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:

(B.Sc.) Timothy Jessen Fuqua

Born in: Naperville, USA

Oral examination: 17th of September 2021

The Evolutionary Landscape

of a Developmental Enhancer

Referees:

Dr. Steffen Lemke

Dr. Anne Ephrussi

SUMMARY

Enhancers are regulatory DNA sequences that control gene spatial-temporal patterning based on their primary DNA sequence. Through the binding of proteins called Transcription Factors (TFs), enhancers turn genes "on" or "off" across fields of cells to express genes in complex patterns throughout development. To this day, we still cannot accurately and precisely synthesize an enhancer *de-novo* based on our best models. These findings suggest that we still have a limited understanding of how much regulatory information is encoded within the primary sequence of an enhancer. Furthermore, it is thought that enhancers are one of the primary drivers of evolution. Yet, we are far from predicting enhancer evolution due to limited technology and sparse experimental data. In this thesis, I review the field of enhancers, their evolution, and the regulation behind the *shavenbaby* locus. I next highlight the high-throughput technology developed to study enhancer mutants at a higher throughput with the help of a custom liquid-handling robot called *Flyspresso* and an adaptive-feedback confocal microscopy plugin. With this automated pipeline, I carry out a mutational scanning experiment on an enhancer at the *shavenbaby* locus called *E3N* to simulate possible paths and modes of evolution. I find that developmental enhancers are densely encoded and highly pleiotropic. I also identified new TF binding sites and examples of developmental biases that either constrain or drive evolution. I then discuss a mutational hotspot that evolves ectopic expression of *shavenbaby* in the developing wing and haltere, which I hypothesize is due to a transcriptional repressor. I additionally create a gene expression atlas for the late *Drosophila* embryo to map fragile and robust components of the *E3N* expression pattern and identify more TF binding sites. Finally, I summarize this thesis with an updated working model for *E3N* and an explanation to what extent we can predict *E3N's* evolution.

ZUSAMMENFASSUNG

Enhancer sind regulatorische DNA-Sequenzen, die das raumzeitliche Muster eines Genes anhand von DNA kontrollieren. Durch das Binden von Proteinen, sogenannten Transkription-Faktoren, schalten Enhancer Gene über ganze Felder von Zellen an und aus. Wodurch Gene in komplexen Expressionsmustern während der Entwicklung exprimiert werden. Bisher, können wir Enhancer nach wie vor nicht de-novo synthetisieren. Diese Fakten legen nahe, dass wir noch ein begrenztes Verständnis haben wieviele regulatorische Informationen in Enhancer-Sequenzen verschlüsselt sind. Zusätzlich wird angenommen, dass Enhancer die Haupttreiber von Evolution sind, obwohl wir noch weit weg davon entfernt sind, die Evolution der Enhancer vorherzusagen auch wegen begrenzter vorhandener Technologien und minimalen experimentellen Daten. In dieser Doktorarbeit fasse ich das Forschungsfeld von Enhancern, deren Evolution und Regulation des shavenbaby Genlocus zusammen. Als nächstes highlighte ich einen high-throughput Robotor, genannt Flyspresso, den wir speziell zur Untersuchug der Enhacermutanten entwickelten, sowie einen adaptiven feedback Plugin zur Analyse der gewonnenen konfokalen Mikroskopie Daten. Mit dieser Pipeline, führe ich ein mutational-scannings Experiment für einen Enhancer in dem shavenbaby Genlocus, genannt E3N durch, um mögliche Wege und Moden der Evolution, zu simulieren. Ich entdeckte, dass Enhancer dicht mit Informationen kodiert und sehr pleiotropisch sind. Zusätzlich entdeckte ich eine neue Transkriptionsfaktor Bindestelle und andere Beispiele von entwicklungsbiologische Biasen, die die Evolution einschränken oder vorantreiben. Als nächstes diskutiere ich einen mutationalen Hotspot, der ektopische Expression von shavenbaby in den Flügel- und Halterenanlagen der Fliege verursacht, ausgelöst von einem transcriptionalen Repressor, so glaube ich. Zusätzlich erstelle ich einen Atlas für späte Entwicklungsphasen der Drosophila Embryonen, um zerbrechliche und robuste Bestandteile der E3N Expressionsmuster zu identifizieren sowie weitere Transcription-Factor Bindestellen. Abschließend fasse ich diese Doktorarbeit in einem aktualisierten Arbeitsmodell für E3N zusammen und liefere eine Erklärung in welchem Ausmaß wir die E3N Evolution vorhersagen können.

ACKNOWLEDGEMENTS

I would first like to thank Justin Crocker, my advisor, who took a chance on me as one of his first Ph.D. students and made this thesis possible. Justin's incredible mentorship not only taught me how to be a better scientist but also how to be a better mentor myself. Justin has created a *killer* and motivating atmosphere, where laughter is always heard in the hallways of EMBL (even during a global pandemic). Justin taught me how to think bigger and two steps ahead, advice that will continue to shape how I think and do science for the rest of my career.

Next, I would like to thank everyone – past and present - in the Crocker group. Your feedback, support, and collaboration were essential to making these projects possible. I hope that our collaboration will continue throughout our careers, and that we meet again at different conferences or later stages of life. In particular, I would like to thank Kerstin Richter, who is not only an excellent technician but a great friend who I will dearly miss.

Most of this thesis is dependent on *Flyspresso,* the custom syringe-based microplate washer, which was built by Jeff Jordan and assisted by Peter Polidoro and Jon Arnold. I cannot thank Jeff enough for creating this incredible machine and for entrusting it with me to publish its schematics. Jeff went above and beyond for me and even flew to Germany to look at *Flyspresso* when it broke, and was always available to answer the questions I had.

I next would like to thank all of my collaborators that helped me along the way. These collaborators include Christian Tischer and Aliaksandr Halavatyi for spending hours staring at my semi-functional Fiji scripts and sitting in the dark debugging automated microscope codes for me. I would like to thank Chaitanya Rastogi for running NRLB algorithms. I also need to thank Albert Tsai for always having an answer regarding microscopy and linguistics.

I would next like to thank David Stern, who hosted me in his lab at Janelia for the first four months of my Ph.D., paid for reagents, and let us take *Flyspresso* to Germany. It was a joy meeting and working with everyone in the Stern lab, especially Aishwarya, Emily, and Elizabeth.

Next, I need to thank the students Marlize van Breugel and Noa Borst for giving me the chance to mentor them and be their daily advisor. I hope that you both learned something while you

were here in the Crocker lab. You both certainly taught me a lot about Biology and helped me develop an upbeat mentoring style which I plan to use and continue to improve upon throughout my career.

During these short four years, I made many valuable friendships. I would like to thank Jakob Wirbel for being a great friend and roommate. Jakob, you helped me assimilate into German culture, forced me to speak German when I did not want to, checked my statistics, and always pushed me to reach for the next hold in the Boulderhaus. I cannot thank you enough for all you have done for me. Next, I need to shout out to Martine Ballinger, who always managed to stay positive, forced me to speak German, and made me do one more pull-up in the gym. Martine, you made doing a Ph.D. enjoyable. Next, I would like to thank Holly Giles, who planned ahead, listened to all of my drama, and brought a little piece of England to Heidelberg. It would also be a crime not to thank Philipp Walch, who was not only a good friend, but also helped me assimilate into Germany and constantly shuttled me to IKEA, Frankfurt Flughafen, or to the best ramen noodle restaurant in Germany. Philipp, I owe you a frozen margarita.

Finally, I need to thank my parents John and Barb Fuqua for encouraging me to move abroad, sending me macaroni and cheese in the mail, and paying for flights home. I also want to thank my partner, Tobias Zirr, for loving, supporting, and encouraging me to be the best person I can be.

For John and Barb

CONTENTS

TABLE OF FIGURES

1   Introduction

"Evolution of form is very much a matter of teaching very old genes new tricks!"

–Sean B. Carroll, 2006

The cis-regulatory hypothesis states that the evolution of biological form is primarily driven by changes in how a gene[1] is expressed, rather than changes to the coding sequences of genes themselves (Prud'homme et al., 2007; Stern and Orgogozo, 2008). Evolution tinkers[2] with gene expression levels and localization to create new forms, such as thicker beaks (Abzhanov, 2004), wing patterns (Brakefield et al., 1996; Gompel et al., 2005), the number of segments on a centipede or snake (Arthur, 2002; M. Woltering, 2012), and whether a lizard can grow feet (Carroll, 2006; Kvon et al., 2016; Leal and Cohn, 2016).

Cis-regulatory elements encode the levels of gene product expressed and the spatial and temporal components to where and when the gene is expressed (Jindal and Farley, 2021). In this chapter, I will introduce terms and concepts used throughout this thesis: *The Evolutionary Landscape of a Developmental Enhancer.* I will start with the working model of a developmental enhancer, then discuss a supposed grammar for how enhancer sequences relate to the patterns they produce. I then review what we know about enhancer evolvability. Finally, I provide the background information for one of the primary genes for this study: *shavenbaby*, including its evolution and regulation.

---

[1] For all intents and purposes, I am using the ontology that a gene is the coding sequence responsible for creating a protein product. Different definitions of genes are useful for different questions. I use this definition because it helps to differentiate between cis-regulatory sequences and coding sequences.
[2] François Jacob was the first to describe evolution as a "tinkerer" (Jacob, 1977).

1.1   DEVELOPMENTAL ENHANCERS

During development, enhancer sequences are constantly switching genes "on" and "off" across fields of cells in a highly concerted and orchestrated manner, resulting in complex and precise gradients, stripes, or other complex gene expression patterns (Banerji et al., 1981) (reviewed in Long et al., 2016; Wittkopp and Kalay, 2012). Mutations to developmental enhancer sequences contribute primarily to the impressive diversity of form within our Biosphere (Carroll, 2006), and due to their importance in development, are the key for predicting evolution (Stern and Orgogozo, 2008) and understanding disease (Corradin and Scacheri, 2014). The primary genetic sequence of an enhancer encodes a spatial-temporal pattern based on an apparent enhancer grammar. How exactly does this work?

The nucleotide sequence of an enhancer contains many short motifs recognized and bound by regulatory proteins known as Transcription Factors (TFs) (Arnosti and Kulkarni, 2005; Jones and Tjian, 1985). Even though we have a solid foundation about which motifs TFs bind to in enhancers (Stormo et al., 1982), it is still unclear how many binding sites enhancers encode. This problem is because many homologous TFs share similar motifs (Karin, 1990), and the affinity of TF binding sites can vary (Crocker et al., 2015; Farley et al., 2015). Upon binding to an enhancer, TFs either repress or activate transcription of a target gene (Wang et al., 2000), and through the combination and pairing of TFs, an enhancer is able or unable to activate transcription within the cell (reviewed in Long et al., 2016; Wittkopp and Kalay, 2012).

To better illustrate what enhancers are and how they are thought to operate, I will give an example using what we know from *Drosophila melanogaster,* where some of the most canonical developmental enhancers have been characterized (Small and Arnosti, 2020). During early *Drosophila* development, a pair-rule gene known as *even-skipped* (*eve*) is responsible for subdividing the early *Drosophila* embryo into seven stripes across the anterior-posterior axis (Macdonald et al., 1986; Nüsslein-volhard and Wieschaus, 1980; Wieschaus and Nüsslein-Volhard, 2016). See **Figure 1A**.
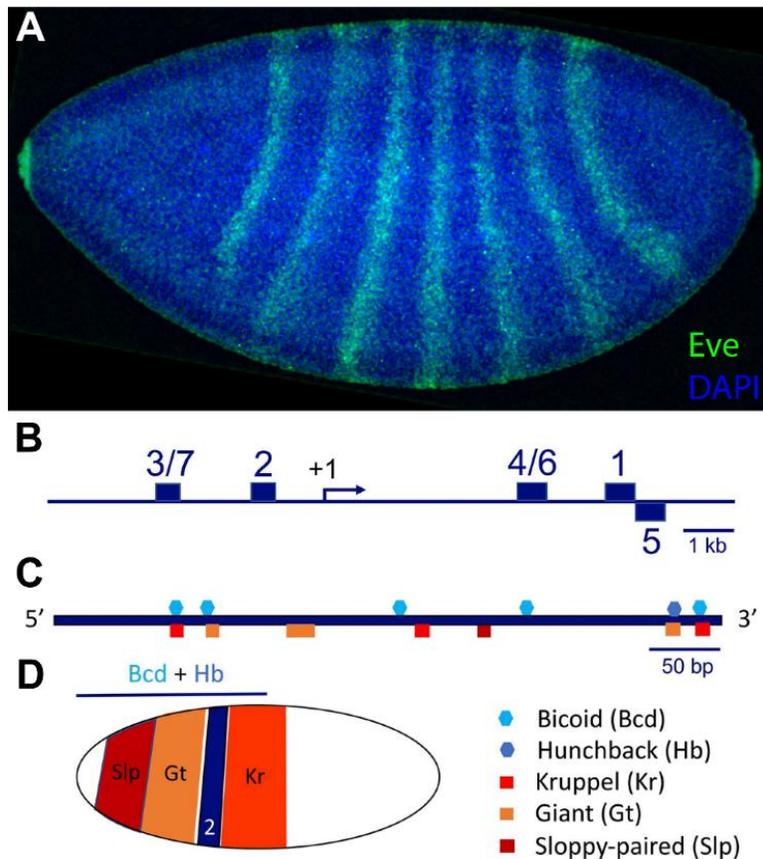
FIGURE 1 THE *EVE* STRIPE 2 ENHANCER

(A) Stage 5 *Drosophila* embryo antibody-stained for Eve (green) and DAPI (blue). (B) Cartoon schematic of the *eve* regulatory locus. Each box indicates the location of each minimalized enhancer element. Arrow indicates the transcription start site. (C) Cartoon schematic of the *eve stripe 2* enhancer and its different Transcription Factor binding sites. (D) Cartoon schematic showing where each Transcription Factor regulating the *eve stripe 2* enhancer is located and the output stripe expression.

Panels **B-D** adapted and modified from Small and Arnosti, 2020. Reproduced with permission License Number: 1108269-1. Panel **A** fixed, stained, and imaged by Timothy Fuqua.

Eve is regulated by five developmental enhancers (Goto et al., 1989; Harding et al., 1989), some of which encode information for multiple stripes (3+7, 4+6), and others for only a single stripe (1, 2, and 5) (**Fig. 1B**). The *eve stripe 2* enhancer has been modeled to be controlled by five TFs: two which activate (Bicoid and Hunchback) and three which repress the pattern (Giant, Kruppel, and Sloppy-paired) (Arnosti et al., 1996; Ludwig et al., 1998) (**Fig. 1C**). The output expression of the

eve stripe 2 enhancer is the sum of where both activators are expressed, and absent either where repressors are bound or where no activators are present. Integration of this information results in a solid and narrow stripe (**Fig. 1D**).

This toy model relating the spatial-temporal patterns of TFs to DNA sequences spurred the notion that enhancers follow a universal and straightforward enhancer grammar (Arnone and Davidson, 1997). However, as biologists characterized more enhancers across different developmental stages and model systems, a master regulatory code, unlike the genetic code for protein sequences, remains to be cracked (Jindal and Farley, 2021). With the current models available today, we are still unable to predict or replicate the *eve stripe 2* enhancer *de-novo* (Vincent et al., 2016). These results ultimately suggest that we are oversimplifying the amount of information and complexity encoded within developmental enhancers.

1.2 ENHANCER GRAMMAR

Many biologists once speculated that one day, we could read a primary enhancer sequence and predict the output expression pattern using an empirically deduced binding site lexicon, gene expression atlases, and grammatical rules (see **Figure 2**) (Arnone and Davidson, 1997; Fowlkes et al., 2008; Thanos and Maniatis, 1995). We have identified through many case studies, evidence, and examples for and against this alleged "enhancer grammar" (Jindal and Farley, 2021). Because we still cannot accurately predict enhancer function, it is still unclear to what extent the primary sequence of an enhancer is even responsible for encoding expression patterns. Furthermore, we still do not know what other components of grammar are missing in our models. Here, I briefly review what we currently know about enhancer grammar and speculate on what could be missing.

Billboards and enhanceosomes

Two of the most well-supported models for enhancers and how they encode regulatory information are the Billboard and Enhanceosome models (Arnosti and Kulkarni, 2005). The two models are not mutually exclusive, and can be thought of as two different ends of a continuous spectrum (Arnosti and Kulkarni, 2005; Spitz and Furlong, 2012). At one end of the spectrum, the Billboard model describes enhancers as regions for Transcription Factors to bind. Like the elements on an (analog) billboard, TF binding sites and regulatory regions can be swapped around without breaking expression. Billboarding explains a phenomenon called binding site turnover (see section 1.3) and can be used to generate enhancer toy models using sigmoidal activation/repression functions (Crocker et al., 2016a; Davis et al., 2020; van Dijk et al., 2017). Billboard models additionally work well for strong activating factors such as Pax6, Gal4, and synthetic TFs to modulate enhancer levels (Crocker and Stern, 2013; Fischer et al., 1988; Giniger and Ptashne, 1988; Sheng et al., 1997).

Enhancer Sequence	TF binding site "lexicon"	Atlases	Grammar	Predicted output:
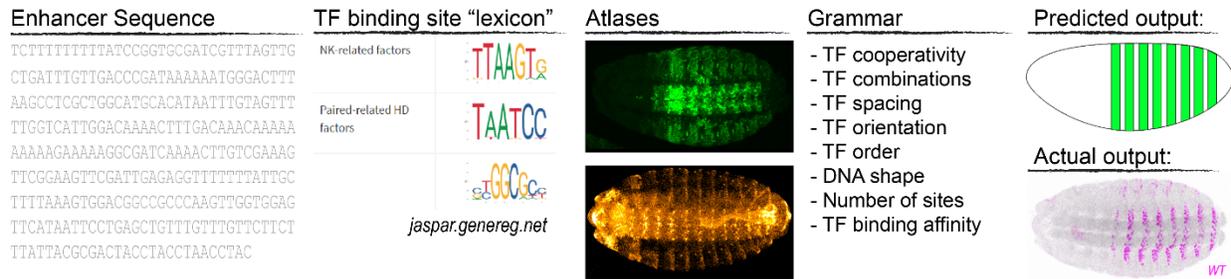
FIGURE 2 PREDICTING ENHANCER FUNCTION

One of the main goals of enhancer biology is to predict how an enhancer DNA sequence correlates with output gene expression patterns. It was once believed that the expression patterns could be easily predicted based off of a binding site "lexicon" for different Transcription Factor (TF) binding sites (jaspar.genereg.net) (Fornes et al., 2019). Binding site motifs could reveal potential regulatory inputs within the DNA sequence. Additionally, if we know the spatial-temporal locations of these TFs with different atlases, we could pair the logic together with a so-called "enhancer grammar" to decode how the TF inputs are interpreted. These models would generate expression patterns that would hopefully match the actual pattern.

At the other end of the spectrum, the Enhanceosome Model focuses primarily on the interactions Transcription Factors have with each other. The enhancer sequence serves as a scaffold for inducing cooperative and steric interactions between Transcription Factors, which form a cooperative complex called an Enhanceosome (Thanos and Maniatis, 1995). This model is based on a single putative Interferon-Beta enhancer. Responsible for eliciting an immune response, the enhancer recruits Transcription Factors into a highly organized and conserved complex (Panne et al., 2007).

The TF collective model merges the Billboard and Enhanceosome models, stating that some enhancer components require TF:TF interactions like an enhanceosome, while others less so (Spitz and Furlong, 2012). Regions within an enhancer and enhancers themselves are likely on a spectrum where some follow billboarding characteristics and others enhanceosome-like characteristics. Regardless of the structure, both models highlight the importance of spacing, DNA shape, TF orientation, binding site number, and affinity. These are terms associated with "enhancer grammar", which I will now describe in further detail.

Cooperative and combinatorial logic

Cooperative pairs of TFs help increase enhancer specificity (Shively et al., 2019). One particularly well-characterized example of cooperative TFs is the proteins Scalloped and Vestigial. The Scalloped TF is responsible for patterning across all of development (Campbell et al., 1992, 1991). To increase the specificity of regulation, Scalloped binds with its co-factor Vestigial to activate expression in the developing *Drosophila* wing (Halder et al., 1998). This cooperative binding has been extensively studied in the developing *Drosophila* wing disc. When combined with the binding of other TFs such as Su(H), Mad/Medea, and Ci, Scalloped can even further precisely pattern within regions of the developing wing (Bray, 1999).

Another example of cooperative TFs contributing to specificity is cofactors for the Homeobox (Hox) genes. The Hox genes are expressed in different segments during development for determining cell identity. Hox genes are homologous and have arisen from a common ancestral gene (de Rosa et al., 1999). The Hox protein binding domain also remained functionally conserved throughout Hox evolution, and all Hox genes bind to a similar motif (Ekker et al., 1994; Krumlauf, 1994; McGinnis and Krumlauf, 1992). So how did the Hox genes subfunctionalize after duplicating to not bind promiscuously to other homeobox sites? This problem is referred to as the Hox Paradox (Prince, 2002). One proposed solution to the Hox Paradox is that each Hox gene recruits a unique binding partner to regulate the proper genes (Mann et al., 2009). Cooperative binding pairs have been identified, such as the cofactors Homothorax and Extradenticle for Ultrabithorax (Jaw et al., 2000; Merabet et al., 2007). The other proposed solution to the Hox Paradox is binding specificity at lower affinity, which I will discuss later in this chapter.

Cooperative interactions are also crucial for activating transcription. Enhancers communicate with the promoter through the use of intermediary proteins. A particularly well-studied intermediate is the Mediator complex, which physically interacts with both the promoter and enhancer (Soutourina, 2018). Enhancer function and specificity are additionally dependent on the spacing between cooperative pairs and partners.

The spacing between binding sites

The spacing between Transcription Factors has proven to be very important for enhancer output. There are at least two reasons why TF spacing is essential for enhancer function. The first reason is that TFs need to be physically close to one another to interact. The *Drosophila Neurogenic Ectoderm Enhancers* (NEEs) are independently evolved enhancers that generate the same expression pattern for non-homologous genes (Crocker et al., 2010; Erives and Levine, 2004). The NEEs are strong evidence for enhancer grammar since the different sequences converged to similar expression patterns using similar enhancer logic. In the NEEs, a particular set of cooperative Transcription Factors: Dorsal and Twist, have a conserved spacing between each other in different *Drosophila* species and different NEEs. Upon changing the spacing of Dorsal and Twist, the width of the NEE stripe expression pattern changes (Crocker and Erives, 2013). These results demonstrate that the spacing between Dorsal and Twist contributes to enhancer function. Another speculated reason for TF spacing is a phenomenon called Helical Phasing, which refers to the orientation of the TF on the double-helix in relation to other bound TFs (Zinkel and Crothers, 1987).

Studies using massively parallel reporter assays (MPRAs) can test tens of thousands of DNA sequences as reporter constructs in cell culture to reveal enhancer grammar and chromatin structure (Kheradpour et al., 2013; Maricque et al., 2019; Melnikov et al., 2012; Mogno et al., 2013). The DNA structure[3] reveals that a single complete turn on the helix is 10-10.5 base pairs (WATSON and CRICK, 1953). MPRA analysis can footprint binding sites in enhancers and test helical phasing between factors by placing 5-mer and 10-mer spacers between enhancer sequences. These insertions puts TFs "in" or "out" of phase (Davis et al., 2020; Melnikov et al., 2012). When both the 5-mer and 10-mer insertions change expression, one can conclude that the spacing between the factors is necessary for expression. When only the 5-mer mutants change expression, but not the 10-mer, the conclusion suggests that the helical phasing of the TFs is necessary for proper expression.

---

[3] The structure of DNA would not have been solved without the work of Rosalind Franklin, who did not receive credit for her contributions until after her death (KLUG, 1968).

For the classic Interferon-Beta Enhanceosome, some regions were found to be affected by both 5 and 10-mer mutants, suggesting the importance of spacing. In the same study, regions within a *c-AMP* enhancer were found to not be sensitive to spacing or helical phasing (Melnikov et al., 2012). However, a follow-up study on the *c-AMP* enhancer, conversely, found a periodicity between activating TFs on the *c-AMP* enhancer, where expression levels oscillated every 10 base pair spacers. The effect size differences between these helically-phased variants were also larger the closer the enhancer was to its promoter (Davis et al., 2020). Changes in effect-sizes to the *c-AMP* enhancer, however, could be ~90% explained with a simple activation function independent of this phasing or spacing. To what extent spacing and helical phasing play a role in enhancer grammar may be case-specific. This study highlights the importance of the spacing between binding sites for helical phasing (enhanceosome model), but that these cooperative interactions may only contribute to a small portion of enhancer grammar compared to simple activation / repression functions (billboard model) (Spitz and Furlong, 2012).

The orientation and order of binding sites

The location where TFs are bound to an enhancer is also an important aspect of enhancer grammar (reviewed in Jindal and Farley, 2021). One clear demonstration of TF binding site order dictating grammar is the *sparkling* enhancer for the *dPax2* gene. In 2011, Swanson and Barolo swapped the binding sites for the *sparkling* enhancer, which subsequently lost expression (Swanson et al., 2011). Additionally, Liu and Posakony swapped binding sites on two different enhancers: *ASE5* and *mα.* The group found no significant change in expression for *ASE5*, but for *mα*, expression was changed (Liu and Posakony, 2012). For the *ASE5* enhancer, these results suggest that binding site order may not be so important and that there is limited to no TF:TF cooperativity, this was further supported when Liu and Posakony also changed the TF binding site spacing in *ASE5* and observed only minimal changes on expression. These experiments suggest that some enhancers work independently of helical phasing, TF cooperativity, spacing, and order (Liu and Posakony, 2012), further supporting the TF collective model (Spitz and Furlong, 2012).

For the individual orientation of a TF binding site, it has been demonstrated that flipping the orientation of a TF binding site can change expression (Passamaneck et al., 2009). I need to be clear here that the terms *orientation* and *order* both insinuate that enhancers have an orientation and are read, for example, from 5' to 3' like a coding sequence. This is simply not the case. The definition of an enhancer is an element that functions independent of its position and orientation (Banerji et al., 1981). Changes to binding site order and orientation are probably classified as grammatical components because their tested phenotypic effects are related to other components, including DNA shape, helical phasing, and TF:TF cooperativity. The issue is that it requires many additional experiments to demonstrate that the effects are caused by DNA shape, helical phasing, and TF cooperativity. Thus, I think *orientation* and *order* are not mechanistic grammatical explanations and are merely umbrella terms for enhancer grammar. In general, let us try to avoid using these terms since they do not directly address the mechanisms of enhancer function. In the future, deep learning models may help us discern different types of grammar (Avsec et al., 2021), which may reveal if orientation and order are fundamental components of enhancer grammar.

DNA Shape

While much attention has been given to the primary structure of DNA (i.e., the sequence itself), how the secondary structure or the physical shape of DNA influences Transcription Factor binding is a large field of study. The shape of DNA itself has distinct features: the Minor Groove Width, Helix Twist, Propeller Twist, and the Roll. DNA sequences can be analyzed on gels to predict shape features (Zinkel and Crothers, 1987). 3D simulations of the DNA can measure the subtle differences of these features (Zhou et al., 2013), which are used to train models capable of predicting Transcription Factor binding independent of the primary sequence, thus reducing the dimensionality of enhancer models themselves (Zhou et al., 2015).

Some proteins can physically change the DNA shape upon binding, such as the bacterial CAP binding protein, which can bend DNA from 90 to 180 degrees (Liu-Johnson et al., 1986). Additionally, binding of the TATA-binding protein (TBP) may be partially dependent on the prior

bending of DNA (Kim et al., 1989; Starr et al., 1995) and is important for forming the Pre-Initiation Complex (Lee and Young, 2000). With transcription factors additionally modifying DNA shape, we may have completely underestimated how much DNA shape controls enhancer logic.

The number of transcription factor binding sites

The number of TF binding sites in an enhancer can control patterning. In *Drosophila,* the maternally deposited TF Bicoid is expressed as a gradient with the highest concentration at the anterior pole, and weakest in the posterior (Driever and Nüsslein-Volhard, 1988). Additionally, the TF Dorsal is expressed along the Dorsal/Ventral axis at the highest concentration in the ventral region (Anderson et al., 1985). Both Bicoid and Dorsal are expressed as morphogen gradients during development (Turing, 1952). Morphogen gradients are classic examples of the French Flag Model (Wolpert, 1969), where the number of binding sites (or affinity) for the morphogen controls where along the developmental axis an enhancer is expressed (Stathopoulos and Levine, 2002).

Fully synthetic enhancer systems have been developed in the early *Drosophila* embryo, where synthetic enhancers respond to synthetic Bicoid gradients. Increasing the number of activator or repressor binding sites within these synthetic enhancers induces corresponding stepwise changes in expression, supporting Wolpert's French Flag model (Wolpert, 1969). Furthermore, the number of synthetic activator binding sites controls the location of expression along the artificial activation gradient (Crocker et al., 2017). For proteins not expressed as morphogen gradients, homotypic clusters (clusters of the same TF binding site) in enhancers can also be used to encode robustness and canalize expression (Crocker et al., 2015; Lifanov, 2003; Payne and Wagner, 2015).

Binding site affinity

Another critical component of cis-regulatory logic is the affinity at which Transcription Factors bind to enhancer sequences (Crocker et al., 2016b). Binding sites can be identified in many different ways, including but not limited to SELEX-Seq (Riley et al., 2014), Chip-Seq (Park, 2009), ATAC-Seq (Li et al., 2019), Electrophoretic Mobility Shift Assays (EMSAs) (Garner and Revzin, 1981), and Protein Binding Microarray Data (Andrilenas et al., 2015). These techniques, while helpful, are biased towards identifying binding sites with the strongest affinity for Transcription Factors, leaving out critical information behind enhancer function (Rastogi et al., 2018). Recent advances in genomics now make it possible to study TF binding at single-cell resolution, giving insights into TF occupancy and how many TFs simultaneously bind to enhancers (Sönmezer et al., 2021).

Low-affinity binding sites play a critical role in the specificity (or lack of specificity) for Transcription Factor binding and total expression levels (Crocker et al., 2015). The *Drosophila E3N* enhancer contains multiple low-affinity binding sites for the Hox factor, Ultrabithorax (Crocker et al., 2015). Changing these binding sites to the highest consensus motif causes ectopic expression associated with the ectopic binding of other Hox proteins (Crocker et al., 2015). The *E3N* enhancer compensates for lower activation levels by encoding homotypic clusters of these low-affinity binding sites to increase the total expression (Crocker et al., 2015; Rastogi et al., 2018). These findings demonstrate that low-affinity Hox binding sites control specificity also provides an explanation to the previously mentioned "Hox-Paradox". Low-affinity binding sites have been identified in other *Drosophila* enhancers (Delker et al., 2019; Parker et al., 2011) and other species to be necessary for proper patterning (Farley et al., 2015). In some cases, they are even conserved in rapidly evolving enhancers, which is the case for a low-affinity Su(H) binding site in the *sparkling* enhancer (Swanson et al., 2011) (see section 1.3).

The affinity of a binding site can also determine whether an activator or a repressor binds to an enhancer. This is the case for the Receptor Tyrosine Kinase (RTK) signaling pathway, where the antagonistic TFs: Pointed (activator) and Yan (repressor) both bind to an ETS binding site. However, it was found that Pointed and Yan have slightly different binding preferences based on

the affinity of the ETS site (O'Neill et al., 1994; Xu et al., 2000). Additionally, many TFs can exist in two forms as either an activator or as a repressor. It has been found that there are binding preference differences between the activator and repressor forms of the Hedgehog signaling pathway Transcription Factor: Cubitus Interruptus (Aza-Blanc et al., 1997). At lower affinities, the repressor form is preferred, and at higher affinities, the activator form (Dessaud et al., 2008; Müller and Basler, 2000; Parker et al., 2011).

Enhancer grammar: summary and open questions

Developmental enhancers have been modeled to create gene expression patterns based on a loose set of parameters which we call "Enhancer Grammar." This grammar includes the encoded cooperativity between Transcription Factors binding sites, the spacing between them, their orientations and order on the DNA sequence, their numbers, their affinity, and the shape of the DNA. Some of this grammar is more thoroughly defined and understood than others, and it is likely that there are still components that we are not considering or underestimating. In their recent enhancer review, Jindal and Farley emphasize, "…that there is a complexity and multidimensionality to enhancers far beyond the linearity of language" (Jindal and Farley, 2021).

I speculate that we are completely underestimating the importance of low-affinity binding sites on enhancer grammar. Additionally, another less understood component of grammar to consider is overlapping and competing TF binding sites (Crocker et al., 2017; Pfeiffer et al., 2008; Stanojevic et al., 1991; Xu et al., 2000). Finally, it is still unclear how much information is encoded within a developmental enhancer (i.e., how many bases are necessary or sufficient to create the expression pattern). I address these questions and provide some answers to them in Chapter 3 of this thesis.

1.3   ENHANCER EVOLVABILITY AND ROBUSTNESS

Binding site turnover

Large-scale genomic studies (Kvon et al., 2014) and case studies mentioned below reveal that many enhancer sequences have undergone rapid sequence turnover. These enhancers lack sequence conservation while maintaining their respective expression patterns. This phenomenon was observed in the *sparkling* eye enhancer of the *Drosophila dPax2* gene (Swanson et al., 2011). An in-silico analysis reveals that the *sparkling* enhancer has lost certain transcriptional regulators throughout its evolutionary course, which were compensated for by the gain of other binding sites. Despite this, the expression patterns across species are highly similar. Swanson and colleagues created chimeras of the *sparkling* enhancers between *Drosophila melanogaster* and *Drosophila pseudoobscura* (**Fig. 3A**). The species-specific enhancers drive identical patterns of expression (**Fig. 3B, C**), but the chimera patterns do not resemble the wild-type patterns (**Fig. 3D, E**) (Swanson et al., 2011).

Similar chimeric experiments were also carried out for the *eve stripe 2* enhancer between *Drosophila melanogaster* and *pseudoobscura* (Ludwig et al., 2000). This study also found that the species-specific enhancers maintain similar expression patterns (**Fig. 3F, G**), but the chimeras change the expression pattern (**Fig. 3H, I**). The chimera studies indicate that the regulatory logic for both enhancers has changed across their sequences over time. However, the fact that the patterns are maintained throughout evolution ultimately suggests that the enhancer expression patterns are under stabilizing selection (Ludwig et al., 1998, 2000). Other speciation studies for *eve stripe 2* also support this claim (Crocker and Erives, 2008; Crocker and Stern, 2017; Martinez et al., 2014).

