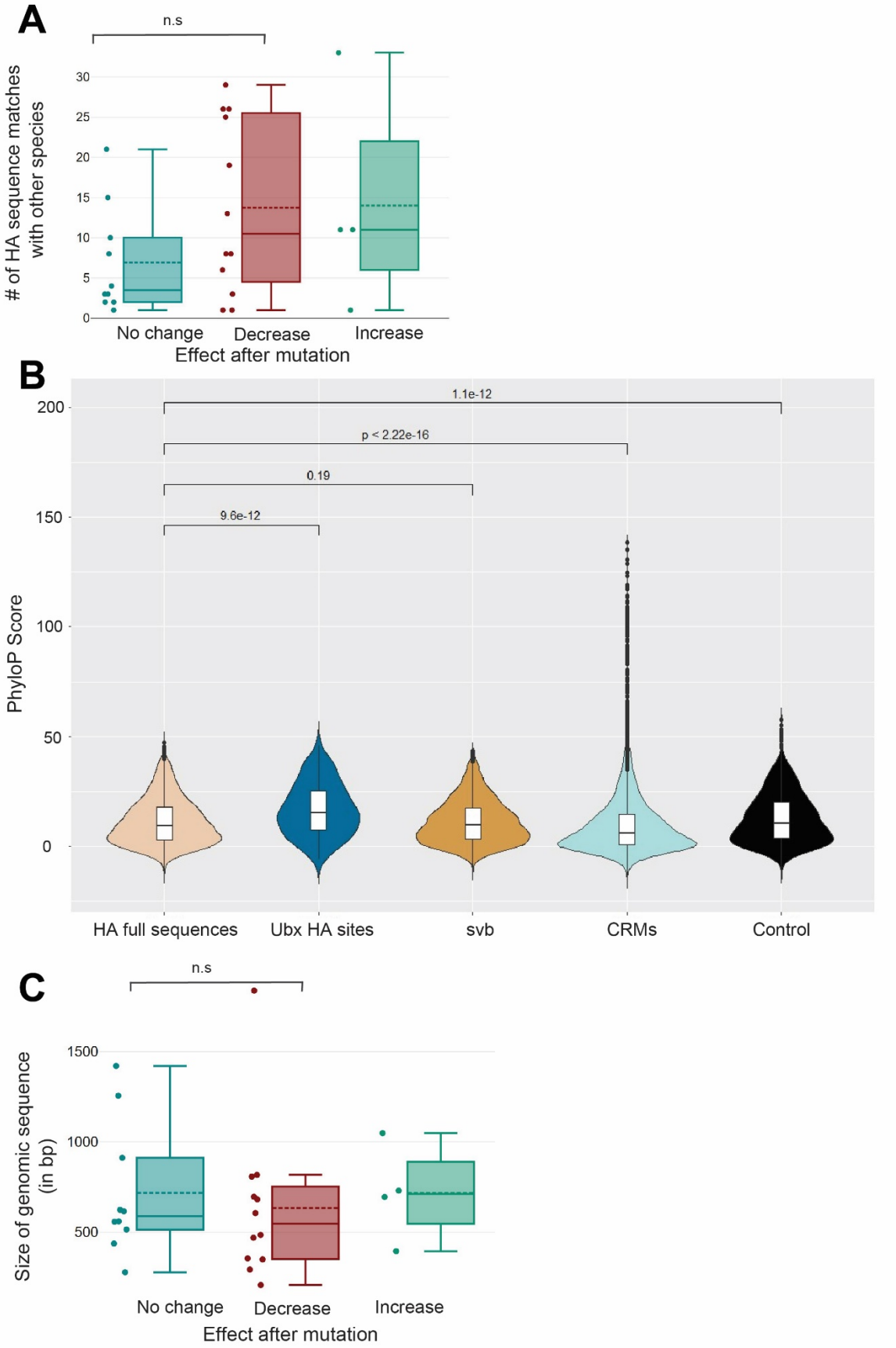


Figure 3.6: **Effects on reporter expression upon mutation do not seem to correlate with conservation or size of wild-type high-affinity-site-containing genomic sequences.**

(A) Number of high-affinity site sequence matches with other species for each condition/group from the effect on reporter expression upon mutation (Figure 3.5). I counted manually the number of high-affinity site sequence matches using the 124 insect species track to the D. melanogaster genome (dm6). The sequences of the Ubx high-affinity library can be found in Supplementary Table 1. Data were plotted using DATAtab. Box plots show mean as the center dashed line, the median is center non-dashed line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers. Each individual point is also shown. Comparisons were made between the "no-change" (n=10) and "decrease" (n=12) groups, and both two-tailed t-test for independent samples and Mann-Whitney U-Test (DATAtab) showed that while the "no change" group had lower values than the "decrease" group, this difference was not statistically significant.
(B) PhyloP scores from the 27 species comparison to the *D. melanogaster* genome (dm6). Conditions are: HA full sequences - The sequences of the Ubx High-Affinity library can be found in Supplementary Table 1; Ubx HA sites - Ubx motifs were predicted inside each of the full genomic regions from the Ubx High-Affinity library; svb- full *svb* locus; CRMs - a sample of *Cis*-Regulatory Modules in the third chromosome downloaded from RedFly DB; Control - Random sampling of sequences in the 3R and 3L chromosomes.
Violin plots show median as the center line, the lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and the whiskers are 1.5*IQR (interquartile range), and frequency as density width.
Wilcoxon tests were done given that the distribution of the data is non-normal (Normality Shapiro tests were done).
This panel was generated and illustrated by Gilberto Alvarez Canales.
(C) Size of genomic sequence (in base pairs) for each condition/group from the effect on reporter expression upon mutation (Figure 3.5). The sequences of the Ubx High-Affinity library can be found in Supplementary Table 1. Data were plotted using DATAtab. Box plots show mean as the center dashed line, the median is center non-dashed line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers. Each individual point is also shown. Comparisons were made between the "no-change" (n=10) and "decrease" (n=12) groups, and both two-tailed t-test for independent samples and Mann-Whitney U-Test (DATAtab) showed that while the "no change" group had higher values than the "decrease" group, this difference was not statistically significant.

3.3 Ubx high-affinity-site-containing genomic regions can also exhibit features of multi-enhancer transcriptional microenvironments

Coming back to the beginning of this chapter, I noted that it remains to be understood whether only low-affinity enhancer contexts can exhibit features of transcriptional microenvironments. In Figure 3.1, I showed that Ubx local enrichment profiles are maintained when low-affinity *svb* sites are substituted with high-affinity sites. To probe this using endogenously originated Ubx high-affinity sites, I chose the pair from the library that exhibits the highest predicted Ubx affinity difference between wild-type and mutated pair. Once the high-affinity motif of HA line #58 is mutated, the predicted Ubx affinity drops dramatically (Figure 3.7 A-D). In fact, the

mutated line HA #60 displays similar levels of predicted Ubx binding affinity than *svb* low-affinity minimal enhancer *E3N* (Figure 3.7 C-D). The profile in Figure 3.7 D shows still putative binding sites for Exd-UbxIV with similar level of affinity values of the wild-type *E3N* enhancer affinities profile shown in Figure 3.7 C. The HA lines #58 and #60 drive reporter expression in different embryonic segments, as seen in Figure 3.7 E.

When looking at the radial Ubx distribution around reporter transcription sites, both wild-type and mutated high-affinity of the pair exhibit a descending relative Ubx concentration as the distance from transcription site increases, characteristic from transcriptional microenvironments (Tsai *et al*., 2017) suggesting relative local Ubx enrichment (Figure 3.7 F). There is not a clear disruption of enrichment when comparing the profiles of the Ubx high-affinity-containing enhancer *#58* and the low-affinity *svb* minimal enhancer *E3N* or the mutated line #60, which has a very low predicted affinity for Ubx as well (Figure 3.7 F). When looking at the relative concentration of Ubx at the centre of the transcription sites, the dashed line indicates that of *E3N*, and it can be seen that the high-affinity line #58 displays a slightly higher starting point, while line #60 displays a very similar level to *E3N* (Figure 3.7 F). The peak of relative Ubx concentration seems to be not exactly at the transcription site centre for both *E3N* and *#60*, with the relative concentration value increasing before entering a descending trajectory, which is not the case for *#58* (Figure 3.7 F). Overall, from the observation of the profile for high-affinity line #58, it is possible to infer that high-affinity contexts can display Ubx local enrichment, similarly to low-affinity contexts (Figure 3.7 F).

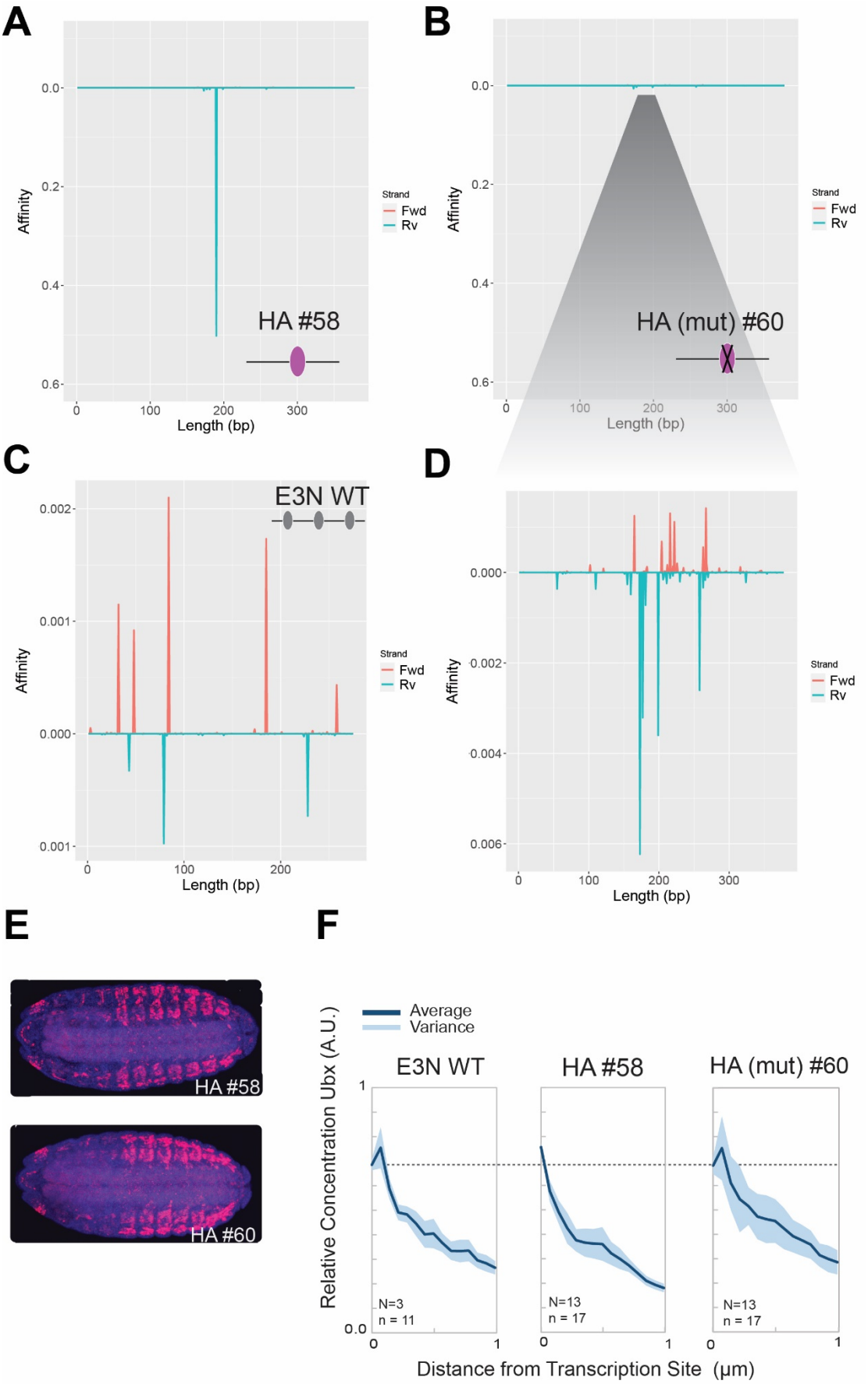


Figure 3.7: **Ubx high-affinity-site-containing genomic region can also exhibit microenvironment-like Ubx local enrichment.**

(A-D) Visual inspection of affinity values for Exd-UbxIV plotted along the sequence of *HA#58* in (A) and respective mutant *HA#60* in (B). (D) is zoom-in of (B) to show comparable scales with *E3N* (C). Reverse (blue) and forward (orange) DNA sequences' peaks are shown on different sides of the axis for better visualization.
(E) Embryo stainings showing reporter expression in lines HA#58 and HA#60. For corresponding sequences, see Supplementary Table 1. Embryo pictures are max projections of 20X Zstacks acquired with Zeiss LSM 880, with nuclear staining in blue and ßGal staining (product from enhancer driving *lacZ*) in magenta.
(F) Average relative concentration of Ubx (in Arbitrary Units of fluorescence intensity) vs distance from the transcription site of reporter driven by sequences *HA#58* and *HA#60*, with *E3N* as a control. Variance is represented by the shaded region. Dashed line represents starting Ubx level in the transcription site centre for the E3N WT condition. Number of embryos (N) and number of transcription sites (n) is indicated in each plot.
Plots in panels A-D were generated by Gilberto Alvarez Canales.

Another feature of Ubx low-affinity microenvironments, as shown in Chapter 2, is the possibility of the occurrence of multi-enhancer interactions. Using another line from the Ubx High-Affinity screen, I explored if transcription sites from a Ubx high-affinity-site-containing genomic region can also co-localize with *svb* on different chromosomes (Figure 3.8). The focus on this line, HA #14, comes following Justin noticing that the reporter expression pattern (Figure 3.8 A) resembled that of *svb*'s ventral enhancers, with some extra stripes. Furthermore, when the high-affinity site is mutated to decrease the affinity 10^3 fold based on NRLB, line HA #16, there is loss of reporter expression (Figure 3.8 A), suggesting that this can be a functional binding site. As all of the sequences from the Ubx High-Affinity library, the HA#14 sequence is located in Chromosome 3L. I observed that transcription sites from a reporter driven by this Ubx high-affinity-site-containing genomic region can co-localize with *svb* transcription sites (Figure 3.8 B-D).

There is a spread of distances, from 0.1 µm (close to the limits of the optical resolution of Airyscan-images, which is around 140 nm) to 2.68 µm (median is 0.39 µm and mean is 0.631 µm) (Figure 3.8 B-D).

It was important to test co-localization between *svb* or *HA#14-lacZ* transcription sites with transcription sites driven by regions in the corresponding chromosomes of the other, to see whether this was something specific of *svb* and the transcription of *lacZ* driven by the Ubx high-affinity-site-containing genomic region HA#14. I did double *in situ* hybridizations between *svb* transcription sites and *alphatubulin-67C* transcription sites (Figure 3.8 E), since *alphatubulin-67C* is on the same chromosome as *HA#14-lacZ* (3L). It is clear by looking at the points that these distances are in general higher than between *svb* and *HA#14-lacZ*

(median is 1.445 µm and mean is 0.611 µm). Comparisons made between these two using both two-tailed t-test for independent samples and Mann-Whitney U-Test showed that this difference was statistically very significant (p=<.001).

Similarly, I did double *in situ* hybridizations between *actin5c* transcription sites and *lacZ* transcription sites driven by HA #14, since *actin5c* is on the same chromosome as *svb* (X). However, out of 12 embryos, there were only four pairs of transcription sites co-expressed in the same nuclei. The distances were 0.429 µm, 0.599 µm, 0.913 µm, and 4.015 µm.

In line with reporter expression being lost when the #14 high-affinity site is mutated (Figure 3.8 A), *lacZ* transcription site signal by *in situ* hybridization for line HA #16 is too weak, not allowing me to find nuclei where it is co-expressed with *svb,* therefore not being possible to measure distances.

To see whether other Ubx high-affinity-site-containing genomic regions could also feature multi-enhancer interactions, I performed double *in situ* hybridizations between *svb* transcription sites and *lacZ* transcription sites driven by HA #58. However, out of a similar number of embryos (10), there were only three pairs of transcription sites co-expressed in the same nuclei. The distances were 1.725 µm, 2.976 µm, and 3.442 µm.

Following the observation of the Ubx microenvironment-typical radial distribution profile that Ubx high-affinity-site-containing genomic region HA #58 exhibited, and the observation of transcription site co-localization between 14 and *svb*, I tested the Ubx radial distribution profile for HA #14. Preliminary data not shown has too much variance and technical aspects have to be fine-tuned/experiments need to be repeated to make sensible conclusions.

Since transcription sites of *svb* co-localize with the ones from *lacZ* driven by Ubx high-affinity-site-containing genomic region 14, and given the observations in Chapter #2 regarding the possible relation between multi-enhancer interaction and phenotypic consequences (trichome number in the case of *svb)*, I also tested whether lines containing Ubx high-affinity-site-containing genomic region 14 could have a different trichome phenotype. They could, for example, exhibit more trichomes, if they were influencing the action of Svb. However, when comparing trichome numbers from the first row of the A1 segment, the number is similar between wild-type,

HA#14, and HA#16 lines (Figure 3.8 F), suggesting it cannot influence endogenous gene expression to the point of a phenotypic change.

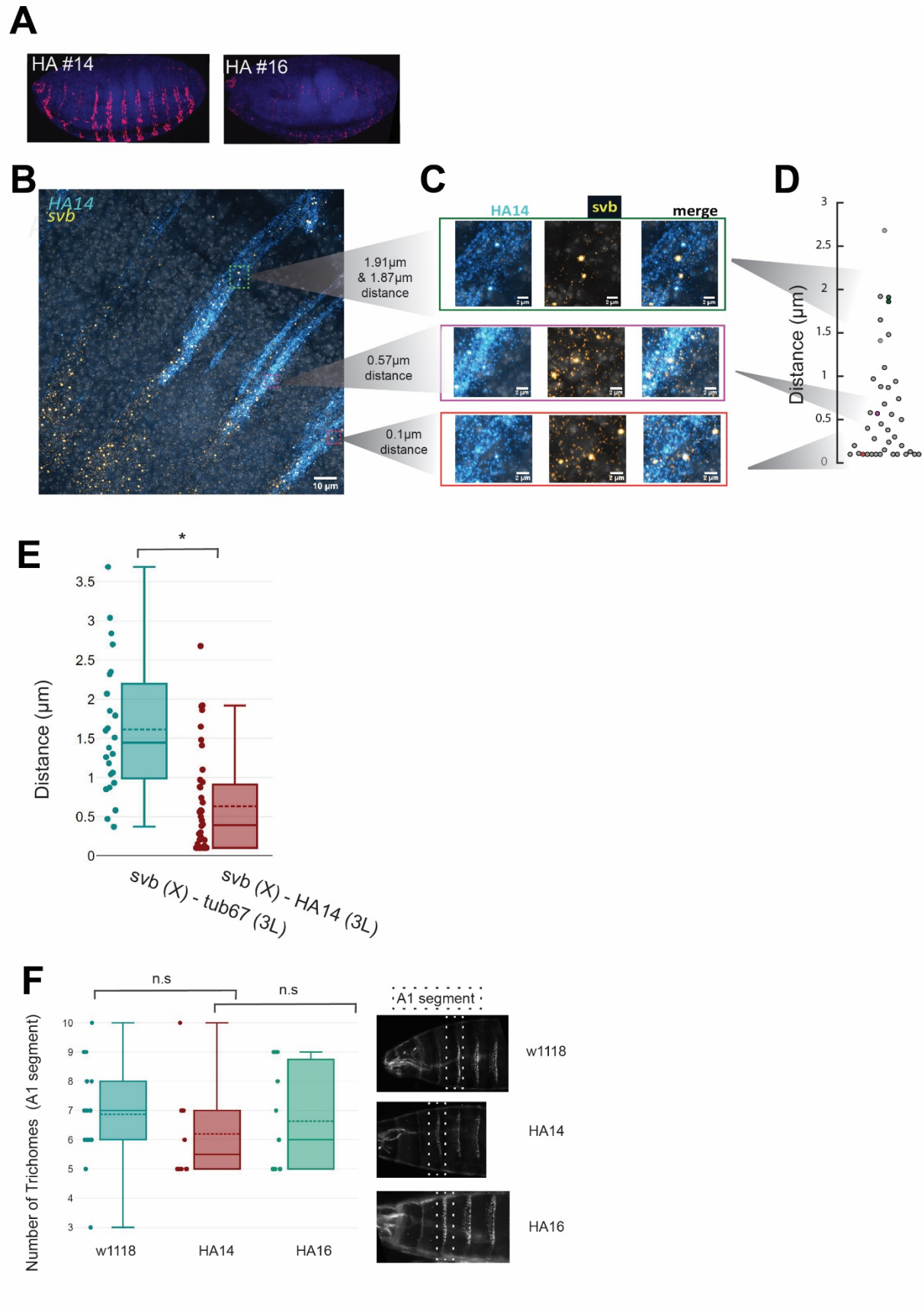


Figure 3.8: **Transcription sites from a high-affinity-site-containing genomic region and *svb* (in different chromosomes) can co-localize.**

(A) Embryo stainings showing reporter expression in lines HA#14 and HA#16. For corresponding sequences see Supplementary Table 1. Embryo pictures are max projections of 20X Zstacks acquired with Zeiss LSM 880, with nuclear staining in blue and ßGal staining (product from enhancer driving *lacZ*) in magenta.
(B) Embryo picture of HA #14 line with double *in situ* hybridization for *svb* and *HA#14-lacZ*. This is a max projection of 63X acquired with Zeiss LSM 880, with nuclear staining in white, *svb*-DIG in yellow and *lacZ*-BIO staining in cyan. Scale bar is in the picture. Boxes are highlighting and connecting to different nuclei that are zoomed in in C.
(C) Boxes in B mark nuclei that are here represented at a higher magnification that are representative of and connect to respective points in (D). Nuclear staining is in white, *svb*-DIG in yellow and *lacZ*-BIO staining in cyan. Scale bar is in the picture.
(D) Distance (in µm) between *svb* and *HA#14-lacZ*. Distances were measured as detailed in Materials & Methods. Points in different colours are highlighted and connected to corresponding representative images in C. Data was plotted using DATAtab.
(E) Distance (in µm) between *svb* and *alphatubulin-67C* (n=24) and *svb* and *HA#14-lacZ* (n=40). Numbers for *svb-HA#14-lacZ* distances are the same as in panel D. Distances for both pairs were measured in HA#14 lines. Distances were measured as detailed in Materials & Methods. The primers for *in situ* probes can be found in Supplementary Table 2. The sequences of the Ubx High-Affinity library can be found in Supplementary Table 1. Data were plotted using DATAtab. Box plots show mean as the center dashed line, the median is center non-dashed line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers. Each individual point is also shown. Comparisons were made between the conditions, and both two-tailed t-test for independent samples and Mann-Whitney U-Test (DATAtab) showed that the difference between them (higher values for *alphatubulin-67C* and *svb*) was statistically very significant (p=<.001).
(F) Number of trichomes counted in the first row of the larval A1 segment (indicated in white dashed line box in pictures) for different indicated conditions, with respective representative images. Data were plotted using DATAtab. Box plots show mean as the center dashed line, the median is center non-dashed line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers. Each individual point is also shown. Comparisons were made between the w1118 (n=16) and HA 14 (n=10) conditions and between the HA14 and HA16 (n=11) conditions, and both two-tailed t-test for independent samples and Mann-Whitney U-Test (DATAtab) showed that there were no statistically significant differences.

3.4 Discussion

These data suggest that sequences containing high-affinity sites can function as transcriptional enhancers across development and can exhibit features of multi-enhancer transcriptional microenvironments such as Ubx local enrichment and transcript co-localization with the Ubx target gene *svb*.

One hypothesis regarding Ubx transcriptional microenvironments is that if low-affinity enhancers use them to fine-tune specificity and robustness, then high-affinity enhancers would perhaps not display microenvironments, since they can be bound by several Hox factors. If this was the case, it would be expected that for sequences naturally containing Ubx high-affinity binding sites, or sequences altered so that sites have high-affinity would present a disruption in the radial distribution profile of Ubx around the centre of transcription sites. With different Hox factors competing for binding to these sequences, then it would be expected that enrichment for Ubx might not be found. However, I have seen that that is not the case, in both Figures 3.1 and 3.7.

Both for Ubx low-affinity sites substituted with high-affinity sites (Figure 3.1) and naturally high-affinity Ubx sites (Figure 3.7), there is a maintenance of the radial distribution profile of Ubx, with a higher peak around the centre of the transcription site that is measured – in this case, mostly reporter genes. It is interesting to observe that profiles are maintained with slight changes in distribution over distance from the transcription site. For example, there can be some differences in the enrichment exactly at the centre of the transcription site, but it is difficult to make clear speculations around that. This is in line with what was seen in Tsai *et al*., 2017, which features one example of an E3N High-affinity sequence – see Figure 1.10 J.

One interesting aspect of these changes is the variability difference across lines. If one considers the order of affinities of the series of Ubx low-affinity sites substituted with high-affinity sites predicted by NRLB where E3N 3-A is the highest affinity, there are some lines with high levels of variance (E3N 3-C and E3N Site 1). It could be speculated that there is more variability in binding events in a “middle”-point or transition between low-affinity and high-affinity binding. Nevertheless, it is important to take into account that the highest affinity reported here is still predicted to be several orders of magnitude lower than the Ubx High-Affinity Library (Gilberto Alvarez Canales, personal communication), so I am looking at a series with

increasing affinity but still zooming into the low-end of a relative binding affinity spectrum.

It is also interesting to observe the similarity of Ubx enrichment profiles between wild-type *E3N*, a high-affinity containing genomic region #58, and the respective region with a mutation in its high-affinity site (Figure 3.7). Indeed, this mutated high-affinity site displays similar levels of Ubx predicted binding affinity that of low-affinity enhancer *E3N*, and they display a very similar Ubx radial distribution profile of microenvironments.

In what concerns the characterization of the screen of genomic regions containing Ubx high-affinity sites, it is important to remember that I used these sequences in reporter assays and that they include more sequence than the high-affinity sites. Therefore, it is important not to forget the possible role or contribution of these neighbouring sequences in all of the observations that have been made. It is also relevant to note that while I observed their function in the context of reporter gene expression, this might not equate in totality to their functionality in their native endogenous location. Moreover, the target genes of each sequence are have not been a focus of this study, but might be of relevance to include in the picture for future studies.

This screen (Figure 3.2) was designed taking into account published Ubx chromatin immunoprecipitation (Choo *et al*., 2011). Most of the data shown in this chapter are based on immunofluorescence or algorithm-based binding predictions. Nevertheless, the binding of Ubx to these reporter genes has not been verified by chromatin immunoprecipitation to these high-affinity library lines. I have tried to do this verification without success due to technical difficulties and time constraints (which include a pandemic and related restricted lab access). This would be a relevant validation to perform in the future.

The finding that the genomic regions containing Ubx-high-affinity sites can drive expression broadly both spatially and across several embryonic stages (Figure 3.3) suggests that these Ubx high-affinity sites can contribute in a varied and broad way for gene expression. If the reporter gene expression was not having a contribution from these sites, one would see no expression changes for all the lines upon mutation of the high-affinity sites. That is not the case for more than half of the lines assayed (Figure 3.5).

The observation of different effects on reporter expression upon mutation of the high-affinity site of these genomic regions is in line with Ubx's

capability of acting as a repressor or activator (Hersh *et al*., 2005; Domsch *et al*., 2019; Loker *et al*., 2021). The analysis that I have done took into account the several stages of embryonic development, scoring overall effects. A finer look at these lines in the future could categorize them in more detail, perhaps dividing effects by specific stages or tissues, or even quantifying effects in tissues, regions, and/or stages of interest. Because this was a screen including various genomic regions that drove reporter expression so broadly, this analysis and scoring was not automated. The possibility to automate this analysis, such as other studies from the lab used (Fuqua *et al*., 2021), could perhaps contribute to this finer and more detailed analysis.

It is important to remember the role of co-factor binding and future experiments could dig more into the presence and contribution of Exd, along with other co-factor binding sites around these sequences and perhaps in their microenvironments. In fact, it has been observed (Tsai *et al*., 2017) that co-factor Hth is co-enriched around Ubx microenvironments. This look at several factors with the affinity predictions is also a reminder of the possible action of other Hox Transcription Factors in these sequences, and that is a direction that future studies could also follow. Finally it should be noted that predictions and *in vitro* studies cannot fully capture the real *in vivo* affinities of transcription factor binding to genomic regions to regulate gene expression; such regions may have high levels of cooperative binding, or other features such as DNA shape that may further modulate affinities. To this end, the use of live imaging could explore local binding kinetics (Tsai *et al*., 2020).

In the future, it would be worthwhile to relate these experiments to additional developmental enhancers. One of these examples is previously described (Ryoo and Mann, 1999) enhancer element *fkh250* (wild-type) which contains a low-affinity Exd-Hox binding site (AGATTAATCG) preferred by Exd-Scr and *fkh250con* (mutant), 2 bp mutation which creates an Exd-Hox consensus site that is also bound by Exd-Antp, Exd-UbxIa, and Exd-AbdA heterodimers (Rastogi *et al*., 2018). The analysis of these lines could serve to further validate the findings of these studies.

One of the most striking findings from this chapter was the co-localization between transcripts of Ubx high-affinity-site containing #14 sequence and Ubx low-affinity target *svb*. It is a very interesting observation (Figure 3.8), that follows the observations described in Chapter 2, and suggests multi-enhancer interactions between genomic regions containing Ubx high-affinity sites and genomic regions containing Ubx low-affinity sites. In addition to

the implications surrounding such interactions, already discussed in Chapter 2 (see discussion section of Chapter 2), this observation could suggest that Ubx is the driving factor of these co-localizations since the only common factor (yet) known between these two genomic regions is that they are bound by Ubx.

In the future, the ability to mark chromosomal locations independent of transcriptional state will be important. To understand the factors driving this co-localization and respective interpretations, it would be important to be able to see the distance profiles between *svb* and reporter transcript from mutated Ubx high-affinity-site containing #16, but that was not possible because there is little reporter expression from the latter. It was also not possible to have enough transcription site pairs measured to see if this phenomenon could be observed with high-affinity enhancer #58, nor for the control of measuring transcription site pairs between #14 and another gene in the same chromosome as *svb*. The ability to mark these enhancer locations could provide more information and perhaps a new mechanistic insight into the questions I have been trying to pursue. For example, the ability to add time resolution to the spatial patterns using live imaging could allow exploring the relationship between enhancer localization, transcriptional stages, and localized transcription factors.

It was, for me, a challenge to detect transcription sites with confidence against the background. This was especially laborious for co-localizations. In fact, for most of the lines, I couldn't use for co-localization analysis the semi-automated way of finding transcription sites that I have used for generating radial distribution profiles (with scripts kindly lent by my colleague Albert Tsai). It would be important to invest in the future in protocol optimization for technically overcoming these difficulties, perhaps with the development of more automated ways of doing the analysis.

In addition, in the future it would be important to assess whether microenvironment-like profiles are found for different Hox Transcription Factors, or other families of transcription factors. Are multiple microenvironments present at the same time around active transcription sites? How do sub-nuclear domains shape transcriptional outputs? In the future, the distribution of microenvironments in zones where multiple Hox factors are co-expressed could be evaluated to lend further insight into these questions.

In sum, I have shown in this chapter that Ubx high-affinity-site-containing genomic regions can also exhibit features of multi-enhancer transcriptional

microenvironments, such as transcripts driven by Ubx high-affinity-site containing region #14 co-localizing Ubx low-affinity target *svb*. In the future, this large collection of enhancers could be used to further explore the relationship of developmental enhancers, binding affinities, and transcriptional microenvironments.

3.5 Contributions

Figure 3.1

Panel A and B are reprinted as indicated in the legend.

This experiment had been discussed between Justin and Rafael and me, but the push to take it forward came from the feedback from Ingrid Lohmann, Eileen Furlong, and Alexander Aulehla in a Thesis Assessment Committee meeting.

The fly lines are from Justin Crocker's (Crocker *et al.*, 2015).

Justin Crocker acquired funding and oversaw the project, providing input.

I maintained a copy of the flies, collected the embryos, did the stainings, did the *in situs*, acquired the confocal images, analysed the data, and illustrated the figure. For *in situs,* I used a probe designed by me and prepared together with Rafael, which we shared.

Albert Tsai provided invaluable help and input regarding image acquisition settings and data analysis, kindly giving and teaching me to use his Fiji scripts for finding transcription sites and calculating microenvironment radial distributions.

Gilberto Alvarez Canales ran NRLB predictions to calculate relative affinities of each sequence.

Rafael Galupa provided invaluable input and supervision.

The Crocker Lab provided insightful comments and discussions.

The binding site scheme in Panel C is inspired by previous papers from the lab.

Figures 3.2, 3.3 and 3.5

Justin Crocker planned these experiments, also known as the Ubx High-Affinity screen, before I arrived at the lab.

Justin Crocker designed the collection of genomic regions and their mutations and created the transgenic Ubx High-Affinity lines.

Justin Crocker acquired funding and oversaw the project, providing input.

I maintained the flies since I arrived at the lab. I did the embryo collection and ßGal staining. Aliaksandr Halavatyi and Sven from Zeiss Support helped me set up the automated imaging of these stainings in the lab's confocal microscope. I did the image acquisition of the fixed embryo library staining.

I analysed the data, including doing the scoring of the mutation effects.

I illustrated the figures and curated respective Supplementary Table 1.

Gilberto Alvarez Canales created scripts for semi-automation of image processing.

Lautaro Gandara provided input regarding the scoring of the Ubx High-Affinity phenotypes.

Rafael Galupa provided valuable input and supervision.
The Crocker Lab, my TAC, and others mentioned in the acknowledgments section provided insightful comments and discussions.

Figures 3.4 and 3.6
These figures relate to data generated from data of Figures 3.2, 3.3 and 3.5 (see respective contributions).
Additionally, I did data analysis and illustrated Figure 3.6 A-C.
Gilberto Alvarez Canales did data analysis and generated panel 3.6B and all panels of Figure 3.4. Gilberto Alvarez Canales provided interesting and relevant input.
Justin Crocker acquired funding and oversaw the project, providing input.
Rafael Galupa provided invaluable input and supervision.
The Crocker Lab and others mentioned in the acknowledgments section provided insightful comments and discussions.

Figure 3.7
Fly lines come from Ubx High-Affinity Library, designed by Justin Crocker (see contributions of Figures 3.2, 3.3 and 3.5).
Justin Crocker acquired funding and oversaw the project, providing input.
Gilberto Alvarez Canales did data analysis, generated the plots of panels A-D, and provided interesting and relevant input.
I maintained the flies, collected the embryos, did the stainings, did the *in situs*, acquired the confocal images, analysed the data, and illustrated panels E and F of the figure. I used a probe designed by me and prepared together with Rafael, which we shared.
Albert Tsai provided invaluable help and input regarding image acquisition settings and data analysis, kindly giving and teaching me to use his Fiji scripts for finding transcription sites and calculating microenvironment radial distributions.
Rafael Galupa provided invaluable input and supervision.
The Crocker Lab and others mentioned in the acknowledgments section provided insightful comments and discussions.

Figure 3.8
The idea for testing the co-localization in the HA14 line came from Justin, who noticed in the spring of 2018 that the reporter expression pattern resembled *svb*'s ventral enhancers. It took almost 4 years to get to this final figure, after much trial and error, progress, and stepbacks about conditions, analysis, and controls. Rafael Galupa guided, supervised, and helped me design the latest version of this experiment, including which controls to use.

Justin Crocker acquired funding and oversaw the project, providing input.
Fly lines come from Ubx High-Affinity Library, designed by Justin Crocker (see contributions of Figures 3.2, 3.3 and 3.5).
I maintained the flies, collected the embryos, did the stainings, did the *in situs* and acquired the confocal images, did the cuticle preps, and illustrated the figure.
I used probes designed and prepared by me, of which some were designed together with, some were prepared together and some were designed and prepared together with Rafael Galupa.
Justin Crocker and I analyzed together the data from *14-svb* condition (Figure 3.8 D).
I analyzed the data from the other conditions.
Justin Crocker taught me how to manually find transcription sites and analyse co-localizations from difficult datasets, provided input, and kindly lent me his monitor (which size helped finding sparse transcription sites) to analyse some of the data.
Albert Tsai provided invaluable help and input regarding image acquisition settings and data analysis, kindly giving and teaching me to use his Fiji scripts for finding transcription sites.
Gilberto Alvarez Canales ran NRLB predictions to calculate relative affinities of each sequence and provided interesting input.
Justin Crocker and I analyzed together the cuticle prep/trichome data (Figure 3.8 F).
Rafael Galupa provided invaluable input and supervision.
The Crocker Lab and others mentioned in the acknowledgments section provided insightful comments and discussions.

4 Recruitment of (modified) Ubx may not be sufficient to drive multi-enhancer clustering

In the two previous chapters, I described how microenvironments composed of high local concentrations of Ubx are observed along with multi-enhancer clustering around both low-affinity enhancers and high-affinity enhancers. However, it remains unclear to what extent Ubx itself is sufficient to drive these phenomena. Here, I outline the development of a synthetic Ubx recruitment system that could be used to explore the role of Ubx protein domains in enhancer clustering and microenvironment formation using high-resolution microscopy.

4.1 Development of a Ubx-ΔDBD-Gal4DBD recruitment system

Together with Rafael Galupa, I designed a Ubx-ΔDBD recruitment system based on the yeast transcription activator protein Gal4/Upstream Activation Sequence (UAS) technology (Brand and Perrimon, 1993). The main aim was to test whether recruiting Ubx protein domains (except its DNA binding domain) to a synthetic platform is sufficient to lead to the formation of microenvironments and of clustering with Ubx-bound sequences (e.g. *svb*). Such a recruitment system has been used before to address the role of other transcription factor protein domains, e.g. of HoxA10 (Bei *et al*., 2007), HoxD9 (Viganò *et al*., 1998), Cdx2 (Taylor *et al*., 1997) and Bicoid (Bellaiche *et al*., 1996; Janody *et al*., 2001). This comes following the hypothesis that microenvironment formation would play a role in multi-enhancer clustering, so consequently one could test whether Ubx protein-protein interactions would be sufficient to drive multi-enhancer clustering.

This system is composed of two components: a Ubx-based fusion protein (Figure 4.1 A) and a recruitment platform (Figure 4.1 B). The fusion protein consists of the Ubx protein sequence where I substituted its DNA binding domain for the one of Gal4 protein (Gal4-DBD; Figure 4.1A, Materials and Methods); I also added a GFP tag, separated by a flexible linker, for detection purposes. I will call it UbxΔGG from now on, standing for Ubx-ΔDBD-Gal4DBD-GFP. UbxΔGG is expressed ubiquitously via an *armadillo* promoter. To serve as a control, a full Gal4 protein fused to GFP, also under *armadillo* promoter (Figure 4.1 A, Materials and Methods). The recruitment platform (Figure 4.1 B, Materials and Methods) consists of a *lacZ* reporter downstream of UAS binding sites and a promoter driving its expression, either the synthetic promoter DSCP (Pfeiffer *et al.*, 2008) or the endogenous *fkh* promoter (Ryoo and Mann, 1999). DSCP is described to be permissive to the enhancers one associates it to (Pfeiffer *et al.*, 2008). The endogenous

fkh promoter was chosen as an alternative (Figure 4.1 B, Materials and Methods) to avoid the possibility that the reporter would be inactive using DSCP alone (Pfeiffer *et al.*, 2008), given that for the read-out for enhancer clustering I rely on an actively transcribed reporter. Importantly, I have shown in Figure 2.2 that endogenous *fkh* does not co-localize with *svb*. It is unclear whether UbxΔGG recruitment will lead to activation of the reporters, as Ubx can function both as an activator and as a repressor (Hersh *et al*., 2005; Domsch *et al*., 2019; Loker *et al*., 2021) – its recruitment might also lead to no transcriptional effect at all. Ubiquitous expression via *armadillo* promoter allows the system to test the action of the fusion proteins without spatial constraints, making it compatible with where the reporters are expressed in the embryo.

After generating transgenic fly lines with each of the components of this recruitment system (Figure 4.1 A-B, Materials and Methods), I started by testing each component separately, before using them in different combinations. In the fusion protein lines (UbxΔGG and Gal4-GFP), I observed ubiquitous UbxΔGG or Gal4-GFP expression, based on GFP stainings (Figure 4.1 C), consistent with their expression being driven by the armadillo promoter. It is also possible to detect UbxΔGG with a Ubx antibody – note the different patterns of Ubx staining between the Gal4-GFP line and the UbxΔGG (Figure 4.1 D). For the Gal4-GFP line (Figure 4.1 D, right panel), the antibody stains endogenous Ubx only and it can be observed the typical Ubx stripe-like pattern at late stages, while in the UbxΔGG line (Figure 4.1 D, left panel), many embryos show strong, ubiquitous staining, indicative that the antibody can recognize the UbxΔGG protein as well (Figure 4.1 D). However, for following experiments, it is the GFP staining that will be critical; given its relatively weak signal, especially when seen under 63X for following experiments, I plan to test different antibodies in the future.

For the embryos carrying the reporter lines alone: in the fkh-lacZ line (Figure 4.1 E/Figure 4.2 A, left panel), there is reporter protein expression in a pattern consistent with *fkh* regulation, at least partially. In the DSCP-lacZ line, the expression pattern is much less localized and more widespread (Figure 4.1 E/ Figure 4.2 A, right panel).

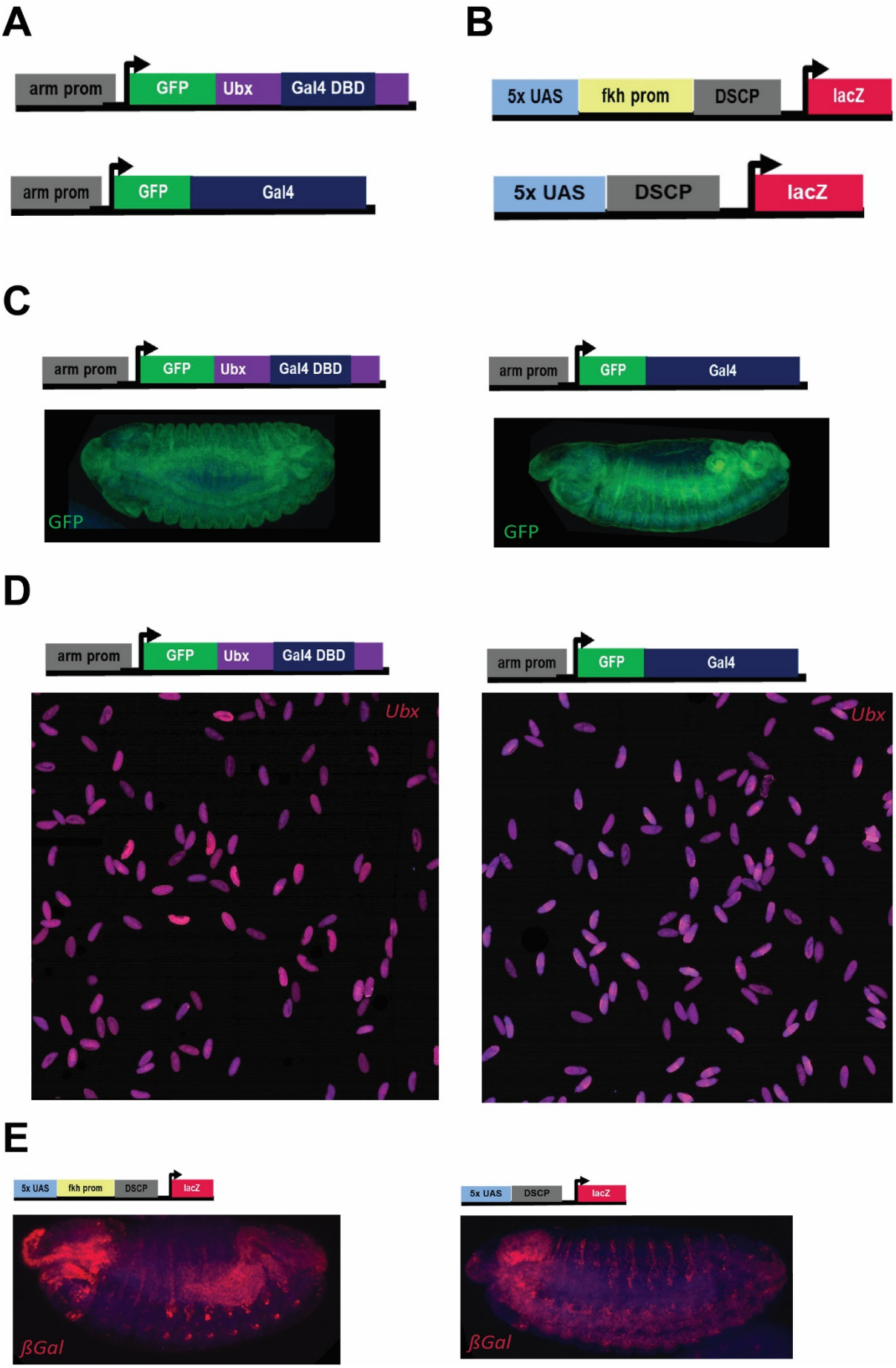


Figure 4.1: **UbxΔGG recruitment system.**

(A) Scheme represents the sequences that were inserted to generate each fusion protein transgenic fly line. 'arm prom' is the *armadillo* promoter. 'DBD' is DNA Binding Domain. For more see Materials and Methods.
(B) Scheme represents the sequences that were inserted to generate each reporter transgenic fly line. 5x UAS are five repetitions of the Upstream Activation Sequence where Gal4 specifically binds. 'fkh prom' is *fkh* promoter. DSCP is the *Drosophila* synthetic core promoter. For more see Materials and Methods.
(C) Representative embryo figures showing fusion protein expression: max projections of 20X Zstacks acquired with Zeiss LSM 880, with nuclear staining in blue and GFP staining in green. Genotypes (fusion protein lines by themselves) are indicated by the schemes above pictures.
(D) Figure showing an overview of several embryos from 5X Tiles acquired with Zeiss LSM 880, with nuclear staining in blue and Ubx staining in magenta. Genotypes (fusion protein lines by themselves) are indicated by the schemes above pictures.
(E) Representative embryo figures showing reporter expression: max projections of 20X Zstacks acquired with Zeiss LSM 880, with nuclear staining in blue and ßGal staining (product from enhancer driving *lacZ*) in magenta. Genotypes (reporter lines by themselves) are indicated by the schemes above pictures.