A *dPax2 sparkling (spa) enhancer*

cone cell expression

D. mel spa		+++
D. pse spa		+++
mel5'+pse3' chimera		−
pse5'+mel3' chimera		++++

B *D. mel spa-GFP*

C *D. pse spa-GFP*

D mel5'+pse3' chimera

E pse5'+mel3' chimera

Adapted from Swanson et al., Current Biology 2011

F *melanogaster (m)*

G *pseudoobscura (p)*

H *p1 - m2 chimera*

I *m1 - p2 chimera*

Adapted from Ludwig et al., Nature 2000

J *Strongylocentrotus purpuratus*

K *Lytechinus variegatus*

Adapted from Romano and Wray, Development 2003

FIGURE 3 ENHANCER CHIMERAS SUGGEST BINDING SITE TURNOVER AND STABILIZING SELECTION

(A) Schematic for the *dPax2 sparkling* enhancer transcription factor binding site compositions for *melanogaster (mel), pseudoobscura (pse),* their chimeras, and cone cell expression. (B-E) Developing eye discs with reporter expression for *melanogaster* (B), *pseudoobscura* (C), *mel-pse* chimera (D), and the *pse-mel* chimera (E). Stage 5 *Drosophila melanogaster* embryos driving reporter expression for the *eve stripe 2*

enhancer from *melanogaster* (**F**), *pseudoobscura* (**G**), *pseudoobscura-melanogaster* chimera (**H**), and the *melanogaster-pseudoobscura* chimera (**I**). Cartoon schematics for sea urchin development, *Endo16* gene expression, and promoter binding site composition in (**J**) *Strongylocentrotus purpuratus* and (**K**) *Lytechinus variegatus.*

Panels **A-E** adapted from Swanson et al., Current Biology 2011. Reproduced with permission License Number: 5038811263810. Panels **F-I** adapted from Ludwig et al., Nature 2020. Reproduced with permission License Number: 5038820063292. Panels **J-K** adapted from Romano and Wray, Development 2003. Reproduced with permission License Number: 1118301-1.

_____

Binding site turnover is a phenomenon extended beyond *Drosophila* enhancers. Another clear example of binding site turnover is in the sea urchin. The *Endo16* gene expression pattern is highly conserved between two species of sea urchin: *Strongylocentrotus purpuratus* and *Lytechinus variegatus*, which diverged from each other over 30 million years ago. A regulatory dissection for the *Endo16* locus in *S. purpuratus* identified 56 different TF binding sites (**Fig. 3J**). However, in *L. variegatus*, only ten of these binding sites remain conserved (**Fig. 3K**). These ten binding sites are insufficient to drive expression in *S. purpuratus*, suggesting that the *L. variegatus* enhancer gained new regulatory logic outside of the conserved region to stabilize expression (Yuh and Davidson, 1996; Yuh et al., 1994, 2001).

Binding site turnover has also been responsible for the evolution of novel enhancer expression patterns and is not limited to stabilizing selection. Maybe one of the most well-characterized examples of this is the gain of a wing spot on *Drosophila biarmipes* (Gompel et al., 2005). This wing spot appeared by gaining a binding site for the TF: Distalless, in an enhancer for the *yellow* gene. Recent dissections of this enhancer reveal that while the gain of a binding site is simple, the enhancer still encodes a dense amount of regulatory information (Le Poul et al., 2020). Other examples include the *shavenbaby E6* enhancer losing expression by gaining a binding site for Abrupt and the loss of multiple Arrowhead sites (Preger-Ben Noon et al., 2016). The *sonic hedgehog ZRS* enhancer also lost Ets1 binding sites for limb patterning in snakes (Kvon et al., 2016; Leal and Cohn, 2016), and the gain and loss of Cortex binding sites is essential for crypsis and mimicry in butterfly wing patterning (Nadeau et al., 2016).

Where do enhancers come from?

So far, I have focused on the evolution of existing enhancer elements. But where do novel enhancers come from? It is suggested that new enhancers can emerge by 1) a duplication event of entire genes or enhancers, 2) spontaneously emerge from the genome, or 3) through the co-option of other elements (reviewed in Long et al., 2016). Below, I briefly review case studies that explore these different forms of emergence.

Duplication events – in proteins or enhancers – can initially create either too much of a product or a temporary redundancy in the genome. When a redundancy emerges, this allows for change with less selective pressure. If a preexisting enhancer is duplicated and creates a temporary redundancy, the duplicated enhancer can either neofunctionalize through the gain or loss of binding sites (again, binding site turnover), or subfunctionalize individual components of the two paralog enhancers. If the duplication creates too much of a product, both paralogs will compensate for the extra expression and lower their activity collectively (stabilizing selection). Furthermore, the most common event following duplication is the loss of function of one of the paralogs. This duplicated and functionless material may be helpful in the spontaneous emergence of new enhancer elements and is called a *proto-enhancer* (reviewed in Long et al., 2016). Despite these mechanisms being well explained (Long et al., 2016), experimental evidence for these modes of evolution is sparse.

The work of Glassford and Rebeiz identified the evolutionary history of an enhancer spontaneously emerging from a proto-enhancer sequence (Glassford and Rebeiz, 2013). Arising from four point mutations, the group tested a combinatorially complete library (16 lines total) between the *Drosophila santomea* and *yakuba,* and traced all possible evolutionary pathways that the common ancestor may have taken. The group found that the enhancer emerged from an initially weak element along many possible evolutionary paths. Furthermore, the enhancer exhibited both cooperative effects and epistasis during its evolution. The study emphasizes how constraint guides the evolution of an enhancer, and how enhancers can spontaneously emerge.

Another option for enhancer genesis is the co-option – or *exaptation* - of transposable elements (Emera et al., 2016; Feschotte, 2008; Gould and Vrba, 1982; Long et al., 2016; Todd et al., 2019).

One study found that *Alu* retrotransposon may be responsible for the de novo birth of enhancers since Alu elements can drive in-vitro expression as reporters, interact across long ranges, are enriched for H3K4me1, and have nucleosomes phased like enhancers (Su et al., 2014). Other transposable elements such as short interspersed repetitive elements (SINEs) have also been identified to have exapted enhancers for the *ISL1* (Bejerano et al., 2006) and *POMC* genes (Santangelo et al., 2007).

Shadow enhancers encode robustness

"A biological system is robust if it continues to function in the face of perturbations" (Wagner, 2013). So, how robust are enhancers to mutations? It was recently discovered that the sequential length of an enhancer increases both the number of TF binding sites and the enhancer's robustness (Barr et al., 2019). Similar findings were also discovered for regions of the *eve stripe 2* enhancer, which become more robust to perturbation when systematically studying larger and larger fragments (Ludwig et al., 2011).

Another way to encode robustness in developmental enhancers is through seemingly redundant *shadow enhancers* – or enhancers that drive overlapping expression patterns for the same gene (Hong et al., 2008). These redundant enhancers usually recruit different regulatory logic from each other (Wunderlich et al., 2015). This diversity in regulatory logic encodes higher levels of robustness important for canalizing expression in the face of environmental or stressful perturbations (Frankel et al., 2010). Shadow enhancers have also been recently demonstrated to suppress transcriptional noise (Waymack et al., 2020), further supporting this claim. I will elaborate more on enhancers canalizing expression in further detail in section 1.4.

Enhancers, *shadow* or not[4], work cooperatively with one another to also contribute to robustness. One model of enhancer cooperativity is the Competition Model, which predicts that low-activity enhancers create additive expression, but more potent enhancers make a sub-additive output (Bothma et al., 2015). These results are probably due to enhancer competition –

---

[4] Scott Barolo has a great Q&A paper on questions such as "What is a shadow enhancer?" and "Is it useful to study an enhancer in isolation?" (Barolo, 2012).

or a lack thereof – at the promoter (Bartman et al., 2016; Bothma et al., 2015; Fukaya et al., 2016). This competition was further demonstrated for a pair of Krüppel shadow enhancers that depend on in which cells in the embryo the shadow enhancers are being expressed (Scholes et al., 2019). These Krüppel enhancers additionally have undergone stabilizing selection by trading regulatory logic between each other (Wunderlich et al., 2015). These experiments additionally stress the importance of the relationship between the distance of an enhancer and promoter to expression strength (Davis et al., 2020; Scholes et al., 2019). Live imaging techniques also demonstrate how enhancers interact with one another (Tsai et al., 2017, 2019).

Summary and open questions

Enhancers exhibit rapid binding site turnover and are constantly changing their structures. Rapid evolution makes enhancer conservation a poor proxy for determining function. In many of these rapidly evolving enhancers, the patterns are still conserved through stabilizing selection. Enhancers evolve from duplication events in the genome, emerge spontaneously from proto-enhancers or neutral sequences, and are exapted from elements such as transposons. Many loci have additionally evolved redundant enhancers, which are called shadow enhancers. Shadow enhancers promote robustness and canalize expression in different environments and have been helpful to study how enhancers cooperate with each other.

One thing that is still unclear at this point is how the regulatory logic underlying an enhancer constrains and promotes its evolvability. I address this in Chapter 3 of this thesis by identifying the critical enhancer logic within the *shavenbaby E3N* enhancer and demonstrate the constraints they impose upon its evolution. Additionally, it is still unclear how novel expression patterns form. I address this in Chapter 4, where we analyze robust and evolvable components of the *E3N* pattern, and explore a pleiotropic "hotspot" associated with novel *E3N* expression in the developing wing and haltere.

1.4   THE *SHAVENBABY* LOCUS

In *Drosophila*, the *shavenbaby* (*ovo/svb*) locus encodes a transcription factor (TF) responsible for regulating over 150 downstream target genes (Menoret et al., 2013; Stern and Frankel, 2013), leading to the concerted formation of hair-like epithelial projections called trichomes (**Fig. 4A**) (Kittelmann et al., 2021; Payre et al., 1999). Trichomes are non-sensory but may be responsible for aiding larval locomotion, hydrophobicity, and aerodynamic flight (van Breugel and Dickinson, 2017; Inestrosa et al., 1996; Kittelmann et al., 2021). Like *tinman/Nkx2.5* for heart formation (Bodmer, 1993; Yin and Frasch, 1998), *eyeless/PAX6* for eye formation (Gehring, 1996; Walther and Gruss, 1991), and Apterous for wing formation (Cohen et al., 1992) – *svb* is a "Master Regulator" gene for trichome formation (Stern and Frankel, 2013). The term "Master Regulator" means that many different signaling pathways converge to regulate *svb*, which in turn regulates many downstream targets (**Fig. 4B**). This is why Eric Wieschaus and Christiane Nüsslein-Volhard were able to identify so many essential developmental genes and pathways in *Drosophila* from studying trichome patterning (Wieschaus and Nüsslein-Volhard, 2016).

Developmental enhancers for the *svb* locus have been an excellent model for studying phenotypic evolution, enhancer redundancy, and enhancer grammar (Kittelmann et al., 2021). Trichomes have also been subsequently lost in various *Drosophila* species (Stern and Frankel, 2013). In particular, substantial differences in *svb* expression between species have been identified between *Drosophila melanogaster* and *Drosophila sechelia*, where a patch of quaternary trichomes is missing (McGregor et al., 2007) (**Fig. 4C**). One source of these differences was found to be caused by multiple mutations to Abrupt and Arrowhead binding sites in the *E6* enhancer. Each of these differences was found to individually lower expression, suggesting that the *sechelia* loss of trichomes was caused by evolutionary selection (Preger-Ben Noon et al., 2016). Additionally, the *svb* network topology has been studied in different contexts (i.e., larvae vs limb development), where significant regulatory differences have been observed. These differences suggest that gene regulatory networks can independently evolve for different developmental contexts (Kittelmann et al., 2018).
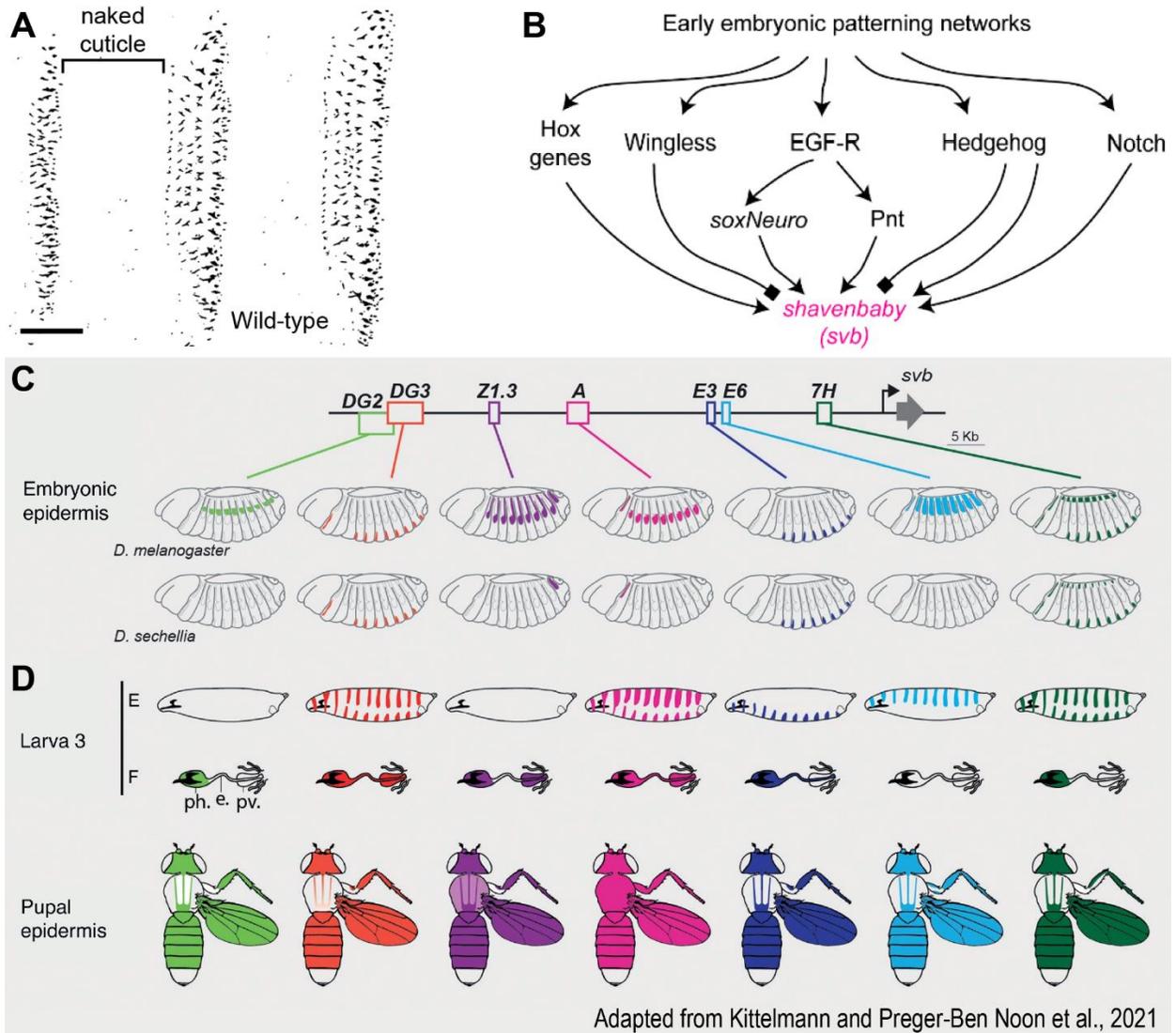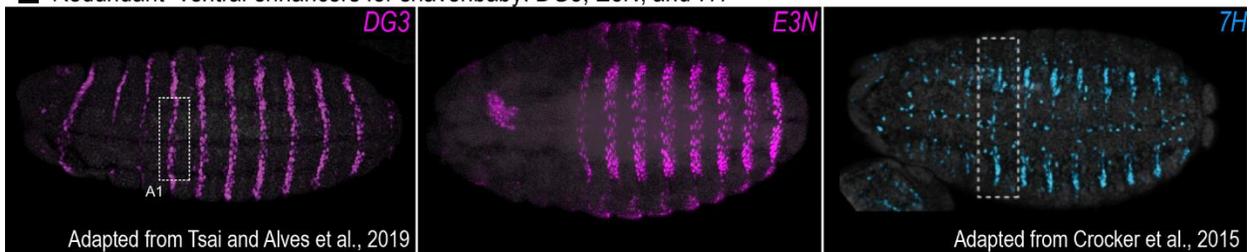
FIGURE 4 SHAVENBABY IS A MASTER REGULATOR FOR TRICHOME FORMATION

(A) Trichomes on the larvae are controlled by the *shavenbaby* gene. (B) Many early embryonic patterning networks feed into *shavenbaby* regulation. (C) There are seven characterized *shavenbaby* enhancers that regulate semi-redundant expression patterns. (D) The patterns are different in *Drosophila sechelia* and are pleiotropic across a wide range of developmental stages and tissue types. (E) The enhancers: *DG3, E3N,* and *7H* regulate the ventral stripe expression pattern.

_____

Enhancer redundancy has also been heavily studied in the *svb* locus. Frankel et al. identified seven semi-redundant *svb* enhancers – or "shadow enhancers" – which confer robust expression in varying environments (Frankel et al., 2010) (**Fig. 4D**). These enhancers are highly pleiotropic since trichomes can be found in larvae and adult flies and generate expression patterns across different developmental stages (Preger-Ben Noon et al., 2018). The ventral *svb* enhancers: *DG3*, *E3N*, and *7H* (see **Fig. 4E**) have also been demonstrated to form transcriptional "hubs" or "microenvironments" with each other (Tsai et al., 2017). When the *DG3* enhancer is removed, subtle trichome differences are observed when the embryo is heat-stressed (Tsai et al., 2019) - further supporting the argument that redundant enhancers canalize robust expression.

Because the putative *svb* enhancers have been fundamental to our overall knowledge of enhancer evolution, redundancy, and grammar, I chose to explore the evolutionary landscape of the *svb E3N* enhancer by mutationally scanning the element and analyzing the changes in patterning. This thesis, *The Evolutionary Landscape of a Developmental Enhancer*, focuses on how the regulatory information in *E3N* constrains and promotes its evolvability and answers several questions: 1) How much regulatory information is encoded within a developmental enhancer? 2) How can I automate and increase our throughput for analyzing enhancers? 3) Can I identify transcription factor binding sites within enhancers using mutational scanning? 4) What are some of the evolutionary constraints acting upon enhancer evolution? 5) How robust and evolvable are developmental enhancers? This work reveals possible and actual modes of enhancer evolution, as well as deepen our understanding of enhancers as a whole.

2 FLYSPRESSO: A CUSTOMIZABLE ROBOT AND AN IMAGING PIPELINE FOR DEVELOPMENTAL SYSTEMS BIOLOGY

> "Thank you very much, Mr. Roboto
> For doing the jobs nobody wants to
> And thank you very much, Mr. Roboto
> For helping me escape to where I needed to"
> —Styx, 1983

2.1 ABSTRACT

Developmental systems biology currently faces the challenge of balancing experimental throughput while controlling biological conditions. Thus, most developmental experiments are considered "low-throughput" in comparison to working with cells. This gap is primarily due to limitations in technology and costs. Here, I overview the pipeline, which I oversaw the development of, that increases the throughput of *Drosophila* embryology. The centerpiece of this pipeline is a prototype liquid handling robot developed by Jeff Jordan, which we call Flyspresso. Flyspresso can be programmed to carry out most protocols shared between devices, is small and can fit inside a fume hood, and is inexpensive compared to other liquid handling stations. In this chapter, I review the technology behind Flyspresso, demonstrate Flyspresso's performance with fixation and staining protocols, and compare the pipeline to traditional methods. With Aliaksandr Halavatyi, I also developed an adaptive feedback confocal microscopy pipeline to image embryos in multi-well slides. My goal with this project is to increase throughput for developmental systems biology to study enhancer grammar and evolution.

2.2 SCIENTIFIC CONTRIBUTIONS

For this project, I oversaw the development of this pipeline. Jeff Jordan is the engineer who developed the Flyspresso liquid handling robot with Jon Arnold and Peter Polidoro, with input from Justin Crocker. I was given the prototype robot and made additional modifications and (many) repairs to the device – including at one point replacing the plungers within the Dispense Manifold with the help of Arthur Milberger. I additionally developed and modified the existing fixation and staining protocols to improve the fixation yield and fluorescence intensities. I additionally met and worked frequently with Aliaksandr Halavatyi, who modified the adaptive imaging pipeline (originally developed and maintained by Antonio Politi) with my input. I also incorporated Christian Tischer into the project to help with the embryo rotation algorithm. I developed the BABB mounting protocol for *Drosophila* embryos shown in Figure 5I. I tested the different components of the pipeline individually and brought them together. To validate the pipeline, I carried out the experiments shown in Figure 5J and Figure 8. In Figure 8, Kerstin Richter dissected the adult *Drosophila* brains and Justin Crocker stained and imaged them. Kerstin Richter additionally fixed the Zebrafish larvae, which I stained with Flyspresso in Figure 8. Kerstin also manually fixed the *Drosophila* embryos for Figure 8D. All of the figures were illustrated by myself with input from Aliaksandr Halavatyi, Justin Crocker, Kerstin Richter, and Jeff Jordan.

Timothy Fuqua (me) organized meetings with different collaborators to develop the different components of the pipeline. Timothy Fuqua repaired and improved upon both the Flyspresso design and protocols. Timothy Fuqua developed the BABB mounting protocol and worked with Aliaksandr Halavatyi who developed the imaging pipeline. Timothy Fuqua bridged the components of the pipeline together and applied them in Chapter 3. Timothy Fuqua validated the pipeline doing most of the experiments in Figure 8.

Jon Arnold worked with Jeff Jordan for developing Flyspresso.

Justin Crocker was the Principle Investigator. Justin Crocker provided feedback to Jeff Jordan and helped during the prototyping phase of developing Flyspresso. Justin Crocker stained and imaged the adult *Drosophila* brain in Figure 8K. Justin Crocker additionally provided feedback and mentorship for me during this project.

Aliaksandr Halavatyi modified the adaptive feedback confocal microscopy plugin to work for *Drosophila* embryos. The plugin: MyPic, was originally developed and is maintained by Antonio

Politi. Aliaksandr Halavatyi also provided feedback and many fruitful discussions during the development of this project with me.

Jeff Jordan was the engineer behind the Flyspresso prototype. Jeff developed multiple prototypes for Flyspresso and worked extensively with Justin Crocker and me to get to the final working pipeline presented here. Jeff Jordan helped me with schematic illustrations and even flew from Canada to Germany to help with a particular robotics problem.

Arthur Milberger worked with me to 3D print and fabricate new stoppers inside the plungers for the Dispense Manifold.

Peter Polidoro worked with Jeff Jordan for developing Flyspresso. Peter Polidoro developed the circuit schematics and maintains their repositories.

Antonio Politi developed the original MyPic toolkit and continues to maintain it (Politi et al., 2018). Thank you Antonio for keeping science as open and accessible as possible.

Kerstin Richter dissected the *Drosophila* brain for Figure 8K and fixed the Zebrafish larvae for Figure 8L. Kerstin manually fixed the *Drosophila* embryos necessary for quantifying the loss of embryos in 8D. Kerstin additionally provided feedback for the figures.

Christian Tischer wrote the embryo rotation algorithm illustrated in Figure 5J for Job 3: Embryo Focus. Christian Tischer also provided feedback and many fruitful discussions during the development of this project.

2.3   INTRODUCTION

Evolutionary developmental biology stresses the importance of changes in regulatory sequences on development and evolution (Wittkopp and Kalay, 2012). Sometimes the subtlest of changes to these regulatory sequences can have a significant impact on a particular phenotype. Quantitatively detecting these differences will be paramount to understanding the relationships between genotypes and phenotypes (Crocker et al., 2016b; Small and Arnosti, 2020). Methods such as antibody staining and reporter gene assays are used to quantify gene expression patterns but require many additional steps that are still carried out manually (Kvon et al., 2014; Weiszmann et al., 2009; Wilk et al., 2016).

High-throughput screens have been integral for our understanding of development and gene regulation (Davis et al., 2020; Melnikov et al., 2012). However, these techniques involve using cell cultures and are spatially limited. Conversely, studying entire embryos and gene expression patterns across different fields of cells during other time points is technologically limited. Improving this technology is integral to further our understanding of regulatory sequences' relationships to their spatial-temporal patterns.

Current liquid-handling robots can carry out protocols on multiple samples simultaneously (C. et al., 2013; Rudnicki and Johnston, 2009). However, the robots are unable to carry out *Drosophila* embryo fixations, a critical step for studying gene expression patterns. Additionally, liquid-handling stations are expensive and often optimized for a single task. These reasons make liquid-handling stations limited and practical for only a subset of laboratories and protocols.

Engineered custom robots are becoming more commonplace in the laboratory setting (Gerber et al., 2017; Wong et al., 2018). To this end, Jeff Jordan designed Flyspresso, a syringe-based microplate washer, to address these issues. Flyspresso can carry out a wide range of biological protocols, including embryo fixation and immunostaining, and is more affordable than commercial devices. I combined Flyspresso with an adaptive feedback confocal microscopy plugin developed by Aliaksandr Halavatyi. This pipeline allows me to streamline embryo fixations,

vitelline membrane removal, antibody stainings, and imaging. I also demonstrate in this chapter that this pipeline increases the number of embryos one can screen without compromising sample quality and frees me from sitting in the dark on a microscope and washing embryos.

Using the Flyspresso pipeline, I carried out a mutational scanning experiment on a *Drosophila* developmental enhancer (Fuqua et al., 2020) (see Chapter 3). I used Flyspresso to fix *Drosophila* embryos for this experiment, and either stained the embryos with an x-gal staining solution (5-bromo-4-chloro-3-indolyl-β-D-galactosidase) or with antibodies for β-galactosidase, and acquired images of the embryos using our imaging pipeline. This chapter provides an overview of this pipeline that I oversaw the development of. The chapter emphasizes how Flyspresso operates for engineers or someone equivalently trained to reproduce or build upon the system. I additionally compare these methods to traditional techniques and provide suggestions on how to improve the system further.

2.4   THE SEMI-AUTOMATED PIPELINE AT A GLANCE

The semi-automated pipeline can be divided into four key steps: 1) Collecting embryos and removing their chorion membranes. 2) Fixing and staining the embryos. 3) Clearing and mounting the embryos in BABB. 4) Adaptive feedback confocal microscopy. For each of these steps, I either oversaw the development or directly worked on them. I provide more details for each step below.

Collecting embryos and removing their chorion membranes

To collect the *Drosophila* embryos, Jeff Jordan fabricated an array of tubes to hold up to 24 different strains of flies at once. (**Fig 5A**). Removable caps with a fine mesh over the surface called Transplates, are placed on both ends of the tubes (**Fig 5B**). Transplates simplify the process of moving flies into and out of the tubes using CO_2. The Transplates also serve as a mesh for the flies to deposit their eggs since the screens rest directly on an apple-juice-agar media with a small amount of yeast paste. I found that it is not necessary to build these chambers, and standard commercially available chambers can also be used. When carrying out egg collections, I observed that allowing the flies to acclimate to the chambers for one to two days increases the number of eggs deposited.

The embryos are washed inside the Transplate meshes using a fly saline solution (see methods), and any dead flies or debris are removed from the Transplates utilizing a pair of forceps. The Transplates are placed in a bleach solution, submerging the embryos within the mesh, and washed with bleach. Afterward, the bleach is washed off the embryos with water. The Transplates with the dechorionated embryos can now be loaded into a custom-fabricated microplate.
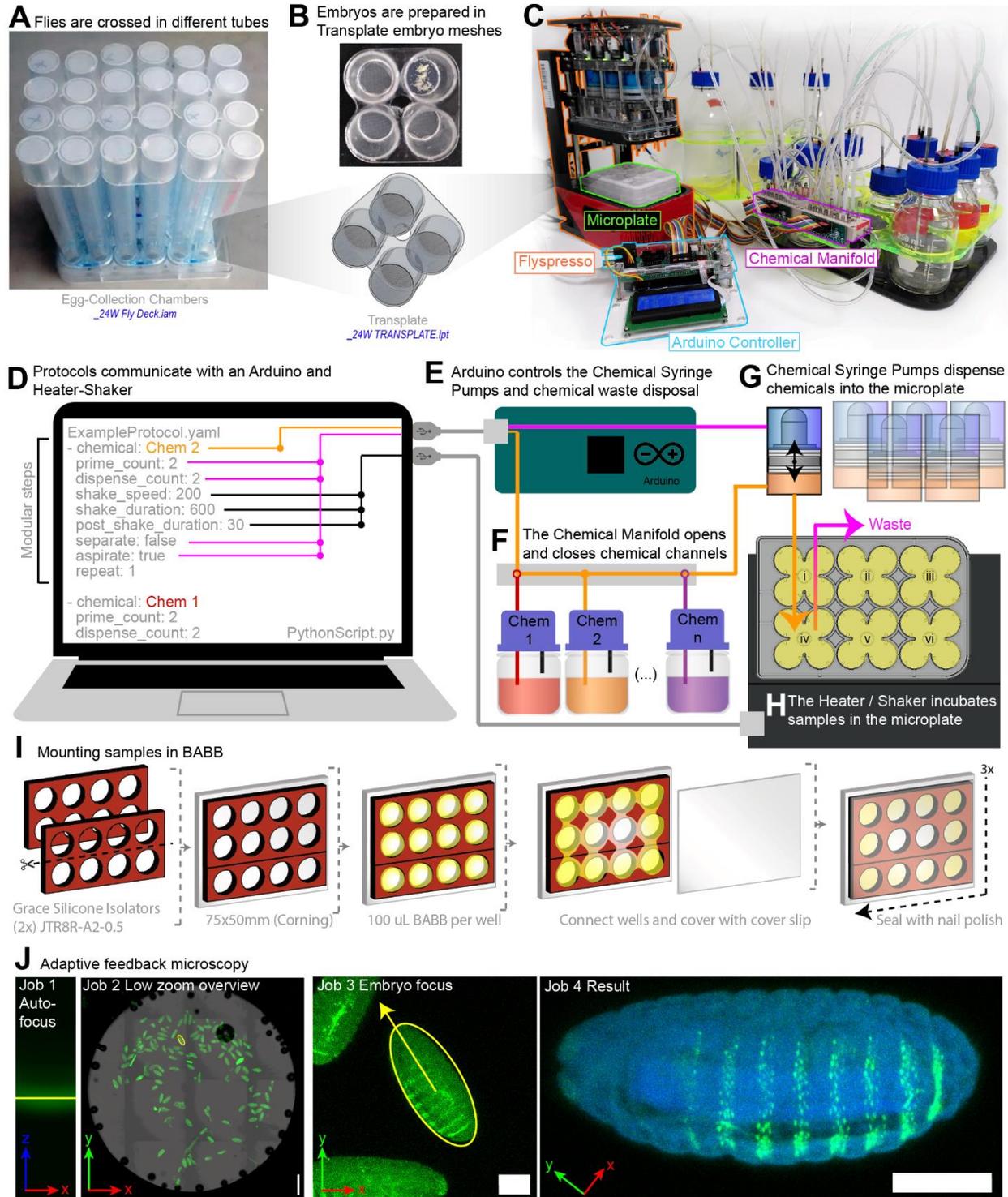
FIGURE 5 AN OVERVIEW OF THE FLYSPRESSO PIPELINE

(A) Custom-built 24-sample egg collection chambers. (B) The embryos are deposited onto detachable Transplate meshes, washed, and bleached manually before being loaded into the microplates. (C) The microplate and Transplates are loaded into Flyspresso, which consists of the device itself, an Arduino

microcontroller, and a chemical manifold. (D) Protocols are written as .yaml files and divided into steps (paragraphs on screen). Each step specifies different conditions (i.e., chemical, shaking speed, duration, etc.). (E) The protocols are interpreted by a python script that communicates with the Arduino microcontroller and the Heater/Shaker device. The Arduino controls the Chemical Manifold, opening and closing valves to select the proper chemical to be sent to the Chemical Syringes. (G) The Arduino also controls the Chemical Syringes, which either prime the chemical to clean the tubing or push the chemical to the microplate. (H) The microplate is secured to the Heater/Shaker device, which can rock the microplate at various speeds and temperatures. (I) After fixing and staining, samples are cleared in Benzyl-Alcohol Benzyl Benzoate (BABB). Silicone isolators are placed on a microscope slide and the embryos are added with BABB to each well. The wells are connected with a small amount of BABB and covered with a coverslip. The slide is immediately sealed in nail polish. (J) The adaptive feedback microscopy is divided into four jobs. Job 1: Autofocus identifies the coverslip by scanning through the slide to find the reflection (marked in yellow). This step focuses on the embryos on the slide. Job 2: Low zoom overview acquires a tile scan of each well to identify embryos (green). Embryos are selected manually or automatically (yellow highlight). Job 3: Embryo focus scans through the selected embryos and identifies their center and angle (yellow arrow). Job 4: Result is the final image. Because the angle of the embryo was calculated, the final image can be acquired as a rectangle instead of a square to double the acquisition time. Scale bars = 100 um.

Figure adapted from Fuqua et al., 2021. (A-H) Flyspresso was primarily designed by Jeff Jordan with input from Peter Polidoro, Jon Arnold, and Justin Crocker. (J) The adaptive feedback confocal microscopy pipeline was developed by Aliaksandr Halavatyi with input from Christian Tischer. Timothy Fuqua illustrated the figure, coordinated the pipeline development, developed the BABB mounting and optimized the protocols.

_____

Fixing and staining embryos

With the microplates loaded, the plate is inserted into Flyspresso for fixation (**Fig 5C**). Flyspresso consists of the robot itself, an Arduino microcontroller, and a Chemical Manifold. Flyspresso was built by Jeff Jordan with input from Jon Arnold, Peter Polidoro, and Justin Crocker. The embryos are fixed using different protocols for antibody and x-gal staining. For the x-gal colorimetric assay, the embryos are fixed in formaldehyde, washed in PBT, incubated in a staining solution, washed again in PBT, and imaged.

To fix and stain the embryos with antibodies, I use paraformaldehyde and heptane. Following the fixation, I isotonically shock the embryos using methanol, removing the vitelline membranes. After washing in methanol, the embryos are washed in PBT, blocking solutions, and antibody solutions, and ethanol. For the protocol itself, see Chapter 6.

Flyspresso operates through the assistance of a laptop to quickly type and edit protocols (**Fig 5D**). The protocols that I either developed or modified are written as .yaml files, where each step in the protocol is divided up into paragraphs. Within each step, different parameters can be programmed such as 1) how often the step is repeated, 2) how long to wait after adding the solution or shaking the samples, 3) where to remove the liquid (upper phase = separate, lower phase = aspirate), 4) how long to shake the samples and at what speed, 5) if the chemical should first clear the tubing (priming), and most important, 6) which chemical is selected. For the protocol, I optimized all of these steps to improve the embryo yield and antibody fluorescence.

A python script interprets the information written within the protocol. The python script sends the information in a simplified version to the Arduino microcontroller (**Fig 5E**). The Arduino 2560 communicates with both the Chemical Manifold and the Chemical syringes. The Chemical Manifold (**Fig 5F**) is an array of solenoid valves that open or close to create a closed circuit between the chosen chemical and the Flyspresso system. The chosen chemical is sent to the Chemical Syringes. Flyspresso has seven Chemical Syringes (**Fig 5G**), one of which is responsible for priming the system, pulling the chemical through the tubing, and sending it to waste. The other six syringes feed directly to the Transplates resting in the Microplate below. The Arduino controls the movements of all the Chemical Syringes. The Microplate is attached to the Heater/Shaker device (**Fig 5H**). This device can be programmed to shake microplates at different temperatures and speeds. For simplicity, the computer controls the Heater/Shaker device using the python interpreter. I do not use the temperature settings for the protocols here.

Clearing and mounting embryos in BABB

I chose to clear and mount my embryo samples in benzyl alcohol benzyl benzoate (BABB) because it allows me to image completely through an embryo with a minimal loss of fluorescence (Dodt et al., 2007). This advantage will enable me to image all embryos and not embryos only in the proper location or orientation. After Flyspresso washes the samples in ethanol, the samples are removed and manually cleared in BABB overnight. The samples are mounted onto microscope slides (**Fig 5I**). Silicone Isolators are cut and divided to put the appropriate number of wells on

the slide. The embryos are added to each isolated well and connected with a thin amount of BABB between the wells. Connecting the wells prevents air bubbles from leaking into the samples. The slides are covered with a cover slip and sealed using clear nail polish.

The adaptive feedback confocal plugin

To image all of the embryos, Aliaksandr Halavatyi built upon a plugin developed by Antonio Politi called MyPic: a Java library that allows the microscope to carry out multiple steps as "jobs" or image parameters and to communicate with the image analysis software: Fiji (Politi et al., 2018; Schindelin et al., 2012). I worked with Aliaksandr Halavatyi to test different versions of the scripts and to find the optimal order and number of jobs. In the end, Aliaksandr Halavatyi modified MyPic so that I can run four different jobs on the embryos (**Fig 5J**), which I describe in detail below.

The first job is called Autofocus, and the microscope scans through the z-axis to identify where the coverslip is within each well using a 5x objective lens. The coverslip is reflected back to the detector. The reflection notifies the microscope at what height to adjust the stage, so all embryos are in focus. The second job is called Low Zoom Overview. With the embryos now in focus, the microscope creates a tile scan image over the well using a 5x objective lens. The embryos are identified either manually or automatically and saved as a series of x-y coordinates. The microscope repeats jobs 1 and 2 for each well. After jobs 1 and 2 have been carried out in every well, the microscope navigates back to the first well and carries out jobs 3 and 4. The third job is called Embryo focus. With a 20x magnification lens and the given x-y coordinates, the microscope quickly scans through the z-axis of the embryo to identify its center and orientation. Once these parameters are determined, the microscope runs the fourth job: Result, which creates the final image, cropped, rotated, and imaged with a 20x objective.

2.5 FLYSPRESSO: THE NUTS AND BOLTS

In the next two sections, I describe Jeff Jordan's prototype design so readers can build the device themselves and improve upon it if they choose. The Flyspresso liquid handling robot can be conceptualized by either its components or the fluidic path. In this section, I describe the features and manifolds (**Fig 6A**) and how they operate together for liquid handling. In the following section, I describe Flyspresso from the hydraulic perspective. The "brain" behind Flyspresso is an Arduino microcontroller (**Fig 6B**), which controls the movement of the Chemical Syringes and the Chemical Manifold. The syringes and manifold are controlled by electronic solenoid valves, where running a small electric current through them opens and closes channels. Solenoid valves allow a hydraulic fluid – or System Liquid – to push and pull the syringes and open and close the channels attached to each chemical on the Chemical Manifold (**Fig 6C**). Upon opening the channels, a vacuum pump and an inert gas push the chemical into the system.

At the top of Flyspresso is a positive-displacement System Piston (**Fig 6D**). This piston is either open or closed using a solenoid valve, switching between a vacuum and the inert gas to displace the internal System Liquid. This System Liquid displacement draws the Chemical Syringes up and down for either priming the system (Priming syringe: 'P') or for adding chemicals to the microplate below (Chemical Syringes: "i-vi"). Priming is done whenever a new chemical is used during the protocol to clear the tubing of residual chemicals. Priming is done using the Priming Syringe, and all of the chemicals from the Priming Syringe go directly to waste.
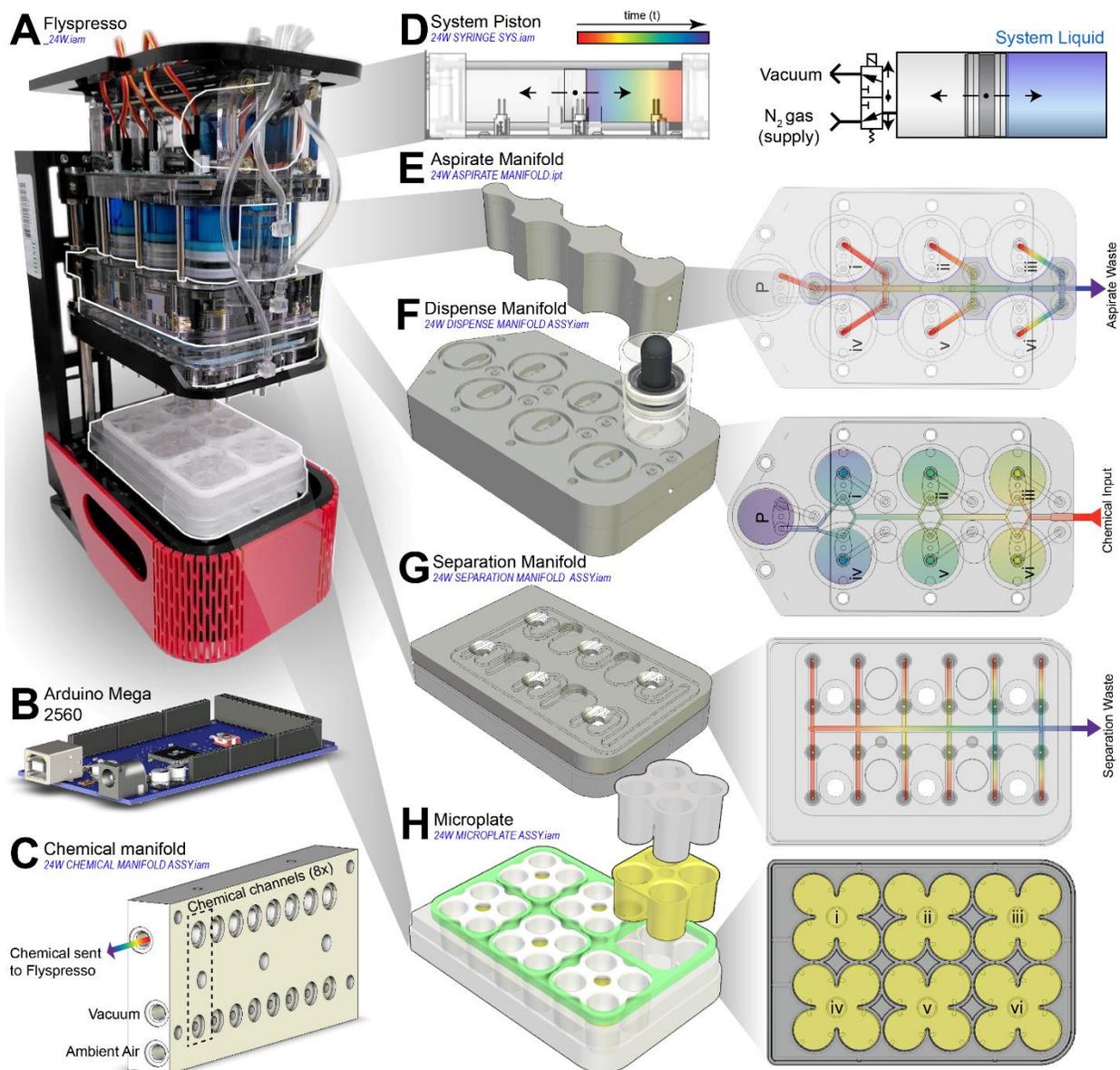
FIGURE 6 THE MANIFOLDS AND COMPONENTS OF FLYSPRESSO

(A) An explosion-view of the Flyspresso robot and its different components and manifolds. From bottom to top: the Heating / Shaking Device (not shown), the Microplate, the Separation Manifold, the Dispense Manifold, the Aspirate Manifold, and the System Piston. **(B)** An Arduino Mega 2560 controls the opening and closing of electronic solenoid valves for controlling the Chemical Manifold and the System Syringes. **(C)** The Chemical Manifold is a modular device with eight individual channels. Each channel connects to a solenoid valve (dotted lines) which switches between a vacuum to hold the Chemical or ambient air to allow the vacuum in Flyspresso to pull the liquid into the robot. **(D)** The System Piston contains a plunger that pushes or pulls an internal System Liquid (blue) through a solenoid valve to switch between an inert gas supply or vacuum. **(E)** The Aspirate Manifold pulls chemicals from the microplate through the syringe tips of the Chemical Syringes and to waste and rests on the Dispense Manifold. **(F)** The Dispense Manifold holds 7 Chemical Syringes: P and i-vi. P stands for the Priming Syringe and is used to clear the tubing.

Syringes i-vi move through the displacement of the System Liquid and pull the Chemical drawn from the Chemical Manifold into their chambers and push it into the corresponding wells in the microplate below. **(G)** The Separation Manifold contains an array of 24 small syringes that remove chemicals from the upper half of the microplate. **(H)** The Microplate houses 6 Transplates (yellow), which sit in wells i-vi corresponding to the Chemical Syringes above. The Transplates each hold four samples and allow up to 24 different embryo pools to be tested. Each Transplate is capped with a Seplate attachment for isotonic shocking.

Figure adapted from Fuqua et al., 2021. Flyspresso was primarily designed by Jeff Jordan with input from Peter Polidoro, Jon Arnold, and Justin Crocker. Timothy Fuqua oversaw the maintenance and repairs of Flyspresso, and additionally fabricated spare components when necessary. Timothy Fuqua illustrated the figure.

Chemical waste from the microplate is vacuumed through the same syringe needles attached to the Chemical Syringes and through the Aspirate Manifold (**Fig 6E**). The Aspirate Manifold sits between the Chemical Syringes, resting on the Dispense Manifold (**Fig 6F**). The Chemical Syringes have an internal cap inside (black semi-sphere) which physically blocks the Chemical Syringe plungers to a controlled volume. This volume is added to the microplate.

The next manifold on Flyspresso is the Separation Manifold (**Fig 6G**). This component contains 24 syringe tips that can separate chemicals from the upper interface for each embryo well in the microplate. The microplate contains six large wells (i-vi), which hold six Transplates (yellow) for 24 testable conditions total. The Transplates are additionally capped with an attachment called a Seplate, which creates an upper and lower interphase within the well. The purpose of the Separation Manifold is for Isotonic Shocking, which I will explain in the following section.

2.6   FLUIDIC PATHWAYS

Another way to conceptualize Flyspresso is through a fluidic perspective, as illustrated in **Figure 7**. As previously described, the System Piston is a positive-displacement syringe that moves an ethanol-based System Liquid through the use of an inert gas supply and a vacuum (**Fig 7A**). The Arduino runs an electric current through a solenoid valve to change whether the System Piston pushes or pulls the System Liquid. If displaced, the System Liquid traverses to move either the Priming Syringe Pump or the Chemical Syringe Pumps. Follow the path of the System Liquid in Figure 6 using the blue lines.

When adding a chemical to the chambers of the Chemical Syringe Pumps (**Fig 7B**), the System Piston pulls the System Liquid, and the Arduino simultaneously opens the solenoid valves attached to the Chemical Syringe Pumps, allowing the negative pressure to pull the Chemical Syringe Pumps. The Chemical Syringe Pumps, in turn, pull the selected chemical into the chambers. The System Piston then pushes the System Liquid, displaces into the Chemical Syringes, pushing the plungers down, and displacing the chemical from the chambers into the microplate. Passive-check valves are located within the tubing, preventing chemicals from flowing the wrong way. Follow the path of the chosen chemical in Figure 6 using the orange lines.

Because all of the tubing is shared in Flyspresso, the tubing needs to be "primed" to prevent cross-contamination between chemicals whenever switching chemicals. The Priming Syringe Pump clears the tubing by pushing and pulling the new chemical like the Chemical Syringe Pumps (**Fig 7C**). The only difference between the Priming Syringe Pump and Chemical Syringe Pumps is when the chemical is displaced, the chemical does not enter the microplate but is sent to waste.
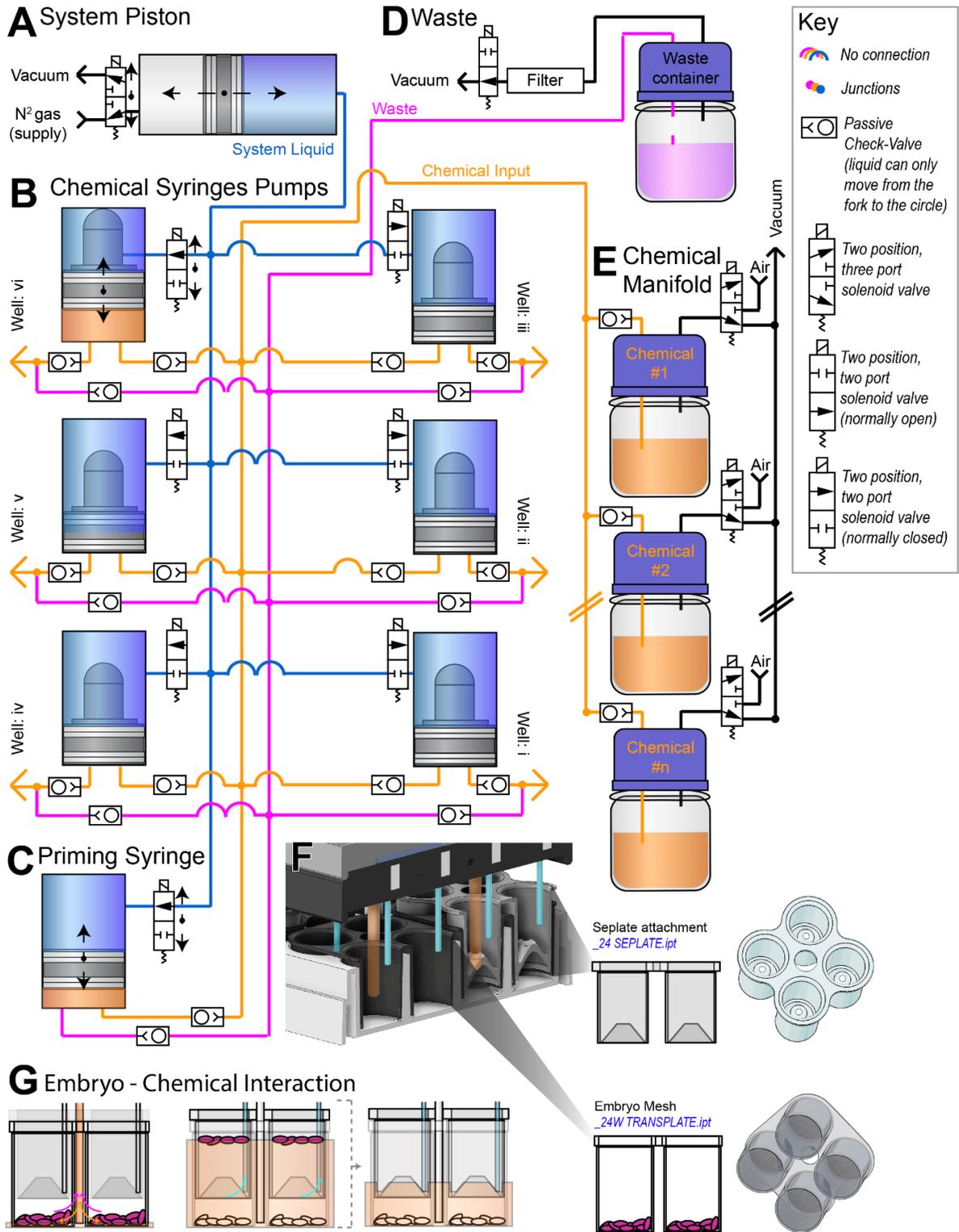
FIGURE 7 FLUIDIC PATHWAYS OF FLYSPRESSO

(A) An ethanol-based System Liquid (blue) is displaced by the System Piston, which operates by switching between an inert gas and vacuum using a solenoid valve (white rectangles, see key). The System Liquid is

displaced to the Chemical Syringe Pumps. (**B**) There are six Chemical Syringe Pumps on Flyspresso. A solenoid valve controls access to each pump. Pulling the System Liquid draws the Chemical Syringe Pumps up, displacing the chosen chemical (orange) and pushing the chemical into the microplate below (not shown). Check valves (small white rectangles with a circle and triangle inside) prevent chemicals from flowing in the wrong direction. (**C**) When switching between chemicals, the Priming Syringe Pump operates like the Chemical Syringe Pumps but dispenses the drawn chemical directly to waste. (**D**) Chemical waste containers collect residual chemicals from aspirating, separating, and priming. Solenoid valves switch from a closed state to a vacuum to pull chemicals out of the system. Charcoal filters are attached to the waste containers to prevent volatile gasses from leaking into the room. Still, the device should sit in a fume hood for the safety of the operator. (**E**) The Chemical Manifold is an array of expandable units capable of controlling eight reagents each. Each reagent bottle is attached to a solenoid valve. The valves alternate switch from a vacuum to ambient air when the Arduino selects the chemical. (**F**) A cross-section of the microplate with the Transplates and Seplates interacting with the Chemical Syringes (orange) tips and Separation Manifold syringe tips (blue). The Transplates rest at the bottom of the microplate wells. Inserted above the Transplates are the Seplate attachments, which have small inverted cones with a hole in the center. These holes allow the removal of chemicals using the Separation Manifold tips. (**G, left**) Chemicals are dispensed into the microplate through the tips of the Chemical Syringe Pumps. These syringes can also aspirate the chemicals to waste. (**G, middle**) Methanol is added to the microplate, and the samples are rapidly shaken in a process known as isotonic shocking. While being shocked, embryos that separate from their vitelline membrane sink to the base of the wells, while embryos or membranes that fail to separate float to the surface through the small one-way openings in the Seplate Attachments (**G, right**). Separation Manifold tips (blue) can remove the debris, leaving only separated embryos on the bottom (white).

Figure adapted from Fuqua et al., 2021. Flyspresso was primarily designed by Jeff Jordan with input from Peter Polidoro, Jon Arnold, and Justin Crocker. Timothy Fuqua illustrated the figure.

_____

Waste from the microplate can be removed using either the Aspiration Manifold (see Figure 6E) or the Separation Manifold (see Figure 6G). If using the Aspiration Manifold, the waste is aspirated from the six wells in the microplate, through the syringes attached to the Chemical Syringe Pumps, and through the Aspiration Manifold. Passive check-valves also prevent waste from flowing back into the microplate. The aspiration occurs when a solenoid valve attached to the waste container is switched (**Fig 7D**), allowing the vacuum to pull the chemicals through. A charcoal filter is installed to prevent volatile gasses from being pulled through the vacuum pump. The Separation Manifold also draws chemical waste to a different waste container with the same setup. The chemical waste path is illustrated as a magenta line in Figure 7.

Finally, Chemical Manifolds are relayed together as modular units (**Fig 7E**). Each unit houses solenoid valves for eight different reagents. The Arduino changes the current in these solenoid valves, so the bottle changes from being under vacuum pressure to ambient air. This pressure change enables the chemical to be drawn into the robot for priming or loading into the Chemical Syringes.

Isotonic shocking

The most considerable nuance of Flyspresso is that it is capable of removing the vitelline membrane of *Drosophila* embryos. The vitelline membrane is a structure that surrounds the outer surface of the embryo plasma membrane. Removing this structure is done through a process called isotonic shocking (Rothwell and Sullivan, 2007). The embryos initially rest at the microplate base in the Transplate baskets (**Fig 7F**). Above the Transplates, Seplate Attachments are added. These attachments contain small a small inverted opening with an inverted cone shape, making them one-way outlets. When the embryos are treated with methanol and rapidly shaken, the vitelline membrane bursts, and the embryos sink to the base of the Transplate (**Fig 7G**). Embryos that do not burst and the remaining membranes float to the surface through the Seplate Attachments. The Separation Manifold then aspirates the fluid above in the Seplate Attachment without disturbing the successfully shocked embryos at the base of the microplate.

2.7 EXPECTED RESULTS WITH THE FLYSPRESSO PIPELINE

I next demonstrate the power of the pipeline and compare it to traditional methodologies. One unique part of the pipeline was mounting the embryos in BABB instead of a standard mounting media such as Prolong Gold. I chose to mount in BABB because it allows deeper imaging by clearing the embryos. I demonstrate this in **Fig 8A** by staining two sets of embryos with a Crumbs antibody (Tepass and Knust, 1993) and mounting one set in Prolong Gold and the other in BABB. A cross-sectional average of each condition is depicted in **Fig 8A,** and I additionally plotted the fluorescence intensity along the depth of the embryo (**Fig 8B**). This simple experiment demonstrates the power of using BABB compared to traditional mounting techniques.

It is also standard in antibody staining protocols to reuse antibody reagents as they can be costly. I reused the Crumbs antibody solution and the secondary antibody solutions twice (three times total) and measured Crumbs intensity (**Fig 8C**). When I did this staining, I found that the fluorescence actually increased after the first use (p=0.0014), and there was no statistically significant difference between the second and third staining (p=0.22). I hypothesize that I see this slight increase in fluorescence because the first staining can block non-specific antibody binding. My results demonstrate that reusing antibody solutions is a cost-effective strategy that additionally improves imaging quality.

I next asked if Flyspresso would lose more embryos during the fixation process compared to traditional methods. Together with Kerstin Richter, we fixed embryos manually (10 pools) and automatically (19 pools) with Flyspresso and counted the number of embryos before and after fixation. I wrote image analysis scripts to count the embryos using Fiji and found that there was no statistically significant difference (p=0.47) between fixing embryos manually (38%) and using Flyspresso (43%).
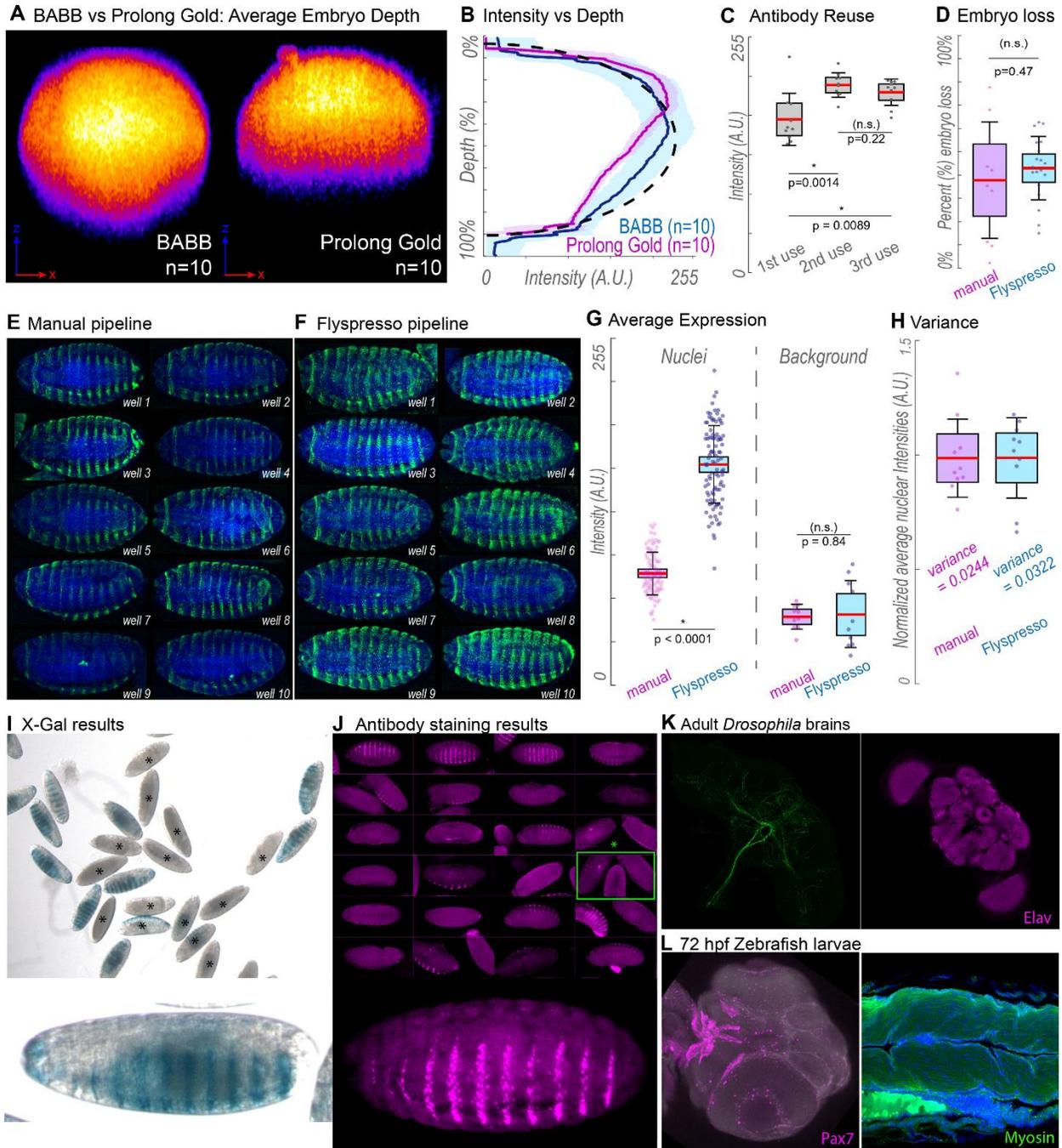
FIGURE 8 RESULTS FROM THE PIPELINE

(**A**) Average composite embryo cross-section through the x-z axis of 10 *Drosophila* embryos in BABB (left) and Prolong Gold (right). (**B**) Intensity vs. Depth for the composited embryos. Blue = BABB and Magneta = Prolong Gold. The solid line is the mean (representative of embryos in (A), the lightly shaded regions are one standard deviation. Black dotted lines are hypothetical embryo without fluorescence decay. (**C**) Box plots measuring Crumbs fluorescence when reusing antibody solutions. Each point represents the average of all nuclear intensities in a single embryo, N=10 embryos each. (**D**) Box plots comparing embryo fixation methods. Each point is the number of embryos lost per fixation. Manual = Magenta (N=10), Flyspresso =

Blue (N=19). (**E**) Example images using manual techniques and (**F**) the pipeline. 10 pools imaged for each condition, 10 embryos per pool (100 embryos imaged per condition). (**G**) Box plots for fluorescence intensity and background intensity. Each point represents the average nuclear intensity in a single embryo. (**H**) Box plots. Nuclear intensities normalized. (**I** and **J**) a variant of the *E3N* enhancer (line *173-2*) stained using x-gal (**I**) and antibodies (**J**). Embryo clusters shown with a green box. (**K**) *Drosophila* brain stained with a reporter and Elav. (**L**) *Danio rerio* larvae (72hpf) stained for Pax7 and Myosin. Box plots: red line = mean, gray box = standard error mean with a 95% confidence interval. Whiskers are one standard deviation. All p-values calculated using a Student two-tailed t-test.

Figure adapted from Fuqua et al., 2021. (**D**) Embryos manually fixed by Kerstin Richter. Automatically fixed embryos done by Timothy Fuqua. Timothy Fuqua wrote the image analysis scripts. (**K**) Brain was dissected by Kerstin Richter. Justin Crocker stained and imaged the brain. Kerstin Richter fixed the Zebrafish larvae in panel **L**. Timothy Fuqua stained and imaged the larvae. All remaining experiments were carried out by Timothy Fuqua. Timothy Fuqua illustrated the figure.

_____

Finally, I compared the imaging quality between manual methods and the automated pipeline. Ten sets of embryos were manually fixed and stained for *shavenbaby*, and another ten using Flyspresso (**Fig 8E, F**). Samples stained with Flyspresso had a much higher amount of *shavenbaby* expression compared to the manual technique (p<0.0001) and insignificant differences in background levels (p=0.84) (**Fig 8G**). I normalized the intensities to their means and demonstrate that the variances are similar (manual = 0.0244, Flyspresso = 0.0322) (**Fig 8H**). My results suggest that the variance does not change, but Flyspresso staining may lead to higher fluorescence levels.

Expected results and troubleshooting

My results from both the x-gal and antibody staining are demonstrated in **Fig 8I and J**. Both sets of embryos come from the same fly line, which carries a variant of the *shavenbaby E3N* enhancer driving *lacZ* as a reporter gene. The eggs were collected after being laid overnight, giving us a range of developmental stages. The resolution of the stainings, however, is very different. X-gal staining is much less refined in comparison to antibody staining. For x-gal staining, the embryos quickly become saturated, and it becomes increasingly difficult to distinguish subtle phenotypic effects. Conversely, for antibody staining, the entire expression pattern can be imaged without oversaturating detectors, and the images are at cellular resolution.

Some problems may occur during the adaptive feedback microscopy. The Embryo Focus Job (job #3) acquires a quick scan through the embryo to identify the angle and center of the embryo. This job has problems when the wells in the microscope slide are overcrowded with embryos. When this happens, sometimes the microscope cannot distinguish between embryos – even though the microscope uses water-shedding and thresholding algorithms to separate embryo images. One example of overcrowding and false segmentation is in **Fig 8J**. A green box outlines an embryo cluster that failed to separate. In general, having less than 100 embryos in a well should prevent this problem from happening too frequently. Additionally, if the staining intensity is too low, there may be rotation problems. I recommend lowering the zoom factor to scan a larger area. One day, I would like to implement more innovative segmentation algorithms such as user-friendly machine learning toolkits to fix these problems (Berg et al., 2019).

I wanted to demonstrate the versatility of Flyspresso by modifying and running the protocols described on different model systems. Justin Crocker and Kerstin Richter carried out an antibody staining using Flyspresso on adult *Drosophila* brains and *Danio rerio* 72hpf larvae (zebrafish). *Drosophila* brains were fixed based on standard protocols (Tito et al., 2016) and stained with Flyspresso (**Fig 8K**). I stained the zebrafish larvae (fixed by Kerstin Richter) with Flyspresso using antibodies for Pax7 and Myosin (**Fig 8L**).

2.8   DISCUSSION

Other commercially available liquid-handling devices

Flyspresso is not the only available liquid-handling device. Liquid handling stations typically come in three flavors: Bulk Liquid Dispensers, Transfer Devices, and Microplate Washers. Bulk Liquid Dispensers are large machines equipped with a syringe pump or peristaltic pumps, usually require more reagents than standard protocols, and dispense liquids at large volumes. Transfer Devices use pipettes but are consequently slower since they need cleaning steps or constantly switch consumable pipette tips. Finally, Microplate Washers can add or remove liquids using an Aspiration Manifold, but are optimized for cell culture and not Systems Biology (C. et al., 2013; Rudnicki and Johnston, 2009). Flyspresso is a hybrid of the different devices since it adds chemicals to microplates like a Bulk Liquid Dispenser and has an Aspiration Manifold.

Two liquid-handling robots in particular come close to Flyspresso; the Insitupro Vsi and the Biolane HTI 16Vx (Intavis). The Transfer Device: Insitupro VSi uses peristaltic pumps and pipettes to run protocols. Like Flyspresso, the Insitupro VSi are equipped with a Heating/Shaking device. The benefit of the Insitupro VSi is that it can work directly with samples on a microscope slide and with 60 samples. Flyspresso is a small device and does not occupy an entire benchtop. Additionally, Flyspresso has expandable Chemical Manifolds, which means the user has more control over the protocol.

Intavis also sells a Microplate Washer called the Biolane HTI 16Vx, which is also peristaltic-based for chemical transfer. The HTI 16Vx has a fixed limit of 16 reagents, can wash samples in microplates containing up to 384 samples, and can wash microscope slides. Again, the device is not expandable and is unable to carry out isotonic shocking. Also, Jeff Jordan designed Flyspresso to hold only 24 samples since it was optimal for rapidly loading the device without cross-contaminating wells.

Who can build Flyspresso?

To build a Flyspresso device, someone familiar with Computer-Aided Design (CAD) files is needed. Having an understanding of microcontrollers and circuitry is also necessary. However, anyone who can complete the protocol manually can operate Flyspresso.

Pipeline limitations

There are some limitations to the Flyspresso design that I encourage others to build upon if they choose to replicate the device. To begin, Flyspresso still requires a priming step, which could waste fewer reagents if a peristaltic pump was used instead. Peristaltic pumps can rotate both clockwise and counterclockwise to rescue reagents (Rudnicki and Johnston, 2009). For expensive reagents such as antibodies, it would be ideal to design a port where the reagent could be directly added to the microplate. To circumvent this problem, I implemented a pause step and manually added antibodies to the microplate.

The current microplates for Flyspresso hold 24 samples that share six reagent wells. This may be problematic since only six conditions can be tested. Modifications to the device and microplate may be necessary to change this design. Next, it is worth considering which chemicals the device will be handling. For our protocol, I still manually bleached the embryos before loading them into Flyspresso. Bleach is a very corrosive chemical, and Heptane and Methanol also corrode the tubing and plastics after about a year of continuous use. Sample mounting could also be improved since this is also still done manually. Integrating the robotics with microfluidic techniques, other robots, or finding a way to image the embryos on the microplate directly may solve this problem (Levario et al., 2013; Shorr et al., 2019).