Then, I proceeded to test the effects on reporter expression when crossing flies carrying each reporter with flies carrying either of the fusion proteins (Figure 4.2). If the embryos derived from these crosses present altered reporter expression, that would suggest that the recruitment of the respective fused protein can regulate the transcriptional activity of the reporter. However, if there are no changes after the crosses, that could mean that the bound fused proteins have no transcriptional effect on the reporters, or that the fused proteins are not binding to start with. At this stage of the project, binding of the fused proteins to UAS binding sites has not been experimentally verified.

Embryos from crosses of the reporter fkh-lacZ line to UbxΔGG do not seem to present changes in reporter expression, as the pattern remains overall similar to the pattern of embryos carrying the fkh-lacZ reporter alone (please compare Figure 4.2 A, left, to Figure 4.2 B, left; Figure 4.2 A is repeated from Figure 4.1 E for clarity purposes). On the other hand, in embryos from crosses fkh-lacZ x Gal4-GFP, the pattern clearly changes (Figure 4.2 C, left): it becomes more ubiquitous and more intense in general (or with higher background).

In embryos from DSCP-lacZ x UbxΔGG crosses, there seem to be slight changes in expression pattern when compared to embryos carrying the DSCP-lacZ reporter alone (please compare Figure 4.2 A, right, to Figure 4.2 B, right), with a slight increase in background levels. Similar to what was observed using the other reporter, in embryos from crosses with Gal4-GFP

fusion protein, the pattern changes (Figure 4.2 C, right), becoming more intense and ubiquitous throughout the embryo.

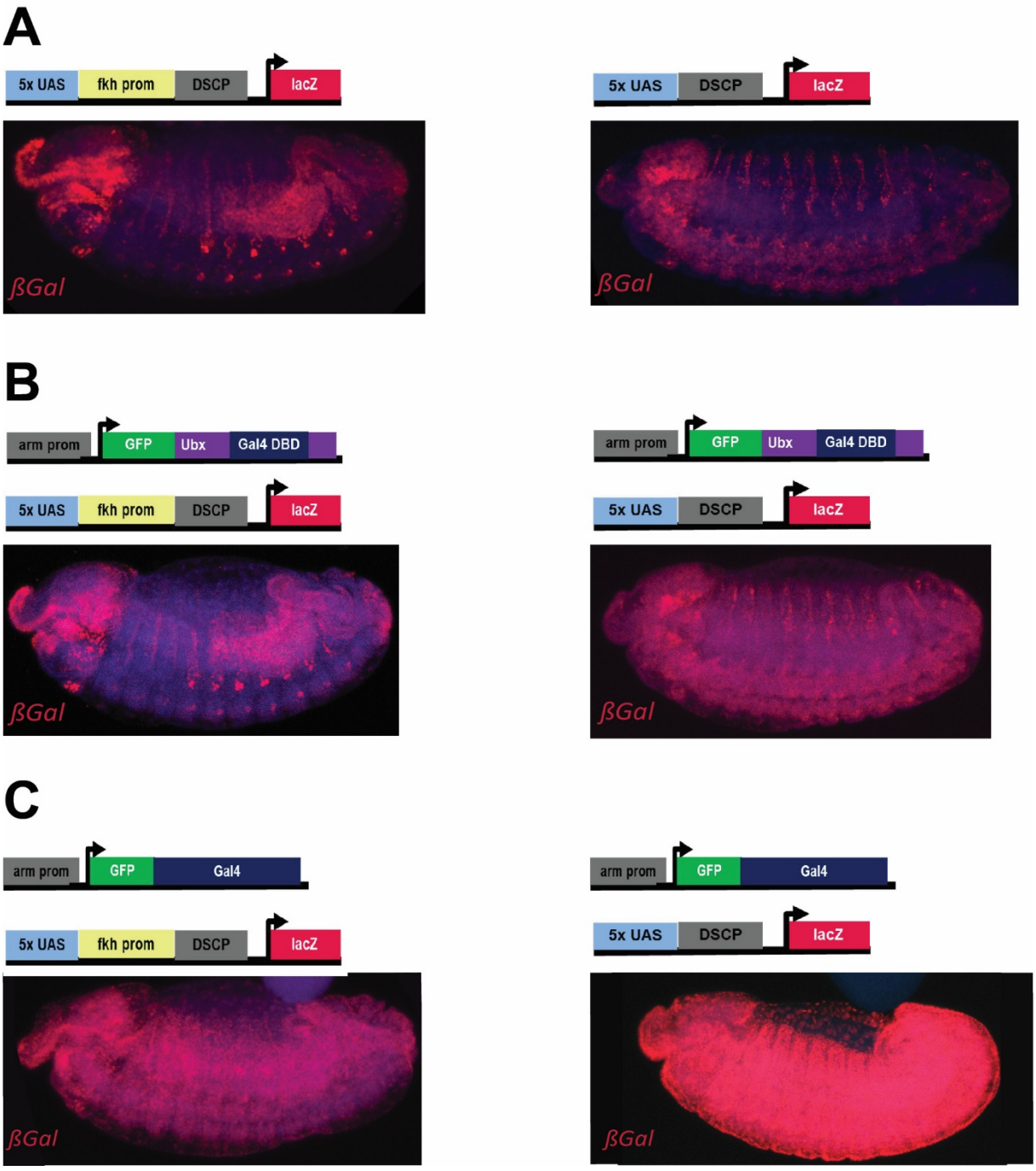


Figure 4.2: **Effects on reporter expression after recruitment of UbxΔGG or Gal4-GFP.**
(A-C) Representative embryo figures showing reporter expression: max projections of 20X Zstacks acquired with Zeiss LSM 880, with nuclear staining in blue and ßGal staining (product from enhancer driving *lacZ*) in magenta. Genotypes are indicated by the schemes above pictures. (A) are reporter lines by themselves (repeated from Figure 4.1 E for clarity purposes), (B) are crosses of

each reporter line with UbxΔGG, and (C) are crosses of each reporter line with Gal4-GFP. For more see Materials and Methods.

These data show that the reporters in the system I designed are expressed on their own and that their expression is augmented in a Gal4-GFP background, as well as in the UbxΔGG (at least for the DSCP-lacZ line). This suggests that the fused proteins are recruited successfully.

4.2 Recruitment of UbxΔGG may not be sufficient to drive co-localization of transcriptionally-active loci

To test whether there is increased clustering of the reporter gene with another Ubx-bound sequence when the fused proteins are present, I did double *in situ* hybridizations for endogenous *svb* and *lacZ* reporter transcripts and measured the distance between them, when co-expressed in the same nucleus, for all the conditions/combinations of the UbxΔGG recruitment system. These were either reporter lines by themselves, crosses of each reporter line with UbxΔGG, and crosses of each reporter line with Gal4-GFP. By order of appearance in Figure 4.3, these are: fkh-lacZ alone, fkh-lacZ crossed with Gal4-GFP, fkh-lacZ crossed with UbxΔGG, DSCP-lacZ alone, DSCP-lacZ crossed with Gal4-GFP, DSCP-lacZ crossed with UbxΔGG. I added the data from *svb-alphatubulin-67C* and *svb*-HA#14-*lacZ* distances from Figure 3.8E for purposes of having a comparison of the distributions and values (the *svb-alphatubulin-67C* distances serve as comparison for the distance between *svb* and a locus not related to Ubx on chromosome 3L – where the reporters from the recruitment system are integrated – and the *svb*-HA#14-*lacZ* distances represent an example of clustering between Ubx-related loci).

If recruitment of UbxΔGG to the synthetic platform was sufficient to lead to the formation of clustering of the reporter gene with *svb*, then a significant decrease in distances between *lacZ* and *svb* transcription sites in embryos carrying the reporter construct and the Ubx-Gal4DBD-GFP would be seen, compared to embryos carrying the reporter construct only. Moreover, the extent of this decrease would be significantly higher than a possible effect from embryos carrying the reporter construct and the Gal4-GFP control. This was not the case (Figure 4.3). The distribution of distances between *lacZ* and *svb* is not significantly different between the conditions (Figure 4.3, more details about statistical tests in the figure legend). Furthermore, this distribution is similar to the one of *svb* and *alphatubulin-67C* and different from the one of *svb* and HA#14-*lacZ* (Figure 4.3), which co-localization was shown and discussed in Figure 3.8.

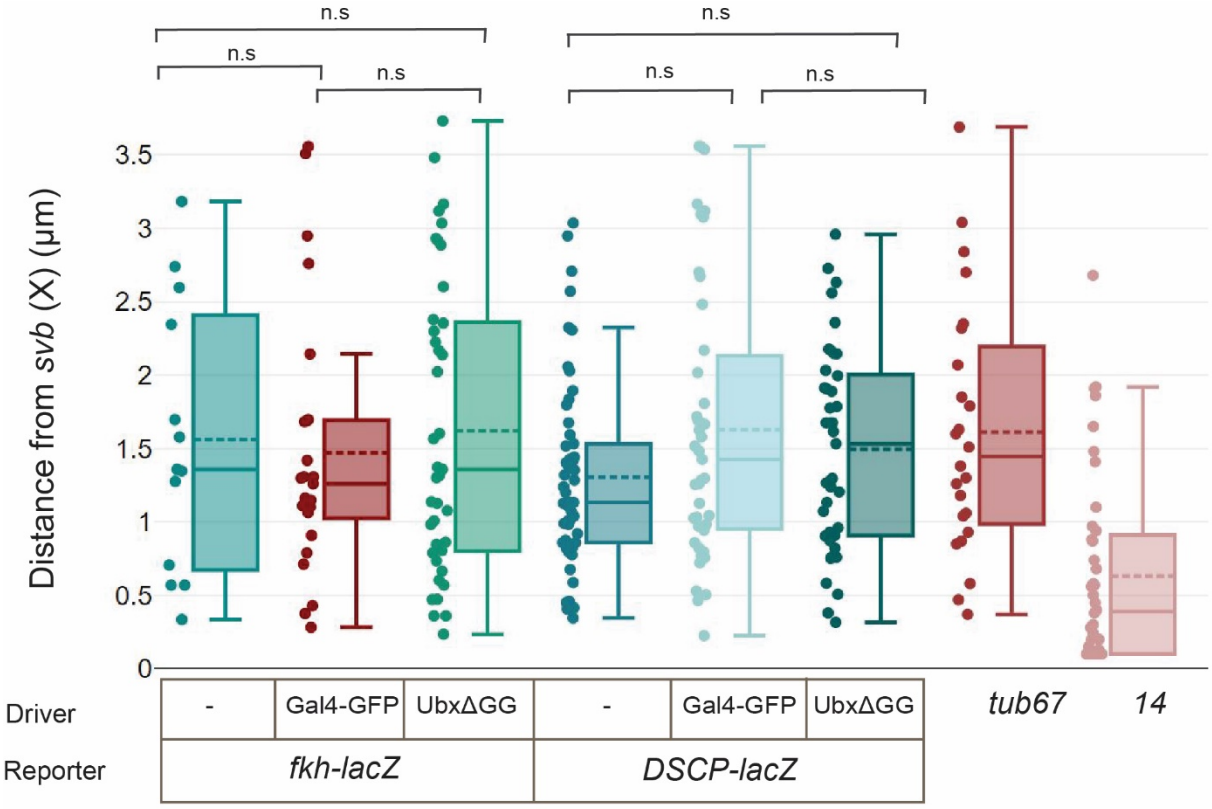


Figure 4.3: **Recruitment of UbxΔGG may not be sufficient to drive co-localization of transcriptionally-active loci.**
Distance (in µm) between *svb* and UbxΔGG -recruitment-system-*lacZ* reporter transcripts in the 3L chromosome, and also *svb* and *alphatubulin-67C* and *svb* and *HA#14-lacZ*. Conditions/genotypes are indicated in the legend (either reporter lines by themselves, crosses of each reporter line with UbxΔGG protein, and or crosses of each reporter line with Gal4-GFP). Numbers for *svb-alphatubulin-67C* and *svb-HA#14-lacZ* distances are the same as in Figure 3.8E. Distances were measured as detailed in Materials & Methods. The primers for *in situ* probes can be found in Supplementary Table 2. Data were plotted using DATAtab. Box plots show mean as the center dashed line, the median is the center non-dashed line, standard deviation as upper and lower limits, and 95% confidence intervals as whiskers. Each individual point is also shown. Comparisons were made between different sets of conditions, and both two-tailed t-test for independent samples and Mann-Whitney U-Test (DATAtab) showed that there were no significant differences between the different UbxΔGG recruitment system conditions. Number of transcription site-pairs counted and respective number of embryos for each condition from left to right: (13,8) (25,7) (41,7) (54,10) (43,10) (41,10) (24,12) (40,12).

These data suggest that UbxΔGG binding alone is not sufficient to induce clustering with a Ubx-bound locus, therefore Ubx protein-protein interactions are not the (at least, only) driving force to form transcription-site co-localization between the sequences that Ubx binds to.

4.3 Discussion

These data suggest that recruitment of Ubx protein domains does not seem to be sufficient to drive co-localization with *svb*, and therefore it is not the (at least, only) driving force to form multi-enhancer clustering between sequences bound by Ubx.

Firstly, it is worth remembering that this test was focused on only one (*svb*) of several Ubx targets. Having that in mind, if recruitment of UbxΔGG is really happening, then this would mean that the previous observations in Chapters 2 and 3 of transcript co-localization are probably regulated by factors/processes other than (or on top of) the influence of Ubx non-DBD protein domains. It could mean that multi-enhancer clustering is not influenced directly by Ubx, or it could also mean that possible latent affinities coming from the Ubx native DBD conformation are crucial for the phenomena. In Figure 3.8, with the *alphatubulin-67C-svb* pairs, I also have excluded the hypothesis that this co-localization is merely something that naturally happens between the X and 3L chromosomes. These phenomena are probably regulated by multiple layers and players.

I found it interesting to observe the differences between conditions regarding the number of transcription-site pairs that I could find co-expressed in the same nucleus, for the same amount of embryos (7-12). For fkh-lacZ alone I found 13, for fkh-lacZ crossed with Gal4-GFP I found 25, for fkh-lacZ crossed with UbxΔGG I found 41, for DSCP-lacZ alone I found 54, for DSCP-lacZ crossed with Gal4-GFP I found 43 and for DSCP-lacZ crossed with UbxΔGG I found 41. Differences in the number of co-expression transcripts could reflect spatiotemporal regulation of reporter expression by UbxΔGG when it is recruited to the platform. But to test that, it would require future counts of all nuclei expressing each transcript and then finding the ratio between co-expressed nuclei and nuclei only expressing *svb* or *lacZ*. As discussed in the previous chapters, tagging DNA instead in the future would also solve this issue of relying on co-expressed transcription sites to evaluate co-localization.

Nevertheless, it would be important to confirm binding to see if the system indeed works. This could be done, for example, by chromatin immunoprecipitation of GFP. From the change of reporter protein expression for Gal4-GFP in Figure 4.2, it seems that for this line it is indeed binding. Since the DNA Binding Domain is the same in this line and the UbxΔGG, it would make sense that UbxΔGG is also being recruited and binding upstream of the reporter. That is, unless the non-DBD Ubx protein domains

in UbxΔGG impair in any way the binding of its Gal4-DBD to the UAS sites. Even if subtle, there seem to be changes in reporter expression after crossing UbxΔGG with the DSCP-lacZ reporter line, so if it is being recruited there, it should also be recruited when crossed with the fkh-lacZ reporter line. Nevertheless, with experimental validation, one could be more certain of this.

After further validating the system, then a following step forward would be to explore whether recruitment of UbxΔGG can induce the formation of microenvironments and how they differ from endogenous Ubx microenvironments. It would also be interesting to assess whether the Gal4 control forms microenvironments around the reporter transcripts. The GFP staining is not great in these lines, but further troubleshooting and optimization could be done in the future, starting by trying out different antibodies. If microenvironments are formed, then it could be interesting to explore the role of UbxΔGG-bound levels and the role of different Ubx protein domains on microenvironment formation. For the first, changes could be made to the UAS sites in the recruitment platform, and see if by modulating binding site number there would be changes in microenvironment formation. For the second, changes could be made to the fusion protein to see if the absence or presence of certain protein domains would affect microenvironment formation.

In sum, in this chapter, I have described a newly developed recruitment system for (modified) Ubx and shown that recruitment of UbxΔGG does not seem to be sufficient to drive transcript co-localization with *svb* (as proxy of multi-enhancer clustering), while also discussing how this system could be a useful tool for future aspects that remain to be tested.

4.4 Contributions

When discussing my motivation to have a more mechanistic look into my Ph.D. research theme, Rafael Galupa proposed me this idea, after which he and I planned the experiments together. Rafael Galupa designed the UbxΔGG recruitment system with my input. Rafael Galupa oversaw and supervised these experiments, providing help whenever needed. Rafael Galupa and I designed, ordered, and built the genomic constructs, except for the DSCP-lacZ construct, which had been already synthesized by Rafael Galupa for another project.

I maintained the fly lines, did the crosses, collected the embryos, did the stainings, did the *in situs*, acquired the confocal images, analysed the data, and illustrated figures.

Albert Tsai provided invaluable help and input regarding *in situ* image acquisition settings and data analysis, kindly giving and teaching me to use his Fiji scripts for finding transcription sites.

Justin Crocker taught me how to manually find transcription sites and analyse co-localizations from difficult datasets, kindly lent me his monitor (which size helped finding sparse transcription sites), and provided input for my analysis of the co-localization data.

Justin Crocker provided valuable input and interesting comments regarding the project and acquired funding.

The Crocker Lab and others mentioned in the acknowledgments section provided insightful comments and discussions.

5 Conclusions and Perspective

This thesis focused on Ubx transcriptional microenvironments, providing a more in-depth understanding of their components, and exploring them in the context of gene regulation during animal development. I have shown that active *svb* enhancers on different chromosomes tend to co-localize and that such observations can coincide with the rescue of defects from a deletion of a redundant enhancer from the *svb* locus at elevated temperatures. I screened a library of short genomic fragments containing endogenous Ubx high-affinity sites and observed that they can exhibit features of multi-enhancer transcriptional microenvironments such as Ubx local enrichment and transcript co-localization with Ubx target *svb*. Lastly, I developed a (modified) Ubx recruitment system to test the role of this Hox Transcription Factor in enhancer clustering.

Throughout the discussions of each chapter, I have discussed open questions that remain. Further studies would be relevant to dissect the interplay between low and high-affinity binding sites, taking most Hox factors and co-factors into account. It is also important to remember that as much as scientists can categorize phenomena, affinities exist in a spectrum (as everything in life). More studies with live embryos will be important to uncover the real *in vivo* affinities of transcription factor binding to genomic regions to regulate gene expression. Such an *in vivo* approach will allow exploring the many layers of gene expression regulation (reviewed in Carnesecchi *et al*., 2018) known to exist alongside the molecular players that have been outlined throughout this thesis. These layers include chromatin domains (reviewed in Furlong and Levine, 2018), the role of transcriptional condensates (Hnisz *et al*., 2017), and how these components contribute towards the understanding of animal development.

To better understand the mechanisms behind microenvironment formation and multi-enhancer clustering, it is essential to identify genome-wide contacts of *svb* enhancers (and others) without the dependence of assays on active transcription sites. It is therefore essential to develop tools that allow visualizing these sites regardless of their transcriptional status. During my Ph.D., in addition to the work I have described, I have also designed and generated different tools for this purpose (such as new plasmids and fly lines), which could be used in the future to implement techniques such as TaDamC or live imaging of marked enhancers. It would be very interesting to follow and compare coordinates of transcriptionally inactive and active sites of the *svb* locus, which could also be extended and

compared with high-affinity enhancers. Interesting measurements include Ubx distributions around the enhancers when they are both active and inactive. Live imaging would allow studying the dynamics of these marked enhancers and inform scientists of the frequency and mobility of independent genomic sites while fixed embryos would allow testing interactions with other proteins and marked sites and compare between active and inactive transcription states. Moreover, these tools would additionally allow comparing contacts across several tissues, interpreting these interactions in the context of development, as well as to test how transient these interactions are and over which length scales they occur. The UbxΔGG recruitment system I developed could be a useful tool for future experiments, such as exploring the role of UbxΔGG-bound levels and the role of different Ubx protein domains on phenomena such as microenvironment formation, or other phenomena related to Ubx developmental regulatory activity or even adapted to look closer at the activity of other transcription factors, such as but not limited to Hox factors.

To conclude, this thesis provides a modest but honest contribution to the building blocks of knowledge towards how animal body plans are shaped through transcriptional regulation. I hope that these findings can be combined with further experiments, including others who may find them of interest. Even if they are not, science is not done in a vacuum (Alves, 2020) - but in a societal context -, and while working on this thesis I believe that I contributed at the same time my share into the scientific environment around me for the better. Increasing technological developments promise an ever-exciting future for the dissection of biological and biochemical mechanisms such as the discussed in this thesis, a bright future to uncover more secrets about life. If this technological development is accompanied by less substantial emphasis in metrics and status optics, and a bigger investment in honest and focused experiments, in supporting people, in scientists giving voice to and surrounding themselves by talented diverse people, then that bright future can be meaningfully realized.

“Within the interdependence of mutual (nondominant) differences lies that security which enables us to descend into the chaos of knowledge and return with true visions of our future, along with the concomitant power to effect those changes which can bring that future into being.”

Audre Lorde, The Master’s Tools Will Never Dismantle the Master’s House (1979)

6 Materials and Methods

6.1 Fly Strains and constructs

Maintenance of D. melanogaster strains was done with standard laboratory conditions, reared at 25°C unless otherwise specified.

w1118 stock is referred to as wild-type.

Chapter 2
Previously described fly strains, by order of appearance:
7H-lacZ, E3N-lacZ (Crocker *et al*., 2015); Df(X)svb108 (Frankel *et al*., 2010); svbBAC-*dsRed* (Preger-Ben Noon *et al*., 2018a); diBAC-gfp is CH322-35A16 EGFP tagged in VK37, covering D (Venken *et al*., 2009); DG3-lacZ (Tsai *et al*., 2017).

Chapter 3
Previously described fly strains, by order of appearance:
E3N-lacZ, E3N3-D-lacZ, E3N3-C-lacZ, E3N3-A-lacZ, E3N3-Site2-lacZ, E3N3-Site1-lacZ (Crocker *et al*., 2015);

High-Affinity Ubx lines: Sequences (Supplementary Table 1) were synthesized and cloned (Genscript) into a placZattB reporter construct and fly strains were generated from w1118 stock, by injection of the constructs (Genetivision) tagged in attP2 sites.