Conclusion

In the next chapter, I will demonstrate the power of this pipeline by mutationally scanning a developmental enhancer.

2.9   DATA AVAILABILITY

All of the CAD files to build Flyspresso can be located on my GitHub repository:

https://github.com/tfuqua95/Flyspresso-CAD-files

The operating software for Flyspresso is located here:

https://github.com/janelia-pypi/hybridizer_python/tree/digital

Schematics for the PCB boards and the Arduino controller are available here:

https://github.com/janelia-modular-devices/mixed_signal_controller

To download the adaptive feedback confocal microscopy plugin, use the following link:

https://git.embl.de/grp-almf/feedback-fly-embryo-crocker

Access to all of the original images and data from the experiments are available for download here:

https://www.embl.de/download/crocker/flyspresso/index.html

3 THE *E3N* ENHANCER IS DENSELY ENCODED, HIGHLY PLEIOTROPIC, AND EVOLUTIONARY CONSTRAINED

"Mutation: it is the key to our evolution. It has enabled us to evolve from a single-celled organism into the dominant species on the planet. This process is slow, and normally taking thousands and thousands of years. But every few hundred millennia, evolution leaps forward."

-Prof. Charles Xavier, 2000

3.1 ABSTRACT

Phenotypic evolution is primarily driven by changes in gene regulation rather than coding sequences. However, the extent to which gene regulation can evolve is unclear because most experimental perturbations are limited to manageable sample sizes, and population genetics has a survivorship bias. I wanted to explore the extent of evolution for the *Drosophila shavenbaby E3N* enhancer, and generated a random mutant reporter library of *E3N.* To screen the *E3N* enhancer library, I developed a semi-automated pipeline to automatically fix, stain, and image *Drosophila* embryos across a range of developmental stages. From this screen, I found that most mutations to *E3N* affected gene expression. These results suggested that *E3N* is densely encoded. I also discovered that most phenotypes were linked with each other such as expression levels, location, and state. Finally, I found that the results from the screen could explain various phenotypes across different Drosophilids. My results suggest that dense and pleiotropic encoding may constrain developmental enhancer evolvability.

3.2 SCIENTIFIC CONTRIBUTIONS

This project was planned prior to my arrival by Justin Crocker, David L. Stern, and Richard S. Mann. For this project I oversaw and also contributed to the development and optimization of the pipeline to automatically screen *Drosophila* embryos (see Chapter 2.2 for further details). The pipeline was developed with Jeff Jordan, Aliaksandr Halavatyi, Peter Polidoro, and Christian Tischer. With the pipeline, I then carried out the majority of the mutational scanning experiment of the *E3N* enhancer library (primarily assembled by Jonathan Sager) and analyzed the data. I additionally supervised a Master's student: Maria Elize van Breugel who worked on the X-gal staining assays and did some confocal microscopy with me for Figure 14. I analyzed Maria Eliza van Breugel's data to produce the EWAC and footprinting plots with input from Jakob Wirbel and Judith Zaugg. Chaitanya Rastogi also predicted total Ubx binding affinity for the enhancers using the NRLB algorithm (Rastogi et al., 2018). David. L Stern carried out the cuticle preps for Figure 16. My advisor, Justin Crocker helped with data analysis and carried out antibody experiments in Figure 11F-L and the experiment in Figure 13 with me. Justin and I developed the figures together with feedback from all contributors mentioned below. Albert Tsai also provided helpful feedback, suggestions, and helped with various data analysis sections.

Timothy Fuqua (me) oversaw and contributed to the development of the automated pipeline (see Chapter 2). Timothy Fuqua maintained the *E3N* stocks and carried out the mutational scanning experiment with antibody stainings for the Figures 9, 10, 11, 12, and 15. Timothy Fuqua also analyzed the data for these figures. For this project, Timothy Fuqua was also the daily supervisor for Master's student Maria Elize van Breugel and partially contributed to the X-gal staining assays in Figure 9F-I, Figure 14A-B and K-L. Timothy Fuqua calculated EWAC and footprinting scores with input from Jakob Wirbel and Judith Zaugg.

Maria Elize van Breugel was a Master's student under my advisory. With my supervision and assistance, Maria Elize van Breugel screened the 274 lines with the X-gal staining assay to make the plots for Figure 9F-I and Figure 14A-B and K-L. With my supervision, Maria Elize van Breugel also stained the pleiotropic mutants in Figure 14C-J.

Justin Crocker was the Principle Investigator. Justin Crocker planned the initial experiment with David L. Stern and Richard S. Mann. Justin Crocker created mutant *E3N* lines with Jonathan Sager and me. Justin Crocker carried out the experiments in Figure 11F-L. Justin Crocker additionally

3.3 INTRODUCTION

Developmental enhancers encode information on the time, locations, and levels of gene expression by binding transcription factors (TFs) at specific DNA sequences (Wittkopp and Kalay, 2012). Therefore, the distribution of transcription factor binding sites within developmental enhancers is the key to understanding enhancer function. TF binding sites are typically identified genetically, biochemically, or through phylogenetic footprinting (Spitz and Furlong, 2012), which has generated a restricted and biased model of the distribution and function of regulatory information in enhancers (Crocker and Ilsley, 2017). Our limited knowledge of enhancer structure is illustrated by the fact that synthetic enhancers that attempt to mimic the regulatory output of even the best-studied developmental enhancers have, so far, consistently failed (Crocker and Ilsley, 2017; Goldwater et al., 2010; Vincent et al., 2016).

Despite the importance of enhancers for proper development, the DNA sequences of orthologous enhancers that retain conserved expression patterns often have divergent sequences. These enhancers have maintained conserved regulatory outputs through the gain of binding sites that compensate for the loss of other sites in a phenomenon known as binding site turnover (Berman et al., 2002; Ludwig et al., 1998, 2000; Swanson et al., 2011). Binding site turnover suggests that there is relatively weak selection on individual binding sites, and a more substantial selection of entire enhancers' regulatory output (Wunderlich et al., 2015).

In many cases, changes in enhancer function have driven phenotypic evolution (Long et al., 2016; Wittkopp and Kalay, 2012). Additionally, there are many examples of individual nucleotide changes altering enhancer function during evolution (Arnoult et al., 2013; Gompel et al., 2005; Kvon et al., 2016; Leal and Cohn, 2016; McGregor et al., 2007; Preger-Ben Noon et al., 2018; Thompson et al., 2018). However, we have almost no idea about what enhancer phenotypes are evolutionarily possible. It is also unclear how the enhancer grammar may constrain enhancer evolvability.

Mutational scanning and saturation mutagenesis experiments provide an unbiased survey of inputs in regulatory sequences (Kircher et al., 2019; Mogno et al., 2013; Patwardhan et al., 2009, 2012; Weingarten-Gabbay et al., 2019). These techniques can pick up potential regulatory logic critical for robust and precise expression (de Boer et al., 2020) and identify mutational effects that contribute to phenotypic plasticity (Duveau et al., 2017). It is still challenging to mutationally scan an enhancer in a developmental context. To mutational scan developmental enhancers in *Drosophila*, I oversaw the development of a semi-automated pipeline that allows quantitative measurement of expression patterns across multiple embryonic stages (Fuqua et al., 2021) (See Chapter 2).

3.3   MUTATIONAL SCANNING FOOTPRINTS *E3N* ACTIVITY

To get an overview of the regulatory information encoded within developmental enhancers, Justin Crocker, David L. Stern, and Richard S. Mann chose to design a study for the *shavenbaby (svb) E3N* enhancer from *Drosophila melanogaster. E3N* drives *svb* in a series of stripes on the ventral face of the embryo to differentiate cells into hair-like structures called trichomes (Crocker et al., 2015, 2016b). Each patch of trichomes is referred to as a denticle belt (**Fig 9A**). The trio decided to study *E3N* because its expression pattern is conserved between different *Drosophila* species, yet the sequence has diverged, possibly through binding site turnover or stabilizing selection. Additionally, *E3N* is a relatively small minimalized element (292 bp) that, despite its size, still integrates information from multiple signaling pathways (Payre, 2004) (**Fig 9B and C**).

Jonathan Sager synthesized a library of mutant *E3N* enhancers with a mutational frequency of 2%. Enhancer variants are cloned upstream of an *hsp70* promoter and *lacZ* reporter gene (**Fig 9D**). It was decided to use the *hsp70* promoter because of experimental contingencies and because it has been demonstrated to cover an extensive range of expression (Lagha et al., 2013). Jonathan Sager isolated 749 unique mutants with an average of seven point mutations each. This distribution of mutations mimics both the 2% mutational frequency and about the number of differences between *melanogaster-E3N* and *simulans-E3N.* Out of the 292 bps, 272 were mutated at least once, and the mutational coverage was Poisson distributed (see Methods, Figure 27). During the entirety of my Ph.D., I maintained these fly lines and synthesized other *E3N* mutants.

To screen the embryos, I developed a semi-automated pipeline. This pipeline incorporates the custom-built liquid handling robot called Flyspresso (**Fig 9E**). Flyspresso automatically fixes and stains *Drosophila* embryos. The embryos are then mounted on microscope slides and automatically imaged using an adaptive-feedback confocal microscope plugin (Fuqua et al., 2021) (See Chapter 2). Flyspresso was primarily built by Jeff Jordan.
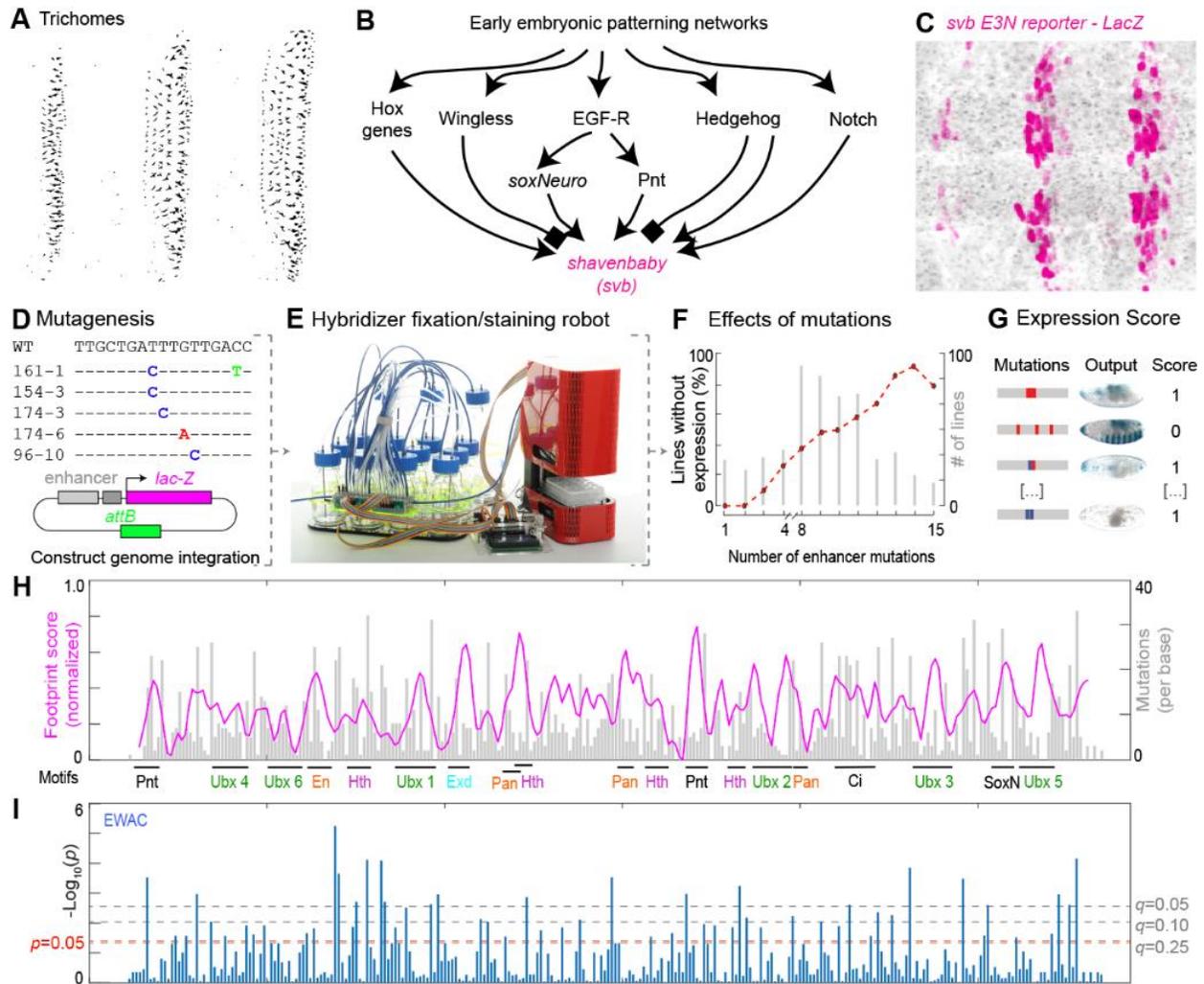
FIGURE 9 *E3N* **IS DENSELY ENCODED AND FRAGILE**

(A) Ventral denticle belts from *Drosophila melanogaster* are composed of hair-like trichome structures. (B) Multiple signaling pathways feed into *shavenbaby (svb)* enhancers. (C) The wild-type *svb E3N* expression pattern on the ventral side of the embryo. (D) Mutant libraries for *E3N* were generated and tested as reporter constructs driving *lacZ*. (E) The Flyspresso liquid handling robot. See Chapter 2. (F) The number of mutations compared with the percentage of lines without detectable *E3N* expression (red). Gray bars mark the number of lines for each category. (G) Example embryos stained with X-gal classified as mutant (1) or WT (0). (H) Footprinting scores along *E3N*. The magenta line represents the score (σ_i, see methods). Higher peaks represent a higher chance that a mutation there will create a gross mutant phenotype. Gray bars indicate the mutational coverage (M_i, see methods). The higher the peak, the more accurate the footprinting score is. (I) Enhancer Wide Association Catalogue (EWAC) scores are the p-values from a log of odds ratio test on every bp of the enhancer. Lines represent p and q values. See methods.

Figure adapted from Fuqua et al., 2020. (G-I) Completed by Timothy Fuqua and Maria Elize van Breugel. Remaining experiments carried out by Timothy Fuqua.

I first wanted to have a broad overview of the effects the mutations were having on the enhancer and asked how much regulatory information is distributed within *E3N*. I advised Master's student Maria Elize van Breugel during this project. With my supervision, Maria Elize van Breugel analyzed 274 lines from the library by screening them with an x-gal staining assay. X-gal staining is a rapid colorimetric assay to mark cells expressing *lacZ*. The method is less sensitive than antibody staining (Fuqua et al., 2021), but helped us understand the gross morphological changes to the expression pattern. From the x-gal stains, Maria Elize van Breugel and I found that as the number of number of mutations increases, so does the percentage of lines without *E3N* expression (**Fig 9F**), suggesting a higher density of regulatory logic than anticipated.

I next asked where the regulatory information was distributed within the *E3N* sequence. To do this, I attempted a "footprinting" technique to statistically identify regions associated with a loss or change to the expression pattern (Belliveau et al., 2018). I classified each line as either WT-like (score = 0) or mutant (score = 1) (**Fig 9G**). This score was applied to every mutated base in each tested line. I divided the sum of these scores by the mutational coverage of the 274 lines, smoothed, and plotted the data **(Fig 9H)** (See Chapter 6). This rudimentary method is far from perfect since it assumes that every base contributes completely non-epistatically to the output expression pattern. Additionally, X-gal staining is limited to gross morphological changes and cannot detect subtle differences in expression. Nevertheless, I discovered that footprinted peaks – or regulatory regions – were scattered across the entire enhancer and many peaks overlapped with previously identified TF binding motifs (Crocker et al., 2015). My colleague Albert Tsai and advisor Justin Crocker helped with some of these calculations.

I wished to treat each base pair independent of the others and carried out a log of odds ratio test on every base of the enhancer after consulting with Judith Zaugg and Jakob Wirbel. I call this test Enhancer Wide Association Catalogue (EWAC), similar to a Genome-Wide Association Study (Ozaki et al., 2002). The EWAC test also identified significant regions that contribute to gross phenotypic changes (**Fig 9I**). 22.9% (67/292) changed the pattern (Q=0.25). Both the footprinting and EWAC analysis revealed that regulatory information is distributed across *E3N,* and critical regions to generating the expression pattern.

3.4   SINGLE POINT MUTATIONS ALTER *E3N* EXPRESSION

The x-gal staining provided a way to quickly analyze hundreds of enhancer variants, gross phenotypic changes, and critical regions necessary for driving *E3N* expression. The assay, however, did not help me quantify subtler phenotypic differences. To this end, I used Flyspresso to fix and stain 117 of the reporter lines and imaged their expression patterns using my adaptive feedback confocal microscopy pipeline (see Chapter 2) (Conrad et al., 2011; Fuqua et al., 2021; Tischer et al., 2014).

I first focused on the 18 lines within the collection that harbored only single point mutations (**Fig 10A-S**). From the lines analyzed, I categorized phenotypic effects into four general categories: levels, state, location up, and location down (**Fig 10T**). "Levels" refer to the nuclear intensity, "State" refers to missing nuclei in the pattern, "Location up / down" refers respectively to the thickening and thinning of the stripes. All 18 of these lines showed a significant decrease in nuclear intensity ranging in effect size ("Levels"). Furthermore, some of the lines exhibited "State" and "Location up/down" effects together (~61%) (**Fig 10U**).

It is frequently assumed that sequence conservation directly correlates with the strength or importance of a regulatory sequence, even though there is an overwhelming amount of evidence to dispute this claim (Kvon et al., 2014; Snetkova et al., 2021). I tested the effect sizes of the single mutations and compared them to two different genomic conservation scores: PhyloP 27 Species (**Fig 10V**) and PhyloP 124 Species (**Fig 10W**). Unsurprisingly, I found that PhyloP estimates did not correlate with effect sizes for either metric (27 species, $R^2 = 0.25$, two-tailed >0.2) (**Fig 10X**) and (124 species, $R^2=0.01$, two-tailed p > 0.9) (**Fig 10Y**) (Kent et al., 2002; Kwasnieski et al., 2012; Pollard et al., 2010). These results suggest that *E3N* has undergone significant binding site turnover and that sequence conservation should not be used to predict effect sizes.
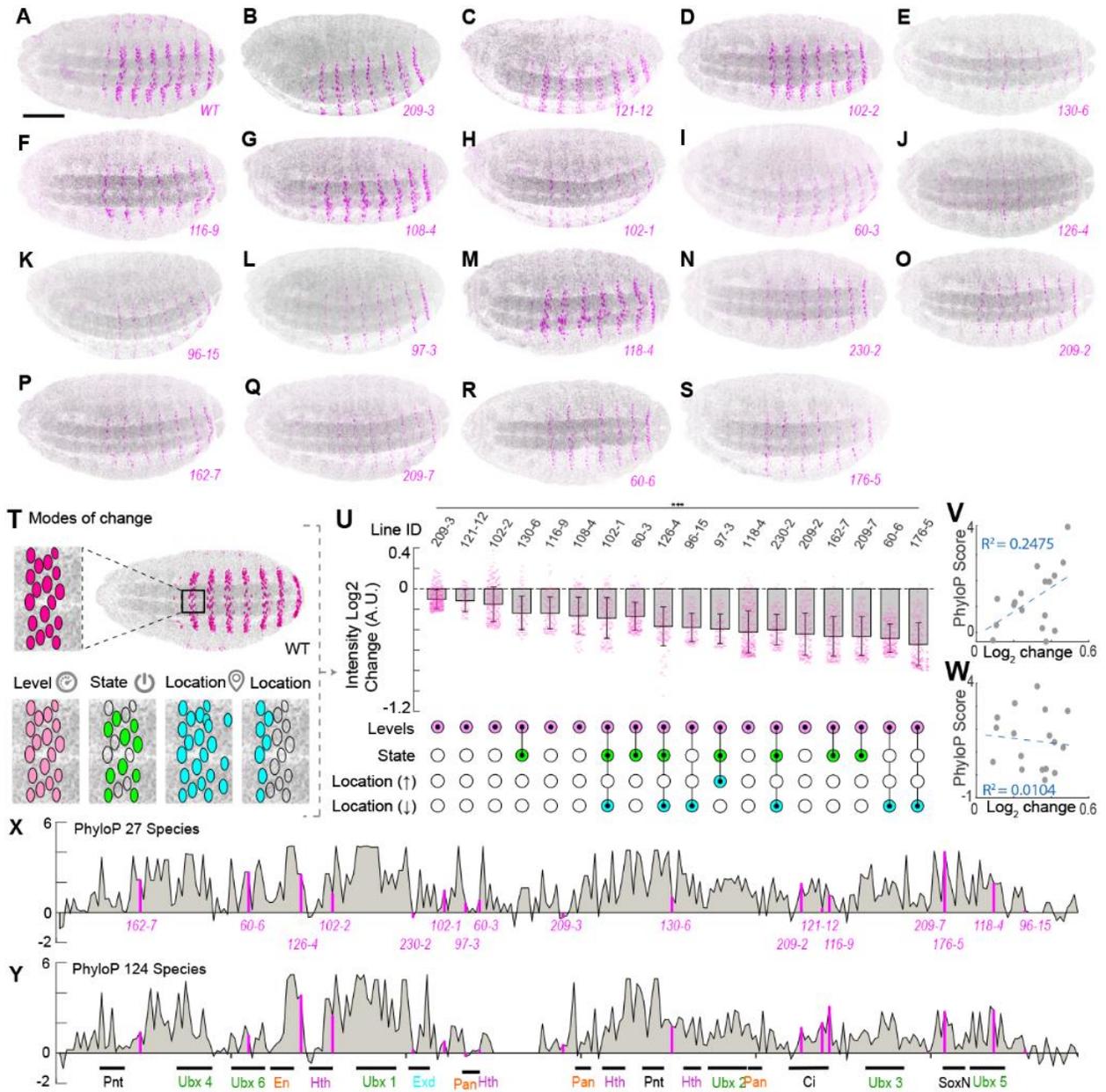
FIGURE 10 POINT MUTATION EFFECT SIZES AND CONSERVATION DO NOT CORRELATE

(A-S) Representative embryos from the *E3N* reporter library carrying single point mutations. The effect sizes greatly vary between lines. (T) Mutant phenotypes are divided into four categories: Levels, State, Location up, and Location down. (U) Phenotypes for the single mutant lines. Top: nuclear intensities. Each point represents a nucleus. Whiskers are +/- one standard deviation. Bottom: filled colored circles indicate the phenotype. (V-W) PhyloP conservation scores and their correlation with the average nuclear change for 27 or 124 species. (X-Y) PhyloP scores and single mutation coordinates marked along the *E3N* enhancer sequence.

Figure adapted from Fuqua et al., 2020. Experiments carried out by Timothy Fuqua.

Based on the evidence provided so far, I concluded that *E3N* is a densely encoded regulatory element. I support my argument for this claim first with the number of mutations correlating with an increased loss of expression (**Fig 9F**). The EWAC and footprinting calculations show peaks throughout the entire enhancer sequence (**Fig 9H and I**). Additionally, all of the single mutations created a significant quantifiable effect on the expression pattern (**Fig 10A-U**), where 61% of these mutants additionally had state and location changes (**Fig 10U**). Thus, most base pairs somehow contribute to the WT expression pattern.

3.5 IDENTIFYING HTH BINDING SITES ASSOCIATED WITH PHENOTYPIC EVOLUTION

I wanted to validate one of the peaks that Maria Elize van Breugel and I had identified from the EWAC and footprinting scores. One particular peak overlapped with a Homothorax (Hth) motif (Choo et al., 2011; Noyes et al., 2008), which I call *Hth-2*. To validate the motif, I selected mutants within my library that contained point mutations within the motif and no more than two mutations elsewhere to minimize confounding results (**Fig 11A and B**). All of the lines exhibited a similar drop in nuclear intensity, and the stripes collapsed to single rows of cells with state-like effects. Justin Crocker also created a controlled knockout of the *Hth-2* motif, which drove the same expression pattern. Hth may be interacting with other binding sites within *E3N* such as Ubx (**Fig 10C**).

I had previously demonstrated that sequence conservation is not correlated with the effect size for single base pair mutations. This *Hth* binding site in *D. melanogaster* is not conserved in *D. virilis* (**Fig 11D and E**) yet is critical for driving WT *E3N* expression. To understand why a critical motif was not conserved in *D. virilis*, Justin Crocker compared the *melanogaster E3N* (**Fig 11F**) and the *E3N Hth-2* targeted knockout (**Fig 11G**) to a *virilis E3N* reporter construct (expressed in *D. melanogaster*) (**Fig 11H**). Justin Crocker found that the *Hth-2* deletion in *mel-E3N* resembled the *vir-E3N* expression pattern. To see if the *Hth-2* motif caused the expression loss in *vir-E3N*, Justin Crocker rescued the site with the matching *melanogaster* sequence (**Fig 11I**). The rescued construct strongly resembled the *melanogaster E3N* and almost fully restored the pattern (**Fig 11J**), suggesting that the *Hth-2* motif contributes to the loss of *E3N* expression in *D. virilis.* Justin Crocker also looked at the trichomes in *D. virilis* to see how the loss of *E3N* would affect the denticle belts, and discovered that *D. virilis* expresses fewer trichomes in the ventral denticle belts than *D. melanogaster* (**Fig11K and L**).
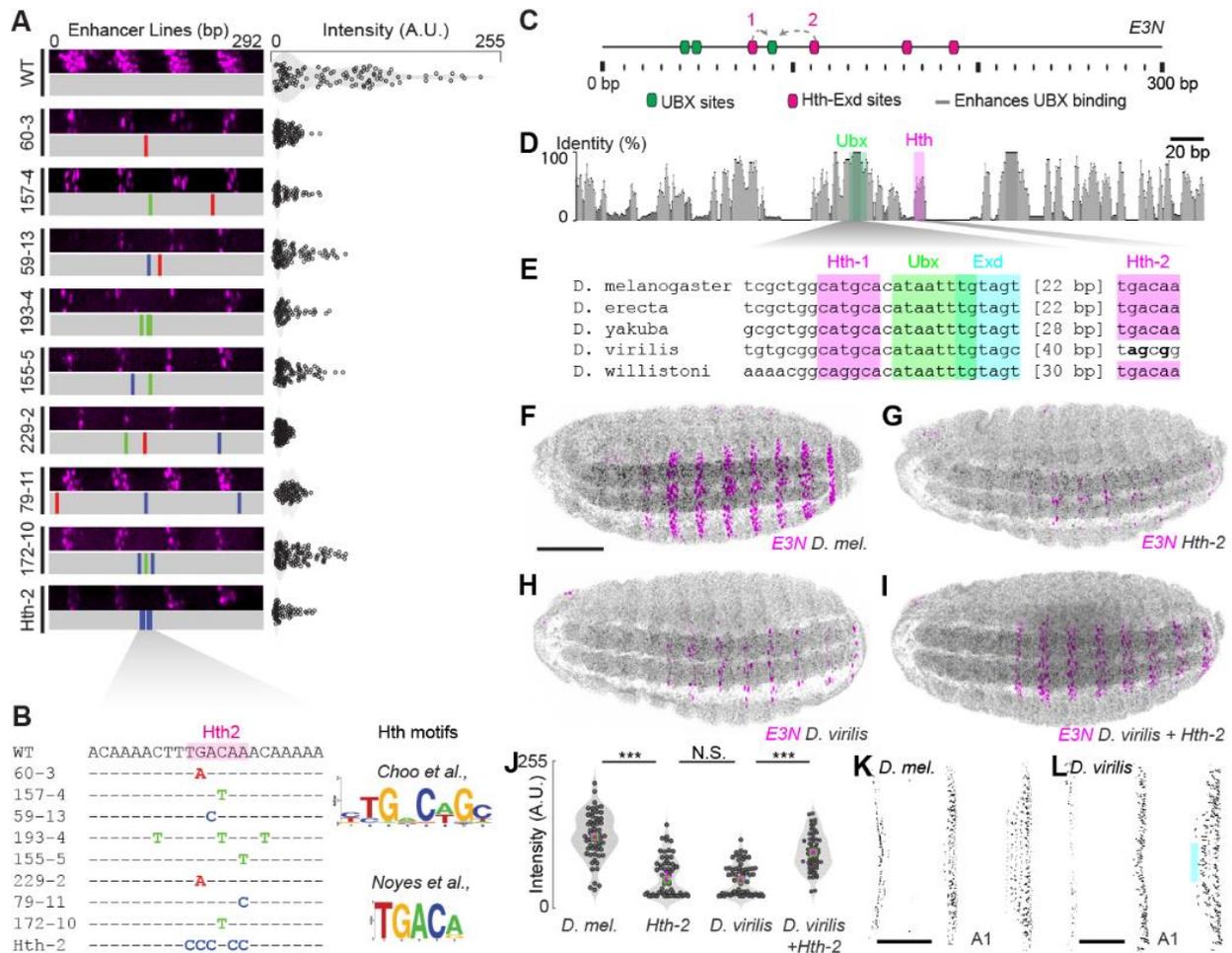
FIGURE 11 IDENTIFYING THE *HTH2* BINDING SITE ASSOCIATED WITH A LOSS OF TRICHOMES

(**A**) Close-up images of the *E3N* expression patterns and the nuclear intensities for individual cells. (**B**) Lines contain point mutations within the *Hth-2* motif. (**C**) The *E3N* enhancer contains multiple Hth and Ubx binding sites that may work cooperatively. (**D and E**) The *Hth-2* motif is not conserved in *D. virilis.* (**F**) *D. melanogaster E3N::lacZ* reporter construct inserted in *D. melanogaster.* (**G**) *D. melanogaster E3N Hth-2* targeted knockout reporter construct inserted into *D. melanogaster.* (**H**) *D. virilis E3N lacZ* reporter construct inserted into *D. melanogaster.* (**I**) *D. virilis E3N + melanogaster Hth-2* rescued site inserted into *D. melanogaster.* (**J**) Nuclear intensities for individual cells with violin plots (n=50, 10 embryos each). Asterisks indicate $p < 0.01$. (**K**) *D. melanogaster* cuticle prep. (**L**) *D. virilis* cuticle prep. Blue box highlights missing ventral trichomes.

Figure adapted from Fuqua et al., 2020. (**F-L**) Completed by Justin Crocker. Remaining experiments carried out by Timothy Fuqua.

3.6 *E3N* MUTANTS REVEAL PLEIOTROPIC RELATIONSHIPS

The dense regulatory information encoded within *E3N* and most nucleotides may create or bias the appearance of novel phenotypes (Arthur, 2002; Gilbert, 2006; Smith et al., 1985; Uller et al., 2018). To further explore potential biases and constraints on *E3N* evolution, I analyze the effects mutations have on Ubx binding sites. It is well established that Ubx binds to *E3N* at homotypic binding clusters at a low affinity (Crocker et al., 2015; Rastogi et al., 2018).

I first compared our findings to a previously characterized high-affinity Ubx mutant (**Fig 12 A-D**) (Crocker et al., 2015). The high-affinity *E3N* variant drives more expression than the WT in the anterior, early stripe (stage 14) and between the stripes themselves (**Fig 12E**). Chaitanya Rastogi calculated the total Ubx affinities computationally for the entire library using a computational tool: NRLB (Rastogi et al., 2018) and I selected lines with minimal mutations and the most extensive range of Ubx affinities (maximum three mutations). Using the automated pipeline, I imaged embryos from these lines across various developmental stages and analyzed the anterior, early stripes, and stripe intensities. Excitingly, these phenotypes also strongly correlated with the total Ubx affinity (**Fig 12F and F'**). This correlation could be explained by the Hox Paradox (see Chapter 1), where higher affinity homeodomains recruit other Hox genes and cause promiscuous binding (Crocker et al., 2016).

Chaitanya Rastogi additionally ran the NRLB algorithm on *Drosophila virilis* for me to identify low-affinity Ubx binding sites (**Fig 13A and B**). I noticed that the total Ubx affinity has not drastically changed, but Ubx binding sites have turned over between the species. This simple analysis reveals that *E3N* may also be undergoing binding site turnover. I wanted to see if changes in total affinity were sufficient to drive ectopic trichomes on the larvae. To this end, Justin Crocker and I created a minimalized construct in which the Ubx high-affinity mutant drives the cDNA of *svb.* This construct exhibits additional trichomes in the stripes and between the segments (**Fig 13C-E**).
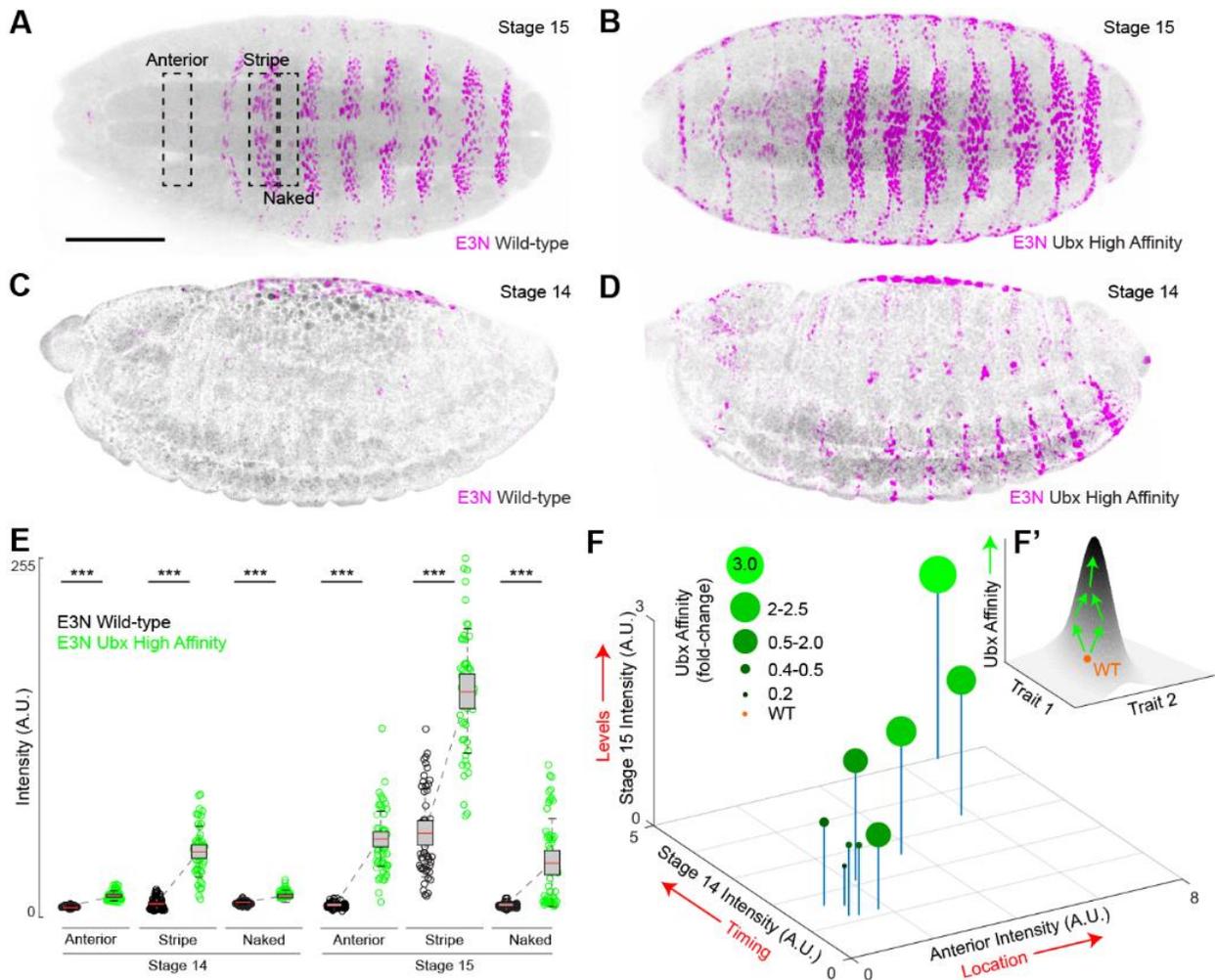
FIGURE 12 UBX AFFINITY, EXPRESSION LEVELS, LOCATION, AND TIMING CORRELATE IN *E3N* MUTANTS

(**A-D**) Stag 14 and 15 *D. melanogaster* embryos with the *WT E3N lacZ* and *Ubx high-affinity E3N lacZ* reporter construct. (**E**) Intensities for the stage 14 and 15 stripes (intra), naked (inter-stripe), and anterior regions between *WT E3N lacZ* and *Ubx high-affinity E3N lacZ.* Box plot whiskers = 95% confidence intervals, the center red line is the mean, and upper and lower limits are the standard deviation. P-values were calculated using a two-tailed t-test, and asterisks indicate p<0.01. (**F**) Nuclear intensities in the anterior region and stripe regions at stages 14 and 15 for lines with different total Ubx affinity. (**F'**) Model of Ubx affinity linking these traits.

Figure adapted from Fuqua et al., 2020. Experiments carried out by Timothy Fuqua. Chaitanya Rastogi ran the NRLB algorithm to predict the total Ubx affinities.
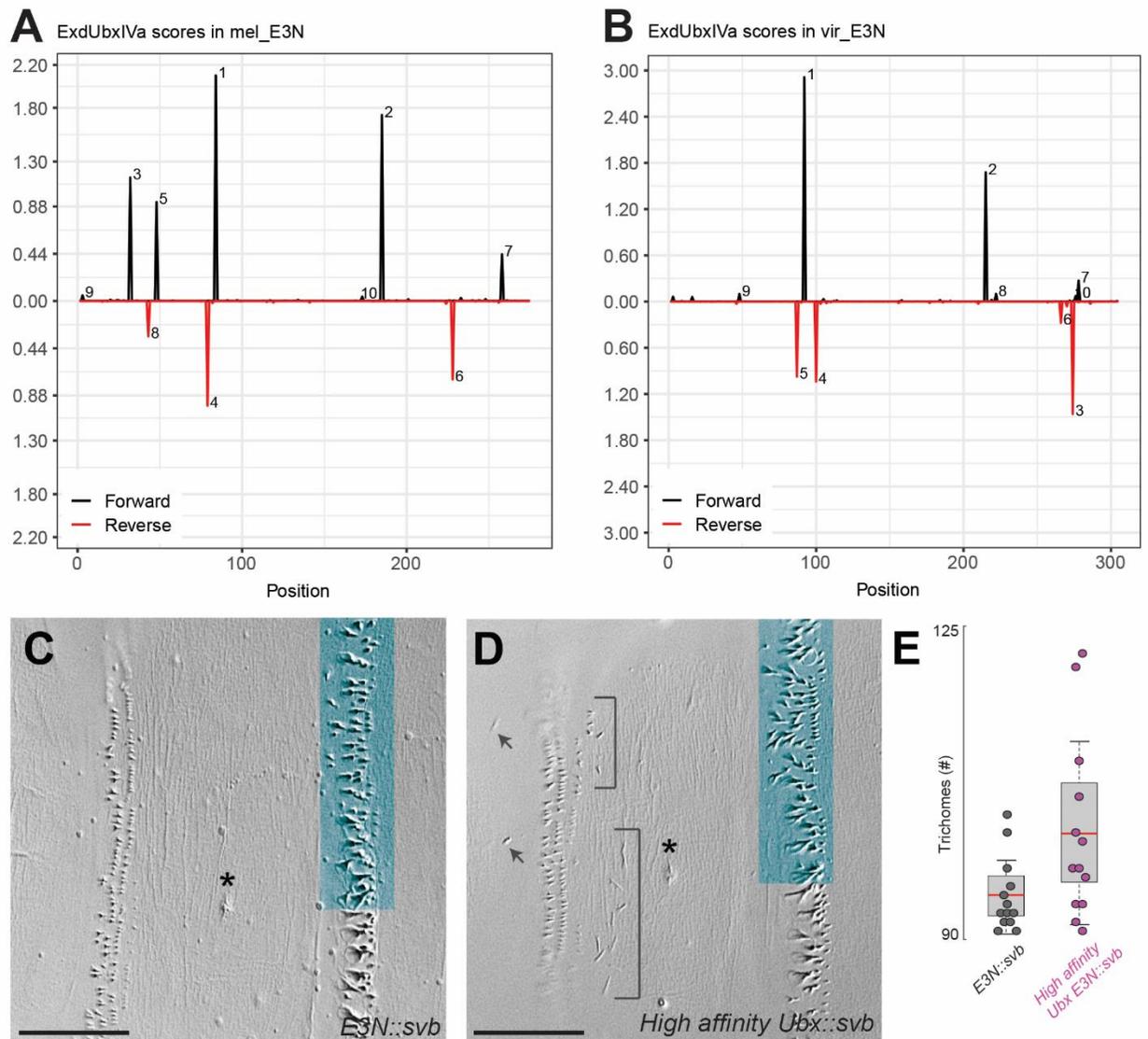
FIGURE 13 UBX BINDING SITE TURNOVER AND ECTOPIC TRICHOMES

(**A and B**) Readout from the NRLB algorithm shows the predicted binding affinities for Ubx on the 5' strand (black) and 3' strand (red). Binding sites and their affinities are illustrated for *D. melanogaster* (**A**) and *D. virilis* (**B**). (**C and D**) *E3N* enhancer variants drive the *svb* cDNA for WT *E3N* (**C**) and *E3N Ubx High Affinity* (**D**). (**E**)Trichome counts from the A1 segment are plotted as boxplots (n=13, p < 0.02). See Tsai et al., 2019b for trichome quantification.

Figure adapted from Fuqua et al., 2020. (**A-B**) NRLB analysis done by Chaitanya Rastogi. (**C-E**) Completed by Timothy Fuqua and Justin Crocker.

The pleiotropic effects of Ubx affinity suggested that other binding sites could also generate linked phenotypes. I went back to the 117 lines that I antibody-stained and examined their activities for ectopic *E3N* expression. I found a general trend that the number of lines with ectopic expression increases with the number of mutations (**Fig 14A**). I also analyzed the 274 lines from the x-gal staining Maria Elize van Breugel's screen, and found a similar trend (**Fig 14B**). From the antibody-stained lines, 32.5% (38/117) created early expression at stage 15, and ~34% (13/38) of these lines drove expression between the stripes in the naked region. The lines also drove ectopic expression in the wing and haltere discs, mouth hooks, and other domains. I imaged these lines again with Maria Elize van Breugel in **Fig 14C-J**. In all cases, the WT *E3N* expression pattern was perturbed, suggesting a dense amount of pleiotropic and regulatory information in *E3N.*

I attempted to map regions associated with extensive pleiotropy by calculating the footprinting and X-gal scores for the 274 lines. This time, each line was scored for being WT-like (0) or pleiotropic (1). Unlike my findings from the first assay (see Figure 9), I did not find clear peaks associated with pleiotropy (**Fig 14K**). This could be because there are too many confounding bases, or too many ways for the enhancer to create pleiotropic expression. I also tried the EWAC analysis, and identified some pleiotropic regions (47/292 bp, ~16%) significantly associated with pleiotropy (q=0.25) (**Fig 14L**).
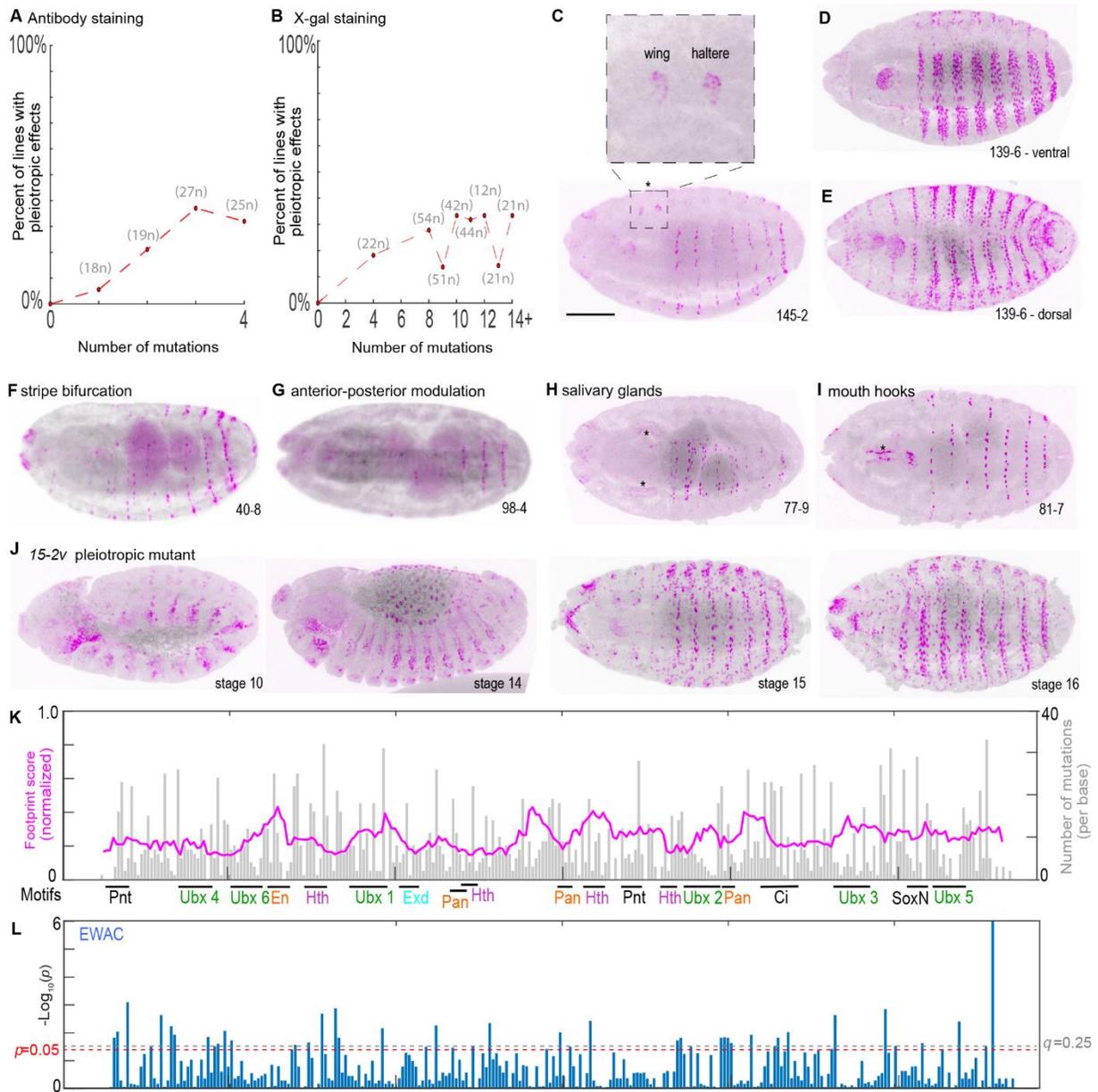
FIGURE 14 *E3N* MUTANTS CREATE PLEIOTROPIC EFFECTS

(**A and B**) The relationship between the mutation number and pleiotropy / ectopic expression for antibody staining (**A**) and x-gal staining (**B**). (**C-J**) Examples of pleiotropic phenotypes. (**C**) *145-2* creates expression in the wing and haltere primordia. (**D and E**) *136-3* creates thicker ventral stripes (**D**) and ectopic dorsal stripes (**E**). (**F**) *40-8* is missing a row of cells within each stripe. Each stripe is "bifurcated". (**G**) *98-4* stripe strength is different along the anterior-posterior axis. (**H**) Savivary gland expression is found in line *77-9*. (**I**) *81-7* creates ectopic expression in the developing mouth hooks. (**J**) *15-2v* activates early and in many other regions of the embryo. (**K**) Footprinting scores for pleiotropic effects. Magenta is the score (σ_i, see methods) and the gray bars are mutational coverage (M_i, see methods). (**L**) EWAC scores. Blue peaks are p-values from a log of odds ratio test on the association of each base creating pleiotropic effects. Dashed lines are for p and q values (Storey and Tibshirani, 2003). See methods.

From the significant EWAC bases, binding motifs for the TF Pangolin (Pan) (Chang et al., 2008) were all significantly associated with pleiotropic expression. I reanalyzed these lines and found that ~46% (20/43) had ectopic expression patterns and ~42% (18/43) reduced expression overall. From this analysis, I noticed that all of the motifs for the TF Pan were significantly associated with ectopic expression. Pan is a TF part of the Wingless or Wnt signaling pathway, and flies with *wingless* mutations drive ectopic levels of Svb, creating lawns of trichomes (**Fig 15A and B**). Pan is a known a repressor of *svb* between the denticle bands in the "naked" region (Bejsovec, 2006) (**Fig 15C**).

I wanted to validate the EWAC scores and selected lines with point mutations in the identified Pan motifs. This included a line: *97-3* which was overlapping with the *Hth-2* motif and created ectopic expression (**Fig 15D and E**). From the 13 lines Justin Crocker and I selected, we found a correlation between lower expression levels within the stripes and higher levels of ectopic expression between the stripes in the naked region (**Fig 15F-G'**). I did not identify any lines which resembled a *wingless* mutant (**Fig 15B**). These results were interesting because it showed that it is phenotypically possible to create a lawn of trichomes, but the enhancer is not able to.
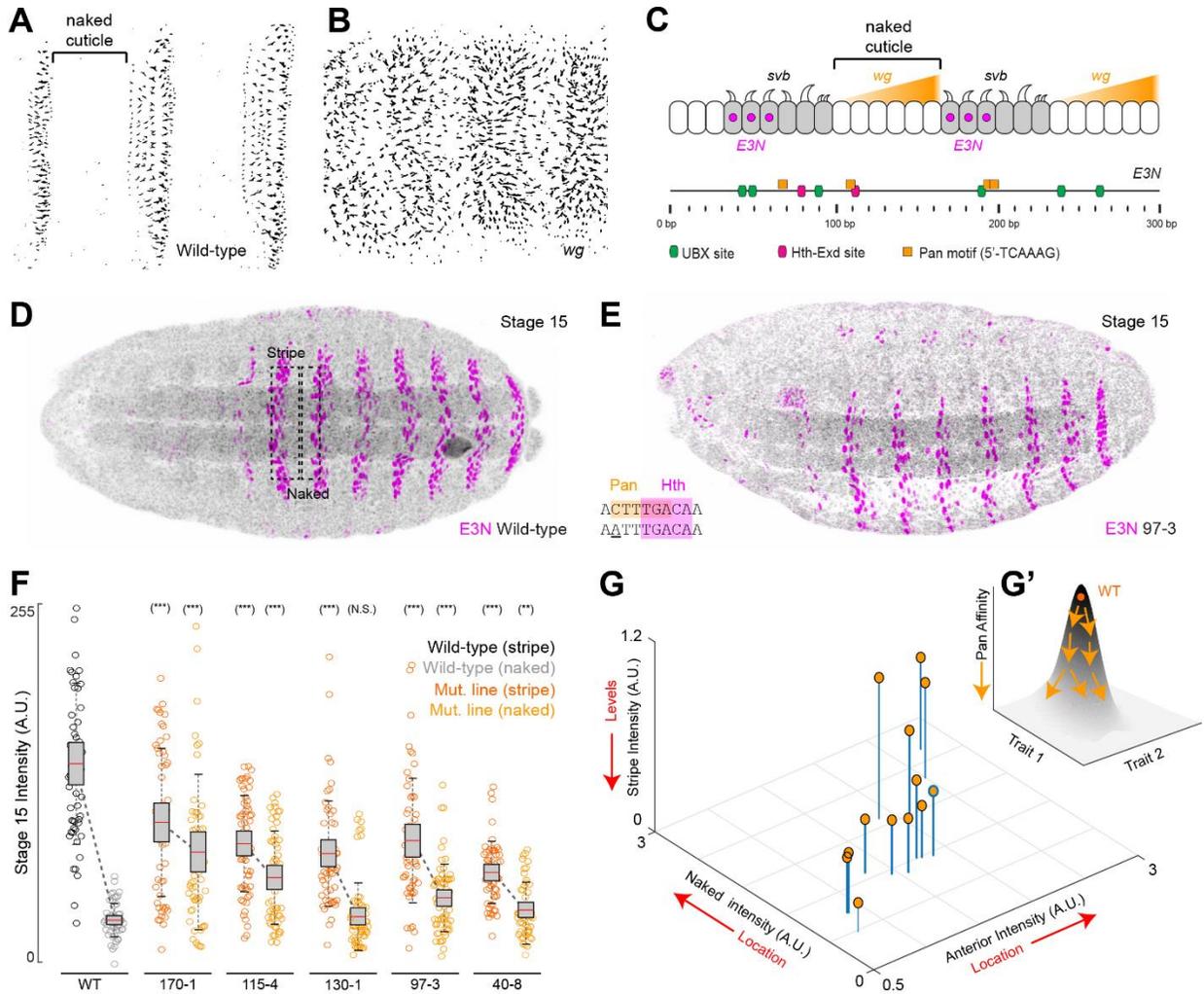
FIGURE 15 PANGOLIN MUTANTS CREATE PLEIOTROPIC EXPRESSION IN *E3N*

(**A and B**) Cuticle preps of the ventral trichomes for a WT fly line and a line deficient for *wingless* (*wg^CX4*). (**C**) Model of Wingless signaling and Pan repressing the naked stripe region. (**D and E**) Stage 15 embryos with *E3N lacZ* reporters for the WT enhancer (**D**) and a mutant (line *97*-3) with a point mutation in a Pan and Hth-2 motif (**E**). (**F**) Boxplots map expression in the stripe and naked regions. Asterisks indicate p < 0.01. Lines with mutations in Pan motifs plotted for stripe intensity, naked intensity, and anterior expression (**G**). The findings show a correlation between Pan affinity and the linkage of these traits (**G'**).

Figure adapted from Fuqua et al., 2020. Experiments carried out by Timothy Fuqua.

3.7 THE MOLECULAR MECHANISMS UNDERLYING *E3N* MAY CONSTRAIN ITS EVOLUTION

The dense encoding and highly pleiotropic regulatory information that creates the precise *E3N* expression pattern may be constraining possible evolutionary paths (Sabarís et al., 2019). I wanted to see what kind of phenotypic variation was available. David L. Stern had previously prepared cuticle preps for 60 different *Drosophila* species (**Fig 16A and B**). Justin Crocker imaged all of them and observed many examples of this in species and highlighted a few examples (**Fig 16C-F**). Throughout this screen, I identified many examples of ways to break the *E3N* enhancer, which would result in a loss of trichomes. The screen also suggested that the molecular underpinnings of *E3N* make it infeasible to create ectopic expression without compromising on stripe intensity. Consistent with these findings, Justin Crocker did not observe any trichome expression between the stripes in the naked region. These findings suggest that maybe the molecular mechanisms underlying *E3N* may constrain the evolution of where new trichomes can and cannot evolve.
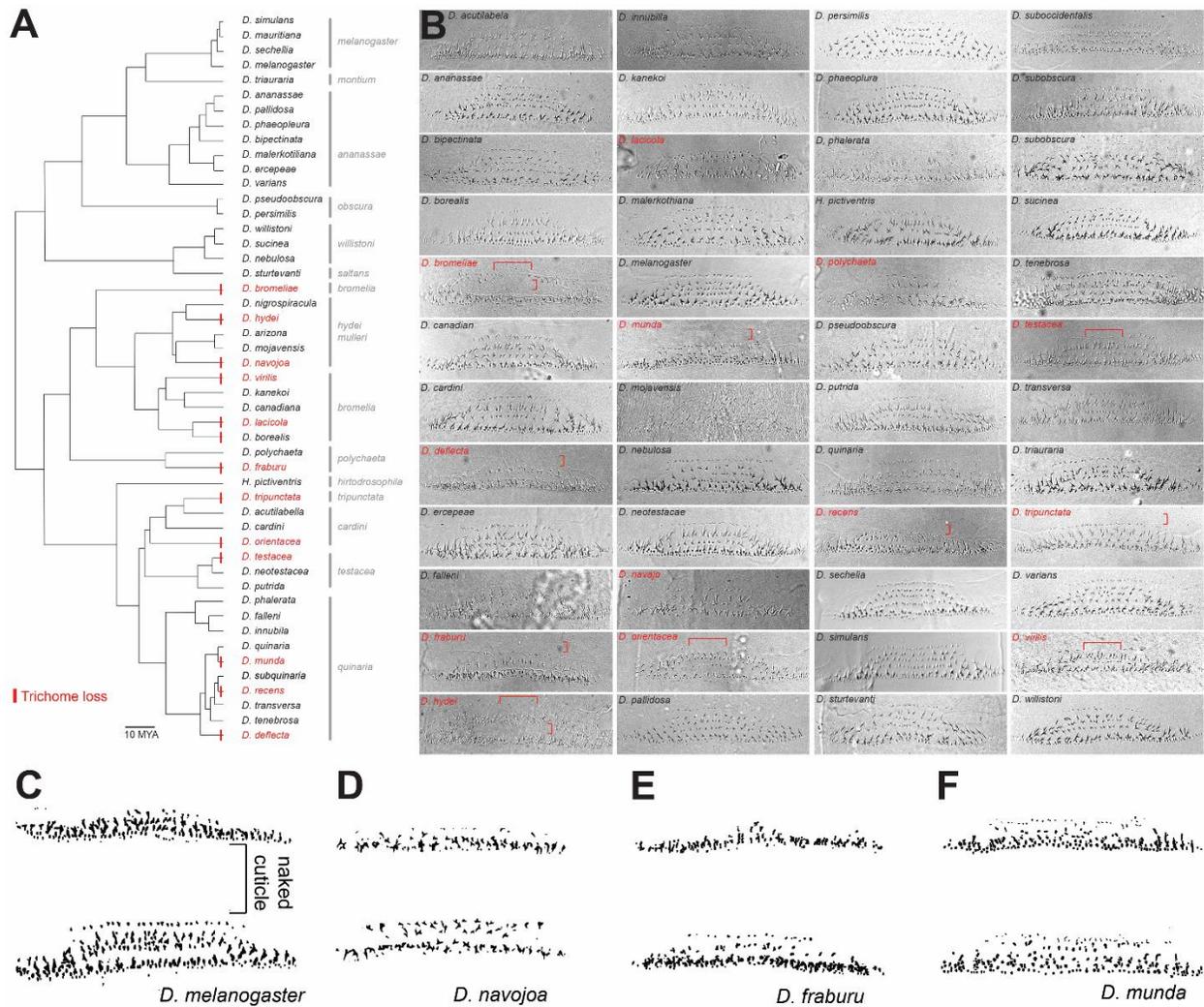
FIGURE 16 CUTICLE PREPS SHOW REDUCED TRICHOMES BUT NOTHING IN-BETWEEN

(A) A phylogenetic tree containing ~150 million years of *Drosophila* evolution. Red marks lines with a loss of trichomes. (B) The cuticle preps for these species. (C-F) Close-up images of cuticles in *D. melanogaster* (C), and missing trichomes in *D. navojoa* (D), *D. fraburu* (E), and *D. munda* (F).

Figure adapted from Fuqua et al., 2020. Experiments completed by David L. Stern and Justin Crocker.

3.8 DISCUSSION

Using a mutational scanning approach, I unbiasedly dissected the regulatory logic and evolvability of a developmental enhancer: *E3N.* Most mutations studied had effects on the *E3N* expression pattern, which suggests that *E3N* is densely encoded throughout the entire sequence. Additionally, many mutants exhibited pleiotropic effects. Together, this dense and pleiotropic information may constrain the evolvability of *E3N.* This is an exciting conclusion as many enhancer sequences are turning over while maintaining their expression patterns (Long et al., 2016), but is congruent with other recently published studies (Kvon et al., 2020; Kwasnieski et al., 2012; Le Poul et al., 2020; Swanson et al., 2011) (see Chapter 5). This is one of the first studies to reveal the constraints acting upon cis-regulatory element evolution, elements that are primarily responsible for phenotypic variation and the evolution of morphology (Prud'homme et al., 2007). The underestimation of regulatory information encoded within enhancers may explain why it is still challenging to synthesize functional enhancers or predict their expression patterns based on their sequences (Crocker and Ilsley, 2017; Vincent et al., 2016).

One way to interpret these results is that *E3N* has evolved along constrained evolutionary paths despite its dense regulatory encoding (Blomberg and Garland, 2002; Swanson et al., 2011; Uller et al., 2018). There are, however, a few things to consider before accepting this conclusion. To begin with, I tested all of the *E3N* variants as reporter constructs integrated on the third chromosome (*attP2*), even though *svb* is on the X-chromosome. Testing all of these elements in an entirely different chromatin environment may have had unintentional consequences on the expression pattern – mainly if regulatory logic adjacent to the insertions also contributes to the final output (Crocker and Stern, 2017; Dey et al., 2015; Kim et al., 2013)

It is also worth considering that enhancers do not function as independent units but cooperatively with one another at the locus (Long et al., 2016). Enhancers within a locus can buffer transcriptional noise (Waymack et al., 2020), expression patterns (Delker et al., 2019; Frankel et al., 2010; Hong et al., 2008; Perry et al., 2010; Tsai et al., 2019), and even compete with each other at the promoter (Bartman et al., 2016; Bothma et al., 2015; Fukaya et al., 2016; Scholes et al., 2019). The *svb* enhancers – *E3N* included – have been shown to physically interact

with each other using high-resolution microscopy and genetics (Tsai et al., 2017, 2019). Thus, minimalized elements are fragile, but their collective expression patterns are likely more robust.

Finally, all of the experiments were carried out under the control of the *hsp70* promoter and not the *svb* promoter. The *hsp70* promoter may have modulated or amplified expression artifacts (Zabidi et al., 2015). Carrying out experiments like this at the native locus are challenging but possible using genetic tools such as CRISPR (Kvon et al., 2020; Lamb et al., 2017).

Studies and genetic dissections of the other *svb* enhancers also reveals many examples of pleiotropy across the larvae and adult tissues (Al Hayek et al., 2021; Kittelmann et al., 2018, 2021; Preger-Ben Noon et al., 2018; Stern and Frankel, 2013). In particular, the *E6* enhancer contains a homotypic cluster of activator sites, which required it to evolve binding sites for very strong repressors to escape the redundancy (Preger-Ben Noon et al., 2018). This study is also evidence for enhancers evolving along constrained evolutionary paths.

From the Pan mutant series (**Fig 15**), I found that ectopic expression between the denticle belts was associated with overall lower expression levels. I demonstrated that it was genetically possible to create a lawn of *svb* expression by removing Pan inputs *in-trans* (**Fig 15B**), but eliminating the inputs *in-cis* leads to a loss of expression. Additionally, Justin Crocker and David L. Stern did not identify any species expressing trichomes in this domain when looking at different *Drosophila* species. These findings could suggest that the molecular underpinnings of *E3N* constrain this expression. Still, it is also possible that having trichomes between the stripes has a lower fitness effect since trichomes serve a variety of purposes (van Breugel and Dickinson, 2017; Inestrosa et al., 1996; Kittelmann et al., 2021).

Other forms of grammar such as low-affinity binding sites, DNA shape, TF spacing, cooperativity, orientation, and the number of sites may also constrain evolvability (Jindal and Farley, 2021; Payne and Wagner, 2014). Low-affinity sites encode precise spatiotemporal gene expression patterns (Crocker et al., 2015; Farley et al., 2015; Gaudet, 2002). *E3N* encodes low-affinity Ubx sites that are highly specific for Ubx but consequently drives low levels of activation. To circumvent this weak activity, *E3N* encodes homotypic clusters of these low-affinity sites to

increase expression and to confer robustness (Crocker et al., 2015). Ubx additionally works cooperatively with Exd and Hth (Jaw et al., 2000; Merabet et al., 2007). Thus, low-affinity Ubx binding sites (Crocker et al., 2015; Farley et al., 2015; Rastogi et al., 2018), the number of activator sites (Preger-Ben Noon et al., 2016), and cooperative Ubx:Hth: Exd interactions (Jaw et al., 2000; Merabet et al., 2007) may also constrain *E3N* evolvability. How other forms of enhancer grammar such as DNA shape, TF spacing, and orientation constrain *E3N* evolution is still unclear.

I also observed binding site overlap between Pan and Hth motifs in *E3N.* Overlapping activators and repressors can encode sharp expression boundaries in synthetic enhancers (Crocker and Stern, 2017). In *E3N,* Pan represses between the stripes, and Hth activates the stripes. This codependency of the factors themselves may also be contributing to sharp stripe expression. It may be evolutionarily impossible to disentangle these inputs without drastically changing the expression pattern. These different gene expression parameters can explain strongly conserved elements (Bejerano et al., 2004; Cande et al., 2009; Plessy et al., 2005). This experiment suggests that enhancers are densely encoded and highly pleiotropic. Dense encoding and pleiotropy may limit evolvability and challenge the view of enhancer modularity itself (Sabarís et al., 2019).

I hope in the future to apply my pipeline to other developmental enhancers and cis-regulatory elements. Studying enhancers using mutational scanning reveals possible paths of evolution, and pairing the findings with actual standing variation in the wild may help us begin to predict evolution.

3.9 DATA AVAILABILITY

Access to all original images, files, and data can be downloaded here:

https://www.embl.de/download/crocker/Dense_and_pleiotropic_regulatory_information_in_a_developmental_enhancer/index.html

4 PREDICTABLE ROBUST, FRAGILE, AND EVOLVABLE ZONES OF *E3N* EXPRESSION

"I argue that the origin of novel characters and novel body plans is one of the most important but least researched questions in evolutionary biology"

- Günter P. Wagner, 2014

4.1 ABSTRACT

Cis-regulatory elements are rapidly evolving DNA sequences which control gene expression patterns and the evolution of morphology. To better understand the potential and limitations of cis-regulatory evolution, random mutagenesis experiments are carried out to reveal potential developmental biases and trends in cis-regulatory evolution. Here, I further explore the phenotypic evolution of a developmental enhancer: *E3N*, by screening mutant variants of the enhancer and using advanced bio-image analysis and gene expression atlases to identify trends and developmental biases. I find that the wild-type *E3N* enhancer encodes a robust underlying expression pattern, as well as an overlying fragile pattern. In addition to these robust and fragile zones of expression, I, Noa Borst, and Justin Crocker identify permissive zones of ectopic expression, but only in the ectoderm. These evolvable phenotypes can emerge through both the gain of an activator or the loss of a repressor binding site. I argue that these robust, evolvable, and fragile expression zones reveal developmental biases on *E3N.* These biases may exist because the expression components themselves are under different selective pressures.

4.2   SCIENTIFIC CONTRIBUTIONS

I planned this project with my advisor Justin Crocker. I carried out the screening of the 100 lines and wrote automated image analysis scripts to analyze the expression patterns. I additionally screened the antibodies for the gene expression atlas. I advised Master's student: Noa Borst and together with Justin, we went through all 100 of the lines to identify lines with ectopic expression. Together with Noa Borst, Noa Borst and I screened the lines in Figure 21G-I for ectopic wing/haltere expression. I also screened the entire combinatorial library for Figure 22.

Timothy Fuqua (me) planned the initial experiment with Justin Crocker. Timothy Fuqua collected and fixed the 100 lines and controls, stained the embryos, and imaged them using the imaging pipeline described in Chapter 2. Timothy Fuqua wrote the image analysis scripts for Figure 17 and analyzed the data to come to the conclusion in Figure 17 and 18. Timothy Fuqua fixed, stained, imaged, and analyzed the antibodies from the DSHB in Figure 19 and Figure 20. Timothy Fuqua, Noa Borst, and Justin Crocker analyzed the lines manually in Figure 21A-F. Together, Noa Borst and Timothy Fuqua screened the lines in Figure 21G-I for ectopic wing/haltere expression. Timothy Fuqua created the combinatorially complete library, fixed, stained, imaged, and analyzed the mutants in Figure 22.

Noa Borst was a Master's student under my advisory. Under my supervision, Noa Borst and I screened the lines in Figure 21G-I for ectopic wing/haltere expression. Noa Borst, Justin Crocker, and I also analyzed all of the data manually to identify the ectopic mutants in Figure 21A-F.

Justin Crocker was the Principle Investigator. Justin Crocker planned the initial experiment with me. Justin Crocker, Noa Borst, and I analyzed the data manually to identify the ectopic mutants in Figure 21A-F. Justin Crocker additionally provided feedback and mentorship.