Chapter 4
The Ubx recruitment system constructs were synthesized and cloned (GenScript) and can be shared upon request to be accessed at https://drive.google.com/drive/folders/10zgQrny42jaOl7ye0Wq1Is1Dj53Fh5Dl?usp=sharing. The fly strains were generated from w1118 stock by injection of the constructs (Genetivision) tagged in either attP2 or attP40 sites.

6.2 Embryo collection

Unless latest otherwise, flies were loaded into egg collection chambers, and embryos were collected after an overnight lay, usually after waiting 2 days for flies to “warm-up” and start laying more.

6.3 Heat-shock experiments

For heat-shock experiments, flies that had already warmed up in collection chambers laid eggs at 25°C for 5 hr on new apple-juice agar plates, these plates were then incubated at 32°C degrees for 7 hr, and then embryos were fixed. For cuticle preparations, dechorionated embryos were kept at 32°C until they emerged as larvae.

6.4 Cuticle preparation and counting of trichomes

Cuticle preparations of collected larvae were done using a published protocol (Stern and Sucena, 2011), imaged with phase-contrast

microscopy, and ventral trichomes in larval A1 or A2 segments were counted using Fiji/ImageJ's find maximum function (Schindelin *et al.*, 2012; Schneider *et al.*, 2012).

For Df(X)svb108 allele crosses, a homozygous marker for the deletion locus was used: the only larvae imaged were the ones without T1 segment trichomes (Tsai and Alves *et al.*, 2019), as any wild-type *svb* allele will produce T1 trichomes in larvae (Tsai and Alves *et al.*, 2019).

6.5 Embryo Fixation for Immuno-fluorescence stainings and *in situ* hybridizations

Apple juice agar plates of egg-laying cages were swapped, ≈1 ml water was squirted onto an apple juice plate, between yeast paste and rim, a brush was used to loosen embryos and flush into cell strainer with water. Embryos were dechorionated for 2 mins with bleach diluted 50% in water in a glass petri dish, kept at a level halfway to the rim of the strainer, by moving the strainer into the dish with bleach. Embryos were washed thoroughly with water to remove bleach. Embryos were washed with embryo wash buffer (6g of NaCl and 2mL of TritonX-100 for a total of 1L of autoclaved water) and dried on top of a paper towel. With a brush, embryos were transferred into scintillation vials with 700µl PFA 16%, 1.7ml PBS/EGTA (1x PBS and 25mM EGTA pH8.0), and 3.0ml Heptane. Vials were shaken for 25 mins at 250rpm in a MaxQ shaker. After that, with a glass pipette, bubbles were popped at the interface and the lower (aqueous phase) was removed completely. Around 5mL of MeOH 100% were added, the vial was capped and vortexed for 30s. After waiting for phases to be separated, non-devitellinized embryos were removed from the interphase, and then the upper (heptane) phase was removed completely. Embryos were rinsed and stored in MeOH 100%, kept at -20°C.

6.6 Immuno-fluorescence stainings and *in situ* hybridizations

Embryos were rehydrated by rocking in EtOH/MeOH 50% for 5 minutes, 3 washes with EtOH 100%, 1h in Xylenes/EtOH 10:1, 3 washes with EtOH 100%, 5 minutes in EtOH 100%,3 washes with MeOH 100%, and 5 minutes in MeOH 100%. Then, embryos were washed 2-3 times quickly with PBT (0.5mL Tween20 in 500mL PBS), rocked for 25 minutes in PFA 5% in PBT, washed 2-3 times quickly with PBT, followed by four 10 minute washes in PBT. Embryos were rocked for 10 minutes in PBT/Hyb buffer (50% Formamide, 4X SSC, 100 µg/mL Salmon DNA, 50 µg/mL Heparin and 0.1% Tween-20) 50%, then rocked for 2 minutes in Hyb, then incubated for at least 1.5h in Hyb at 55°C, washing embryos in between with at least three washes. Then, embryos were incubated overnight at 55°C with 100 µL probe solution (diluted probes in Hyb were first denatured for 5 minutes at 80°C, then put on ice for 5 minutes, then at 55°C for 10 mins). The next day, embryos were washed 3-4 times over 2-3h in Hyb at 55°C, then were rocked for 10 minutes in PBT/Hyb 50% at room temperature, and the following protocol continued at room temperature. The following part of the protocol is similar for stainings without *in situ* hybridization, which is just

preceded by 10 minutes of rocking in PBT/MeOH 50%. Embryos were washed 4 times in PBT for 15 minutes each, were rocked for thirty minutes in PBT/blocking buffer from Roche (5:1), and then incubated in primary antibody/PBT/blocking dilution (see information below) overnight at 4°C. The next day, embryos were washed 4 times in PBT for 15 minutes each and incubated in secondary antibody/PBT/blocking dilution (see information below) for 2 hours at room temperature. Then, embryos were washed times in PBT for 15 minutes each and mounted and imaged as described in the following section 'Mounting and imaging of fixed embryos'.

Proteins were detected by using primary antibodies, which in turn were detected by fluorescently labelled secondary antibodies (Alexa Fluor dyes 1:500, Invitrogen). These secondary antibodies were also used to detect primary antibodies against in situ hybridizations labels (DIG, FITC, or biotin) of antisense RNA-probes.

The following primary antibodies were used (dilution used in brackets):

Ubx: Developmental Studies Hybridoma Bank, FP3.38-C (1:20)
DsRed: MBL anti-RFP PM005 (1:100)
DIG: Thermofisher, 700772 (1:100)
Biotin: Thermofisher, PA1-26792 (1:100)
FITC: Thermofisher, A889 (1:100)
ß-Gal: Promega anti-ß-Gal antibody (1:250)
GFP: Thermofisher, A11120 (1:250)

In situ hybridization antisense RNA-probes were made against reporter construct RNAs (*lacZ*, *dsRed*, *gfp*) or endogenous RNA (*svb*, *fkh*, act5c, *atub67C*). Dilutions were in the 1:100 range.
The primer sequences to generate these probes are listed in Supplementary Table 2, except for probes for *dsRed* reporter construct RNA and *fkh* RNA, for which existing probes in the lab were used.

6.7 Mounting and imaging of fixed embryos

After staining and/or in situ hybridize, *Drosophila* embryos were mounted in ProLong Gold+DAPI mounting media (Molecular Probes, Eugene, OR). To image the fixed embryos, a Zeiss LSM 880 confocal microscope with FastAiryscan (Carl Zeiss Microscopy, Jena, Germany) was used, using 405, 488, 561, and 633 nm wavelength excitation lasers depending on the fluorescent dyes. When used (Chapter 2), Airyscan mode was applied as previously described (Tsai *et al*., 2017).
Fiji/ImageJ (Schindelin *et al*., 2012; Schneider *et al*., 2012) and Matlab (MathWorks, Natick, MA, USA) were used to process the images.

For Df(X)svb108 allele crosses, a homozygous marker for the deletion locus was used: the only embryos imaged were the ones without T1 segment *svb* mRNA expression (Tsai and Alves *et al*., 2019), as any wild-type *svb* allele will produce T1 trichomes in larvae (Tsai and Alves *et al*., 2019).

6.8 Microenvironment analysis

In the second chapter, the find maximum function of Fiji/ImageJ was used to identify the center of *svb* transcription sites. The local Ubx concentration was computed through the fluorescence intensity of the Ubx channel inside a region of interest (ROI) consisting of a 4-pixel diameter circle centered on the transcription site (170 nm which is about the lateral resolution limit of 3D mode AiryScan). The transcriptional output was computed through the fluorescence intensity of the RNA channel inside a region of interest (ROI) consisting of a 4-pixel diameter circle centered on the transcription site (170 nm which is about the lateral resolution limit of 3D mode AiryScan).
Intensity is the per-pixel average intensity with the maximum readout of the sensor normalized to 255. Plots were generated in Matlab (MathWorks, Natick, MA, USA).

In the third chapter, the find maximum function of Fiji/ImageJ was used to identify the center of transcription sites. Radially averaged distributions of Ubx centered around transcription sites were calculated as described in Tsai *et al*., 2017. Plots were generated in Excel (Microsoft Corporation, USA).

6.9 Transcription site co-localization analysis

In Figures 2.2 and 2.4, nuclei were chosen that co-expressed *svb* and *dsRed*/*gfp*. The find maximum function of Fiji/ImageJ was used to identify the center of transcription sites, and their coordinates were used to calculate distances between pairs of transcription sites. In Figure 2.4, co-localization was considered for distances lower or equal to 360 nm. Plots were generated in Matlab (MathWorks, Natick, MA, USA).

In Figures 3.8 and 4.3, nuclei were chosen that co-expressed *svb* and *lacZ*, *svb* and *atub67C,* or *lacZ* and *act5c*. Transcription sites were manually identified, using a high threshold in channel adjustments for making sure that background would not appear and risk be confused with a transcription site. Length between a line drawn manually between the centres of each transcription site was used to calculate the distance between them. Plots were generated in and statistical tests were made in DATAtab Team (DATAtab e.U., 2021. Graz, Austria).

6.10 Manual conservation analysis

The number of high-affinity site sequence matches were counted manually using the 124 insect species track to the D. melanogaster genome (dm6). More information is in the legend of Figure 3.6A.

6.11 PhyloP analysis

As stated in the contribution section, my colleague Gilberto Alvarez Canales did this analysis. Violin plots were generated from the PhyloP scores from the 27 species comparison to the D. melanogaster genome (dm6). Conditions were as described in legend from Figure 3.6B. Wilcoxon tests

were done given that the distribution of the data is non-normal (Normality Shapiro tests were done).

6.12 Binding-affinity predictions and visualizations

As stated in the contribution section, my colleague Gilberto Alvarez Canales did this analysis. Affinity values were estimated with the NRLBtools package in R (Rastogi *et al*., 2018). Affinities were obtained for each base pair in the forward and reverse DNA sequences for each enhancer. Total affinities for each of the Hox genes were estimated with the sum over all the base pair affinity values for each sequence. These values were normalized by the length of each sequence. Individual examinations of position-specific affinity changes in the mutants were done with the normalized difference between each enhancer and its respective mutant and through visual inspections.

Affinity distributions were generated to see how the mutations can affect the binding affinities for the Hox genes in all the Ubx High-Affinity sequences. Affinities were predicted for each of the sequences. Density distributions were generated for comparison between wild-type and mutant sequences.

Changes in total affinity were estimated for each sequence for each of the Hox genes. The changes were obtained by taking the difference between the total affinity of each WT sequence with its respective Hox gene. To make visually comparable estimations, the values were normalized with the median affinity values from all the sequences for each factor and sequence length.

6.13 Data availability

Figure 2.2:
Data is available at Dryad Digital Repository under a CC0 Public Domain Dedication at http://dx.doi.org/10.5061/dryad.q96g6 (Tsai *et al*., 2017).
Figures 2.3, 2.4, and 2.6:
Imaging data is available at https://www.embl.de/download/crocker/svb_enhancer_co-localization/index.html (Tsai and Alves *et al*., 2019).

7 Appendix

Supplementary Table 1: List of Ubx High-Affinity library sequences

Line#	Sequence (In red: HA motif, mutations in non-capital letters)	Size (bp)
>HA2	AATGATTTATTACTGTATTTGCTCAAGTGCACTTGTGGAACACCAGCACCCCCCATCCAAGGTGTTTTTAAAAGCGCCTAAAAAAACGTTTTCACCACAAGCTCAAACTGCTTAACCTTCAG AGCTCTGAAGTATCACGCAGTTGACGGCAATTCATGGCCATCAAAATCTATTGGGGGCAAACTTGAAATCTGTTCCGTGGCACAACTGAAGTTTTAACCCCCGAAAAGTGCCACCTGGA AAATGCCACCCCCCATTGTCACCTTTGGCAGTCGAACTCATAATCGGGACACGTGCTGGCTTCGTTATTTCTCTCTTATTCTTGGCTGCTTGGCTGGCTTTCATTTGCGGAACTTTCCTT CGTCCTTCCGCGTCTTTTCAAATTGTCAAAGTTTTTTCTTCAGCTTTTTTGGCCggcttggtttggtttggtttggtttgtttcggtttggttggttGGTGGGGGGCTAACTACTTTGCCGTTTTTATGGATGCTCT TTATTTAGTCAATGACTTTGACGTCGACAACTTTAATCGAAAGCAATCTGGTGGAGAGTCCATGAGGTGAGGTGTGCGCACGCGTTTCGCTTAATGCCGCGTAAATTGTTTCGCAGCCA AAGTTTG	624
> HAmut4	AcTGcTTTcTTcCTGTATTTGCTCAAGTGCACTTGTGGAACACCAGCACCCCCCATCCAAGGTGTTTTTAAAAGCGCCTAAAAAAACGTTTTCACCACAAGCTCAAACTGCTTAACCTTCAG AGCTCTGAAGTATCACGCAGTTGACGGCAATTCATGGCCATCAAAATCTATTGGGGGCAAACTTGAAATCTGTTCCGTGGCACAACTGAAGTTTTAACCCCCGAAAAGTGCCACCTGGA AAATGCCACCCCCCATTGTCACCTTTGGCAGTCGAACTCATAATCGGGACACGTGCTGGCTTCGTTATTTCTCTCTTATTCTTGGCTGCTTGGCTGGCTTTCATTTGCGGAACTTTCCTT CGTCCTTCCGCGTCTTTTCAAATTGTCAAAGTTTTTTCTTCAGCTTTTTTGGCCggcttggtttggtttggtttggtttgtttcggtttggttggttGGTGGGGGGCTAACTACTTTGCCGTTTTTATGGATGCTCT TTATTTAGTCAATGACTTTGACGTCGACAACTTTAATCGAAAGCAATCTGGTGGAGAGTCCATGAGGTGAGGTGTGCGCACGCGTTTCGCTTAATGCCGCGTAAATTGTTTCGCAGCCA AAGTTTG	
>HA6	AGTCCTGGTTGGCCACAAAAATTCTATTTTGCATAAGAATTCTTCTGACACCGAATATCAAAGGTGGCACACAGATGCAGTACGAGTAATGCAAAGTATTTCAATAACCCAGCAAACACCT GACGCAGTAACTCTTGAAATGAAAATTCAACTCAAGCTTCAAATGGAACAATGTTATATATTGTTTCTTTTCCAAAAGCAGAAAGGAAAATATAAATGGATTTTCTCAAATATTAAAAGGTTTA CGTAGTAAATTAAGTAACTAGTCTAATACTGACGCGTAATAAATCATAATTTACTTAGATCATAGAAAACTAATAACTAGAGAATCAGACCAACAGCCTAAAAGTTTGTTATATATTTAATGCC ATTTTTAATTTTCGCCCGCTATGCATCATCGTATTGCAACAAACTTTGCCATATAGAACATAAAGTTCGCCGTTCTAAGAGCGGCTCGTGTTCGCATAGAAAGTGGGTGGCTGTTT	485
> HAmut8	AGTCCTGGTTGGCCACAAAAATTCTATTTTGCATAAGAATTCTTCTGACACCGAATATCAAAGGTGGCACACAGATGCAGTACGAGTAATGCAAAGTATTTCAATAACCCAGCAAACACCT GACGCAGTAACTCTTGAAATGAAAATTCAACTCAAGCTTCAAATGGAACAATGTTATATATTGTTTCTTTTCCAAAAGCAGAAAGGAAAATATAAATGGATTTTCTCAAATATTAAAAGGTTTA CGTAGTAAATTAAGTAACTAGTCTAATACTGACGCGgAAgAAAgCAgAATTTACTTAGATCATAGAAAACTAATAACTAGAGAATCAGACCAACAGCCTAAAAGTTTGTTATATATTTAATGCC ATTTTTAATTTTCGCCCGCTATGCATCATCGTATTGCAACAAACTTTGCCATATAGAACATAAAGTTCGCCGTTCTAAGAGCGGCTCGTGTTCGCATAGAAAGTGGGTGGCTGTTT	
>HA10	TAAGATTAAAGCAACTGAAATTAAAGCTAGTCATAAATAATTGCTTAAATCTACCATTATCACTAAGCTAGTTTGTAAATAAGCAAGaattattaaaataaaaataatattaatttaatCTTACTATGCAAATAC ATATTTTATTCATAAGATTAAAGCAACTGAAATTAAAGCTAGTCATAAATCATTGCTTAAATCTACAATTATCACTAAGCTAGTTTGTAAATAAGCAAGaaatattaaaataaaaataatattaattattattattat tattattatatattatattaACTTATTTTCTTGCAGTGATCATTCAGGTCGCCAAAGCGCTTAGGGTAAATAATATTCAACAAAGGGAAAAAATAAACGAAAAATCCAAGTCATACACACAGAAAAACAC CGGCGGTGGCCACCTTCTGCTAAGCCGGGTGAAC	437
> HAmut12	TAAGATTAAAGCAACTGAAATTAAAGCTAGTCATAAATAATTGCTTAAATCTACCATTATCACTAAGCTAGTTTGTAAATAAGCAAGaattattaaaataaaaataatattaatttaatCTTACTATGCAAATAC ATATTTTATTCATAAGATTAAAGCAACTGAAATTAAAGCTAGgCAgAAAgCAgTGCTTAAATCTACAATTATCACTAAGCTAGTTTGTAAATAAGCAAGaaatattaaaataaaaataatattaattattattattat tattattatatattatattaACTTATTTTCTTGCAGTGATCATTCAGGTCGCCAAAGCGCTTAGGGTAAATAATATTCAACAAAGGGAAAAAATAAACGAAAAATCCAAGTCATACACACAGAAAAACAC CGGCGGTGGCCACCTTCTGCTAAGCCGGGTGAAC	
>HA14	GCTCTGCCCACATTACCCCACTCACACCGAGATTTTCCATGTGGCCGACAGTTTTTCACGCTTTTCCCCACCAATTGTGAGCGGGACAAAGTCTGCTGACTGCGTCGACCCGTTCCAGT CATAAATCATTGGGCATTTGCCATATGCAAAGGGCTATTGCTATCACTAAGTCCGGCATTAAACTCGGCTTTGACTTTCATTCACCCA	207
> HAmut16	GCTCTGCCCACATTACCCCACTCACACCGAGATTTTCCATGTGGCCGACAGTTTTTCACGCTTTTCCCCACCAATTGTGAGCGGGACAAAGTCTGCTGACTGCGTCGACCCGTTCCAGg CAgAAAgCATTGGGCATTTGCCATATGCAAAGGGCTATTGCTATCACTAAGTCCGGCATTAAACTCGGCTTTGACTTTCATTCACCCA	
>HA18	TTTCGCCCGCCCCCTTGTCATAAATCATTTGGGGTTTTCATTTTCGGGGGGGTGGGAGACAGCCTGGTATTTTGAATAATTTTTGTTGTTAATTTCTGTTTTGATTTAGCTCCCAAGAAGC GGCAAACTTGCGTTATTCATATTTTATAGCACATTAACTTGTCTCGCCGATAATTTTGTTTGACATGGTTTTAGTGTCTGCCCGCGGGAAAAGCGAGATTCTCCTTAGTTAGCGCTTCCCG ATCAATCAGTCATTTATCAAAGTTTCTTTCACTTTACTTTTCCCAGCTGCCCCGCCCCCATTCATATTTCTTGTGATAAAAAAGCCCGCTGGATTTTGGGAATTTCTCACACGTTTCGTTACA TATCAAATATGTGGAGTTTTTTCCCCCATCCCCGCTATGTTTGTTTTCTTTTTCCTTCTACGTAATGATTCCGATAACGCCCACATTAACGAGAACTTTTCGTTTTTCCCCCGCACACGAAC AATGAAAACGATTTTCCAGCAGCACAAACAATTATAGCGATGCATAAATTAAGCACAACCCAATGTGAATAACTAATTTCCTGTACGGCAATTACTAAAAAGAGAATTCCACTTTAATTCCGA TTCCAAGAGCAAGACCAAGACCACAACCAAGTGCGCATTAAATAAGCATATTTGGATGCATGATACACTTAATCGCTGGAAAAATGCTTGGAGGAAAACCCGAAATGCACCATAAAAAAC GTGACAAAATGCCAGTTTGTGATAGCGAAGTGTGACCCTGTTTTTTTTCCGAGGCGCAACTCACAAATTTATGCACAGAAATCAACCACAATGAGAGTGTGCTCGATAAGGTGGAACAC GACGACCACCACAGATCGTTGCGTTTTTAATGCACAGCATTTTTGATAGAGGTAGGATCGGGG	912
> HAmut20	TTTCGCCCGCCCCCTTGgCAgAAAgCAgTTGGGGTTTTCATTTTCGGGGGGGTGGGAGACAGCCTGGTATTTTGAATAATTTTTGTTGTTAATTTCTGTTTTGATTTAGCTCCCAAGAAGC GGCAAACTTGCGTTATTCATATTTTATAGCACATTAACTTGTCTCGCCGATAATTTTGTTTGACATGGTTTTAGTGTCTGCCCGCGGGAAAAGCGAGATTCTCCTTAGTTAGCGCTTCCCG ATCAATCAGTCATTTATCAAAGTTTCTTTCACTTTACTTTTCCCAGCTGCCCCGCCCCCATTCATATTTCTTGTGATAAAAAAGCCCGCTGGATTTTGGGAATTTCTCACACGTTTCGTTACA TATCAAATATGTGGAGTTTTTTCCCCCATCCCCGCTATGTTTGTTTTCTTTTTCCTTCTACGTAATGATTCCGATAACGCCCACATTAACGAGAACTTTTCGTTTTTCCCCCGCACACGAAC AATGAAAACGATTTTCCAGCAGCACAAACAATTATAGCGATGCATAAATTAAGCACAACCCAATGTGAATAACTAATTTCCTGTACGGCAATTACTAAAAAGAGAATTCCACTTTAATTCCGA TTCCAAGAGCAAGACCAAGACCACAACCAAGTGCGCATTAAATAAGCATATTTGGATGCATGATACACTTAATCGCTGGAAAAATGCTTGGAGGAAAACCCGAAATGCACCATAAAAAAC GTGACAAAATGCCAGTTTGTGATAGCGAAGTGTGACCCTGTTTTTTTTCCGAGGCGCAACTCACAAATTTATGCACAGAAATCAACCACAATGAGAGTGTGCTCGATAAGGTGGAACAC GACGACCACCACAGATCGTTGCGTTTTTAATGCACAGCATTTTTGATAGAGGTAGGATCGGGG	