Anna Kreshuk and her group developed and maintain Ilastik (Berg et al., 2019), a user-friendly machine learning algorithm that I used for image analysis in Figure 17. Thank you Anna for keeping science as open and accessible as possible.

4.3   INTRODUCTION

Cis-regulatory elements such as developmental enhancers are DNA sequences that integrate information from multiple signaling pathways throughout development to drive gene expression patterns across space and time (Jindal and Farley, 2021; Small and Arnosti, 2020). It was initially postulated in the cis-regulatory hypothesis that enhancers are highly robust and plastic to mutations, evolving faster than coding sequences, and are the primary drivers of phenotypic evolution (Prud'homme et al., 2007; Stern and Orgogozo, 2008; Wagner and Lynch, 2008). The cis-regulatory hypothesis emphasizes less constraint on enhancers compared to coding sequences. The view, however, does not acknowledge to what extent enhancer constraint limits or biases evolution.

In Chapter 3, I recently dissected a developmental enhancer for the *shavenbaby* (*svb*) locus called *E3N* and discovered that its late-stage stripe expression pattern is densely encoded with pleiotropic information that likely constrains its evolution (Fuqua et al., 2020, 2021). Other recent enhancer studies also begin to support this claim (de Boer et al., 2020; Goldwater et al., 2010; Kvon et al., 2020; Le Poul et al., 2020), suggesting that the cis-regulatory hypothesis may be an oversimplified model to explain the evolution of phenotypes and morphology (Sabarís et al., 2019).

Robustness and evolvability are evolutionary terms that define to what extent sequences can tolerate mutations and to what extent phenotypes can change, respectively (Payne and Wagner, 2019; Wagner, 2013). These terms, however, are vague and loosely defined regarding enhancers. Does being robust mean that no components of the pattern can evolve? Does evolvable necessarily mean that any and all possible phenotypes must emerge? If the answer to these questions were true, then nothing is robust or evolvable. Based on this logic, I asked, to what extent is *E3N* robust and evolvable?

Here I carefully analyze the robustness and evolvability of the *E3N* expression pattern using user-friendly machine learning algorithms (Berg et al., 2019) and image registration algorithms (Arganda-Carreras et al.; Schindelin et al., 2012) to define robust and fragile expression zones.

Justin Crocker, Noa Borst, and I additionally identify ectopic zones of expression – but these mutants are only expressed in the ectoderm of the embryo – suggesting a potential constraint on the evolvability of *E3N* enhancers between germ layers. I then pair this information with a gene expression atlas. I additionally identify a pleiotropic hotspot associated with repressing ectopic wing and haltere expression using combinatorial libraries. My results reveal robust, fragile, and evolvable features of the *E3N* expression pattern, provide a phenotypic landscape of *E3N* evolution, and reveals potential evolutionary constraints. These findings may help us predict *E3N* evolution and give insights into which components may be under higher selective pressures.

4.4 ROBUST AND FRAGILE ZONES OF *E3N* EXPRESSION

To identify which components of *E3N* were robust, fragile, and evolvable to mutation, I randomly selected 100 lines, where ten contained single mutations, ten double mutations, ten triple mutations, etc. Lines were fixed and stained using the previously developed automated pipeline (Fuqua et al., 2021) (see Chapter 2). I wrote image analysis scripts that incorporate Ilastik (Berg et al., 2019) (https://www.ilastik.org/), a user-friendly machine learning algorithm that creates reproducible and executable feature selections. Trained features from Ilastik are integrated with the open-source image analysis software Fiji (Schindelin et al., 2012) and Python scripts (Virtanen et al., 2019) to segment and count nuclei, measure their intensities and *E3N* stripe widths. These summary statistics reveal the phenotypic landscape of the *E3N* enhancer evolution (**Figure 17**).

For the 100 lines, I first analyzed the nuclei within the *E3N* stripe region (**Fig 17A**). I found that the WT enhancer expresses an average of 201 nuclei within this domain. The number of nuclei from the library was normally distributed, where the median variant had 121 nuclei in this domain, and the WT was at the upper tail end of the distribution (**Fig 17B**). I identified a few lines that drove extra nuclei within the measured domain compared to the WT (**Fig 17C, D**). Below the WT, I noticed that the stripes immediately begin to collapse to stripes 1-2 nuclei in width (**Fig 17E**), including the median: *231-3* (**Fig 17F**). Lines below the median then begin to lose expression in the anterior (left) while maintaining expression in the posterior (right) (**Fig 17G**). Eventually, even the posterior stripes disappear, and the only remaining component of *E3N* expression are cells marking the distal tips where the WT stripes were (**Fig 17H**). These nuclei are expressed in all of the reporter constructs at low levels. This progressive and predictable loss of nuclei in *E3N* is synonymous with the most fragile to most robust components of *E3N* expression. The WT *E3N* pattern with thick stripes (4-5 nuclei across) is a fragile trait, while single nuclei width stripes corresponding to the median, are a robust pattern underlying WT expression pattern.
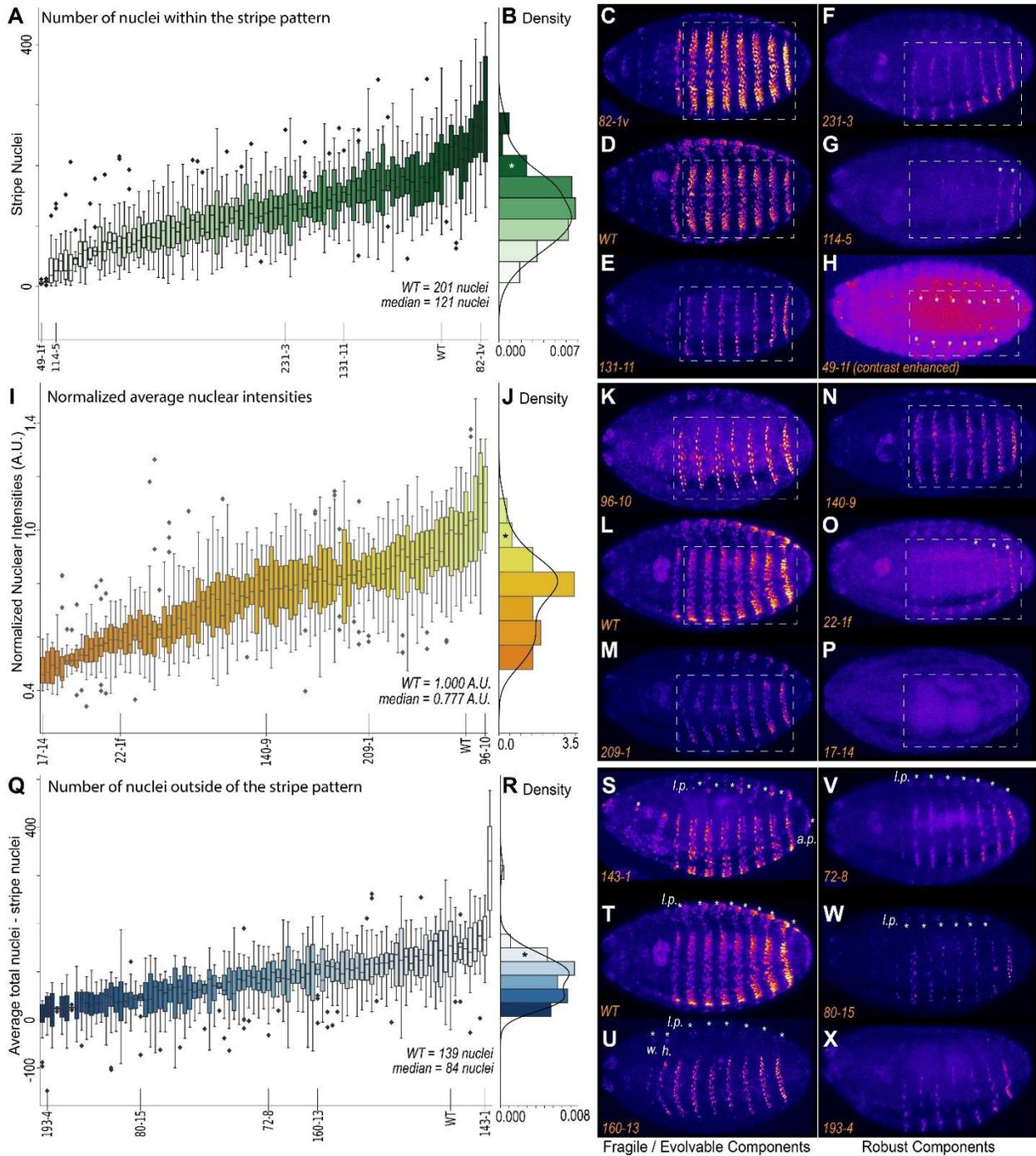
FIGURE 17 THE PHENOTYPIC LANDSCAPE OF *E3N*: ROBUST AND FRAGILE COMPONENTS

(A) Box plots for the number of nuclei in 100 tested variants. (B) Histogram of the average nuclei values per line. Asterisk indicates the bin with WT expression. Black lines are the kernel density estimate (KDE) for the lines. (C-H) Selected representative *E3N* reporter lines across the phenotypic distribution. Dotted boxes illustrate the region analyzed by the code. (C) *82-1v* drives more nuclei than the WT (D) in the stripe domain. (E) *131-11* drives less nuclei and the stripes begin to collapse to the width of a single nucleus across. (F) *231-3* the median variant from the distribution also drives collapsed stripes. (G) Line *114-5,* the stripes fade

Timothy Fuqua | 78

in the anterior (left) before the posterior (right). Asterisks mark weak expression. (**H**) Line *49-1f* drives feeble expression on the distal ends of where the WT stripes would be expressed (asterisks). Brightness and contrast adjusted for clarity. (**K-P**) Selected representative *E3N* reporter lines across the phenotypic distribution. Brightness and contrast were not enhanced for clarity. Dotted boxes illustrate the region analyzed by the code. (**I**) Box plots for the normalized nuclear intensities from the 100 tested variants. (**J**) Histogram of the average nuclei values per line. Asterisk indicates the bin with WT expression. Black lines are the kernel density estimate (KDE) for the lines. (**K**) *96-10* drives the highest nuclear intensity in the stripe region. (**L**) *E3N* WT reporter. (**M**) *209-1* lines begin to drive lower levels of expression. (**N**) *140-9,* the median variant from the distribution drives 77% of expression compared to WT. (**O**) *22-1f* nuclear expression continues to decrease below the robust level, stronger in the anterior than the posterior. (**P**) *17-14* expression is no longer detectable. (**Q**) Box plots for the number of nuclei outside the stripe domain in 100 tested variants. (**R**) Histogram of the average external nuclei values per line. Asterisk indicates the bin with WT expression. Black lines are the kernel density estimate (KDE) for the lines. (**S-X**) Selected representative *E3N* reporter lines across the phenotypic distribution. Brightness and contrast-enhanced for clarity. The following initial correspond to the following: l.p. lateral patches, w. wing disc, h. haltere, a.p. anal pad. (**S**) Line *143-1* drives reporter expression in the *E6* lateral patches (asterisks). (**T**) *WT* also drives lateral patch expression (asterisks). (**U**) *160-13* drives ectopic wing/haltere expression (asterisks). (**V**) *72-8* the median variant from the distribution drives low levels of lateral patches (asterisks). (**W**) The weak lateral patches are maintained below the median in *80-15.* (**X**) The lateral patches are not detectable in *193-4.* Experiments were carried out by Timothy Fuqua.

I next analyzed the normalized intensities of these nuclei from this library (**Fig 17I**). I found that most lines show a decrease in intensity and roughly follow a normal distribution, where the median of all mutants was ~78% as intense as the WT (**Fig 17J**). A handful of lines exhibited small increases in nuclear intensities (**Fig 17K**) than the WT (**Fig 17L**). The intensities begin to lower (**Fig 17M**), reaching the median of intensities at line *140-9* (**Fig 17N**). Afterward, like the trend in the total nuclei, I found that nuclear intensity was more robust in the posterior vs. the anterior below the median value (**Fig 17O**). Eventually, the signal is not detectable (**Fig 17P**). This distribution and progression revealed that most mutations retain ~78% of expression, a robust feature of *E3N* expression, and any additional activation is sensitive to perturbation.

A Overlying fragile expression pattern of *E3N*
B Underlying robust expression pattern of *E3N*

Strong lateral patches →

Weak lateral patches →
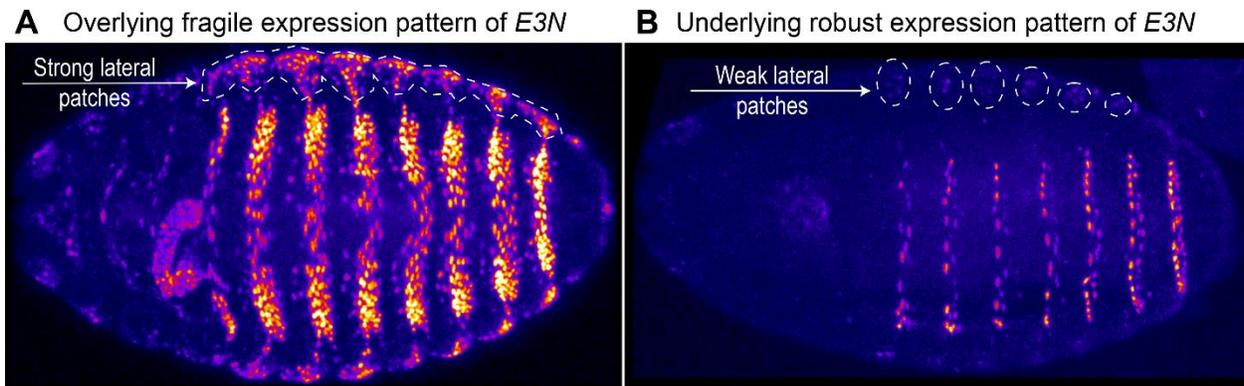
FIGURE 18 FRAGILE AND ROBUST *E3N*

(**A**) WT *E3N* expression pattern drives an additional 22% nuclear expression, thick stripes, and strong expression in the lateral patches. (**B**) Line *172-8* represents each median value from the summary statistics (see Figure 17). The stripes are collapsed to almost a single-nucleus thick stripe. The nuclei are 78% as bright as the WT. Nuclei are also brighter in the posterior than the anterior (posterior bias). Robust *E3N* also drives weak expression in the lateral patches. Experiments were carried out by Timothy Fuqua.

Finally, I wanted to measure the expression outside of the stripe domain. To do this, I subtracted the distributions within the stripe region from the average total nuclei per line. This calculation is a proxy for the number of nuclei expressed outside of the stripes (**Fig 17Q**). Like the other distributions, the WT is near the highest value, expressing 139 nuclei outside of the stripe domain, but the median external nuclei was 84 nuclei (**Fig 17R**). I found that a few lines, including WT *E3N,* drive strong expression in the lateral patches (**Fig 17S and T**). This strong lateral patch expression quickly disappears and is replaced by a weaker and different lateral patch pattern. I identified a mutant driving ectopic expression in the wing and haltere primordia called *160-13* (**Fig 17U**). The line expressing the median number of external nuclei, *72-8,* drives very weak expression in the lateral patches (**Fig 17V**). These lateral patches are maintained even below the median (**Fig 17W**) and eventually vanish in some mutants at the tail end of the distribution (**Fig 17X**).

To summarize, I discovered that *E3N* encodes an underlying robust expression pattern consisting of thin stripes that are ~ one nucleus across, expressing ~78% of expression compared to WT more robustly in the posterior than the anterior, and additionally drives weak expression in the

lateral patches. Overlying this expression pattern is an additional fragile pattern generating thicker stripes 4-5 cells across, a further 0-22% increase in nuclear intensity, expression in the anterior stripes, and lateral patches (**Fig 18A and B**).

4.4 GENE EXPRESSION ATLASES MAP DEVELOPMENT AND POSSIBLE REGULATORY INPUTS FOR *E3N*

The regulatory inputs feeding into the *svb* locus are well characterized (Delon et al., 2003; Payre et al., 1999; Rizzo and Bejsovec, 2017) (see Chapter 5). However, there are limited spatial gene expression atlases for these inputs during late embryogenesis. Gene expression atlases have been helpful for modeling and predicting transcription in the *Drosophila* early embryo (Fowlkes et al., 2008) . I selected antibodies that bind to proteins for various developmental processes such as the Central Nervous System (**Fig 19A**), Tracheal development (**Fig 19B**), cytoskeletal patterning (**Fig 19C**), and segmentation networks (**Fig 19D**). The antibodies were co-stained with DAPI and a bacterial artificial chromosome (BAC) driving *dsRed* expression from a *svb* cDNA as a fiduciary stain. With these stains, I can map the expression patterns from different embryos onto a composite embryo (see methods).

The antibody stains and composite images allowed me to analyze the co-localization or anti-localization of different proteins, such as the core domain of Broad and an isoform of Extradenticle (Exd) not being expressed in the Ultrabithorax domain (**Fig 19A**). I also tested antibodies for the same proteins and found examples of antibodies binding to specific protein isoforms, which was the case for two Ubx antibodies (Fp3.38 and Fp6.87) and a domain of Broad, which is now only expressed in the AbdB segment (**Fig 19B**). I also found a few examples of transcription factors that may form interesting nuclear microenvironments (**Fig 19C**). See (Tsai et al., 2017, 2019).
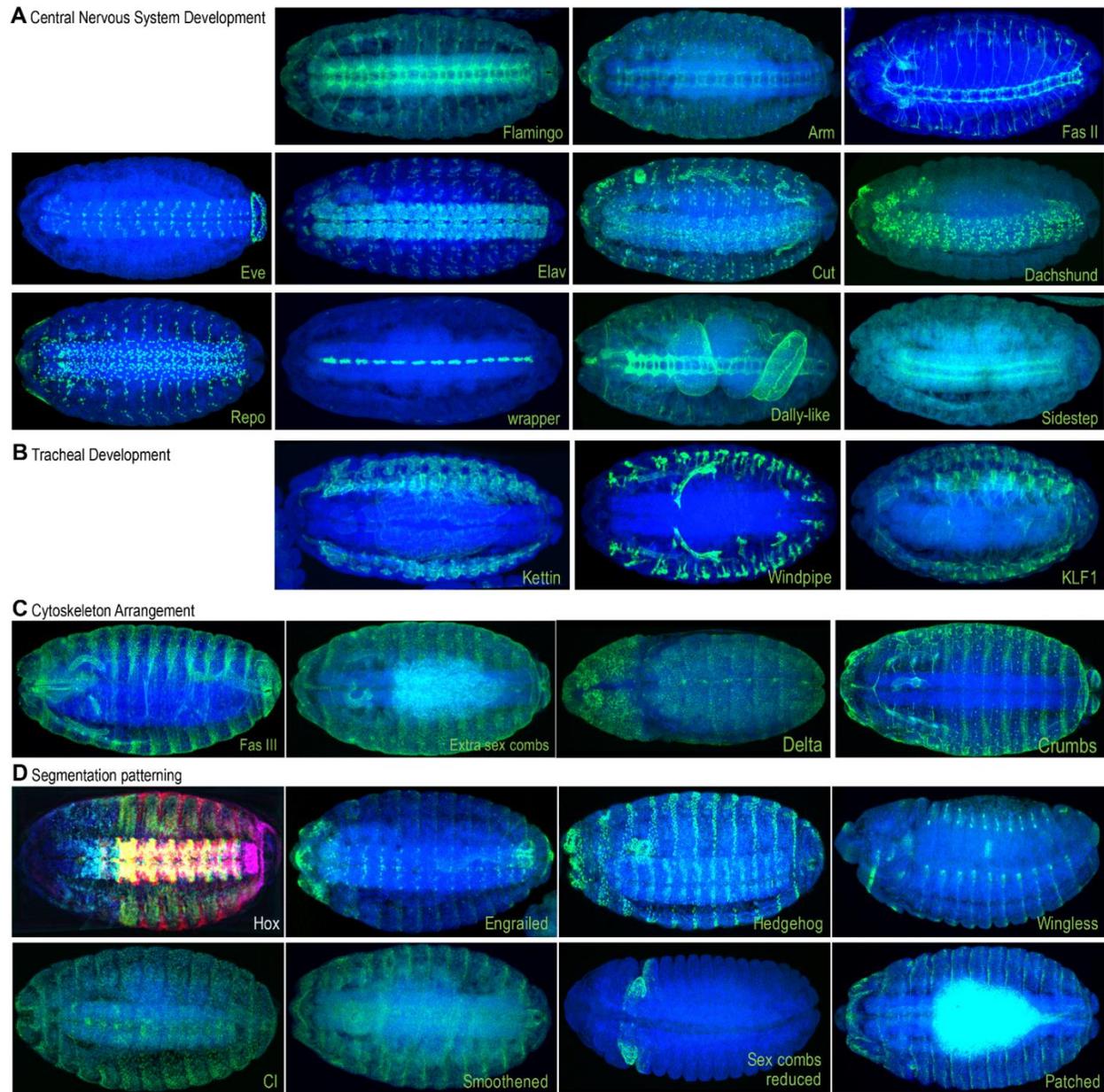
FIGURE 19 CHARACTERIZING LATE EMBRYONIC EXPRESSION PATTERNS

Representative images from the DSHB atlas for the Nervous System (**A**), the Tracheal System (**B**), the Cytoskeleton (**C**) and Segmentation (**D**). Green = antibody expression, Blue = DAPI. Multi-colored embryo in (D) corresponds to multiple HOX genes. Experiments were carried out by Timothy Fuqua.
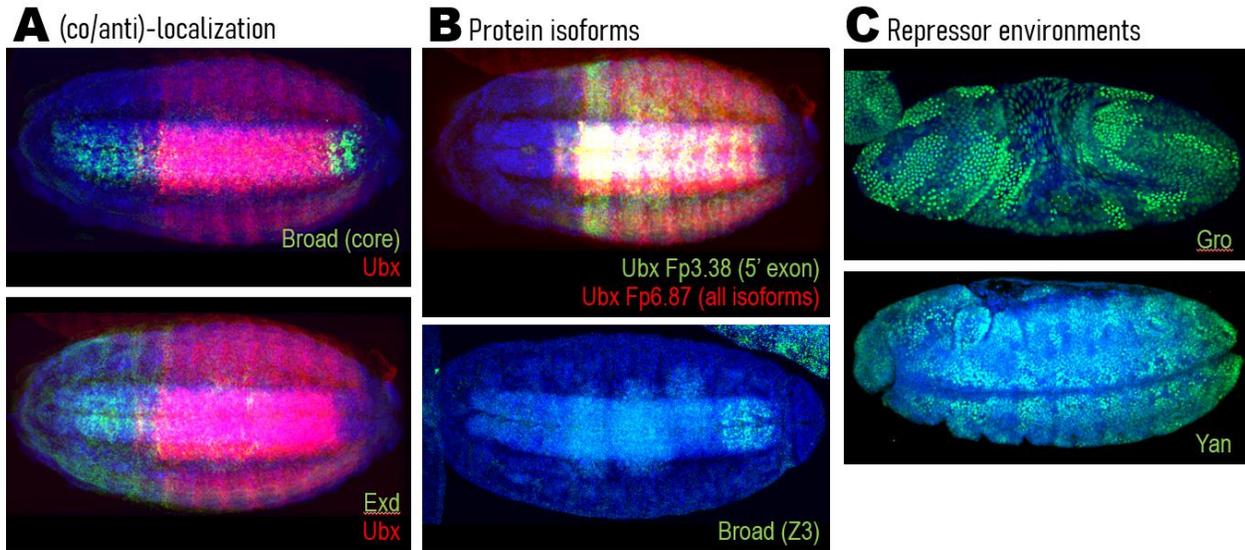
FIGURE 20 REGISTERED IMAGES REVEAL DIFFERENT ASPECTS OF BIOLOGY

 (A) Using a preliminary version of the DSHB atlas viewer, I identify areas of co-localization and anti-localization. (Top) the DSHB antibody Broad, binds to the Core domain of the Broad protein. The core of Broad is not expressed in regions where the protein Ubx is localized. (Bottom) Additionally, the protein Exd is not expressed where Ubx is localized. (B) The DSHB has antibodies that target the same antigen, however, these antibodies bind to different isoforms for Ubx (top) and the Z3 domain of Broad, although the core of Broad is also expressed in the anterior (left side of the embryo, see panel A, top). (C) Antibodies for repressor proteins reveal expression patterns with different concentrations of the repressor. Experiments were carried out by Timothy Fuqua.

4.5   EVOLVABLE ZONES OF *E3N* IN THE ECTODERM

I went through all 100 of the lines (2,256 embryos) and with Noa Borst and Justin Crocker, we analyzed every embryo for potential ectopic expression. From the 100 lines that I tested, Noa Borst, Justin Crocker, and I observed ectopic expression patterns in new tissues including the anal pads (**Fig 21A**), the salivary glands (**Fig 21B**), a series of posterior dorsal stripes (**Fig 21C**), the developing mouth hooks (**Fig 21D**), the developing wing and haltere discs (**Fig 21E**), and in the anal plate ring (**Fig 21F**). Interestingly, *svb* is known to be expressed in some of these tissue types, but not by *E3N.* Furthermore, all of these ectopic tissues are part of the ectoderm, suggesting that it may be easier to evolve novel expression in the same tissue type. However, more experiments need to be executed to support this argument. I decided to explore some of these novel phenotypes further with Noa Borst.

I had previously identified a mutant *E3N* line that drove expression in the wing and haltere called *145-2* (see Figure 15)*.* To see if *160-13* (see **Fig 21E**), the other line that drove ectopic wing and haltere expression, shared any mutations with *145-2*, I aligned the sequences. I identified a ~20 bp region where both lines shared mutations and selected lines with different point mutations in this region and minimal mutations outside (**Fig 21 G and H**). Together with Noa Borst, Noa Borst and I found that 10/18 lines drove ectopic wing and haltere expression (**Fig 21I**). The point mutations were located in different bases of the enhancer. The most parsimonious explanation for the expression gain is through the loss of a binding site rather than the gain of an activator site. I am still working on experiments to find out what repressor binds to *E3N.*
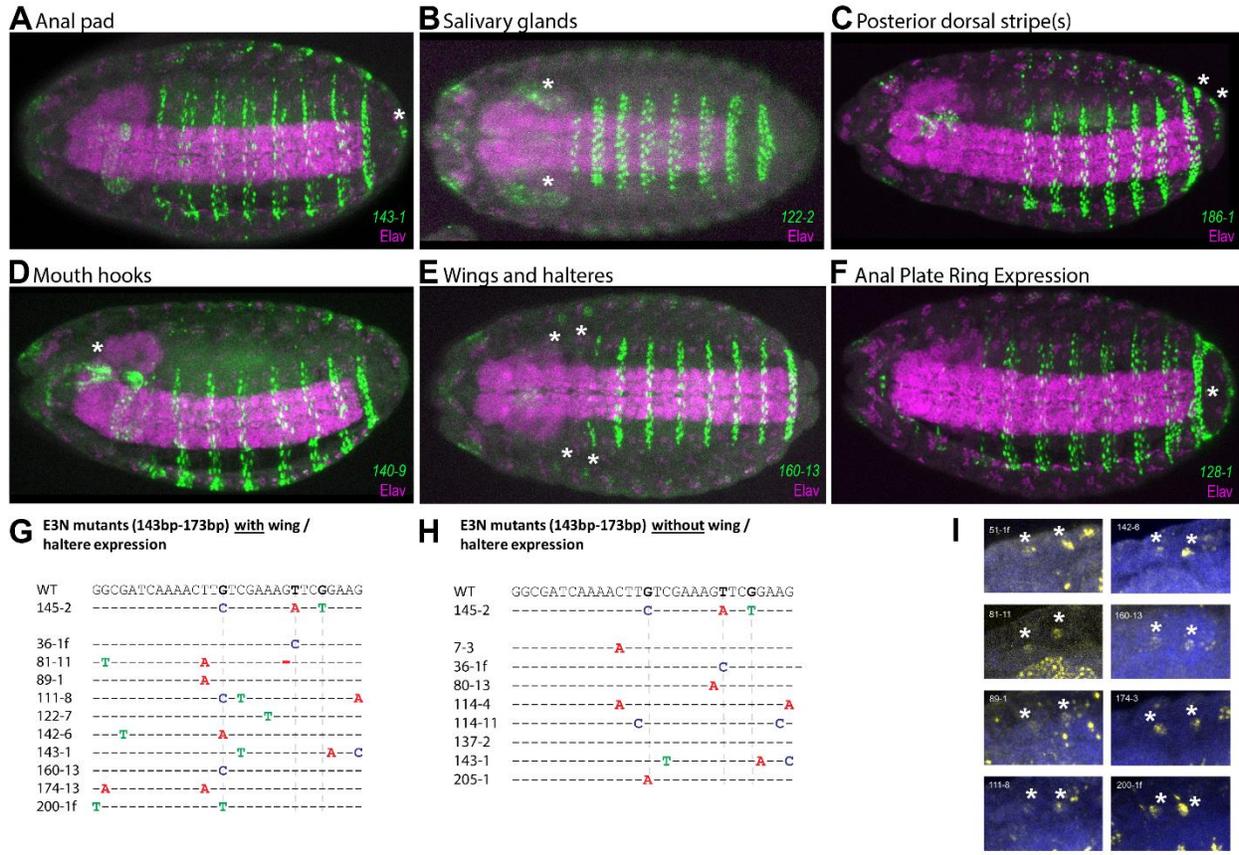
FIGURE 21 ECTOPIC EXPRESSION IS BIASED TOWARDS ECTODERMAL ACTIVITY

Example lines with ectopic expression in the different tissues, marked with an asterisk (*) for the (**A**) Anal pads, (**B**) Salivary glands, (**C**) Posterior dorsal stripes, (**D**) Mouth hooks, (**E**) Wing and haltere discs, (**F**) Anal plate ring. (**G**) List of tested lines that also create wing / haltere expression. (**H**) List of tested lines that do not create wing/haltere expression. (**I**) Subset of images of positive lines with wing/haltere expression. Asterisks indicate wing/halteres.

Panels (**A-F**) completed by Timothy Fuqua, Justin Crocker, and Noa Borst. Panels (**G-I**) completed by Timothy Fuqua and Noa Borst.

4.6   A COMBINATORIAL LIBRARY REVEALS EVOLUTIONARY PATHS TO ECTOPIC EXPRESSION

To further explore this novel expression pattern in *145-2*, I created a combinatorially complete library of reporter constructs between the WT *E3N* expression pattern and the *145-2* line (**Fig 22A**), changing each of the four mutant bases in *145-2*: a,b,c,d in combination (**Fig 22B**). From this library, I identified a range of phenotypic effects where nuclear intensities, stripe widths, lateral patches, and wing and haltere expression changed (**Fig 22C**).

The combinatorial library reveals to us that *mut-b* was the only single point mutation to drive weak levels of wing and haltere expression (**Fig 22D, left**). However, when combining the mutations together, *mut-bc* drives increasing levels of wing / haltere expression (p<0.001) (**Fig 22E**), which was further enhanced by adding the *mut-d* mutation (p<0.001). Mutations to *mut-a* had no effect on wing or haltere expression since *mut-bcd* was similar to *mut-abcd* (p=0.12) and *mut-bc* was similar to *mut-abc* (p=0.12). These results suggested that mutations *b, c,* and *d* create the ectopic wing / haltere expression stepwise, but *a* does not.

I also observed changes to the total nuclear intensities (**Fig 22D**, **right**). I found that independently, *mut-c* does not significantly change the nuclear intensity (p=0.47), and *mut-d* creates a significant decrease (p=0.009). Interestingly, when testing the mutations in combination, the nuclear intensity increases (p=0.047), suggesting positive epistasis between the bases (**Fig 22F**). I also discovered that *mut-b* (p=0.45)*, mut-c* (p=0.47)*,* and *mut-bc* (p=0.12) alone or in combination do not significantly alter nuclear intensities (**Fig 22G**). In fact, the loss of nuclear intensity can be entirely associated to *mut-a* (p=0.006 and *mut-d* (p=0.009), which additively lower expression in *mut-ad.* The *mut-ad* construct creates the same level of expression as *145-2* (p=0.32).

FIGURE 22 COMBINATORIAL PATHS LEAD TO WING AND HALTERE EXPRESSION

(**A**) The *E3N* variant: *145-2* or *abcd* creates a weaker expression than the WT, and generates novel activity in the wing and haltere primordia (asterisks). (**B**) *145-2* contains four point mutations: *a, b, c,* and *d.* (**C**) Combinatorial library for all constructs. Mutation letter (i.e. *bd*) indicates that *b* and *d* are mutated to match *145-2.* I did not analyze *acd* yet. Asterisks indicate wing/haltere (W/H) expression. (**D**) Paths illustrating how *E3N* could mutate and acquire wing/haltere expression (left) and lose nuclear intensity (right). (**E**) Boxplots for wing/haltere (W/H) intensities of *b, bc, abc, bcd,* and *abcd.* (**F**) Boxplots for nuclear intensities of *WT, c, d,* and *cd.* (**G**) Boxplots for nuclear intensities of *WT, b, c, bc* (left) and *WT, a, d, ad,* and *abcd* (right). In box plots, the center red line is the mean, and the upper / lower limits are one standard deviation. Whiskers show a 95% confidence interval. Experiments were carried out by Timothy Fuqua.

Altogether, these results suggest that *mut-b, mut-c*, and *mut-d* are primarily contributing to this hypothetical repressor binding site, while *mut-a* and *mut-d* activate *E3N* expression. The positive epistasis would potentially be explained by the repressor at *mut-b* and *mut-c* overlapping and competing with the activator at *mut-d.* It is still unclear which repressor could be binding to *E3N* to block this expression. Because mutations *b* and *c* do not change the WT *E3N* stripe expression pattern and only create wing/haltere expression, the wing/haltere phenotype is not linked to the WT pattern. Thus, it is possible for *E3N* to evolve this ectopic wing/haltere expression pattern without pleiotropic consequences on the ventral stripe pattern.

4.7 DISCUSSION

Using a mutational scanning approach, I found fragile and robust components encoded within the *E3N* enhancer. Strong nuclear expression is fragile in *E3N*, but a lower robust level at ~78% expression is maintained. The number of nuclei is also both fragile and robust. Strong lateral patch expression is quickly reduced to a weak lateral expression. This is followed by the stripes collapsing to the width of a single nucleus. Then, the stripes begin to vanish but are more robustly maintained in the posterior. Finally, the most robust component of *E3N* expression is the cells flanking the ends of where the stripes would be located.