Line#	Sequence (In red: HA motif, mutations in non-capital letters)	Size (bp)
>HA22	TTTCACTACAAAACGAGTCTTCAAATTCTGCACTTGCTCTGCTTTTACTCGCGCCGAAGCCGATAAATATAAAAAGCAGCTTTTAATTAAGGCAATGAAAAGCTAAGCCAAGAAGCTGAGG AAAAGTGGACTGGCGTGTTTTTTGGCTTTGCTTGTCTCAGCGAGGGTTGCCATTGAGCTTGGGAAAATCGCAATAAATCATCTGGAGCTCGAAGTTCCCGTGGCTACAGATTATCTTTT CTTTTTAAACATAATCCAAATTGGTTCAAGCCCTGTATGTTGCCGGAGTTTAAAGCTTCTGGATTGTTGACATTGAAATTTTTTTTTTATCTAAGCTATTGTCTGCTGGC	349
>HAmut24	TTTCACTACAAAACGAGTCTTCAAATTCTGCACTTGCTCTGCTTTTACTCGCGCCGAAGCCGATAAATATAAAAAGCAGCTTTTAATTAAGGCAATGAAAAGCTAAGCCAAGAAGCTGAGG AAAAGTGGACTGGCGTGTTTTTTGGCTTTGCTTGTCTCAGCGAGGGTTGCCATTGAGCTTGGGAAAATCGCAAgAAAgCAgCTGGAGCTCGAAGTTCCCGTGGCTACAGATTATCTTTT CTTTTTAAACATAATCCAAATTGGTTCAAGCCCTGTATGTTGCCGGAGTTTAAAGCTTCTGGATTGTTGACATTGAAATTTTTTTTTTATCTAAGCTATTGTCTGCTGGC	
>HA26	AATAAGCGATATTTCTCGAGGGATGAATGTGCTTTGGTGCGTGTTTTATTTTATTTCGTTGAGTTACAATCCACCATCTAACGGTACACGCTTTTCCCTTCCGCCCCCTCGAAGGGTGGT GCTCGCCGGTTCAGTGGGCGGGGTTGTTTTGGATGGGGTGGTGCTGTGCGCATGGCCGCCACCGAACGGAGGTGCAAGGGAAAATGTCGAATGCGAAATTAAGAAAACATGGCAG TAAAATTGGGCAATAAATCATTGGGGGGCCTCAAAGTGTTACTTCCACAATGTACACAGTTGGAATTTGTTTAGCTACAAATTGTATAGATTTTGCAACACACTGCGACATATGCCGTTGG AGGAAATGGCGTGGAAAACTGATTGGCTTTATTTATAACCGGGAAAAATTGACTTTGCGGGAAAATGCGGTTCAAACGGCGGTAGTTAATTAAAATCAAAATAATTCATTGTATTTTTTTAC AGGTGCTAAAAATGTTATATATGTTGTTGTTTTTTTTTTTAGCTTAACAAGGACTTTTGGCCTTCTCCATTCACAGCTTGTAAGTCTTGTGGCTCCACTGGTCAATCGAAAGGCAATCCCCG TAATCCCATATCGCTCACAGCTTTCCCATCCTCCGATGCGTGTGTGAGTGTGATTAAAATGGCGATATGAATGCATAATAATACGAAAGTTATGA	695
>HAmut28	AATAAGCGATATTTCTCGAGGGATGAATGTGCTTTGGTGCGTGTTTTATTTTATTTCGTTGAGTTACAATCCACCATCTAACGGTACACGCTTTTCCCTTCCGCCCCCTCGAAGGGTGGT GCTCGCCGGTTCAGTGGGCGGGGTTGTTTTGGATGGGGTGGTGCTGTGCGCATGGCCGCCACCGAACGGAGGTGCAAGGGAAAATGTCGAATGCGAAATTAAGAAAACATGGCAG TAAAATTGGGCAAgAAAgCAgTGGGGGGCCTCAAAGTGTTACTTCCACAATGTACACAGTTGGAATTTGTTTAGCTACAAATTGTATAGATTTTGCAACACACTGCGACATATGCCGTTGG AGGAAATGGCGTGGAAAACTGATTGGCTTTATTTATAACCGGGAAAAATTGACTTTGCGGGAAAATGCGGTTCAAACGGCGGTAGTTAATTAAAATCAAAATAATTCATTGTATTTTTTTAC AGGTGCTAAAAATGTTATATATGTTGTTGTTTTTTTTTTTAGCTTAACAAGGACTTTTGGCCTTCTCCATTCACAGCTTGTAAGTCTTGTGGCTCCACTGGTCAATCGAAAGGCAATCCCCG TAATCCCATATCGCTCACAGCTTTCCCATCCTCCGATGCGTGTGTGAGTGTGATTAAAATGGCGATATGAATGCATAATAATACGAAAGTTATGA	
>HA30	CGAAGACGAAATCAATAAAACCGGCAAGAGTCCACGTGTTCTGGCCACGAGCGGTTCGCATTAACGGATTTTCACTGCACCCGAAAGGAGGCAGCCCCTAGCATTTATAAAATCCAAAT GGAAATGACATTTATTAGAAAACCGATGGCTGATCAATGGATTAGATATCACTGTCATTATTACTCCAATTATACGGCCACATTGAGCGTTGATTAAAGCTGCCATTCGACGAAAGAATAAT GATGGAAAAATATTGCCTTAAATACGATACCCACTTGACAGATTATGCATATATTATGAATCATATTAAATTAATTTTAATCATAATGCTTACATAAGGCAGCGTGTGATTTGATTTATTACCA GCGCATAAGGTTTTAATTAACTTGATTTTGCTTCGTCGATGTTATGCATGGCGGAATTTTACAAACCATAATTAAATTTCAAAACAAAAACGGAACAGTATCCCTGCAATGCATTTATTTGAA TAATTTGCTTGCAGTTATGTTTACTAGGCAATTACATATTGTAATTAAATGTATTTGCATAATTTACCCAACGTAGCACAATTAATTTTACATAGCTAATAAAATATGTTGGTTTTAATCAGATT GCA	616
>HAmut32	CGAAGACGAAATCAATAAAACCGGCAAGAGTCCACGTGTTCTGGCCACGAGCGGTTCGCATTAACGGATTTTCACTGCACCCGAAAGGAGGCAGCCCCTAGCATTTATAAAATCCAAAT GGAAATGACATTTATTAGAAAACCGATGGCTGATCAATGGATTAGATATCACTGTCATTATTACTCCAATTATACGGCCACATTGAGCGTTGATTAAAGCTGCCATTCGACGAAAGAATAAT GATGGAAAAATATTGCCTTAAATACGATACCCACTTGACAGATTATGCATATATTATGAATCATATTAAATTAATTTTAATCATAATGCTTACATAAGGCAGCGTGTGATTTGcTTTcTTcCCAG CGCATAAGGTTTTAATTAACTTGATTTTGCTTCGTCGATGTTATGCATGGCGGAATTTTACAAACCATAATTAAATTTCAAAACAAAAACGGAACAGTATCCCTGCAATGCATTTATTTGAAT AATTTGCTTGCAGTTATGTTTACTAGGCAATTACATATTGTAATTAAATGTATTTGCATAATTTACCCAACGTAGCACAATTAATTTTACATAGCTAATAAAATATGTTGGTTTTAATCAGATTG CA	
> HA34	TGAATGAACCAATGTTGAagaagcagcacaagcagcagcagcaacaacaacacagcggcagcTAGGAGAGCGAAAGAGAGCGAAGGCAAACAAACGAAAATACAATAATAAAAAAGAAGGAAAAACAAC AAAAGGTAAGTGCGTGTAATGCGTGTACGTGTGTGTGTGTGTATGGGTCTGGAGGCCGCACCTCCCAATCAGAAAATTAGTGCTGTGTGGGTGCGCGTCACACTTATTTGTACTTATTT TTTCCGGCTTTAACTGTTATTATTATTTAAGTAGCACGAACGAGAGAATTACATAATTTCGGTATTTTTTTTGTGTCTCTCTTTTGAGACTCCGAATAGAAAGCGAAAGTATTCTTATGGGAA CACACTAGCGTGTGCGTGTGTGTTTGTTTTATTTATGAATGGTGATTATGAGACTTGATAATTAATTGCCAGCATGCGTTGTGATTTGATTTCTGTGCGTACACCAGTTGAGCGGCTGTCT GTCTCTTTTGCACCCACAGCTGGTCATAAATCAAATAAGAATTTCACGCCACACAAATGAAAAGCAACAAACAAAAGAGACGAAGCGATGAACTCCCTCGCGCTCTGATCAGAACGGGAA GAAGAAGCAGCCGCATTTGCAGCTGCGCCGAGCGATCATTTTCGCATACTTCTTTCTTAATTTGTTTTTGCCGAGCCATTTTTAATGTGAAAATCTTGCGGGCGAAAATTGGTCGCCGCA GTTGGCCAAAAACATAACGCATAAACAACACATTTTATGGGACGCCGGTGGCGGCGCACTTACACAAGAAGAAATT	808
> HAmut36	TGAATGAACCAATGTTGAagaagcagcacaagcagcagcagcaacaacaacacagcggcagcTAGGAGAGCGAAAGAGAGCGAAGGCAAACAAACGAAAATACAATAATAAAAAAGAAGGAAAAACAAC AAAAGGTAAGTGCGTGTAATGCGTGTACGTGTGTGTGTGTGTATGGGTCTGGAGGCCGCACCTCCCAATCAGAAAATTAGTGCTGTGTGGGTGCGCGTCACACTTATTTGTACTTATTT TTTCCGGCTTTAACTGTTATTATTATTTAAGTAGCACGAACGAGAGAATTACATAATTTCGGTATTTTTTTTGTGTCTCTCTTTTGAGACTCCGAATAGAAAGCGAAAGTATTCTTATGGGAA CACACTAGCGTGTGCGTGTGTGTTTGTTTTATTTATGAATGGTGATTATGAGACTTGATAATTAATTGCCAGCATGCGTTGTGATTTGATTTCTGTGCGTACACCAGTTGAGCGGCTGTCT GTCTCTTTTGCACCCACAGCTGGaCAaAAAaCAAATAAGAATTTCACGCCACACAAATGAAAAGCAACAAACAAAAGAGACGAAGCGATGAACTCCCTCGCGCTCTGATCAGAACGGGAA GAAGAAGCAGCCGCATTTGCAGCTGCGCCGAGCGATCATTTTCGCATACTTCTTTCTTAATTTGTTTTTGCCGAGCCATTTTTAATGTGAAAATCTTGCGGGCGAAAATTGGTCGCCGCA GTTGGCCAAAAACATAACGCATAAACAACACATTTTATGGGACGCCGGTGGCGGCGCACTTACACAAGAAGAAATT	
> HA38	CGATTAGAAAGCTAAGTGCAATTTTGAAGAGATCATGTAAATGAGTTCTTTTAATTGCAATTTTTTGTACACGAAAACCACCATACCTTCTTATCTGTACTGACCTACATCTAATTCTTCTCAG CTGATAGATAGATACCTCATAGATCATATAAGTCCCAGTACTAATATAACCCACTTCTTGCTCATAAATCATTCGTTTCGAAAATTGAGTAAACGTACACCATGCGTGCGAGCTCATTTACT CCAGTCAGCTCCGCCAGAAATTGATTTATTCT	278
> HAmut40	CGATTAGAAAGCTAAGTGCAATTTTGAAGAGATCATGTAAATGAGTTCTTTTAATTGCAATTTTTTGTACACGAAAACCACCATACCTTCTTATCTGTACTGACCTACATCTAATTCTTCTCAG CTGATAGATAGATACCTCATAGATCATATAAGTCCCAGTACTAATATAACCCACTTCTTGCgCAgAAAgCAggCGTTTCGAAAATTGAGTAAACGTACACCATGCGTGCGAGCTCATTTACT CCAGTCAGCTCCGCCAGAAATTGATTTATTCT	

Line#	Sequence (In red: HA motif, mutations in non-capital letters)	Size (bp)
>HA50	ACAATTTAAAATTTTTGCTGGCACATTTTATTATTACAAATGGCACTTTGGAGAATTGTGTACCACGGATTTTGTAAGCGGCGCTCTAATGATTTATGAATTTATTCAATGACAATTGGCAAG TTGAGCCAAAGTGCTTTTAGCTGGTGAAGCGGACCGGTCATTGAGCCTCGGAACCACTTCATTCGTGGCTGGGAATTGAAATGCACAGAGGAAGAGTTTGAGTTCGCAGTAGTTCACA GTTAGTTCACTGGTGGCCGTCATAGGCGCCAACATATTTTGTTCTTGCATTTAAATTATGAGCTCTGCCATTATGCATCAATCGTTAATCCAGCGGCGGTTAAAAGCAGCACAA	355
>HAmut52	ACAATTTAAAATTTTTGCTGGCACATTTTATTATTACAAATGGCACTTTGGAGAATTGTGTACCACGGATTTTGTAAGCGGCGCTCTccTGcTTTcTGcATTTATTCAATGACAATTGGCAAGT TGAGCCAAAGTGCTTTTAGCTGGTGAAGCGGACCGGTCATTGAGCCTCGGAACCACTTCATTCGTGGCTGGGAATTGAAATGCACAGAGGAAGAGTTTGAGTTCGCAGTAGTTCACAG TTAGTTCACTGGTGGCCGTCATAGGCGCCAACATATTTTGTTCTTGCATTTAAATTATGAGCTCTGCCATTATGCATCAATCGTTAATCCAGCGGCGGTTAAAAGCAGCACAA	
>HA54	TTTCATAATCCCTCTTAGAATCTCCTTGGTACCACGCCCACAAATGTCAGAGGGGGCTGGTGCGTGGGTAAGTGTGTGTGTGTGTCAGTGATGTTCAAGTGCAAAATGTGAATAAAATA CCTGAATTGCTAAGAAATAAGCGGAATTGCATTAAGTGCATCCGCTGCAGTGGGGTAAACCCCCCACAAAAAGGGGGCTAAAAGCCACTGAAATGTGAGGAAATCCGAGGATTTAAGC TGCTTGGTGTGGCCAGCAATTTCAAGGATATGCAGGAGATGGAAATTAGTTTAGTTATGCAGCGATGGCAGAGGATTTGCACTGCAAGGTTAGAATAACGAACAAGAAAGCAGTGCTTA GTCTTTTCTAGAAATCAATGAATTTGAGAAACACGTATAAAAACAGCTGTCTAAAATAATATTATTTTTAGACCTTTAAGTGCACTTTGCTCCGACTTTTGACTAAATTCACCACGTTTACGAG GTCATTGTTTCTCGCAGCTCCTCGTGTGCTTGTGACATCCTGTTGGAtgcgtgtgtgtgtgtgtgtgtgcgttttgtgtctctgtatgcgtgtTGTCCATTTTAATGACAAACATAACCATGAAAACATGAGCATTA ATGGACAGCCATCCATAATGGACATTAACTGCTGGGCGGAGTGTCGGTGGGGGCCAACAATTGGTTGACATTTGAAGTCATGTCTCGGACTAAATTAGACAGTCGATCCCCGAAAACC CCACACTTACTGCACAACCCCCCCCCCCCCCTCTCCACAATAGATTGTATGTAATCGGTAATATATAATATCAATGTATGAGCGCAATTGGTCAACTGAGCCGAGGATTTGTCCGAGTC TCGAGCTGTGGGCATGCATAAAATTGATATCTTAATGGCTGGAATGGGTGTTCCAACCAAAGCATTCAATGGTCAGGTTAATTTGGTGCGTGCCCAACCCACCGTTAAGGGAGTTATCT TTTGGGGGAATGGGTCGCCAACTGCACATGAATCCTAATGGAGTGAGAGCAGTTGGTGGACGGAAGTCAACACAATGGGCTACAGAATATTGGAACATAGCGCTTATCCTCAAAATGG CACTTATAATGCATGAATCGATGAGATTTCATTAAAAGTACTTGGTATTACTTTACTTCTATATACTGTTAATCCCTAAAATCCATTTCTAATGAATATGTTCAATGGAAATGTTGCTATCACTC CACACACGATATTAAAATCCGCTCCAACTAATTCAATCAAGAAACGCTTTTATCTACTCAAACGACACCCAGGAAAATGTAAAAAAAAAAAATCCCTAGAAACGCCATTAAATCGAGCTCAC ATCGACTTCAAAGCTCATTAAAACGATTCTCCAGCGGACAGAAAGGGAGAGAACAAGCGGAAAAGCGGAAAAGTGGGGAGCCAACTGCCCCCAAAAGAATCATAAATCATTGCTGAAAT TTTCATAAATTTAATGAACATTAAGTGCGCCCCAAGGCTTGAGGGTGTGTGTGGAACGCAGGGGACGCGGCGGCAATCGTTACATGACTTCCACTGGACACCAGGACTACCAGGACG GCGAAGGAGGAAACGAGAGGCAAGAAAGGACCCGAAATGGCAGACATCAGTCAAGCGTAATTCAATGTAAAGCGATATACCGATTTCCACACCCACTTCTCACATGCGATTTTTACAGT GCAGGAAAACATCTTGTTATTAATGCTTAAAACCTACGATTTGGATGCCTATGGAAAAGTAATTCGAATTTTTCTTTTAAATAATTATTGCTTGTTTTCTATACTTTAAACCACTATAAATGCTC CAAATGACTATGACTTTTT	1838
>HAmut56	TTTCATAATCCCTCTTAGAATCTCCTTGGTACCACGCCCACAAATGTCAGAGGGGGCTGGTGCGTGGGTAAGTGTGTGTGTGTGTCAGTGATGTTCAAGTGCAAAATGTGAATAAAATA CCTGAATTGCTAAGAAATAAGCGGAATTGCATTAAGTGCATCCGCTGCAGTGGGGTAAACCCCCCACAAAAAGGGGGCTAAAAGCCACTGAAATGTGAGGAAATCCGAGGATTTAAGC TGCTTGGTGTGGCCAGCAATTTCAAGGATATGCAGGAGATGGAAATTAGTTTAGTTATGCAGCGATGGCAGAGGATTTGCACTGCAAGGTTAGAATAACGAACAAGAAAGCAGTGCTTA GTCTTTTCTAGAAATCAATGAATTTGAGAAACACGTATAAAAACAGCTGTCTAAAATAATATTATTTTTAGACCTTTAAGTGCACTTTGCTCCGACTTTTGACTAAATTCACCACGTTTACGAG GTCATTGTTTCTCGCAGCTCCTCGTGTGCTTGTGACATCCTGTTGGAtgcgtgtgtgtgtgtgtgtgtgcgttttgtgtctctgtatgcgtgtTGTCCATTTTAATGACAAACATAACCATGAAAACATGAGCATTA ATGGACAGCCATCCATAATGGACATTAACTGCTGGGCGGAGTGTCGGTGGGGGCCAACAATTGGTTGACATTTGAAGTCATGTCTCGGACTAAATTAGACAGTCGATCCCCGAAAACC CCACACTTACTGCACAACCCCCCCCCCCCCCTCTCCACAATAGATTGTATGTAATCGGTAATATATAATATCAATGTATGAGCGCAATTGGTCAACTGAGCCGAGGATTTGTCCGAGTC TCGAGCTGTGGGCATGCATAAAATTGATATCTTAATGGCTGGAATGGGTGTTCCAACCAAAGCATTCAATGGTCAGGTTAATTTGGTGCGTGCCCAACCCACCGTTAAGGGAGTTATCT TTTGGGGGAATGGGTCGCCAACTGCACATGAATCCTAATGGAGTGAGAGCAGTTGGTGGACGGAAGTCAACACAATGGGCTACAGAATATTGGAACATAGCGCTTATCCTCAAAATGG CACTTATAATGCATGAATCGATGAGATTTCATTAAAAGTACTTGGTATTACTTTACTTCTATATACTGTTAATCCCTAAAATCCATTTCTAATGAATATGTTCAATGGAAATGTTGCTATCACTC CACACACGATATTAAAATCCGCTCCAACTAATTCAATCAAGAAACGCTTTTATCTACTCAAACGACACCCAGGAAAATGTAAAAAAAAAAAATCCCTAGAAACGCCATTAAATCGAGCTCAC ATCGACTTCAAAGCTCATTAAAACGATTCTCCAGCGGACAGAAAGGGAGAGAACAAGCGGAAAAGCGGAAAAGTGGGGAGCCAACTGCCCCCAAAAGAAgCAgAAAgCAggGCTGAAAT TTTCATAAATTTAATGAACATTAAGTGCGCCCCAAGGCTTGAGGGTGTGTGTGGAACGCAGGGGACGCGGCGGCAATCGTTACATGACTTCCACTGGACACCAGGACTACCAGGACG GCGAAGGAGGAAACGAGAGGCAAGAAAGGACCCGAAATGGCAGACATCAGTCAAGCGTAATTCAATGTAAAGCGATATACCGATTTCCACACCCACTTCTCACATGCGATTTTTACAGT GCAGGAAAACATCTTGTTATTAATGCTTAAAACCTACGATTTGGATGCCTATGGAAAAGTAATTCGAATTTTTCTTTTAAATAATTATTGCTTGTTTTCTATACTTTAAACCACTATAAATGCTC CAAATGACTATGACTTTTT	
>HA58	AGCCTTATGTAACGGAGAGTCGCAGGGGCATGCGCACATGGCAACAGTGGATCGTGGGGCCATAAAAATAGAAATCTATACTAAGTACCCACTACGAACGACTCTGACTAATCGCCGT ATTTCTATTTTAGTAGCTAGGGGTCAGGTTGCAGAAGGATTCCATCAAATATTTTACGTCtataaataaataaatatcataaatcatttaaattataaaattatttaaatatattgatatgaataaaaacacattaaatgtggttcataa aatatttaatatttttacatttaaaaataGTTTTAGATGACTTCCAATTTTAGTTTAAATACTTAAATTCCTACTAGTAAAATGATATGACCTTAAGTTTTGGGTCATGCTTTTCGAAATCAGTAAG	395
>HAmut60	AGCCTTATGTAACGGAGAGTCGCAGGGGCATGCGCACATGGCAACAGTGGATCGTGGGGCCATAAAAATAGAAATCTATACTAAGTACCCACTACGAACGACTCTGACTAATCGCCGT ATTTCTATTTTAGTAGCTAGGGGTCAGGTTGCAGAAGGATTCCATCAAATATTTTACGTCtataaataaataaatagcagaaagcaggtaaattataaaattatttaaatatattgatatgaataaaaacacattaaatgtggttcat aaaatatttaatatttttacatttaaaaataGTTTTAGATGACTTCCAATTTTAGTTTAAATACTTAAATTCCTACTAGTAAAATGATATGACCTTAAGTTTTGGGTCATGCTTTTCGAAATCAGTAAG	
>HA62	ACTTAAGGAATTCATGAGTTCGCTGCTCATTCTCTTGGTCTCGAATTCGGTATATTCGTCTATAAGGGTGACTTTTCTGTCATGAAACCGTCAAATGTTTTTATGATCTGCCATATTTAATTC TGAAATAATTTTGGACATTTCACAGCGTTCATAAACGTTGACACTTGGCGAGCAGGAAAAAAAAAAGGCCAATCGGAGGCGAGTGAAAAATAGCAAAACGAAAATACCTCGGCACATATAT TTTTACATAAAACACCTATATACATATATCAATATGGAAAACTTGTGCATCAGGTGGTTTGGATTTTTAAGTATCTGCTAAAGAACATTTAATATTGCCACATAATTAAACGTTTCTTTCAATTC TCCTTGACGTCCATGAAAATCGAGGCCAATAAAAATGTACTTCTTTGTAAACAAAATAGAAAAGCAATCATTTCGAGGTTTTTGAAGCTTAAACTAATCAATGTGCTTAATTAAACGCAAAAG TGCAAATGGAATGTCCCAAAGCTAGGAAAACCTAATTGCATTTTTGCATTTACCTCATAAATCATTCGGTGAACCGTAGACAAAACAAGGAAAATTGCACTTTGAAAATCTCGCG	606
>HAmut64	ACTTAAGGAATTCATGAGTTCGCTGCTCATTCTCTTGGTCTCGAATTCGGTATATTCGTCTATAAGGGTGACTTTTCTGTCATGAAACCGTCAAATGTTTTTATGATCTGCCATATTTAATTC TGAAATAATTTTGGACATTTCACAGCGTTCATAAACGTTGACACTTGGCGAGCAGGAAAAAAAAAAGGCCAATCGGAGGCGAGTGAAAAATAGCAAAACGAAAATACCTCGGCACATATAT TTTTACATAAAACACCTATATACATATATCAATATGGAAAACTTGTGCATCAGGTGGTTTGGATTTTTAAGTATCTGCTAAAGAACATTTAATATTGCCACATAATTAAACGTTTCTTTCAATTC TCCTTGACGTCCATGAAAATCGAGGCCAATAAAAATGTACTTCTTTGTAAACAAAATAGAAAAGCAATCATTTCGAGGTTTTTGAAGCTTAAACTAATCAATGTGCTTAATTAAACGCAAAAG TGCAAATGGAATGTCCCAAAGCTAGGAAAACCTAATTGCATTTTTGCATTTACCgCAgAAAgCAggCGGTGAACCGTAGACAAAACAAGGAAAATTGCACTTTGAAAATCTCGCG	

Line#	Sequence (In red: HA motif, mutations in non-capital letters)	Size (bp)
>HA66	atatgtatatatgtttaaatatatatatatTTTCGATAATTGCTTGACAAATCATTAGTTGCAGATACTTGGGAACGGAACTCACCTGAAACGAGGGCAGATAAAAGCAATGatttattaattataattttttaaaCAAACT ATTGATTTATGCAAAAATAATGAACTGCCCAATTTATGACTCTGGTTGGTCAATTAATCAAATACAATCTATGGGTAATTAAAATTGCACCAATATTAATGCATTTCAAATATTAATAACAATCG AGTGTTAATTAACATTTTTAAACACTTATGAAACTAAATTTTGTTGTGAAATAATTTTTTCATACTCATCAAAATCAAATGTTATATATCTATATTTCTATTTTTCTATCTAACTAATATATTTATTT GCTGCatttaaatattaaaaatatatatgaaaatatCTCTTTATAAAATTGCGATTTTTCTCAAAGACAAGAACT	469
>HAmut68	atatgtatatatgtttaaatatatatatatTTTCGATAATTGCTTGACAAATCATTAGTTGCAGATACTTGGGAACGGAACTCACCTGAAACGAGGGCAGATAAAAGCccTGctttcttcattataattttttaaaCAAACT ATTGATTTATGCAAAAATAATGAACTGCCCAATTTATGACTCTGGTTGGTCAATTAATCAAATACAATCTATGGGTAATTAAAATTGCACCAATATTAATGCATTTCAAATATTAATAACAATCG AGTGTTAATTAACATTTTTAAACACTTATGAAACTAAATTTTGTTGTGAAATAATTTTTTCATACTCATCAAAATCAAATGTTATATATCTATATTTCTATTTTTCTATCTAACTAATATATTTATTT GCTGCatttaaatattaaaaatatatatgaaaatatCTCTTTATAAAATTGCGATTTTTCTCAAAGACAAGAACT	
>HA70	CCGCTTTCCCCATCTCGAAATCACGAACACAAACCGGAGGACAAACTTTTTCCtatttttttttttttttttttGCAAAAATAAGTAAAGTTTGAATTTGGCAAACATAAATTACTTTTGCGGCTCTATTCGC TGTCAGGACCACCTAATACCTTGTTTCAGAAAGGCATAGATCGTTTAACCTATAATGTAGGATGAAATATTAAAAATTCTATATCTAGACATCTCATTTAATAGTACAATTTTTAAATACATTCA ATACGTAAATTGAGAAGGCTTTAAAAACACATAGAACTTATTTCTAATAAATCATTTAAATCTAAATAAATATCTTTAAAGCCACAATTTTTATTAAAATCATCAAAGTAATAAATACTTTCAGCT CAACAGATGTTCTGGATATATTCGGTTCTATATTTACCAGaaaaaatttaataaatattttcaatattttaGAAAACTTAAGTAAATAGTTTTTTTCTGTAAGAAAAAGTCTTGTATGTCGGAAAAAAATG	515
>HAmut72	CCGCTTTCCCCATCTCGAAATCACGAACACAAACCGGAGGACAAACTTTTTCCtatttttttttttttttttttGCAAAAATAAGTAAAGTTTGAATTTGGCAAACATAAATTACTTTTGCGGCTCTATTCGC TGTCAGGACCACCTAATACCTTGTTTCAGAAAGGCATAGATCGTTTAACCTATAATGTAGGATGAAATATTAAAAATTCTATATCTAGACATCTCATTTAATAGTACAATTTTTAAATACATTCA ATACGTAAATTGAGAAGGCTTTAAAAACACATAGAACTTATTTCgAAgAAAgCAggTAAATCTAAATAAATATCTTTAAAGCCACAATTTTTATTAAAATCATCAAAGTAATAAATACTTTCAGCT CAACAGATGTTCTGGATATATTCGGTTCTATATTTACCAGaaaaaatttaataaatattttcaatattttaGAAAACTTAAGTAAATAGTTTTTTTCTGTAAGAAAAAGTCTTGTATGTCGGAAAAAAATG	
>HA74	atatttGGAATTAACGATTATCTAAATATTAAAATTCTATACCGTATATTTTGTATTTCGCATATAAAAAGTGTACCGAAAATTCGCTGATATCAAAACTTTTTCCTGGCAGTTTCGTTTCACTTTG ATTTTTGTTGCGATTTATTGTGTGACAAAACCAAGTTTTATCCCTGACATTTTTTCCTTACATCGACCATTTTTTTTGTTATTGTTGCGATTATTATGATTCCTCAGTTGACTGCCGGCCCTTT TTTTAATAAATCATTTCCGCTGCTTGAGCTGGGAATATTATTTCCGCTTCCGCTCGAAATCGTCCGTTGTGCCCCGTTCCGGCCAGAATTTTGCATGGCTGCCTTTATCCAACAGTTTTCA GTTTTCAACTTTCCCGACTCGGCCACAGCCGGGAAACGAAGAAAGCAGTTGTGAAAGCCAAGCGGAGAGAAACCAAACGGTTTGGAATCATTGAAATAATTGGCAAATCGATTGAAGCT CCAACGACAAAAGGATAGCGCCATGGGGCCATTAGGAGATTAAAAACTACCCAAAGCTAGCAATTAACCAAATGGAGAAAAGGAAATAGCCGAAGTCATCCAGATGCCATACTCATCCC CGATCTATCGGTATCTGATACTTATATCTACAGTATCTTCCATCGTTGGGAAATAAACAACTCGGGCGGAGCTCAGCATGCGGTGGTAAATGGCAGTAAAGCAATAACTGTAACCATTTA TTTTTTGGCAACTACCCTCACACACACGCGCACAGTGGGTGTCCTGCACTAATGTCACCACAAAATAAAAAAAAAGAAACCCGTAAAACGCAATGCGATCTGACCTCAATGAGCGCTGG GATTTGTTCTCCGCTGGCCACAAAAAAGAGATGCGGAAACAGAAAGGATATAAATAGTTCAAGTGGGTTGCCGTCTGCTGGAAATTTAATTCGTTTGTTTGGCAACAATTTCTGAGAAATT CCACCGAAGAATTGGAGCGCCAGCGCGCACTCGAGACAATGGGAAAAAAAAGGGCACGAAAACTCCGACAAGACAATCG	1049
>HAmut76	atatttGGAATTAACGATTATCTAAATATTAAAATTCTATACCGTATATTTTGTATTTCGCATATAAAAAGTGTACCGAAAATTCGCTGATATCAAAACTTTTTCCTGGCAGTTTCGTTTCACTTTG ATTTTTGTTGCGATTTATTGTGTGACAAAACCAAGTTTTATCCCTGACATTTTTTCCTTACATCGACCATTTTTTTTGTTATTGTTGCGATTATTATGATTCCTCAGTTGACTGCCGGCCCTTT TTTgAAgAAAgCAggTCCGCTGCTTGAGCTGGGAATATTATTTCCGCTTCCGCTCGAAATCGTCCGTTGTGCCCCGTTCCGGCCAGAATTTTGCATGGCTGCCTTTATCCAACAGTTTTCA GTTTTCAACTTTCCCGACTCGGCCACAGCCGGGAAACGAAGAAAGCAGTTGTGAAAGCCAAGCGGAGAGAAACCAAACGGTTTGGAATCATTGAAATAATTGGCAAATCGATTGAAGCT CCAACGACAAAAGGATAGCGCCATGGGGCCATTAGGAGATTAAAAACTACCCAAAGCTAGCAATTAACCAAATGGAGAAAAGGAAATAGCCGAAGTCATCCAGATGCCATACTCATCCC CGATCTATCGGTATCTGATACTTATATCTACAGTATCTTCCATCGTTGGGAAATAAACAACTCGGGCGGAGCTCAGCATGCGGTGGTAAATGGCAGTAAAGCAATAACTGTAACCATTTA TTTTTTGGCAACTACCCTCACACACACGCGCACAGTGGGTGTCCTGCACTAATGTCACCACAAAATAAAAAAAAAGAAACCCGTAAAACGCAATGCGATCTGACCTCAATGAGCGCTGG GATTTGTTCTCCGCTGGCCACAAAAAAGAGATGCGGAAACAGAAAGGATATAAATAGTTCAAGTGGGTTGCCGTCTGCTGGAAATTTAATTCGTTTGTTTGGCAACAATTTCTGAGAAATT CCACCGAAGAATTGGAGCGCCAGCGCGCACTCGAGACAATGGGAAAAAAAAGGGCACGAAAACTCCGACAAGACAATCG	

Line#	Sequence (In red: HA motif, mutations in non-capital letters)	Size (bp)
>HA78	AAGATTGCTCTGAAAATTGTTAATAACAATGATAGAAAAAGATGACTAATACTAAGCATTCATCGGTATTTAGGGGAAGTATTAGTTATATTGCCAAACTCACCTTGCCATCATCACAGCTG CTTTGTTTAGTCTTGTGCGATTCATTGGGACCCTGCGGAAACAATTTACCCTGCCAAGTATCCTGTGTGGTGGACTTCCCCGACTTGGTTTTGTTATCGACCAGCGGTTTTTTCTTCACC AAGTGCGGCTCTCCGGGAGCTTGTGGAAACCACTTGCCCTGCCAGCCATCCTTGGTTGTGAACTCCAGCGGAGGAACTTCCTCCGTGGTGCTCTTCATTTTTAATTCAATCGCTTCCT GCTGTTGTCCAGGGGCATGGGGAAACCACTTGCCCTGCCAGCCATTCTTGCCGGGCGTCGCACCTAATGGAACCTGCACACGTTGCACATCATTGGAAATGTCGGGTGGGGACTTC CCAATCTTCCCTAGCTTCTCATCAGGATCAAGAGTTGGCTCCAACAGATTGGCCTGCGACTTCTTGATGGATTTCGCATCGATGAGGGCGAACCCAAGGCACAAGAGGGTCAAGTAAT GCCaaataaacataataattgcaagcaaaagaaaaacaaaagcgcaaaataagcaaacaaaaatataattaagaaCTTGGCAAAACAGCTGTGGAGCTGCGTTTTTCAACACTGGTCAAAGCATGACTTTGTTGCAG GGTAGCAACAGGTTGTTTGGAGTTTGCATAAATCTCCGATGATAATTATGTGTTGTTTTAATaactaaattaaaattatttacgccaaattaaagtaaataaagttaatttaaaatgactacttaaatcaagcagatttaaaAGG AAATCAATGTATATTTGAAATATTCAGACTGGATTCTAAGGTTAAATGACAAATTACTCACCAGAAATCGGCTCATAATAATAAATCATTTTATTTTCAAATGCCCCCTGCGAACGCTCCCCC GCATTGCACACACCAGGGGGTATTTTACGAAGTGCCCACACCTGGCCACACTGGTCGCAGCTACAAAAAGAAGAAGAACAAGGGCCAACTTGATGGTCACTTGCTACTGCGCTTTATT AATTCATTATTTTCCAAAATAATGCACATATATACCCGCCCACAGCCGCTCTGCGCCCCCGCTAATCGAAAAATGTGAGTGGTAAGCGTTTTTACCAGAGTCTTATCAACAATTTAAGCCT CAATTCCCTTCCGTTTCCGTGATTCGCCACCCAGAAAACAATAACAAAGGGGTGTTGTTGTTGCTGTCGCTGCTGCTGCGCTGGAAAAACAACTCGCATAAAATTTTAAAACGCACAAAG AAGGCGAACGTGGAAAGTTAGGCCAACAACCCAGTCGCGATGTCGTGGTTCAATCCCTGGGAT	1421
>HAmut80	AAGATTGCTCTGAAAATTGTTAATAACAATGATAGAAAAAGATGACTAATACTAAGCATTCATCGGTATTTAGGGGAAGTATTAGTTATATTGCCAAACTCACCTTGCCATCATCACAGCTG CTTTGTTTAGTCTTGTGCGATTCATTGGGACCCTGCGGAAACAATTTACCCTGCCAAGTATCCTGTGTGGTGGACTTCCCCGACTTGGTTTTGTTATCGACCAGCGGTTTTTTCTTCACC AAGTGCGGCTCTCCGGGAGCTTGTGGAAACCACTTGCCCTGCCAGCCATCCTTGGTTGTGAACTCCAGCGGAGGAACTTCCTCCGTGGTGCTCTTCATTTTTAATTCAATCGCTTCCT GCTGTTGTCCAGGGGCATGGGGAAACCACTTGCCCTGCCAGCCATTCTTGCCGGGCGTCGCACCTAATGGAACCTGCACACGTTGCACATCATTGGAAATGTCGGGTGGGGACTTC CCAATCTTCCCTAGCTTCTCATCAGGATCAAGAGTTGGCTCCAACAGATTGGCCTGCGACTTCTTGATGGATTTCGCATCGATGAGGGCGAACCCAAGGCACAAGAGGGTCAAGTAAT GCCaaataaacataataattgcaagcaaaagaaaaacaaaagcgcaaaataagcaaacaaaaatataattaagaaCTTGGCAAAACAGCTGTGGAGCTGCGTTTTTCAACACTGGTCAAAGCATGACTTTGTTGCAG GGTAGCAACAGGTTGTTTGGAGTTTGCATAAATCTCCGATGATAATTATGTGTTGTTTTAATaactaaattaaaattatttacgccaaattaaagtaaataaagttaatttaaaatgactacttaaatcaagcagatttaaaAGG AAATCAATGTATATTTGAAATATTCAGACTGGATTCTAAGGTTAAATGACAAATTACTCACCAGAAATCGGCTCATAAgAAgAAAgCAggTTATTTTCAAATGCCCCCTGCGAACGCTCCCCC GCATTGCACACACCAGGGGGTATTTTACGAAGTGCCCACACCTGGCCACACTGGTCGCAGCTACAAAAAGAAGAAGAACAAGGGCCAACTTGATGGTCACTTGCTACTGCGCTTTATT AATTCATTATTTTCCAAAATAATGCACATATATACCCGCCCACAGCCGCTCTGCGCCCCCGCTAATCGAAAAATGTGAGTGGTAAGCGTTTTTACCAGAGTCTTATCAACAATTTAAGCCT CAATTCCCTTCCGTTTCCGTGATTCGCCACCCAGAAAACAATAACAAAGGGGTGTTGTTGTTGCTGTCGCTGCTGCTGCGCTGGAAAAACAACTCGCATAAAATTTTAAAACGCACAAAG AAGGCGAACGTGGAAAGTTAGGCCAACAACCCAGTCGCGATGTCGTGGTTCAATCCCTGGGAT	
>HA82	CGAAAGAAACTCGTAAATTTTTTATATGAACAACTTCAGAATGCGCTTGGATTTCCACTTGAGTTATCAATTTTTGTAGCCCCAATTTTTTGGGAAACGTTGCTCGCAACATTTTACGACAAA GAAAAGTTTTTGTACCCCCTTGATTCCGTCTCATCATCAGTCTGCTAATTTTATTTCGTACGAACACCCGCAAATGTTTTACACTGAGAAAAGAGAGTGTCAAAATAATAGAATTATCTATGA TTTTTTTTTTTTTGGCGAACATACACAGTTTTTTATTAGAGGTTGCAAAAACCATTATACaaataatttcaaaataatttatttttttaggaatttattttattttgttaaatttCTTTACTGTGCAGACTCATGAAAGATTTTC GCAGATGTCGCCGCTGCGACAAGCTCATCAAATGTTTCCAAATCGATTCCATCATTTAAAGCGTATTAGAACAGTCAACAAATATGTGTGTGAGGGAAAAGGAAAGGCGATCAAAATAGA AAGAGAGCAAAGGAAATCCTTGGCAGGATGTGCGACACATTGCGCTTTGAACTTGTCAGGAAATGCTGAAAAAAAGCGCTGAAGAAGTGTCGTAATTTAATTAAGTTTCGCACCAGATG ACAAGCGAAACAAAAGTGGAAATGGTGGAGGAAAAGCCAAAGAAAAGCATAAAAATCAACGGGAAAATTGCATAGAATCGAGTGCAGGATAAAGAGCAATGAAAAGGACATTGATTACA GCCAACCTTTTCGCGATTTCAATGACATTTACTCAATTTAGCAAATTGCCATCGTCATAAATATCAATATTGAGTGCGATATAGCTTGTTAAGGTTGCGGATTAGGAAATGATGGCTTTTTA TATCGAATCGGTTGAAACGGTAATTGAGTAATTAAGTGGAAATCTAAACGGGGCACCATCAAAGTAAACGCTGCTAATTGAGTTTATTGTAACGCATAAAACCTCCTTGATACGTTTAAAT GGGtataaaaatatattttatttattatttGCCTACTCATTTTGCAATATCTCTGTTCTTCCATAAAAATAAGCACATTTATGTGCTTGGAAAATATTTAATTAAAACGTAAACTGCCAGTATGAAACGAAGC CCGCATGACTGTAATCCAATTAATAAATCATTGCGAAGTAATCAATCAATACGAGGTGCAAGCCCAAAGGCCCGCACACAACAATCGACCAGTTAGCATTACAGGACACAATGGAGCACC GGAAAATACAC	1256
>HAmut84	CGAAAGAAACTCGTAAATTTTTTATATGAACAACTTCAGAATGCGCTTGGATTTCCACTTGAGTTATCAATTTTTGTAGCCCCAATTTTTTGGGAAACGTTGCTCGCAACATTTTACGACAAA GAAAAGTTTTTGTACCCCCTTGATTCCGTCTCATCATCAGTCTGCTAATTTTATTTCGTACGAACACCCGCAAATGTTTTACACTGAGAAAAGAGAGTGTCAAAATAATAGAATTATCTATGA TTTTTTTTTTTTTGGCGAACATACACAGTTTTTTATTAGAGGTTGCAAAAACCATTATACaaataatttcaaaataatttatttttttaggaatttattttattttgttaaatttCTTTACTGTGCAGACTCATGAAAGATTTTC GCAGATGTCGCCGCTGCGACAAGCTCATCAAATGTTTCCAAATCGATTCCATCATTTAAAGCGTATTAGAACAGTCAACAAATATGTGTGTGAGGGAAAAGGAAAGGCGATCAAAATAGA AAGAGAGCAAAGGAAATCCTTGGCAGGATGTGCGACACATTGCGCTTTGAACTTGTCAGGAAATGCTGAAAAAAAGCGCTGAAGAAGTGTCGTAATTTAATTAAGTTTCGCACCAGATG ACAAGCGAAACAAAAGTGGAAATGGTGGAGGAAAAGCCAAAGAAAAGCATAAAAATCAACGGGAAAATTGCATAGAATCGAGTGCAGGATAAAGAGCAATGAAAAGGACATTGATTACA GCCAACCTTTTCGCGATTTCAATGACATTTACTCAATTTAGCAAATTGCCATCGTCATAAATATCAATATTGAGTGCGATATAGCTTGTTAAGGTTGCGGATTAGGAAATGATGGCTTTTTA TATCGAATCGGTTGAAACGGTAATTGAGTAATTAAGTGGAAATCTAAACGGGGCACCATCAAAGTAAACGCTGCTAATTGAGTTTATTGTAACGCATAAAACCTCCTTGATACGTTTAAAT GGGtataaaaatatattttatttattatttGCCTACTCATTTTGCAATATCTCTGTTCTTCCATAAAAATAAGCACATTTATGTGCTTGGAAAATATTTAATTAAAACGTAAACTGCCAGTATGAAACGAAGC CCGCATGACTGTAATCCAATgAAgAAAgCAggGCGAAGTAATCAATCAATACGAGGTGCAAGCCCAAAGGCCCGCACACAACAATCGACCAGTTAGCATTACAGGACACAATGGAGCACC GGAAAATACAC	
>HA86	TGGCTGTGACTCAAAGGGCCGGGATATAATCACCTTCTCAGGGAGTGAAGTTTCCGCGAAACGTCTGGCTTTGGAATTTAAGGTGTATTTACAAAGTCACAGAGTAGGAAATGGTTGGA TTTAGCATGAAATCGGCTAGAAGGTCTTTATGAATATTTTATTATCGAACATCAATATTTCAATGATTTATTATTTTTAATTGCTACATTTCTTACACATGCTTTCTTAAAGATTTGATTGATTTG AATTTTACTTAGCTAGGATACCATAACTCAAATCAGTATTTATATATTATATGATTTGACAAGGTCAGATATCTTAGATCTCTAACTCTATACCACCCATCTTTCAGGGAAGGCAAATTTACCC ACATCTTTCGACTGCTTAAATAAGACATCAAAATCATATTGAACATCTTTAGATAGATCACCTTAGATAGAAACAGTCTTTTGTTCCTCTCGATACTGATCGTTCAAATGAAAAAGGATGGAA AAGGATTAAGCGGCTTAGGGGTGTGTGCCCGTATGCATGATCCTGATGATGTTTTGTCATATCTAGCGACTGGCAGTCAAAGGTTAGCCAAACGAAGAACAATGCCCAGAATATTCGGT GGAAAACCACAAATAGCGGTACCCTAACATTTTTCGGCTTCCAGCTGTTTTCCACGGCGGTGATGAAGTTG	682
>HAmut88	TGGCTGTGACTCAAAGGGCCGGGATATAATCACCTTCTCAGGGAGTGAAGTTTCCGCGAAACGTCTGGCTTTGGAATTTAAGGTGTATTTACAAAGTCACAGAGTAGGAAATGGTTGGA TTTAGCATGAAATCGGCTAGAAGGTCTTTATGAATATTTTATTATCGAACATCAATATTTCccTGcTTTcTTcTTTTTAATTGCTACATTTCTTACACATGCTTTCTTAAAGATTTGATTGATTTGA ATTTTACTTAGCTAGGATACCATAACTCAAATCAGTATTTATATATTATATGATTTGACAAGGTCAGATATCTTAGATCTCTAACTCTATACCACCCATCTTTCAGGGAAGGCAAATTTACCCA CATCTTTCGACTGCTTAAATAAGACATCAAAATCATATTGAACATCTTTAGATAGATCACCTTAGATAGAAACAGTCTTTTGTTCCTCTCGATACTGATCGTTCAAATGAAAAAGGATGGAAA AGGATTAAGCGGCTTAGGGGTGTGTGCCCGTATGCATGATCCTGATGATGTTTTGTCATATCTAGCGACTGGCAGTCAAAGGTTAGCCAAACGAAGAACAATGCCCAGAATATTCGGT GGAAAACCACAAATAGCGGTACCCTAACATTTTTCGGCTTCCAGCTGTTTTCCACGGCGGTGATGAAGTTG	