There are a number of reasons why *E3N* may encode robust and fragile components of gene expression. One reason is that these components are a developmental bias of the molecular encoding of *E3N* itself. For example, homotypic clusters of TF binding sites encode robustness (Crocker et al., 2015; Payne and Wagner, 2015). Different bioinformatics analysis and experiments reveal that *E3N* contains at least six binding sites for the TF: Ultrabithorax (Ubx) (Crocker et al., 2015; Fuqua et al., 2020; Rastogi et al., 2018), at least four binding sites for the TF: Homothorax (Hth) (Fuqua et al., 2020), and at least two binding sites for the Pointed (Pnt) activator (Fuqua et al., 2020; Al Hayek et al., 2021). Perturbing one of these sites will likely not completely break the expression pattern since there is still a threshold of activation from the other signaling inputs (Crocker et al., 2017). Thus, fragile components may exist simply because there are less TF binding sites encoding the pattern compared to the robust pattern (See Chapter 5 for more information on how the *E3N* pattern is formed).

Robust and fragile components may also exist because they are under different selective pressures. For example, the ventral trichomes are primarily responsible for taxis (Inestrosa et al., 1996). By alternating between segments with and without trichomes, larvae have high levels of traction for not only movement but also adhesion to different surfaces. Having a minimal core number of trichomes absolutely essential for movement would be under a higher selective pressure compared to accessory trichomes, which undoubtedly contribute to taxis, but to a lesser extent. I hypothesize that robust thin stripes are the most essential component for larval locomotion.

The same explanation would also explain why the enhancer mutants have a median expression loss of 22%. *E3N* may encode additional activation than necessary, and driving 22% less *shavenbaby* product likely has minimal effects on the formation of a trichome or not. It has been demonstrated that weaker enhancers integrate regulatory information additively at a promoter through the Competition Model (see Chapter 1.3) (Bothma et al., 2015). If this is also true for *E3N*, then an entire deletion of one out of the three ventral enhancers would drive only 66% of WT expression. In fly lines carrying an entire deletion of the *svb DG3* ventral enhancer, only a slight loss of trichomes is observed when the fly was under high levels of heat stress (Tsai et al., 2019). Thus at 66% ventral expression, trichomes are still produced. If most mutations reduce *E3N* to only 78% activity, then ~ 93% (78% + 100% + 100% / 3) of *svb* expression would still be canalized by both *E3N* and the other ventral shadow enhancers, and the ventral trichome expression would likely be unchanged. Thus, I hypothesize that the additional fragile 22% of *E3N* intensity likely canalizes expression in stressful environments and is encoded robustness (Sucena et al., 2003).

Noa Borst, Justin Crocker, and I also observed ectopic expression in various tissue types, including the anal pads, salivary glands, posterior dorsal stripes, mouth hooks, wing and haltere discs, and the anal plate ring. The only commonality between these tissue types is that they are all expressed in the ectoderm. These results suggest that *E3N* may be developmentally biased to drive expression in the ectoderm. What is the underlying mechanism for this? It could be that *E3N* requires an ectodermal-specific pioneer factor to open the heterochromatin surrounding itself (Zaret and Carroll, 2011). In the ectoderm, the TF Grainyhead is known to act as a pioneer-like factor to activate ectodermal tissue (Sundararajan et al., 2020). However, the gain of expression could be either due to the loss of a repressor, such as the wing and haltere lines, or the gain of an activator.

Overall, these results demonstrate the power of mutational scanning to identify trends and biases in enhancer function and evolution, identify molecular mechanisms regulating these biases, and may in the future help us predict enhancer evolution.

5 DISCUSSION AND CONCLUSIONS

> "Work it harder, make it better
> Do it faster, makes us stronger
> More than ever, hour after hour
> Work is never over"
> – Daft Punk

5.1 INTRODUCTION

In this thesis I coordinated and participated in the development of a semi-automated pipeline to streamline the analysis of gene expression patterns in developmental systems. I applied this pipeline to study the evolvability and logic of a developmental enhancer called *E3N* by screening hundreds of reporter assays. I followed up on a recurrent phenotype that adds gene expression to the developing wing and haltere on the fly and mapped this to a hotspot that Noa Borst and I hypothesize to be a repressor binding site. Finally, I explore late *Drosophila* embryogenesis by creating a gene expression atlas to understand better the regulatory inputs controlling the *E3N* enhancer. In this discussion, I contextualize our findings from the *E3N* enhancer screen to other results in the literature to make some generalized statements on enhancer encoding and evolution. I then review the regulatory information for *E3N* that I and other research groups have found, and culminate it into a simplified working model. This working model is helpful for explaining to what extent one can predict *E3N*'s evolution. Finally, I speculate on the impact of this work within the community and future research directions.

5.2   DEVELOPMENTAL ENHANCERS ARE DENSELY ENCODED

From the *E3N* screen, I found that the *E3N* enhancer encodes a large amount of regulatory information since most mutations significantly affect expression. Additionally, the phenotypes were complex and pleiotropic. Mutations would frequently affect multiple components of the enhancer expression pattern. I argue that this pleiotropy likely constrains the evolution of *E3N* expression, which may partially explain why ventral trichomes have remained essentially unchanged throughout evolution. Do other experiments also suggest dense enhancer encoding and pleiotropy? Three studies, in particular, stand out and are worth describing in further detail.

The *yellow spot[196]* enhancer

The *yellow spot[196]* enhancer was first characterized in Sean Carroll's group. The enhancer recruits the TFs: Distalless and Engrailed, to drive *yellow* (black pigmentation) in a well-defined spot on the wings of *Drosophila biarmipes* (Gompel et al., 2005). Recently, the *spot[196]* enhancer was dissected at a higher throughput (Le Poul et al., 2020). Le Poul and colleagues created a mutational series for *spot[196]* by replacing segments of the enhancer with poly-adenine tracts and tiled these tracts across the sequence in different constructs. The lines were tested as reporter constructs, and their expression patterns were registered to one another as composite statistical representations. This registration allowed direct comparison of regions within the wing to analyze the expression patterns unbiasedly. Despite the enhancer's model to contain only five TF binding sites (four Distalless and one Engrailed) (Arnoult et al., 2013; Gompel et al., 2005), changes within and outside of these binding sites significantly changed the expression pattern. The group concluded shortly after my manuscript was published, that the *spot[196]* enhancer also encodes a dense amount of regulatory information like *E3N* (Le Poul et al., 2020).

The *dPax2 sparkling* enhancer

The work of Swanson and colleagues on the *dPax2 sparkling* enhancer has been not only seminal for understanding enhancer grammar and binding site turnover, but it was also the first evidence to suggest the dense-encoding of enhancers (Goldwater et al., 2010; Swanson et al., 2011). The TFs regulate the *spa* enhancer: Lz, Pnt, and Su(H) to drive reporter expression in cone cells (Flores et al., 2000). Deleting the gaps between these TF binding sites also caused changes to the *spa* expression pattern, concluding that almost the entire *spa* enhancer contains regulatory information (Goldwater et al., 2010). Attempts at constructing synthetic versions of the *spa* enhancer with these 12 binding sites failed to drive appropriate expression (Goldwater et al., 2010). These findings were similar to the attempts made to build a synthetic *eve stripe 2* enhancer (Vincent et al., 2016) and suggest additional regulatory information between the known binding sites.

The *Sonic Hedgehog ZRS* enhancer

Another critical study on enhancer encoding is the *ZRS* enhancer dissection, published shortly before the release of my manuscript (Kvon et al., 2020). This study focuses on mutations within the *ZRS* enhancer associated with the enhanceropathy: polydactylism (Hill and Lettice, 2013). To identify new variants within *ZRS* that cause polydactyly, the group created *ZRS* reporter libraries with mutations ranging from 2-5%. Kvon and colleagues found that at a 5% mutation frequency (~40 point mutations each), 18% of the *ZRS* mutants drove a reduced expression pattern, and 82% had no expression. At a 2% mutational frequency (~16 point mutations each), 44% were reduced, 26% were lost, 7% showed a gain of expression, and 23% were classified as "normal."

For the *ZRS* study, reporter constructs were stained using the X-gal staining system. For part of the *E3N* analysis, Maria Elize van Breugel and I also carried out X-gal staining. We found that at a 5% mutation frequency in *E3N* (~15bp), ~90% of our mutants showed no expression, which is similar to the *ZRS*, where 82% showed no expression. Additionally, at a 2% mutation frequency in *E3N* (~6bp), mutants are ~30% complete loss and 20% with a gain of expression. These results are again similar to the *ZRS* enhancer screen, suggesting that both enhancers' density of

regulatory information is similar, even though *E3N* is 292 bp and *ZRS* 789 bp. It is important to remember that these numbers were acquired using a colorimetric X-gal staining approach, which is not as precise compared to other histological techniques like antibody staining (Fuqua et al., 2021).

Evidence against dense regulatory encoding

A recent publication carried out an enhancer screen similar to our own method studying the role of encoding within ultraconserved enhancers. Surprisingly, the expression patterns were essentially unchanged by mutations. (Snetkova et al., 2021). What does this mean? One issue with this study was that it assumes that mutations must have a large effect size since all expression patterns were qualitatively and manually quantified, using low-resolution X-gal staining. Additionally, the phenotypes were tested in laboratory conditions, where the embryos were minimally stressed. Many of these bases may likely be necessary for canalizing expression in different environments (Frankel et al., 2010; Osterwalder et al., 2018).

Other large-scale enhancer mutagenesis experiments also suggest that enhancers can tolerate a large number of mutations. These studies, however, are done using massively parallel reporter assays (MPRAs) (Melnikov et al., 2012). I argue that MPRA experiments depict a higher amount of robustness because the reporters do not represent a field of cells with different types, chromatin states, TFs, and are not measured across a developmental timescale. The only information drawn from MPRAs is the fluorescence of the reporter assay, which makes it challenging to analyze pleiotropic effects or changes to the expression pattern parameters. This is one of the most significant advantages our screening technique has over traditional methodologies. In Chapter 4, I also identified mutants with insignificant changes in expression levels compared to the WT, but also had changes to the stripes and lateral patches. Had these mutants been tested in a MPRA, they would have been classified as neutral mutations.

5.3 HOW DOES *E3N* MAKE ITS EXPRESSION PATTERN?

"This more complex pattern of ventral expression suggests that there might be multiple inputs into the activity of this enhancer" – (Rizzo and Bejsovec, 2017)

The *E3N* enhancer encodes a dense amount of regulatory information that I am only beginning to appreciate (Fuqua et al., 2020). Based on the results from this thesis and previous works, I will review the known regulators of *E3N* to explain how I hypothesize *E3N* generates its expression pattern. This model is both far from complete and oversimplifies how much regulatory information is feeding into the enhancer. Nevertheless, I will begin with the well-characterized regulatory inputs. I will additionally discuss sites in need of validation.

From the *E3N* enhancer screen, I identified mutational hotspots using a footprinting and GWAS-like method called Enhancer-Wide Association Catalog (EWAC) (Fuqua et al., 2020). These assays identified regions important for regulating *E3N* and creating pleiotropic expression. Many of these regions overlap with previously identified TF binding sites for *E3N* which I discuss below, and others overlap with motifs for TFs which could be regulating *E3N.*

Ubx and AbdA activate *E3N* with Hth as cofactors

The transcription factor Ultrabithorax (Ubx) is responsible for patterning the T3 and A1 segments of the *Drosophila* embryo (Fernandes et al., 1994). Like all Hox genes, the protein encodes a homeodomain that binds to the canonical `TAAT` sequence (McGinnis et al., 1984). It was proposed that low-affinity binding sites can explain Hox binding specificity, where the lower the affinity, the more posterior-expressed Hox gene can precisely bind to enhancers (Crocker et al., 2015). This phenomenon was demonstrated by changing the affinities of low-affinity Ubx binding sites in *E3N* (Crocker et al., 2015; Fuqua et al., 2020; Rastogi et al., 2018). Increasing the affinities

caused the gain of ectopic expression in anterior regions and increased overall expression levels. This phenotype was due to stronger binding of Ubx and AbdA and the additional binding of other anterior-expressed Hox genes such as Antp. Respectively, further lowering the affinity decreases overall Ubx binding and, with it, reduced expression. However, the expression from this mutational series was primarily maintained in the posterior stripes (Crocker et al., 2015) – possibly due to the preserved specific low-affinity binding of AbdA. This result may also explain why the posterior stripes of *E3N* are more robust than the anterior stripes (see Chapter 4).

Ubx and AbdA are recruited to activate *E3N* expression by binding at a low affinity to minimize the ectopic binding of other homeodomain-containing factors. Based on the work of Crocker et al. and Rastogi et al., there are approximately six low-affinity Ubx / AbdA binding sites in *E3N* (Crocker et al., 2015; Rastogi et al., 2018). I tested *E3N* variants with mutations to these six sites and found that the total predicted Ubx affinity correlates with stripe expression, anterior expression, as well as the timing of activation (Fuqua et al., 2020). These sites additionally showed up in the EWAC and footprinting analysis. Because the total affinity can be used to predict *E3N* expression independent of the enhancer sequence, I would hypothesize that Ubx / AbdA binding follows a billboarding mechanism. However, follow-up experiments such as adding Ubx sites throughout the enhancer while maintaining the total affinity would need to be done to support this claim.

Justin Crocker, Namiko Abe, Richard S. Mann, and I demonstrated that the TF Homothorax (Hth) binds to the *E3N* enhancer at four sites through mutagenesis, mutant crosses, and EMSAs. Together, we found that mutations to the second Hth site: Hth2, caused lower nuclear intensity and state-like effects on expression. I think that mutations to Hth binding sites cause this specific phenotype because Hth may be acting as a pioneer-like factor to facilitate the opening of heterochromatin (Fuqua et al., 2020). If this is true, without Hth2, the probability of the heterochromatin opening is lower, thus the speckled state-like phenotypes. Hth is also a known co-factor for Ubx, and it has been demonstrated that co-factors may also help with Hox binding specificity (Jaw et al., 2000; Merabet et al., 2007). Cooperative binding of Ubx and Hth could also

explain lower expression levels when the Hth2 site is mutated. Further experiments need to be done to confirm this cooperative binding and the pioneer-capabilities of Hth.

It was additionally found that the *Hth2* binding site is not conserved in *Drosophila virilis*, which drives feeble *E3N* expression. Changing this binding site back to the *melanogaster* motif recapitulates more of the *melanogaster* pattern, and may explain why *Drosophila virilis* have less ventral trichomes (Fuqua et al., 2020).

Stripe width is potentially modulated by CI binding affinity

The Hedgehog (Hh) signaling pathway uses a transcription factor called Cubitus Interruptus (CI) to either activate or repress transcription. In the absence of the Hh signaling ligand, CI acts as a repressor, but as an activator in cells receiving the Hh signal (Aza-Blanc et al., 1997). The Hh ligand itself is expressed as a gradient in the embryo and in the developing imaginal tissue. This Hh gradient thus creates a corresponding gradient for CI in both repressor and activator forms (Müller and Basler, 2000). It was found in Parker et al. that the binding affinity of CI can determine how wide a stripe is expressed in CI-target enhancers in the *Drosophila* wing disc (Parker et al., 2011). A CI motif in *E3N* showed significant EWAC and footprinting scores. While I currently lack biochemical and genetic validation, I hypothesize that CI may be responsible for modulating *E3N* stripe width. I am currently testing variants of CI binding sites at different affinities and carrying out genetic crosses to validate this site with Noa Borst and Gilberto Alvarez Canales.

Pointed binding sites regulate *E3N* expression

Pointed (Pnt) is an ETS transcription factor activated by EGFR signaling (O'Neill et al., 1994; Xu et al., 2000). Based on a motif search, I identified two strong Pnt binding sites that correspond with sharp peaks in the footprinting and EWAC analysis. I call these sites "Pnt1" and "Pnt2". Since then, Pnt1 was recently identified to regulate stripe expression and posterior midgut expression in *E3N* (Al Hayek et al., 2021). Noa Borst and I are currently testing the Pnt sites with additional reporter assays.

_E3N_ is repressed by Pan and a wing/haltere repressor

When crossed into mutant Wg backgrounds, _E3_ (a larger enhancer fragment that includes _E3N_) drives ectopic expression between its stripes (Rizzo and Bejsovec, 2017). This naked stripe is repressed by the wingless (Wg) transcriptional repressor: Pan / Tcf (Chanut-Delalande et al., 2006). I identified motifs for Pan in our enhancer screen that significantly associated with creating pleiotropic expression in _E3N_. When mutated, _E3N_ begins to create ectopic expression between the stripes; however, this phenotype was linked with lowered expression (Fuqua et al., 2020). One reason for this decreased expression may be because one of the Pan sites overlaps with the previously described Hth2 activator site. I speculate that the other Pan sites also share overlapping signatures with other transcriptional activators to constrain ectopic naked expression. An additional cluster of Tcf binding sites was proposed towards the 3' end of the enhancer, where mutants may create ectopic expression between the stripes (Rizzo and Bejsovec, 2017). However, the given embryo presented in the supplemental figure is much older than the WT embryo, and the motif does not match the consensus sequence. These results may be slightly misleading. Nevertheless, the consensus is that _E3N_ is repressed by Pan.

Noa Borst and I additionally found a mutational hotspot associated with ectopic wing and haltere expression (Chapter 4). Interestingly, I found no significant change between the _E3N_ expression patterns other than the gain of expression in the new tissue. Noa Borst and I identified multiple independent mutations leading to this ectopic expression and concluded that the most parsimonious explanation for our results was that the site recruits a repressor to silence activity in wing and haltere cells. However, _E3N_ is a highly pleiotropic enhancer expressed across a range of developmental stages and tissue types (Fuqua et al., 2020; Kittelmann et al., 2021), and I do not want to exclude the possibility that this repressor may also be an activator in different developmental contexts.

Speculative binding sites

Likely other transcription factors are binding to *E3N* at various affinities. Some of which could correspond to the footprinted activity peaks Maria Elize van Breugel and I identified. Based on the *E3N* expression pattern, *E3N* is highly localized with the TF Engrailed (Chanut-Delalande et al., 2006). Engrailed (En) binds to a homeodomain (Draganescu and Tullius, 1998; EKER, 1929), like Ubx (McGinnis et al., 1984), and there may be a level of shared regulation or competition between the activators. Additionally, the transcription factor Sox-Neuro is a well-established regulator of the *svb* locus, and could be binding to *E3N* as well with various affinities (Rizzo and Bejsovec, 2017). Notch signaling is another known regulatory input for *shavenbaby*, and It would not surprise me if Su(H) binding sites were later identified in *E3N* as well (Chanut-Delalande et al., 2006). I present a model of all described regulatory information in **Figure 23**.
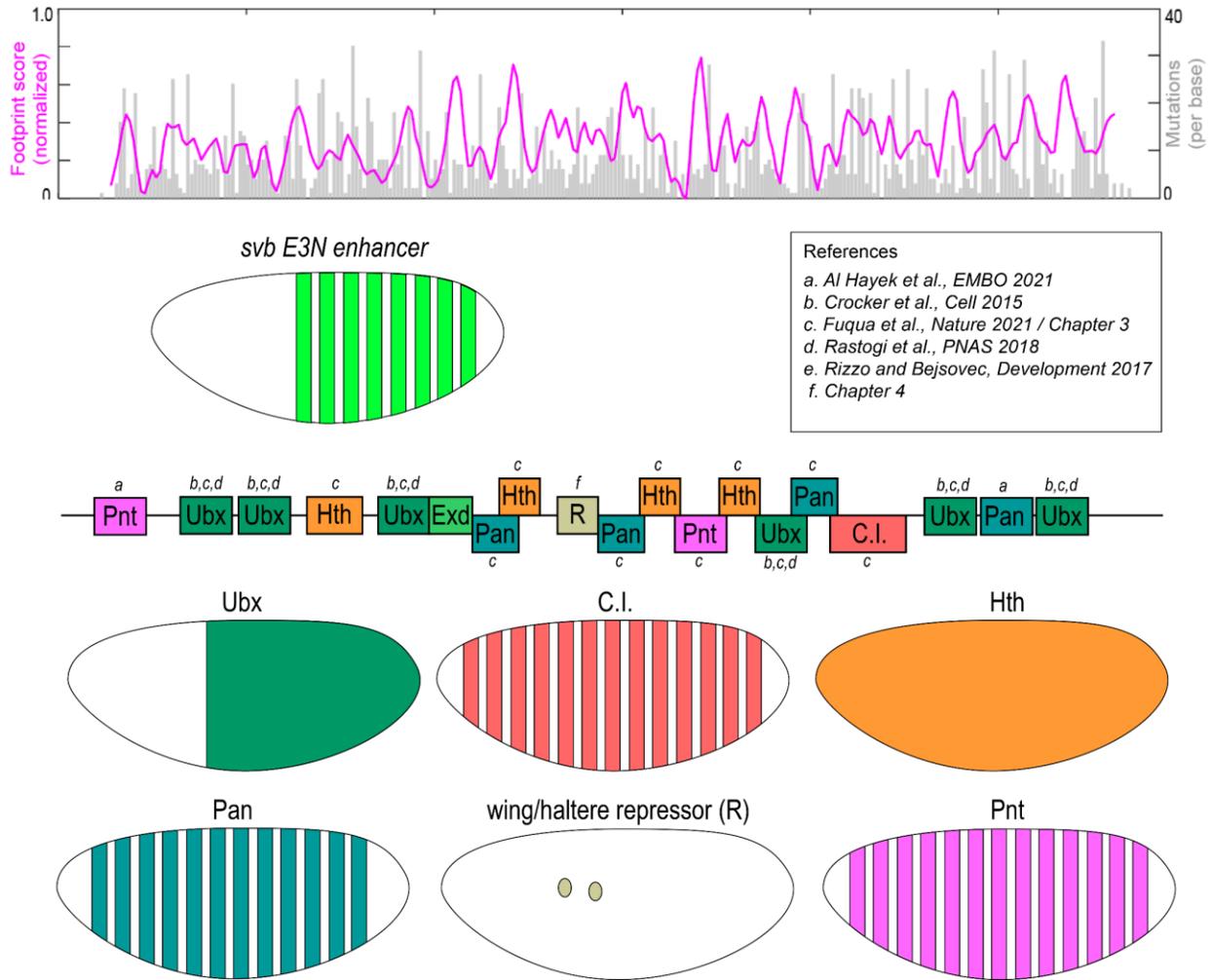
FIGURE 23 AN OVERSIMPLIFIED MODEL FOR E3N EXPRESSION

(**Top**): plot of the *E3N* sequence spanning across the x-axis from 5' → 3'. Gray bars correspond to the mutations tested for each base. The magenta line is the normalized footprinting score. Peak height corresponds to the probability that a mutation at that base will change the expression pattern. (**Bottom**): a working model for *E3N* with cartoon embryos representing input and outputs.

Figure partially adapted from Fuqua et al., 2020. Figure created by Timothy Fuqua.

5.4   CAN WE PREDICT *E3N* ENHANCER EVOLUTION?

Many evolutionary-developmental biologists, including myself, hope that one day we can predict the evolution of morphology. One of the most obvious ways to test evolution is to start randomly mutating DNA sequences. This random-walking approach is used for deep mutational scanning (Fowler and Fields, 2014), saturation mutagenesis (Kircher et al., 2019; Patwardhan et al., 2009; Reetz and Carballeira, 2007), and predicting phenotypes for neighboring mutants (Mighell et al., 2018). So what if I wanted to test multiple mutations or explore the "deep mutational space" of a developmental enhancer? For a 292 bp enhancer such as *E3N*, mutating every individual base in a unique combination comes out to be 4^{292} unique genotypes. That is a staggering sixty-three septenquinquagintillion (6.3×10^{175}) combinations to explore[5]. The lab would need a few more Ph.D. students and robots to tackle this experiment.

Developmental bias

Before giving up immediately on predicting *E3N's* evolution, there is a glimmer of hope. I do not need to look at all sixty-three septenquinquagintillion mutations of *E3N* to predict how it could evolve. I already begin to find correlations in the phenotypes shown in Chapters 3 and 4. These types of correlations are referred to as developmental biases (Smith et al., 1985). Developmental biases, despite the negative connotation, are trait correlations with either positive (developmental drive) or adverse effects on fitness (developmental constraint) (Arthur, 2002).

When describing developmental bias and constraint, a helpful representation is a phenotypic landscape (Aguilar-Rodríguez et al., 2017). Such landscapes are multidimensional spaces that highlight the relationships between expression pattern components. In **Figure 24A,** a phenotypic morphospace is illustrated between two arbitrary traits, 1 and 2. The wild-type phenotype is in the center in magenta, and the mutant phenotypes are in white. These mutant phenotypes are isotropically scattered without a visible correlation in the plot, telling us there is no apparent bias

---

[5] A similar anecdote is presented in Andreas Wagner's book, *Robustness and Evolvability in Living Systems* (Wagner, 2013) to explain the problem of studying evolvability.

between the two traits. Conversely, in **Figure 24B,** the traits are strongly correlated with one another and follow a trend (magenta dotted line). There are regions where the traits are biased towards evolving and areas where they are absent (red).

When analyzing a morphospace, the question is always, why do we see the trends and a lack of phenotypes? The answer to this question is not entirely clear. There are two different evolutionary schools of thought to approach it: Functionalism and Structuralism (Wagner, 2014). The Functionalist would look at this correlation – particularly for allometric relationships (Bolstad et al., 2015) - and claim the traits are selected to be linked. The Structuralist would argue that this is not necessarily true because this linkage could be due to interdependencies such as sharing an underlying gene regulatory network (Schaerli et al., 2018).
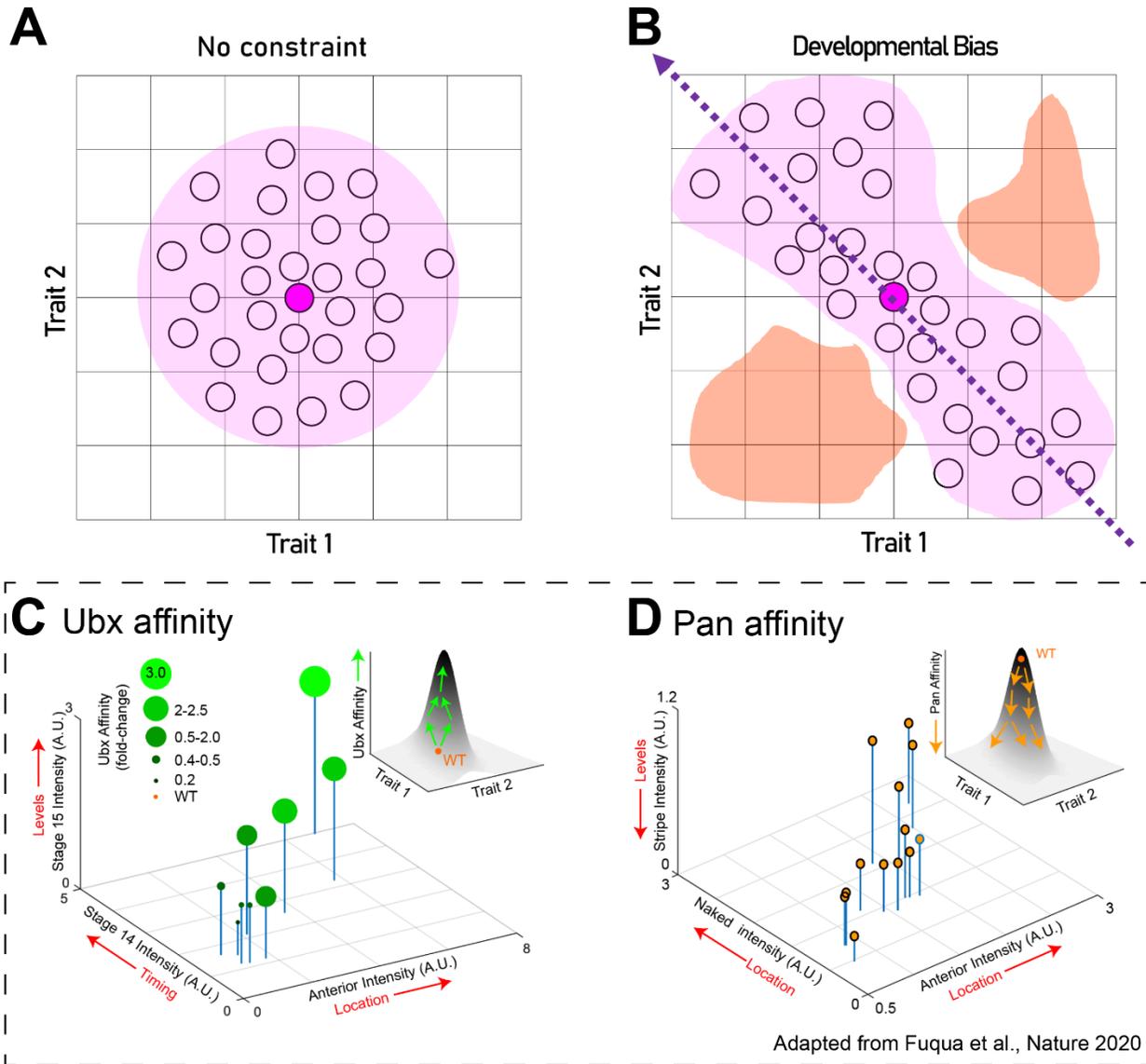
FIGURE 24 THE PHENOTYPIC MORPHOSPACES OF *E3N*

(A-B) A phenotypic landscape/morphospace that depicts the relationship between two arbitrary traits. The dark magenta circle is the WT sequence, and surrounding circles mutants or variants. (A) Phenotypic landscape not under apparent constraints or developmental biases. (B) Phenotypic landscape with a developmental bias: possibly due to constraint or selective pressures. (C-D) Phenotypic landscapes for *E3N* reveal pleiotropic relationships and developmental biases. (C) Total affinity of the TF Ubx is correlated with the levels of stripe intensity at stage 15, early activation at stage 14, and ectopic activity in the anterior.(D) The affinity of the TF Pangolin is correlated with the stripe levels, the expression between the stripes, and anterior expression.

Panels C and D adapted from Fuqua et al., 2020. Figure created by Timothy Fuqua.

Predicting *E3N:* a Structuralist approach

One advantage of our *E3N* study is that I can disentangle the Structuralist and Functionalist arguments. The mutant *E3N* lines generated had a Poisson distribution of mutations, and almost every base pair was mutated. The lines were tested as reporter constructs and had likely insignificant effects on fitness or selective pressures. Using reporter constructs allows me to remove fitness from the analysis and look at the phenotypes from a purely Structuralist viewpoint. Meaning, I assume that the correlations and lack of phenotypes are due to molecular constraints and not due to fitness. The caveat of this is that the developmental biases identified do not tell us anything about the selection acting upon the enhancer. For the *E3N* enhancer, I specifically analyzed the pleiotropic relationships for the transcription factors: Ultrabithorax and Pangolin, which developmentally bias possible *E3N* phenotypes.

For Ultrabithorax (Ubx), I found that the total affinity of Ubx binding is correlated with developmental timing, in addition to stripe intensity, and anterior expression (**Figure 24C**) (Crocker et al., 2015). This linkage of traits suggests a relative developmental bias (for definitions see Arthur, 2002) for possible phenotypes when changing the total Ubx affinity. Ubx affinity is thus a tunable and predictable evolutionary parameter that can be used to predict the final expression pattern for *E3N* mutations. I additionally found another relative developmental bias between the total affinity of Pan, stripe intensity, anterior expression, and ectopic expression between the stripes (**Figure 24D**). Like Ubx, total Pan affinity is a tunable and predictable evolutionary pattern for predicting phenotypes independent of selection.

Say a hypothetical selective pressure was favoring the increase of a single trait such as earlier activation while maintaining (stabilizing) the stripe intensity. To achieve this phenotype, evolution would have to tune multiple parameters: 1) raise Ubx affinity to activate earlier expression and consequently increase stripe and anterior intensity, and 2) decrease Pan affinity to lower stripe intensity and consequently create ectopic naked stripe expression and anterior expression. **In other words: evolution is likely not tuning single parameters due to the highly pleiotropic regulatory information, but rather many parameters simultaneously to account for pleiotropy and dense encoding.** The result is that multiple traits are changed to activate earlier

and stabilize stripe intensity. It is likely that each of the TF binding sites in the oversimplified model of *E3N* in **Figure 19** is a tunable parameter to stabilize or generate new phenotypes. Requiring multiple tunable parameters would explain rapid sequence turnover in enhancers and why sequence conservation is a poor metric of enhancer function.

A clear demonstration of evolution using multiple TF binding sites as tunable parameters through stabilizing selection was done with the *eve stripe 2* enhancer (Martinez et al., 2014). Martinez et al. created functional models for the *eve stripe 2* enhancer across different *Drosophila* species. The enhancer makes the same expression pattern across species (Ludwig et al., 1998, 2000). Using a combined modeling and phylogenetic approach, the group could reconstruct functional ancestral forms of the enhancer. The group found that binding sites for the activators and repressors were being tuned to compensate for the maintained expression pattern. In particular, the total amount of Bicoid and Hunchback activation was positively correlated with the amount of Giant repression (Martinez et al., 2014). A previous finding also supports these results at the *eve* locus, that functional divergence and convergence can be attributed to changes in activation levels (Ludwig et al., 2005).