Line#	Sequence (In red: HA motif, mutations in non-capital letters)	Size (bp)
>HA90	AGACTTCCTCTATTTAGTCTGCATTATTAATCCCCGAATGAGATGAAGGTTTTCAAAAACAGATTTTGGGATTAGAGGTGTTCAATTTTGTCTTTTGGCTACTTAGCGGAAATTGGAAAATT ATAAAGTTTAAAAATCAGCAAGGTTATAATAAATATTTAATCTAGTATTTATCCTTTGGATACAAATTTATGTACCATACTTGTACTGAACGAACTATTGAAGTTCGTAATGAAAAAGTCATACT TTCAGAACCGGTTGTGCTCGTTTTCATATTAATAAATCATTGCAACTTGGCAAATATTTGAGTATCACCATAGGTAAACGAATTATAAGAGGAATTTTAATCTCTTGCTTAATCATAAGTTATA TATTTTGTTATATTTGAGTTGTTGTTATACTTCCAACTGGTTACTCATTAATTTAAACAACACGGGCGTGAAATCGGGTGCAAGAGACACGCGACAATTATTAATTATTTAAGGCTACGAATT TTATCAGCCAGATAGGCTCAACGACAAAATTATATGAAAATTTCATGCGATTAATTAGAAAAATAT	560
>HAmut92	AGACTTCCTCTATTTAGTCTGCATTATTAATCCCCGAATGAGATGAAGGTTTTCAAAAACAGATTTTGGGATTAGAGGTGTTCAATTTTGTCTTTTGGCTACTTAGCGGAAATTGGAAAATT ATAAAGTTTAAAAATCAGCAAGGTTATAATAAATATTTAATCTAGTATTTATCCTTTGGATACAAATTTATGTACCATACTTGTACTGAACGAACTATTGAAGTTCGTAATGAAAAAGTCATACT TTCAGAACCGGTTGTGCTCGTTTTCATATgAAgAAAgCAggGCAACTTGGCAAATATTTGAGTATCACCATAGGTAAACGAATTATAAGAGGAATTTTAATCTCTTGCTTAATCATAAGTTATA TATTTTGTTATATTTGAGTTGTTGTTATACTTCCAACTGGTTACTCATTAATTTAAACAACACGGGCGTGAAATCGGGTGCAAGAGACACGCGACAATTATTAATTATTTAAGGCTACGAATT TTATCAGCCAGATAGGCTCAACGACAAAATTATATGAAAATTTCATGCGATTAATTAGAAAAATAT	
>HA98	GCTTATCAAAGTTCAGTTCCCGAGGGCCTTAAAATGTTGTGGACCCGATGAAGTCGATAATATTTTTGGTGATTTGAAAATGAAAAAAAAAACGCTGTTTTTGCCGTTTCTGGTCCTTTGT TTGTCTAGAGACTTTATTGTGTGAAATTTAGGTTTTTATTTGAGCGGTTTTTGTTTTATTTAAATTTACTGAGCTTGCATATGATTAATTAACCGACTCATGTCGCCATCTATCAATGGGGGC TTAAATATAATTTTAATTTCAAAGGCAAGTCAAACAAGTTTTCTTTGTGGTGGTCGGCTGTTTTCGTTGAAGCTTTTTGTTGGATTCGTTTTATTAACAACAATTTCGTTTACTTTTGTTGCTA AGTTTTGGGCTTTCGGCCGAATCTCAGAGGTCGTGATTAATGTTTTTACAAAAGAGTTGTCGCCATTGGTGTGCTGATGTATAAATCTTTTTGCTGCTTTGCCATGAGTGTGAACTTGATT AAGAATCCGTTTAAAAGGAAGTGGTTGCTCATGTATAAGGGTCTAAATATGTTAGTAAAGAATTGTAGGACTTAAACAAAAATAGTCTAGGGAAATGTGCGTTTTCATACAAATTTTGAATAT TTTCCCCCCATTTCTCCCAATTGCATCGGTGTTTTTTCTGTCTGTCACTCGGCCATAAATCATAAAAGTTTTTGCCGGGCAAACAGAACGGAGTTTATGATTTTGTCAGCTGTGTAACATTT CCAGCTGATAAAATCAATAAATTATGAGAAAATAATTTCATTTGATGCCATCTATTTGATTTTGCAAAAATCGAAGCTGGCAA	817
>HAmut100	GCTTATCAAAGTTCAGTTCCCGAGGGCCTTAAAATGTTGTGGACCCGATGAAGTCGATAATATTTTTGGTGATTTGAAAATGAAAAAAAAAACGCTGTTTTTGCCGTTTCTGGTCCTTTGT TTGTCTAGAGACTTTATTGTGTGAAATTTAGGTTTTTATTTGAGCGGTTTTTGTTTTATTTAAATTTACTGAGCTTGCATATGATTAATTAACCGACTCATGTCGCCATCTATCAATGGGGGC TTAAATATAATTTTAATTTCAAAGGCAAGTCAAACAAGTTTTCTTTGTGGTGGTCGGCTGTTTTCGTTGAAGCTTTTTGTTGGATTCGTTTTATTAACAACAATTTCGTTTACTTTTGTTGCTA AGTTTTGGGCTTTCGGCCGAATCTCAGAGGTCGTGATTAATGTTTTTACAAAAGAGTTGTCGCCATTGGTGTGCTGATGTATAAATCTTTTTGCTGCTTTGCCATGAGTGTGAACTTGATT AAGAATCCGTTTAAAAGGAAGTGGTTGCTCATGTATAAGGGTCTAAATATGTTAGTAAAGAATTGTAGGACTTAAACAAAAATAGTCTAGGGAAATGTGCGTTTTCATACAAATTTTGAATAT TTTCCCCCCATTTCTCCCAATTGCATCGGTGTTTTTTCTGTCTGTCACTCGGCCAgAAAgCAgAAAAGTTTTTGCCGGGCAAACAGAACGGAGTTTATGATTTTGTCAGCTGTGTAACATTT CCAGCTGATAAAATCAATAAATTATGAGAAAATAATTTCATTTGATGCCATCTATTTGATTTTGCAAAAATCGAAGCTGGCAA	
>HA102	CGCAGCCCAAGCAAGCTGTAATTAATTTTGCAAAATTTATTCATTTCGACAAACATTTGGCAGACGACAGCCGCAGTCAGTTAACTGCGTTTGCATATGTTTTTGCGGCTTTCATGGCCAT GTCCGAAATGTCGGCCGACACGCAACATGGCCACATCCACATCAAAATCCGCATGTGGGTCACCAGTTATGGTTCGAGGGTGGCTTTTTAAGATGATTTATGGCCCGAGAAATTGGGC AGCATATGCCGGGAATTAGGCTGGCCGAGAATCTGGGAACATGGCTTCAATTTG	293
>HAmut104	CGCAGCCCAAGCAAGCTGTAATTAATTTTGCAAAATTTATTCATTTCGACAAACATTTGGCAGACGACAGCCGCAGTCAGTTAACTGCGTTTGCATATGTTTTTGCGGCTTTCATGGCCAT GTCCGAAATGTCGGCCGACACGCAACATGGCCACATCCACATCAAAATCCGCATGTGGGTCACCAGTTATGGTTCGAGGGTGGCTTTTTAAGcTGcTTTcTGGCCCGAGAAATTGGGCA GCATATGCCGGGAATTAGGCTGGCCGAGAATCTGGGAACATGGCTTCAATTTG	
>HA106	GCGCCTGCGACTGCAGTTTTTCGATCGGATCGATCCTCGGATATATCTATATAGTATATGTATATATACGTGATGACAAATGAGAGTTTTTAAGCCAAGAGGCACGAGGAAGAGGCATG GCAATAGAAAATTAAGTTGAGGACGCATGTTCGTCTATTTTGGACTTGAAAGTTTTCCGCGCATTTGTGGCCATAAATCATAGCCCAAAGAGACGGAGACGGCGGCTGGAAAACGGATC GGAAAATGCTCAGAAAATCAATTACGTATCGCTTTCGGCACATACCGTACGCACACACAATTGAGTTGTCAATCAGATGCTGATTTTTGTCGCCATTTGCTGTCAATCAGCAGCAACAATT GCAACCGGCCATCAATTAtttgttttgactctgttttatttgaattttgtcttggttttcgatcttAGCTATTTGCTGCACGTGCGTCGCAGTTGTTGCATCGCATGTTGCCGGCTCATTAGTTTGGCTTTTATTTTT GGAAAAGCCAACGCGACGCATTTGCCGGCCTGTCTGCCTCTGATTTTCCTCCGATAT	558
>HAmut108	GCGCCTGCGACTGCAGTTTTTCGATCGGATCGATCCTCGGATATATCTATATAGTATATGTATATATACGTGATGACAAATGAGAGTTTTTAAGCCAAGAGGCACGAGGAAGAGGCATG GCAATAGAAAATTAAGTTGAGGACGCATGTTCGTCTATTTTGGACTTGAAAGTTTTCCGCGCATTTGTGGCCAgAAAgCAgAGCCCAAAGAGACGGAGACGGCGGCTGGAAAACGGATC GGAAAATGCTCAGAAAATCAATTACGTATCGCTTTCGGCACATACCGTACGCACACACAATTGAGTTGTCAATCAGATGCTGATTTTTGTCGCCATTTGCTGTCAATCAGCAGCAACAATT GCAACCGGCCATCAATTAtttgttttgactctgttttatttgaattttgtcttggttttcgatcttAGCTATTTGCTGCACGTGCGTCGCAGTTGTTGCATCGCATGTTGCCGGCTCATTAGTTTGGCTTTTATTTTT GGAAAAGCCAACGCGACGCATTTGCCGGCCTGTCTGCCTCTGATTTTCCTCCGATAT	
>HA110	GATTATAAGTGCATATTGGAACTATTATCAGTGGGAAAATATTTATATAAACATATGGCACGTTTTTTAATTAGAATCTGCTTCAGGTTGAGTGTTTTCCTAGCTTGAATTACGACTGCAGTT ATGAGTTCTGAAGAATTTGGGTGGGGTACAAAAACCCAAGAGTTCTGATGAGTTCCGCAAGTGTGAGGTACGAAAAGTCAGGGAAATGTCACAAATATATAAAAAACTAGATAATGATTT ATGGCTGAGTGATCACGGTTTATTTTTATTGGGGACGAGTATCTAGTATTAGTAAGTATAAGTATTAGTGCCAGATAAAAGGGGGCTAATTGCTTTTTCTCTTTTTTAAGAATCTAGATATA TTATTGACAAACACGGATTATTTAAAACTTTGTAGATACAACAAAAAATACAAAACTATAACTTTTCTCTTAATAAACAAAAATTCGTCATTAATTACAAAATATCATATTGGCATAAAAATTCGA GTTTTTTATATAAAAATTCATATCTTCGCCACACTTGTTGCAAGCTGTGAAAATGTGCGTCAAAATTGGGCCTACGCATGACCTTCTCTCAAAGTCAAATGCGGTTGTTCCACTTTACTTGG GCGGTGCGACGTCCCCCCGTCGGAAAATGGGGAAGTCAGCTGTTTGCGACATGGCGCATGCTCCCGTTGAAAACGAAAGTGAAAGTCAATTGTACGCCCAATTGCCACCCCCACACCC	730
>HAmut112	GATTATAAGTGCATATTGGAACTATTATCAGTGGGAAAATATTTATATAAACATATGGCACGTTTTTTAATTAGAATCTGCTTCAGGTTGAGTGTTTTCCTAGCTTGAATTACGACTGCAGTT ATGAGTTCTGAAGAATTTGGGTGGGGTACAAAAACCCAAGAGTTCTGATGAGTTCCGCAAGTGTGAGGTACGAAAAGTCAGGGAAATGTCACAAATATATAAAAAACTAGATAcTGcTTTc TGGCTGAGTGATCACGGTTTATTTTTATTGGGGACGAGTATCTAGTATTAGTAAGTATAAGTATTAGTGCCAGATAAAAGGGGGCTAATTGCTTTTTCTCTTTTTTAAGAATCTAGATATATT ATTGACAAACACGGATTATTTAAAACTTTGTAGATACAACAAAAAATACAAAACTATAACTTTTCTCTTAATAAACAAAAATTCGTCATTAATTACAAAATATCATATTGGCATAAAAATTCGAGT TTTTTATATAAAAATTCATATCTTCGCCACACTTGTTGCAAGCTGTGAAAATGTGCGTCAAAATTGGGCCTACGCATGACCTTCTCTCAAAGTCAAATGCGGTTGTTCCACTTTACTTGGGC GGTGCGACGTCCCCCCGTCGGAAAATGGGGAAGTCAGCTGTTTGCGACATGGCGCATGCTCCCGTTGAAAACGAAAGTGAAAGTCAATTGTACGCCCAATTGCCACCCCCACACCC	
>HA114	TCGGGCAGAACCGGGCAGCCATTCCCATTGCCATCTGCGTCTCACCTGATTGGTGTCCCATTAGGGCAACAAATTAGCTCTTGGCCAGCTCTCTTGGCCATTCCAGGCTTCCGGGCC CTGGGTTATTTACGACCAGGAGATGCAATCCCCGAGCAGATCGAACAGCATTTATGCATTTTGCTTTGCCATAAATCATAATGGCAGTGCAAGTCCCCCACTTTACTACTTTTATGGAACT GTGCACAGTAGCAAAAGTAGTCGTTTTATGACACAAAAAAGTTAAGAACTCAGAACATGAAAAATAATCACTATATTTAAGTTATTTTCATAAGTTCTTGTACATATTATAATAGTAGTTTCTTT TGAGGATAAATAGTTAATAGATACATTTTCTATATTTTAGTAACCTTATTTAATGCCGGATAGACATTGGTTTACTTTGGTATCTTTTAAATACACAAATATTTTTGTAGTACTAGTATGATAAG TACTCTACATTTTTCTGTGTGTAATACGTAGTTAACTGTCTTACTTTCAGCTGGCGGCGAAGGCTTTTGCCTAATTGATTTACAACACTTTCCAACCTTACTCAATGTTGGAAGTGACAAAT GCAGGGCTCTTGTCGCTGGCTTCTCttattattattattattgttgttattactattatCTTTATCTGGCGACTTGCGGGAGATGT	696
>HAmut116	TCGGGCAGAACCGGGCAGCCATTCCCATTGCCATCTGCGTCTCACCTGATTGGTGTCCCATTAGGGCAACAAATTAGCTCTTGGCCAGCTCTCTTGGCCATTCCAGGCTTCCGGGCC CTGGGTTATTTACGACCAGGAGATGCAATCCCCGAGCAGATCGAACAGCATTTATGCATTTTGCTTTGCCAgAAAgCAgAATGGCAGTGCAAGTCCCCCACTTTACTACTTTTATGGAACT GTGCACAGTAGCAAAAGTAGTCGTTTTATGACACAAAAAAGTTAAGAACTCAGAACATGAAAAATAATCACTATATTTAAGTTATTTTCATAAGTTCTTGTACATATTATAATAGTAGTTTCTTT TGAGGATAAATAGTTAATAGATACATTTTCTATATTTTAGTAACCTTATTTAATGCCGGATAGACATTGGTTTACTTTGGTATCTTTTAAATACACAAATATTTTTGTAGTACTAGTATGATAAG TACTCTACATTTTTCTGTGTGTAATACGTAGTTAACTGTCTTACTTTCAGCTGGCGGCGAAGGCTTTTGCCTAATTGATTTACAACACTTTCCAACCTTACTCAATGTTGGAAGTGACAAAT GCAGGGCTCTTGTCGCTGGCTTCTCttattattattattattgttgttattactattatCTTTATCTGGCGACTTGCGGGAGATGT	

Supplementary Table 2: List of primers for *in situ* probes

Target	Pair #	F/RT7	Sequence (5'-3')
svb	MA_1	Fwd	GTTCAGCGTTCTTTTGGCGT
		Rev+T7	GAAATTAATACGACTCACTATAGGGACTTGGTTGGCTTGGCGATA
	MA_2	Fwd	TGCGTTTGTCAAGCCGAAAG
		Rev+T7	GAAATTAATACGACTCACTATAGGGCAGGCATGCATGATACGCAC
	MA_3	Fwd	GGGGGCAGATATCGAAAGGG
		Rev+T7	GAAATTAATACGACTCACTATAGGGTCGAGACCGATAGTGTGGGT
	MA_4	Fwd	GCCTGCGATCTTGATCTCGA
		Rev+T7	GAAATTAATACGACTCACTATAGGGCTCAGAACCGTCCTTTCGCT
	MA_5	Fwd	GACTGCAACAGTTGGCCATG
		Rev+T7	GAAATTAATACGACTCACTATAGGGAGTGCAGCGAAAAAGGCAAG
	MA_6	Fwd	CGAATGCGTGATCGGCATTT
		Rev+T7	GAAATTAATACGACTCACTATAGGGAGTGCAGCGAAAAAGGCAAG
	MA_7	Fwd	CTGCACCCACGACTACAGTT
		Rev+T7	GAAATTAATACGACTCACTATAGGGAACTCGGCGGCAAAGTTTTC
	MA_8	Fwd	GAAAACTTTGCCGCCGAGTT
		Rev+T7	GAAATTAATACGACTCACTATAGGGTATAGAATCGTGGGCGTGGC
	MA_9	Fwd	GCGGTAATCCCTCAGCCTAC
		Rev+T7	GAAATTAATACGACTCACTATAGGGCGGACAGCTGCTCCAGTAAA
	MA_10	Fwd	GTTCGGGTCAGTGTCCCAAT
		Rev+T7	GAAATTAATACGACTCACTATAGGGGCCCCGGACTATATTGTGGG
lacZ	MA_1	Fwd	CGGGTAAACTGGCTCGGATT
		Rev+T7	GAAATTAATACGACTCACTATAGGGCTGTTGACTGTAGCGGCTGA
	MA_2	Fwd	AAAAACAACTGCTGACGCCG
		Rev+T7	GAAATTAATACGACTCACTATAGGGCGGTAGGTTTTCCGGCTGAT
	MA_3	Fwd	GAACTGCCTGAACTACCGCA
		Rev+T7	GAAATTAATACGACTCACTATAGGGGCCAACGCTTATTACCCAGC
	MA_4	Fwd	GGCGGTGATTTTGGCGATAC
		Rev+T7	GAAATTAATACGACTCACTATAGGGGCGTACTGTGAGCCAGAGTT
	MA_5	Fwd	TCACGAGCATCATCCTCTGC
		Rev+T7	GAAATTAATACGACTCACTATAGGGGTGGCCTGATTCATTCCCCA
	MA_6	Fwd	ATGGGTAACAGTCTTGGCGG
		Rev+T7	GAAATTAATACGACTCACTATAGGGAGTGCAGGAGCTCGTTATCG
gfp	MA_1	Fwd	ATCATGGCCGACAAGCAGAA
		Rev+T7	GAAATTAATACGACTCACTATAGGGGGACTGGGTGCTCAGGTAGTG
	MA_2	Fwd	TTCTTCAAGGACGACGGCAA
		Rev+T7	GAAATTAATACGACTCACTATAGGGGCTCGATGTTGTGGCGGATCT
	MA_3	Fwd	AGGAGCGCACCATCTTCTTC
		Rev+T7	GAAATTAATACGACTCACTATAGGGGTTCTTCTGCTTGTCGGCCAT
	MA_4	Fwd	TCTTCTTCAAGGACGACGGC
		Rev+T7	GAAATTAATACGACTCACTATAGGGGTCTCGTTGGGGTCTTTGCTC
	MA_5	Fwd	ATCATGGCCGACAAGCAGAA
		Rev+T7	GAAATTAATACGACTCACTATAGGGGAACTCCAGCAGGACCATGTG
	MA_6	Fwd	ATGGCCGACAAGCAGAAGAA
		Rev+T7	GAAATTAATACGACTCACTATAGGGGTCTCGTTGGGGTCTTTGCTC
	MA_7	Fwd	TCAAGGAGGACGGCAACATC
		Rev+T7	GAAATTAATACGACTCACTATAGGGGTCTCGTTGGGGTCTTTGCTC
	MA_8	Fwd	GAGCTGAAGGGCATCGACTT
		Rev+T7	GAAATTAATACGACTCACTATAGGGGAACTCCAGCAGGACCATGTG

Target	Pair #	F/RT7	Sequence (5'-3')
atub67C	MA_1	Fwd	CCGACGGTCATCGATGATGT
		Rev+T7	GAAATTAATACGACTCACTATAGGGGAGGAATCCCTGCAAGCTGT
	MA_2	Fwd	CAGCCTGAAGACCAAGGAGG
		Rev+T7	GAAATTAATACGACTCACTATAGGGACATCATCGATGACCGTCGG
	MA_3	Fwd	ATGGAAAACAAGTGCCACGC
		Rev+T7	GAAATTAATACGACTCACTATAGGGTGAGGAATCCCTGCAAGCTG
	MA_4	Fwd	CCCCAGAATCCACTTTCCCC
		Rev+T7	GAAATTAATACGACTCACTATAGGGCCTCGAGGTCTTGGCCAAAT
	MA_5	Fwd	CACTCCACCATGGACCACTC
		Rev+T7	GAAATTAATACGACTCACTATAGGGGGGGAAAGTGGATTCTGGGG
	MA_6	Fwd	CCCCAGAATCCACTTTCCCC
		Rev+T7	GAAATTAATACGACTCACTATAGGGAAAGCATGCAACAGGCCATG
	MA_7	Fwd	CAGCCTGAAGACCAAGGAGG
		Rev+T7	GAAATTAATACGACTCACTATAGGGAGGAATCCCTGCAAGCTGTC
	MA_8	Fwd	ATGGAAAACAAGTGCCACGC
		Rev+T7	GAAATTAATACGACTCACTATAGGGCACCTCCTTGCCAATGGAGT
	MA_9	Fwd	CATGGCCTGTTGCATGCTTT
		Rev+T7	GAAATTAATACGACTCACTATAGGGCCTCGAGGTCTTGGCCAAAT
	MA_10	Fwd	AGGTGTGGGAAGTGAACATGA
		Rev+T7	GAAATTAATACGACTCACTATAGGGAGGTTAGGTTAGGTCCTCCCTT
	MA_11-13	Fwd	GAAATTAATACGACTCACTATAGGGAGGTTAGGTCCTCCCTTTGT
		Rev+T7	GAAATTAATACGACTCACTATAGGGGCGCCATTGTTTAAGGTTAGGT
		Rev+T7	GGTGTGGGAAGTGAACATGAA
		Rev+T7	GAAATTAATACGACTCACTATAGGGAGGCGCCATTGTTTAAGGTT
act5c	MA_1	Fwd	CGGTATCGTTCTGGACTCCG
		Rev+T7	GAAATTAATACGACTCACTATAGGGGCGGTGGTGGTGAAAGAGTA
	MA_2	Fwd	CGGTATCGTTCTGGACTCCG
		Rev+T7	GAAATTAATACGACTCACTATAGGGGGCCATCTCCTGCTCAAAGT
	MA_3	Fwd	AAGTACCCCATTGAGCACGG
		Rev+T7	GAAATTAATACGACTCACTATAGGGCGGAGTCCAGAACGATACCG
	MA_4	Fwd	CGGTATCGTTCTGGACTCCG
		Rev+T7	GAAATTAATACGACTCACTATAGGGAGGGCAACATAGCACAGCTT
	MA_5	Fwd	TACTCTTTCACCACCACCGC
		Rev+T7	GAAATTAATACGACTCACTATAGGGAGCCTCCATTCCCAAGAACG
	MA_6	Fwd	AGGCCAACCGTGAGAAGATG
		Rev+T7	GAAATTAATACGACTCACTATAGGGGGCCATCTCCTGCTCAAAGT
	MA_7	Fwd	GGCCAACCGTGAGAAGATGA
		Rev+T7	GAAATTAATACGACTCACTATAGGGGCGGTGGTGGTGAAAGAGTA
	MA_8	Fwd	AACACACCCGCCATGTATGT
		Rev+T7	GAAATTAATACGACTCACTATAGGGGCGGTGGTGGTGAAAGAGTA
	MA_9	Fwd	AACACACCCGCCATGTATGT
		Rev+T7	GAAATTAATACGACTCACTATAGGGTGGCCATCTCCTGCTCAAAG
	MA_10	Fwd	AAGTACCCCATTGAGCACGG
		Rev+T7	GAAATTAATACGACTCACTATAGGGACATACATGGCGGGTGTGTT
	MA_11	Fwd	AAGCTGTGCTATGTTGCCCT
		Rev+T7	GAAATTAATACGACTCACTATAGGGATTCCCAAGAACGAGGGCTG
	MA_12	Fwd	AGAAGCTTCTCCCCTCCCTT
		Rev+T7	GAAATTAATACGACTCACTATAGGGTTGCGTGGTTTCCTTGGACT
	MA_13	Fwd	TGCTCTCTTTGCGGCTTTCT
		Rev+T7	GAAATTAATACGACTCACTATAGGGTTGCGTGGTTTCCTTGGACT
	MA_14	Fwd	AGAAGCTTCTCCCCTCCCTT
		Rev+T7	GAAATTAATACGACTCACTATAGGGACTGCGCTGGTCTACCAAAG
	MA_15	Fwd	CCGGCATTCGTTAAGTGTCG
		Rev+T7	GAAATTAATACGACTCACTATAGGGTTGCGTGGTTTCCTTGGACT
	MA_16	Fwd	GCACCGGCATTCGTTAAGTG
		Rev+T7	GAAATTAATACGACTCACTATAGGGCTGCGCTGGTCTACCAAAGT
	MA_17	Fwd	GCTTCTCCCCTCCCTTTTCG
		Rev+T7	GAAATTAATACGACTCACTATAGGGTTGCGTGGTTTCCTTGGACT
	MA_18	Fwd	TGCTCTCTTTGCGGCTTTCT
		Rev+T7	GAAATTAATACGACTCACTATAGGGGACTGCGCTGGTCTACCAAA

8 References

Abolaji, A.O., Fasae, K.D., Iwezor, C.E., Aschner, M., and Farombi, E.O. (2020). Curcumin attenuates copper-induced oxidative stress and neurotoxicity in Drosophila melanogaster. Toxicol. Reports *7*, 261–268.

Alfred, J., and Baldwin, I.T. (2015). New opportunities at the wild frontier. Elife *4*, 1–4.

Alves, M.R. (2020). The Natural Fallacy in a Post-Truth era. EMBO Rep. *21*, e49859.

Bellaiche, Y., Bandyopadhyay, R., Desplan, C., and Dostatni, N. (1996). Neither the homeodomain nor the activation domain of Bicoid is specifically required for its down-regulation by the Torso receptor tyrosine kinase cascade. Development *122*, 3499–3508.

Bender, W., Akam, M., Karch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B., and Hogness, D.S. (1983). Molecular genetics of the bithorax complex in Drosophila melanogaster. Science (80-.). *221*, 23–29.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. Cell *133*, 1266–1276.

Bhimsaria, D., Rodríguez-Martínez, J.A., Pan, J., Roston, D., Korkmaz, E.N., Cui, Q., Ramanathan, P., and Ansari, A.Z. (2018). Specificity landscapes unmask submaximal binding site preferences of transcription factors. Proc. Natl. Acad. Sci. *115*, E10586–E10595.

Biosa, A., Sanchez-Martinez, A., Filograna, R., Terriente-Felix, A., Alam, S.M., Beltramini, M., Bubacco, L., Bisaglia, M., and Whitworth, A.J. (2018). Superoxide dismutating molecules rescue the toxic effects of PINK1 and parkin loss. Hum. Mol. Genet. *27*, 1618–1629.

Boisclair Lachance, J.-F., Webber, J.L., Hong, L., Dinner, A.R., and Rebay, I. (2018). Cooperative recruitment of Yan via a high-affinity ETS supersite organizes repression to confer specificity and robustness to cardiac cell fate specification. Genes Dev. *32*, 389–401.

Branco, M.R., and Pombo, A. (2006). Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. PLoS Biol. *4*, e138.

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

Briscoe, J., and Thérond, P. (2005). Hedgehog Signaling: From the Drosophila Cuticle to Anti-Cancer Drugs. Dev. Cell *8*, 143–151.

Brookes, M. (2001). Fly: The Unsung Hero of Twentieth-Century Science. (The Ecco Press).

Carnesecchi, J., Pinto, P.B., and Lohmann, I. (2018). Hox transcription factors: an overview of multi-step regulators of gene expression. Int. J. Dev. Biol. *62*, 723–732.

Carnesecchi, J., Sigismondo, G., Domsch, K., Baader, C.E.P., Rafiee, M.-R., Krijgsveld, J., and Lohmann, I. (2020). Multi-level and lineage-specific interactomes of the Hox transcription factor Ubx contribute to its functional specificity. Nat. Commun. 2020 111 *11*, 1–17.

Carnesecchi, J., Boumpas, P., Sanchez, P. van N. y, Domsch, K., Pinto, H.D., Pinto, P.B., and Lohmann, I. (2021). The Hox transcription factor Ultrabithorax binds RNA and regulates co-transcriptional splicing through an interplay with RNA polymerase II. BioRxiv 2021.03.25.434787.

Cavalheiro, G.R., Pollex, T., and Furlong, E.E. (2021). To loop or not to loop: what is the role of TADs in enhancer function and gene regulation? Curr. Opin. Genet. Dev. *67*, 119–129.

Cho, W.-K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.-H., Lavis, L.D., Lionnet, T., et al. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife *5*, e13617.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Choo, S.W., White, R., and Russell, S. (2011). Genome-Wide Analysis of the Binding of the Hox Protein Ultrabithorax and the Hox Cofactor Homothorax in Drosophila. PLoS One *6*, e14778.

Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-darzacq, C., and Hajj, B. (2013). Polymerase II Clustering in. Science (80-.). *245*, 664–667.

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

Crocker, J., Preger-Ben Noon, E., and Stern, D.L. (2016a). The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. Curr. Top. Dev. Biol. *17*, 455–469.

Crocker, J., Noon, E.P., and Stern, D.L. (2016b). The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. - PubMed - NCBI. *117*, 455–469.

Crow, J.F., and Bender, W. (2004). Edward B. Lewis (1918-2004). Genetics *168*, 17773–1783.

Curtis, J. (1833). British entomology, being illustrations and descriptions of the genera of insects found in Great Britain and Ireland: containing coloured figures from nature of the most rare and beautiful species, and in many instances of the plants upon which they are f (Privately printed. London, UK.).

Delker, R.K., Ranade, V., Loker, R., Voutev, R., and Mann, R.S. (2019). Low affinity binding sites in an activating CRM mediate negative autoregulation of the Drosophila Hox gene Ultrabithorax. PLOS Genet. *15*, e1008444.

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*.

Edgar, B., Domsch, K., Carnesecchi, J., Disela, V., Friedrich, J., Trost, N., Ermakova, O., Polychronidou, M., and Lohmann, I. The Hox transcription factor Ubx stabilizes lineage commitment by suppressing cellular plasticity in Drosophila.

Ella Preger-Ben Noon, A., Sabarís, G., Ortiz, D.M., Sager, J., Liebowitz, A., Stern, D.L., Preger-Ben Noon, E., and Frankel, N. (2018). Comprehensive Analysis of a *cis*-Regulatory Region Reveals Pleiotropy in Enhancer Function. Cell Rep. *22*, 3021–3031.

Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. Science (80-.). *350*, 325–328.

Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., and Levine, M.S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proc. Natl. Acad. Sci. *113*, 6508–6513.

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*, 490–493.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.

Fuqua, T., Jordan, J., Halavatyi, A., Tischer, C., Richter, K., and Crocker, J. (2021). An open-source semi-automated robotics pipeline for embryo immunohistochemistry. Sci. Reports 2021 111 *11*, 1–16.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science (80-.). *361*, 1341–1345.

Garcia-Fernàndez, J. (2005). The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. *6*, 881–892.

Gaudet, J., and Mango, S.E. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science *295*, 821–825.

Gemkow, M.J., Verveer, P.J., and Arndt-Jovin, D.J. (1998). Homologous association of the Bithorax-Complex during embryogenesis: consequences for transvection in Drosophila melanogaster. Development.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96–100.

Gilbert, S.F. (2000a). Developmental Biology, Chapter 7, Fertilization: Beginning a new organism. (Sunderland (MA): Sinauer Associates).

Gilbert, S.F. (2000b). Developmental Biology, The Circle of Life: The Stages of Animal Development (Sunderland (MA): Sinauer Associates).

Goodman, F.R. (2002). Limb malformations and the human HOX genes. Am. J. Med. Genet. *112*, 256–265.

Gould, A., Morrison, A., Sproat, G., White, R.A.H., and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: Two mechanisms for specifying overlapping Hox expression patterns. Genes Dev. *11*, 900–913.

Hersh, B.M., Carroll, S.B., and Carroll, S.B. (2005). Direct regulation of knot gene expression by Ultrabithorax and the evolution of *cis*-regulatory elements in Drosophila. Development *132*, 1567–1577.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*.

Hsiao, H.-C., Gonzalez, K.L., Jr., D.J.C., Jordy, K.E., Matthews, K.S., and Bondos, S.E. (2014). The Intrinsically Disordered Regions of the Drosophila melanogaster Hox Protein Ultrabithorax Select Interacting Proteins Based on Partner Topology. PLoS One *9*, e108217.

Hueber, S.D., and Lohmann, I. (2008). Shaping segments: Hox gene function in the genomic age. BioEssays *30*, 965–979.

Janody, F., Sturny, R., Schaeffer, V., Azou, Y., and Dostatni, N. (2001). Two distinct domains of Bicoid mediate its transcriptional downregulation by the Torso pathway. Development *128*, 2281–2290.

Johnston, R.J., and Desplan, C. (2014). Interchromosomal communication coordinates intrinsically stochastic expression between alleles. Science (80-.). *343*, 661–665.

Kaufman, T.C., Seeger, M.A., and Olsen, G. (1990). Molecular and genetic organization of the antennapedia gene complex of drosophila melanogaster. Adv. Genet. *27*, 309–362.

Kittelmann, S., Preger-Ben Noon, E., McGregor, A.P., and Frankel, N. (2021). A complex gene regulatory architecture underlies the development and evolution of cuticle morphology in Drosophila. Curr. Opin. Genet. Dev. *69*, 21–27.

Kribelbauer, J.F., Rastogi, C., Bussemaker, H.J., and Mann, R.S. (2019). Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. Https://Doi.Org/10.1146/Annurev-Cellbio-100617-062719 *35*, 357–379.

Leonelli, S., and Ankeny, R.A. (2013). What makes a model organism? Endeavour *37*, 209–212.

Lewis, E.B. (1963). Genes and Developmental Pathways. Am. Zool. *3*, 33–56.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Lewis I. Held, J. (2017). Deep Homology? Uncanny Similarities of Humans and Flies Uncovered by Evo-Devo (Cambridge University Press).

Lim, B., Heist, T., Levine, M., and Fukaya, T. (2018a). Visualization of Transvection in Living Drosophila Embryos. Mol. Cell *0*.

Lim, B., Heist, T., Levine, M., and Fukaya, T. (2018b). Visualization of Transvection in Living Drosophila Embryos. Mol. Cell.

Liu, Y., Matthews, K.S., and Bondos, S.E. (2008). Multiple Intrinsically Disordered Sequences Alter DNA Binding by the Homeodomain of the Drosophila

Hox Protein Ultrabithorax *. J. Biol. Chem. *283*, 20874–20887.

Liu, Z., Legant, W.R., Chen, B.-C., Li, L., Grimm, J.B., Lavis, L.D., Betzig, E., and Tjian, R. (2014). 3D imaging of Sox2 enhancer clusters in embryonic stem cells. Elife *3*, e04236.

Loker, R., Sanner, J.E., and Mann, R.S. (2021). Cell-type-specific Hox regulatory strategies orchestrate tissue identity. Curr. Biol.

Maass, P.G., Barutcu, A.R., Weiner, C.L., and Rinn, J.L. (2018). Inter-chromosomal Contact Properties in Live-Cell Imaging and in Hi-C. Mol. Cell *69*, 1039-1045.e3.

Maass, P.G., Barutcu, A.R., and Rinn, J.L. (2019). Interchromosomal interactions: A genomic love story of kissing chromosomes. J Cell Biol *218*, 27–38.

Manchester Fly Facility (2017). Why fly? | droso4schools.

Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Chapter 3 Hox Specificity: Unique Roles for Cofactors and Collaborators. Curr. Top. Dev. Biol. *88*, 63–101.

Markow, T.A. (2015). The secret lives of Drosophila flies. Elife *4*, 1–9.

Marta Vicente-Crespo (2015). Can the humble fruit fly help create a flourishing African scientific community? The Guardian.

Martín-Bermudo, M.D., Gebel, L., and Palacios, I.M. (2017). DrosAfrica: Building an African biomedical research community using Drosophila. Semin. Cell Dev. Biol. *70*, 58–64.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283–302.

McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A., and Gehring, W.J. (1984). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. Nat. 1984 3085958 *308*, 428–433.

Merabet, S., Saadaoui, M., Sambrani, N., Hudry, B., Pradel, J., Affolter, M., and Graba, Y. (2007). A unique Extradenticle recruitment mode in the Drosophila Hox protein Ultrabithorax. Proc. Natl. Acad. Sci. *104*, 16946–16951.

Mir, M., Reimer, A., Haines, J.E., Li, X.-Y., Stadler, M., Garcia, H., Eisen, M.B., and Darzacq, X. (2017). Dense Bicoid hubs accentuate binding along the morphogen gradient. Genes Dev. *31*, 1784–1794.

Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nat. 2019 5657740 *565*, 448–453.

Morisaki, T., Müller, W.G., Golob, N., Mazza, D., and McNally, J.G. (2014). Single-molecule analysis of transcription factor binding at transcription sites in live cells. Nat. Commun. *5*, 4456.

Nelson, D., and Cox, M. (2013). Lehninger Principles of Biochemistry (New York: W.H. Freeman).

Nobel Foundation The Nobel Prize in Physiology or Medicine 1933 (NobelPrize.org).

Nobel Foundation The Nobel Prize in Physiology or Medicine 2017 (NobelPrize.org).

Nobel Foundation The Nobel Prize in Physiology or Medicine 1995 (NobelPrize.org).

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

Oyetayo, B.O., Abolaji, A.O., Fasae, K.D., and Aderibigbe, A. (2020). Ameliorative role of diets fortified with Curcumin in a Drosophila melanogaster model of aluminum chloride-induced neurotoxicity. J. Funct. Foods *71*, 104035.

Palacios, I., Vicente-Crespo, M., and Martín-Bermudo, M.D. (2020). The humble fruit fly is helping the African science community to thrive. Nat. Rev. Mol. Cell Biol. 2020 2110 *21*, 558–559.

Pascual-Anaya, J., Kuratani, S., and Garcia-Fernàndez, J. (2013). Evolution of Hox gene clusters in deuterostomes. BMC Dev. Biol. *13*, 26.

Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S., and Aggarwal, A.K. (1999). Structure of a DNA-bound Ultrabithorax–Extradenticle homeodomain complex. Nat. 1999 3976721 *397*, 714–719.

Patel, S., and Prokop, A. (2017). The Manchester Fly Facility: Implementing an objective-driven long-term science communication initiative. Semin. Cell Dev. Biol. *70*, 38–48.

Patel, S., DeMaine, S., Heafield, J., Bianchi, L., and Prokop, A. (2017). The droso4schools project: Long-term scientist-teacher collaborations to promote science communication and education in schools. Semin. Cell Dev. Biol. *70*, 73–84.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. *6*, 893–904.

Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol. *20*, 1562–1567.

Petsko, G.A. (2011). In praise of model organisms. Genome Biol. *12*.

Pfeiffer, B.D., Jenett, A., Hammonds, A.S., Ngo, T.-T.B., Misra, S., Murphy, C., Scully, A., Carlson, J.W., Wan, K.H., Laverty, T.R., et al. (2008). Tools for neuroanatomy and neurogenetics in Drosophila. Proc. Natl. Acad. Sci. *105*, 9715–9720.

Postika, N., Metzler, M., Affolter, M., Müller, M., Schedl, P., Georgiev, P., and Kyrchanova, O. (2018). Boundaries mediate long-distance interactions between enhancers and promoters in the Drosophila Bithorax complex. PLOS Genet. *14*, e1007702.

Prokop, A. (2016). Fruit flies in biological research. Biol. Sci. Rev. *28*, 10–14.

Prokop, A. (2018). Why funding fruit fly research is essential for the biomedical sciences.

Ramos, A.I., and Barolo, S. (2013). Low-affinity transcription factor binding sites

shape morphogen responses and enhancer evolution. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *368*, 20130018.

Rastogi, C., Rube, H.T., Kribelbauer, J.F., Crocker, J., Loker, R.E., Martini, G.D., Laptenko, O., Freed-Pastor, W.A., Prives, C., Stern, D.L., et al. (2018). Accurate and sensitive quantification of protein-DNA binding affinity. *115*.

Reiter, L.T., Potocki, L., Chien, S., Gribskov, M., and Bier, E. (2001). A Systematic Analysis of Human Disease-Associated Gene Sequences In Drosophila melanogaster. Genome Res. *11*, 1114–1125.

Roberts, D.B. (2006). Drosophila melanogaster: the model organism. Entomol. Exp. Appl. *121*, 93–103.

Rodríguez-Carballo, E., Lopez-Delisle, L., Willemin, A., Beccari, L., Gitto, S., Mascrez, B., and Duboule, D. (2020). Chromatin topology and the timing of enhancer function at the HoxD locus. Proc. Natl. Acad. Sci. *117*, 31231–31241.

Rodríguez-Martínez, J.A., Reinke, A.W., Bhimsaria, D., Keating, A.E., and Ansari, A.Z. (2017). Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. Elife *6*, 1–29.

Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. Nat. 2002 4156874 *415*, 914–917.

Ryoo, H.D., and Mann, R.S. (1999). The control of trunk Hox specificity and activity by extradenticle. Genes Dev. *13*, 1704–1716.

Saadaoui, M., Merabet, S., Litim-Mecheri, I., Arbeille, E., Sambrani, N., Damen, W., Brena, C., Pradel, J., and Graba, Y. (2011). Selection of distinct Hox–Extradenticle interaction modes fine-tunes Hox protein activity. Proc. Natl. Acad. Sci. *108*, 2276–2281.

Sagie Brodsky, A., Jana, T., Mittelman, K., Chapal, M., Krishna Kumar, D., Carmi, M., Brodsky, S., and Barkai, N. (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity.

Scardigli, R., Bäumer, N., Gruss, P., Guillemot, F., and Le Roux, I. (2003). Direct and concentration-dependent regulation of the proneural gene Neurogenin2 by Pax6. Development *130*, 3269–3281.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676–682.

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

Scott, M.P., Tamkun, J.W., and Hartzell, G.W. (1989). The structure and function of the homeodomain. Biochim. Biophys. Acta - Rev. Cancer *989*, 25–48.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., et al. (2011). Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. Cell *147*, 1270–1282.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Stern, D.L., and Sucena, E. (2011). Preparation of cuticles from unhatched first-instar Drosophila larvae. Cold Spring Harb. Protoc. *2011*, pdb.prot065532-pdb.prot065532.

Takahashi, Y., Osumi, N., and Patel, N.H. (2000). Body patterning. PNAS 1–2.

Taylor, J.K., Levy, T., Suh, E.R., and Traber, P.G. (1997). Activation of enhancer elements by the homeobox gene Cdx2 is cell line specific. Nucleic Acids Res. *25*, 2293–2300.

Tsai, A., Muthusamy, A.K., Alves, M.R., Lavis, L.D., Singer, R.H., Stern, D.L., and Crocker, J. (2017). Nuclear microenvironments modulate transcription from low-affinity enhancers. Elife *6*, e28975.

Tsai, A., Alves, M.R., and Crocker, J. (2019). Multi-enhancer transcriptional hubs confer phenotypic robustness. Elife *8*.

Tsai, A., Galupa, R., and Crocker, J. (2020). Robust and efficient gene regulation through localized nuclear microenvironments.

UCMP Virtual Museum of Paleontology Metazoa Index.

Viets, K., Sauria, M.E.G., Chernoff, C., Rodriguez Viales, R., Echterling, M., Anderson, C., Tran, S., Dove, A., Goyal, R., Voortman, L., et al. (2019). Characterization of Button Loci that Promote Homologous Chromosome Pairing and Cell-Type-Specific Interchromosomal Gene Regulation. Dev. Cell.

Herewith I declare that I, Mariana Rama Pedro Alves, prepared this Ph.D. thesis:

Interrogating the complex role of Ubx and multi-enhancer transcriptional hubs throughout development

on my own and with no other sources and aids than quoted.