Predicting *E3N:* a population genetics approach

"Predicting the genetic basis of evolution requires a comprehensive synthesis of molecular developmental biology and population genetics" (Stern and Orgogozo, 2008)

Experiments such as the *eve stripe 2* reconstruction are a great way to understand and predict regulatory evolution since it is an intersection between evolutionary-developmental biology and population genetics. To predict the evolution of *E3N,* I have already discussed the molecular components of *E3N* and now focus on population genetics.
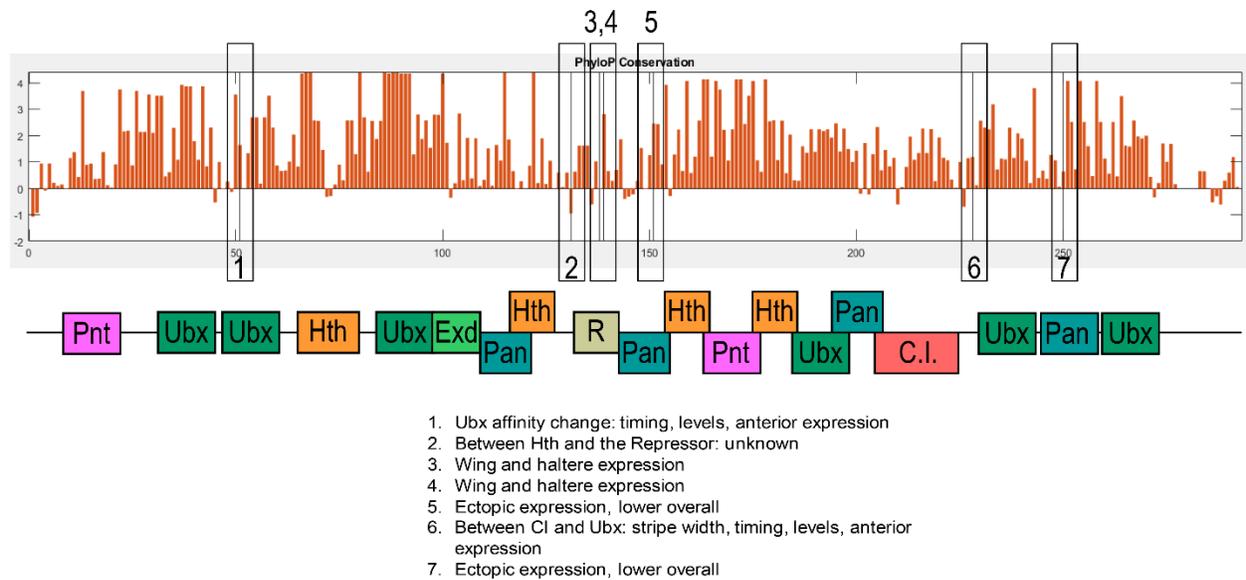
FIGURE 25 STANDING VARIATION OF *E3N* IN *DROSOPHILA MELANOGASTER*

(Top) PhyloP conservation scores for the *E3N* enhancer. The x-axis is the *E3N* enhancer sequence from 5' → 3'. The y-axis is the rate of evolution, where the higher the peak, the slower the evolution is occurring. Black boxes highlight regions of standing variation. Fabian Ruperti and I identified 7 SNPs within *E3N.* **(Bottom)** Hypothesized phenotypic effects of the 7 SNPs based on their location within the sequence. Figure created by Timothy Fuqua and Fabian Ruperti.

When considering the evolution of a developmental enhancer, looking into populations can give great insights into predicting evolution (Fisher, 1919). This approach includes measuring the variation within and across species at the genotypic and phenotypic levels. Using the *Drosophila* NEXUS and DGRP datasets (Huang et al., 2014; Lack et al., 2016; Mackay et al., 2012), Fabian Ruperti and I identify seven polymorphisms in the *melanogaster svb E3N* enhancer (**Figure 25**), which I call SNPs for simplicity. A closer look at Figure 26 reveals where the seven SNPs align with the *E3N* sequence conservation and mapped TF binding sites at the genotypic level. Based on the results from the *E3N* screen, these mutations will likely affect *E3N* expression since most mutations tested showed phenotypic effects on the minimalized element (Fuqua et al., 2020).

The specific effects of these mutations can also be predicted based on where they are located. For example, SNP 1 will likely change the affinity of Ubx, which will cause pleiotropic effects on timing, levels, and anterior expression. SNPs 3 and 4 are located near the wing and haltere

repressor site and may exhibit ectopic expression in this region. There are additionally SNPs in multiple Pan sites, which will likely pleiotropically lower expression but drive expression in-between the stripes. SNPs lying between the binding sites may exhibit effects expected from both binding factors due to potential binding competition. In the future, I would like to test these SNPs to see how they affect both minimal *E3N* reporter expression and how they affect cuticle and trichome patterning to begin making Functionalist inferences.

Structuralism, Functionalism, and Population Genetics reveal selection and constraint

Structuralist approaches reveal developmental biases and nothing about selection. Functionalist approaches can reveal potential biases which may or may not be selective but say nothing about the developmental constraints. To what extent can we learn about evolution from either approach? By approaching a biological problem from both perspectives, we can gather evidence for evolutionary constraint and selection.

Taking a Functionalist approach, we can compare patterns across species to identify trends and correlations. However, these trends and correlations cannot tell us anything about the molecular mechanisms underlying the correlations. For example, Justin Crocker and David Stern analyzed the ventral trichomes of over 60 different *Drosophila* species in Chapter 3 (see Figure 17). The two found examples of trichome loss from these cuticle preps and an absence of trichomes between the stripes. Functionally, this suggests to me that selection does not favor stripes between the cuticles. Structurally, this indicates that it is not possible to form trichomes between the stripes. Based on the previously identified pleiotropic relationship between stripe intensity, the ectopic expression between the stripes, and Tcf affinity, I speculate that I see these results because of the molecular constraints acting upon the system.

5.5   SUMMARY AND FUTURE DIRECTIONS

In this thesis, *The Evolutionary Landscape of a Developmental Enhancer,* I developed a semi-automated robotics pipeline to mutationally scan a mutant reporter library for the *svb E3N* enhancer. The experiment found that most base pairs contribute to the *E3N* expression pattern and that many mutations were pleiotropic, suggesting that *E3N* and other enhancers are densely encoded with regulatory information. The experiment further identified pleiotropic relationships, developmental biases, and limits on ectopic mutants within *E3N* which constrain enhancer evolution. I also analyzed the mutant expression patterns and found that the WT *E3N* is fragile, but the enhancer also encodes an underlying robust expression pattern. All of this experimental information was culminated into an up-to-date working model of *E3N* logic and evolution.

There are many different research directions to pursue from this body of work. From an engineering perspective, there are many improvements that I can make to the automation and robotics to increase screening throughput. From a *shavenbaby* perspective, I could screen more *E3N* mutants with a larger – or smaller – mutational frequency to explore the phenotypic space further and tease apart the enhancer's grammar. It would also be interesting to see how the mutant phenotypes respond to different temperatures or stressors to study further how *svb* canalizes robust expression (Frankel et al., 2010). To better understand the evolution of *E3N*, I could test the standing variation within *E3N* and those seven SNPs mentioned previously, and attempt to trace the evolutionary paths between species like Glassford and Rebeiz did for the *Nep1 optic lobe* enhancer (Glassford and Rebeiz, 2013). I could also create ancestral reconstructions like what was done with *eve stripe 2* enhancer (Martinez et al., 2014).

I could also screen other developmental enhancers or cis-regulatory elements such as promoters, polycomb response elements, or insulators. It would also be interesting to carry out these screens at the native locus using CRISPR or to mutate multiple enhancers at once to understand enhancer crosstalk. One could dedicate a lifetime of research to enhancer regulation and evolution. This thesis is hopefully only the beginning of high-throughput enhancer screening.

6. MATERIALS AND METHODS

Software

You can find the Computer-Aided Design (CAD) files for Flyspresso using the link below:

https://github.com/tfuqua95/Flyspresso-CAD-files

You can find the codes for Flyspresso and schematics for the circuit boards can be found using the link below:

https://github.com/janelia-pypi/hybridizer_python/tree/digital

You can find the instructions for installing and operating the adaptive feedback microscopy pipeline using this link as well as in Fuqua et al., 2021:

https://git.embl.de/grp-almf/feedback-fly-embryo-crocker

Solutions

- **Antibody Fixative**: 4.6% Paraformaldehyde (Electron Microscopy Sciences 15710) and 25 µM EGTA in PBS.
- **BABB**: 1 part benzyl alcohol (Sigma-Aldrich 305197-1L) and 2 parts benzyl benzoate (Sigma-Aldrich B6630-1L).
- **Blocking Solution**: 1:5 Western Blocking Reagent (Roche SKU 11921673001) and PBT solution.
- **Fly Saline**: 0.1 M NaCl and 0.04% Triton X-100 (Sigma –Aldrich X100-100ML) in sterile water.
- **PBT**: PBS and 0.1% Triton X-100.
- **X-gal Fixative**: - 2% Formaldehyde (Sigma F8775 – 25ML) and 0.2% Glutaraldehyde (Sigma G5882-50ML) in PBS.
- **X-gal Staining Solution**: - 5-bromo-4-chloro-3-indolyl-ß-D-galactosidase (Invitrogen B1690 – 1G) [20 mg/mL DMF (Sigma D4551-250ML)], 400 mM potassium ferricyanide (III) (Sigma-Aldrich 244023-100G), 400 mM potassium ferrocyanide (Sigma-Aldrich P3289), 200 mM magnesium chloride, H_2O.

Egg collections (30 minutes)

Add flies to the collection chambers with some yeast paste. Let the flies acclimate to the chambers for at least 24 hours. To do an overnight egg collection, move the chambers to 25°C around 18:00 and collect the following morning at 09:00. Swap the Transplate attachments and remove dead flies from the screens with tweezers. Next, wash the embryos in with Fly Saline Solution and bleach the embryos for 90 seconds in a 50% bleach solution. Wash the embryos with water and load the embryos into the Flyspresso microplates. Cap the Transplates with Seplate attachments. Continue to either **X-gal staining** or **Antibody staining**.

X-gal staining (3 hours)

Using Flyspresso, add 2 mL of the X-gal Fixative and 2 mL of Heptane to each well of the microplate. The heater/shaker device shakes the microplate for 20 minutes at 200 RPM. After the fixation, program a pause step to remove the Transplates and blot their screens on a paper towel to remove excess heptane. Have Flyspresso carry out three consecutive ten-minute PBT washes. Add 4 mL of the X-gal staining solution to each of the wells and incubate the embryos for 2 hours at 37°C. Have Flyspresso carry out three consecutive ten-minute PBT washes. Image the embryos in PBT.

Antibody staining (7 hours)

Using Flyspresso, add 4 mL of the Antibody Fixative Solution and 4 mL of Heptane to each microplate well. The heater/shaker device shakes the microplate for 25 minutes at 250 RPM. Program a pause step after the fixation to blot the Transplates dry with a paper towel. This step removes excess heptane and increases isotonic shocking efficiency. Have Flyspresso carry out three consecutive ten-minute methanol washes.

Serially wash the embryos in PBT in ten-minute intervals until fully rehydrated. Have Flyspresso add 4 mL of the Blocking Solution to each well and block for 25 minutes, shaking at 250 RPM. Pause Flyspresso and add 4 mL of primary antibody solution to each well. For the experiments in this thesis, antibodies were stained at the following concentrations: Beta-Galactosidase (1:500, abcam ab9361), Crumbs (1:10, DSHB Cq4 Supernatant), ELAV (1:20, DSHB Elav-9F8A9), Pax7 (1:4

DSHB PAX7), Myosin (1:4, DSHB F59), RFP (1:500, MBL PM005). Allow samples to incubate for 2 hours to overnight in the solution, shaking at 250 RPM. Zebrafish larvae and *Drosophila* brains were incubated for 16 hours at 4°C.

Wash the antibody solution out of the microplate with two consecutive ten-minute PBT washes. Block the samples again as previously described. Pause Flyspresso again and add 4 mL of the antibody staining solution to each well with AlexaFluor 488, 633, and 647 (1:500, Invitrogen) secondary antibodies. Cover the samples during the 2-hour incubation, shaking at 250 RPM. Wash the samples twice consecutively for ten minutes in PBT. Mount the samples in water or glycerol-based media. If mounting in BABB continue to the **BABB** section below.

BABB (48 hours)

Flyspresso serially dehydrates the embryos in ethanol in ten-minute increments. Using a Pasteur pipette, transfer the embryos to 1.5 mL Eppendorf tubes. Remove as much ethanol as possible and add 1 mL of BABB to each of the tubes. The samples will become transparent and very difficult to see. Incubate the samples in BABB for ten minutes each and wash again in BABB overnight. Aspirate at the 100-µL mark to avoid removing the samples.

To mount the samples, cut the Grace Silicone isolators to include the number of wells for your samples. To each well, add 100 µL of the BABB/embryo mixture to each of the wells. The solution will be raised above the well held by surface tension. Allow the embryos to sink after adding the BABB/embryos to each well. Using a pipette tip, connect the BABB between wells and slowly lower the coverslip from left to right over the samples. Seal the slide with three coats of clear nail polish. Let the samples rest for 24 hours before imaging.

BABB and Prolong Gold depth

I stained the embryos with the Crumbs antibody (1:10, DSHB Cq4 supernatant) and Alexa Fluor 488 (1:500, Invitrogen). I split the samples into two pools, where one was mounted in Prolong Gold and the other in BABB. Ten embryos per pool were imaged and measured the intensities in

Fiji. The index of refraction had to be corrected for BABB (HELL et al., 1993). I used the following formula for the correction (Diel et al., 2020; VISSER et al., 1992):

$$\frac{d'}{d} = \frac{\tan\left(\sin^{-1}\frac{0.5\ NA}{n_1}\right)}{\tan\left(\sin^{-1}\frac{0.5\ NA}{n_2}\right)}$$

E3N library assembly

The enhancer library was assembled with Genscript (Genscript, Netherlands) primarily by Jonathan Sager using degenerate PCR with a 2% mutational frequency (**Fig 26A**). Genscript then cloned the variants into the *pLacZattB* vector. Plasmids were integrated into the *attP2* site, and positive transformants were sequenced. The *E3N* enhancer is 35% T, 21% G, 17% C, and 27% A (**Fig 26B**). I calculated mutation rates using these numbers to find mutational biases (**Fig 26C**) and found that the library was normally distributed with a slight bias of mutating Gs at a higher frequency compared to other bases. The mutations are scattered across the entire *E3N* sequence (**Fig 26D**).
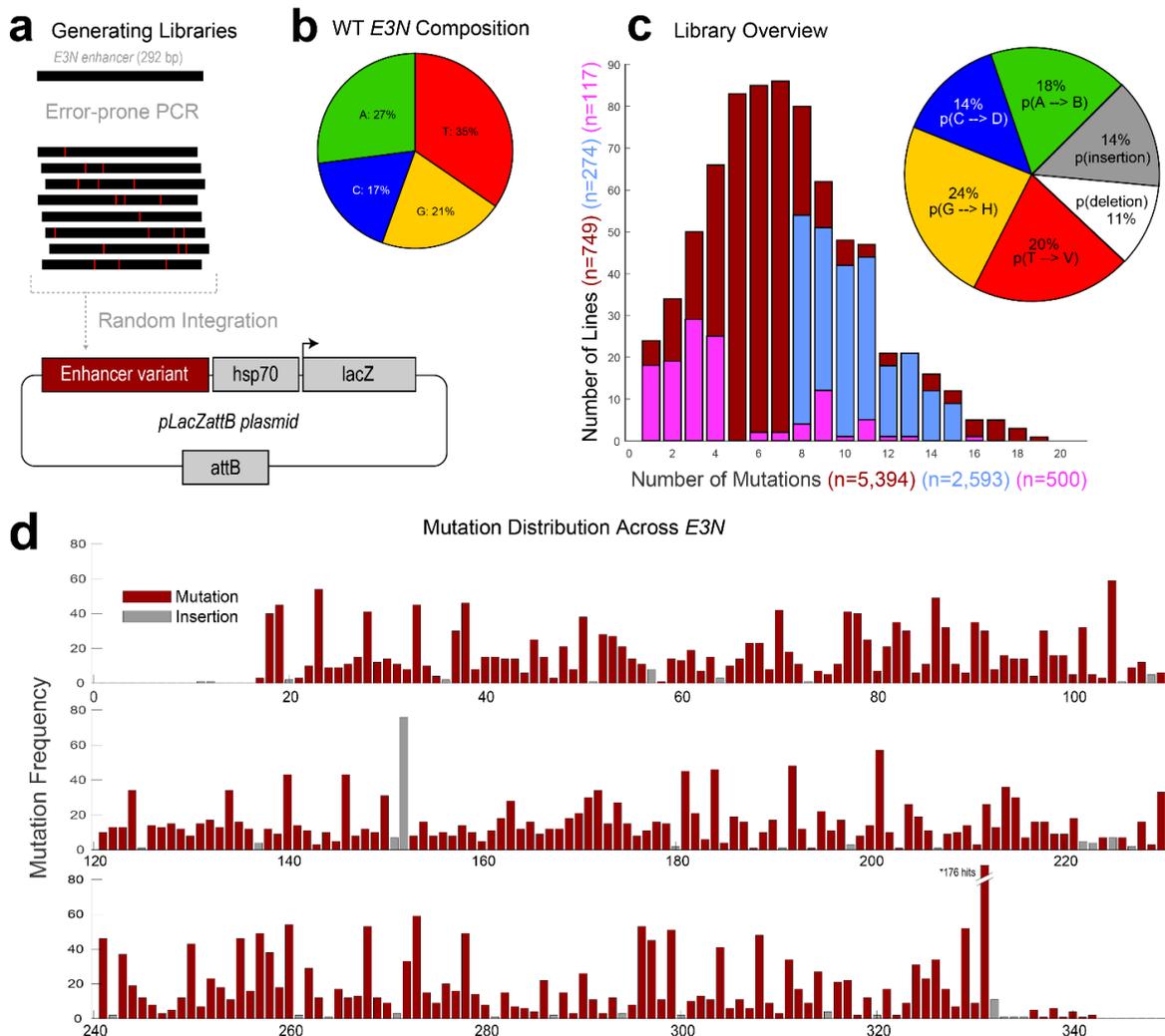
FIGURE 26 DISTRIBUTION OF MUTATIONS

(**A**) Generating a mutant library for *E3N* using degenerate PCR and the *placZattB* plasmid. The plasmid was integrated into the *attP2* landing site. (**B**) Composition of *E3N.* These values were used to normalize the mutation rates in (**C**). (**C**) Distribution of mutations tested using antibody staining, x-gal staining, and the total library distribution. Pie chart shows the probability (p) of a mutation occurring, given the WT base pair. (**D**) A Manhatten plot for all mutations within *E3N.* Gray bars are insertions and crimson bars are marked in crimson.

Figure adapted from Fuqua et al., 2020. Library assembled by Jonathan Sager, Justin Crocker, and Timothy Fuqua. Figure created by Timothy Fuqua and Jonathan Sager.

Footprinting scores

I aligned the enhancers using *pairwise2* alignment in *Biopython* (Cock et al., 2009). I treated deletions as mismatches and removed insertions from the analysis, so all sequences were 292 bps. Binary alignments were made where matches = 0 and mismatches = 1. For every base (i = 1...292) in the WT *E3N* and each mutant line (j = 1...274), I assigned a score $a_{i,j}$ = 0 to every base where the line was not mutant, and $a_{i,j}$ = 1 for a mutant base. The Mutation Coverage M_i is the sum across all 274 lines:

$$M_i = \sum_{j}^{274} a_{i,j}$$

For every base (i = 1...292) in the WT *E3N* and each mutant line (j = 1...274), a score $s_{i,j}$ = 0 was given if the base is either not mutated or did not significantly change *E3N* expression. For mutant bases that changed *E3N*, $s_{i,j}$ = 1. The total score S_i at every base i is equal to the sum across all lines:

$$S_i = \sum_{j}^{274} s_{i,j}$$

I then thresholded the values, where every base mutated less than 5 times was removed ($S_i \equiv$ NaN). We normalized the footprinting scores over the total coverage:

$$\sigma_i = \frac{S_i}{M_i}$$

The data were smoothened in Matlab with the *smoothdata* function (Gaussian-weighted moving average window = 5 bases) and plotted. Albert Tsai provided discussion and input for writing these formulas.

EWAC Scores

For every base (i = 1...292) in the WT *E3N* sequence, the total score A_i at every base i is the total coverage M_i subtracted from S_i (see Calculated Footprinting Scores):

$$A_i = S_i - M_i$$

If a value was not available for S_i (S_i = NaN) A_i was set to = 0.5.

The total score C_i at every base (i = 1…292) is the score A_i subtracted from all lines without expression (Q = 129), subtracted from the total number of lines (J = 274):

$$C_i = J - Q - A_i$$

The total score D_i at every base (i = 1…292) is the score S_i (see **Footprinting scores**) subtracted from the total number of lines without expression (Q = 129):

$$D_i = Q - S_i$$

I generated a 2x2 contingency table for every base (i = 1…292) and carried out a chi-squared test:

	X+	X-
Mut +	A_i	S_i
Mut -	C_i	D_i

I used *chi2_contingency* from SciPy (Virtanen et al., 2019) and combined the data using *Pandas* (McKinney, 2010). We calculated the Q-values utilizing the approach in Storey et (Storey and Tibshirani, 2003). Jakob Wirbel, Judith Zaugg, and Albert Tsai helped with the math for calculating EWAC scores.

Image and analysis

For Chapter 3, Z-stacks of the embryo images were max-projected, rotated, cropped, and concatenated into a montage to view the variation within each mutant line (**Fig 27A**). For some of the images of embryos mounted in BABB, Christian Tischer rotated the embryos computationally (**Fig 27B**). Using fiduciary markers, I carried out elastic transformations of the images (see Figure 28) (**Fig 27C**) and composited the deformed embryos into a single statistical embryo (**Fig 27D**). When these registration methods were not possible, patterns were analyzed by using three methods: Sliding Window, State Method, and Plot Profiles.

For the Sliding Window (**Fig 27E**), I drew a box the size of the naked region and centered it over the stripe, and measured the average intensity within the box, dragging it over the entire embryo. Data were plotted as box plots using *NotBoxPlot* in Matlab (Campbell, 2020), where each data point is a single measurement. For the State Method (**Fig 27F**), I drew an oval ROI and traced over each nucleus – active or not – and measured its nuclear intensity. Each point on the plot is a single nucleus. Gray bars are violin plots representing the data (Jonas, 2020). For the Plot Profiles (**Fig 27G**), I drew a box across the striped expression pattern and averaged the y-values. Dark gray values are the average, and gray is +/- one standard deviation.

For Chapter 4, I used a more complex image analysis on the *E3N* expression pattern. I trained a machine-learning classifier to identify nuclei based on a wild-type *E3N* expression pattern using the Ilastik toolkit (Berg et al., 2019). The classifier did not take nuclear intensities into account. The classifier automatically selected the nuclei from either the entire embryo or nuclei within the stripe region, measured their intensities and counted them. For stripe widths, I used the *find_peaks* function in matplotlib (Hunter, 2007). Plots were generated using Seaborn (Waskom, 2021). Christian Tischer provided many fruitful discussions on image analysis during my Ph.D.
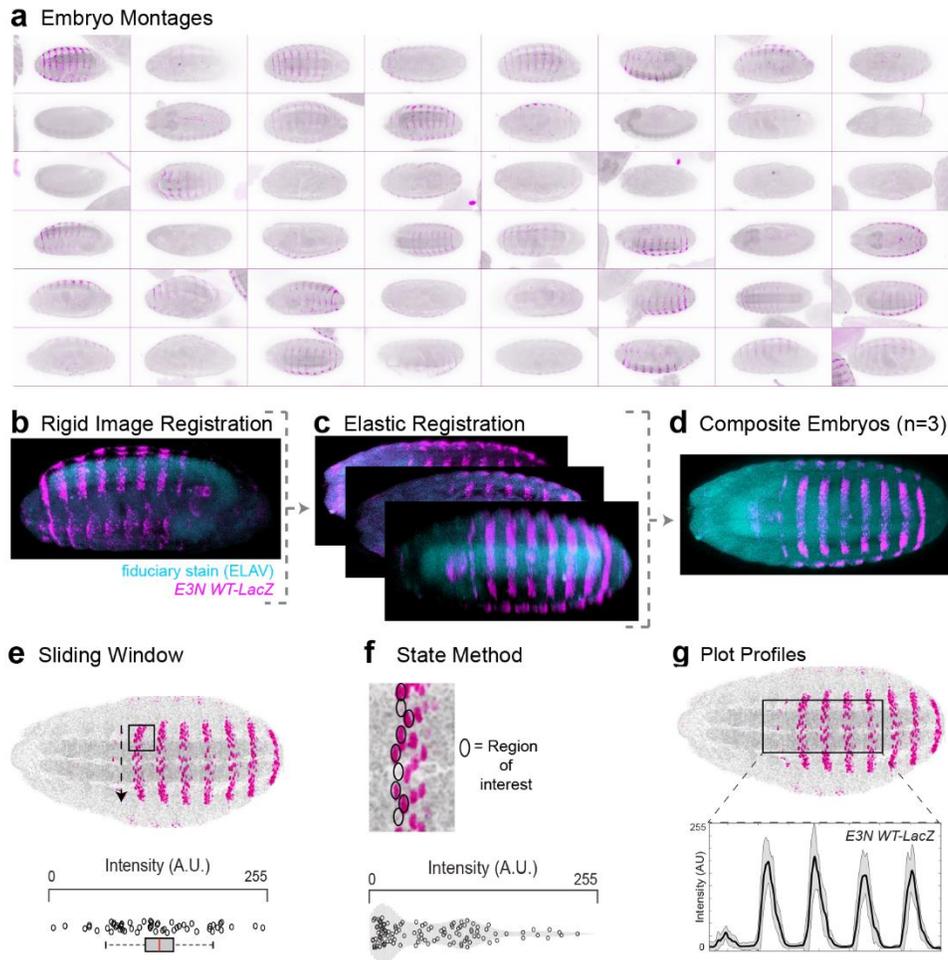
a Embryo Montages

b Rigid Image Registration

fiduciary stain (ELAV)
E3N WT-LacZ

c Elastic Registration

d Composite Embryos (n=3)

e Sliding Window

Intensity (A.U.)
0   255

f State Method

() = Region of interest

Intensity (A.U.)
0   255

g Plot Profiles

E3N WT-LacZ

255

Intensity (AU)

0

FIGURE 27 IMAGE ANALYSIS

(**A**) Montage of projected embryos. (**B-D**) Registering images with fiduciary stains. (**B**) The embryo is computationally rotated based off the fiduciary stain. (**C**) Images are registered to each other based off the fiduciary stain. (**D**) The 2D projections are averaged together. (**E**) Sliding Window technique. A square ROI is slid across the stripe, and measurements are taken. Multiple measurements are gathered and plotted as single points on a boxplot. The line in the middle is the mean, upper and lower limits are the standard deviation and the whiskers are a 95% confidence interval. (**F**) The State Method. An oval ROI is connected across an entire stripe, going over individual nuclei in the A2 stripe. Each measurement is a single point on the plot. (**G**) Plot Profiles. A rectangle ROI is drawn between the A1 and A5 segments. The mean is acquired for each pixel column and plotted. The shaded regions is +/- standard deviation.

Figure adapted from Fuqua et al., 2020. Experiments carried out by Timothy Fuqua. Christian Tischer helped with embryo rotations in panel b.

DSHB atlas

The DSHB antibody atlas project is still incomplete. All of the antibodies are co-stained with a *svb BAC* driving *dsRed* expression and DAPI (**Fig 28A**). I trained a classifier using Ilastik to identify the vertical stripes from Svb and horizontal stripes from the ventral nerve cord – converting the pattern into a coordinate "grid" system. A master template image was generated this way (**Fig 28A**) and for each experimental image (**Fig 28B**). With the two sets of coordinates, the experimental "grid" is registered to the template "grid" using *bUnwarpJ* (Arganda-Carreras et al., 2006) (**Fig 28C**). The mathematical transformation calculated is also applied to the channel with the antibody staining (**Fig 28D**). The registered images are averaged together for different stains (**Fig 28E-H**) and can be viewed together in a digital composite image (**Fig 28I**).
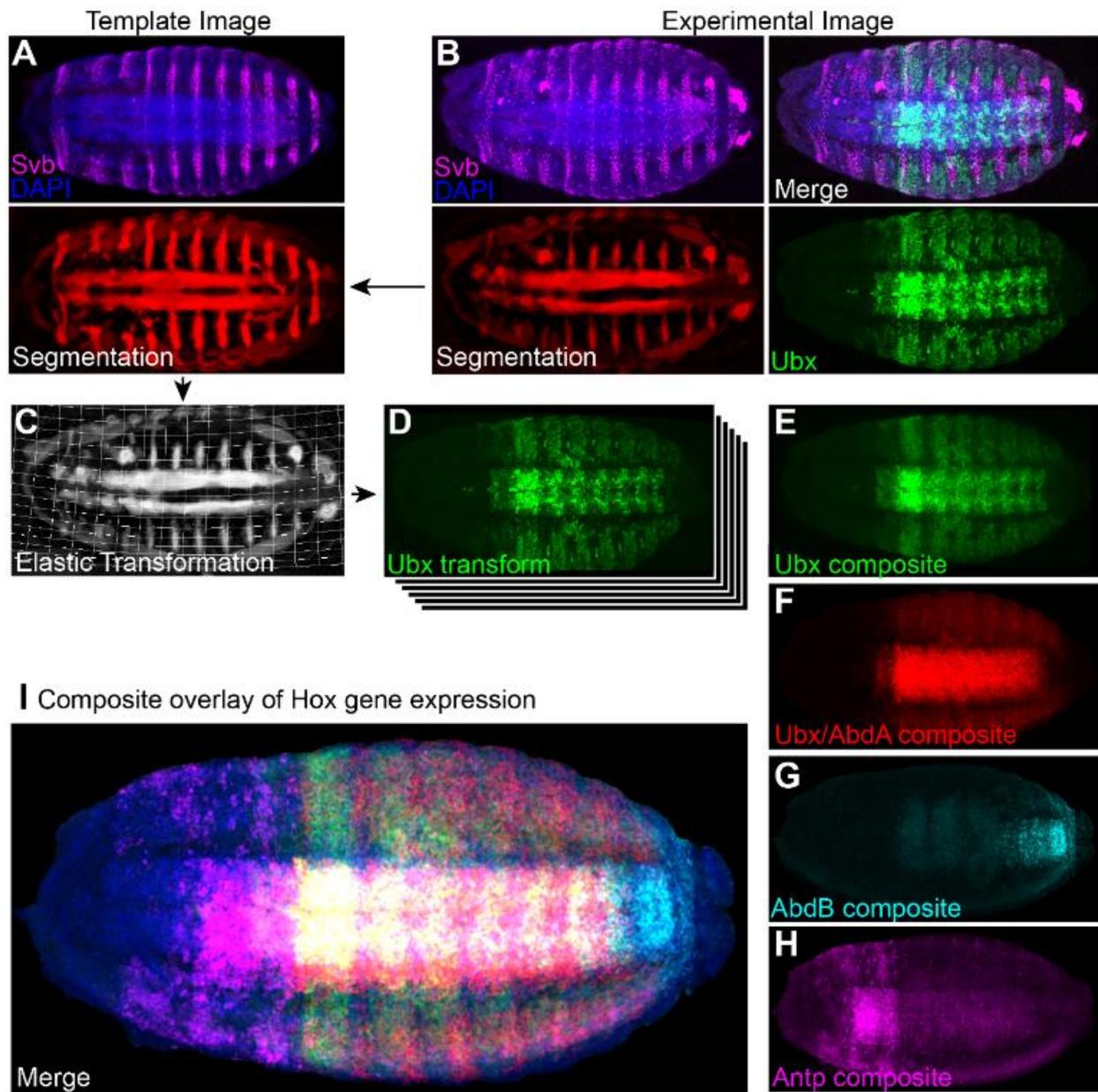
FIGURE 28 CREATING A GENE EXPRESSION ATLAS USING INTERNAL FIDUCIARY STAINS

(A) A template image that all stains were aligned to was co-stained with Svb expression (svbBAC::dsRed) and DAPI. Svb and DAPI expression were segmented using a neural network to mask the central nervous system (two horizontal stripes) and the Svb stripes (vertical stripes). (B) Experimental images were additionally co-stained with antibody or stain of interest (Ubx FP3.38 Antibody). (C) The segmented mask from the experimental image was elastically transformed to match the segmented mask from the template image. (D) The transformation function is applied to the stain of interest. (E) Transformed stains are composited together using the mean or median of all transformed images for Ubx, (F) Ubx/AbdA, (G) AbdB, (H) and Antp. (I) Composites can be loaded as channels to create a representative image. Experiments carried out by Timothy Fuqua.

7. REFERENCES

Abzhanov, A. (2004). Bmp4 and Morphological Variation of Beaks in Darwin's Finches. Science (80-.). *305*, 1462–1465.

Aguilar-Rodríguez, J., Payne, J.L., and Wagner, A. (2017). A thousand empirical adaptive landscapes and their navigability. Nat. Ecol. Evol. *1*, 0045.

Anderson, K. V., Jürgens, G., and Nüsslein-Volhard, C. (1985). Establishment of dorsal-ventral polarity in the Drosophila embryo: Genetic studies on the role of the Toll gene product. Cell *42*, 779–789.

Andrilenas, K.K., Penvose, A., and Siggers, T. (2015). Using protein-binding microarrays to study transcription factor specificity: homologs, isoforms and complexes. Brief. Funct. Genomics *14*, 17–29.

Arganda-Carreras, I., Sorzano, C.O.S., Marabini, R., Carazo, J.M., Ortiz-de-Solorzano, C., and Kybic, J. (2006). Consistent and Elastic Registration of Histological Sections Using Vector-Spline Regularization. In LNCS, Volume 4241, pp. 85–95.

Arnone, M.I., and Davidson, E.H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development *124*, 1851–1864.

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

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

Arnoult, L., Su, K.F.Y., Manoel, D., Minervino, C., Magrina, J., Gompel, N., and Prud'homme, B. (2013). Emergence and Diversification of Fly Pigmentation Through Evolution of a Gene Regulatory Module. Science (80-.). *339*, 1423–1426.

Arthur, W. (2002). The interaction between developmental bias and natural selection: From centipede segments to a general hypothesis. Heredity (Edinb). *89*, 239–246.

Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., et al. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. Nat. Genet. *53*, 354–366.

Aza-Blanc, P., Ramírez-Weber, F.-A., Laget, M.-P., Schwartz, C., and Kornberg, T.B. (1997). Proteolysis That Is Inhibited by Hedgehog Targets Cubitus interruptus Protein to the Nucleus and Converts It to a Repressor. Cell *89*, 1043–1053.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Barolo, S. (2012). Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays *34*, 135–141.

Barr, K., Reinitz, J., and Radulescu, O. (2019). An in silico analysis of robust but fragile gene

regulation links enhancer length to robustness. PLoS Comput. Biol. *15*, 1–22.

Bartman, C.R., Hsu, S.C., Hsiung, C.C.S., Raj, A., and Blobel, G.A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol. Cell *62*, 237–247.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science (80-.). *304*, 1321–1325.

Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., and Haussler, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature *441*, 87–90.

Bejsovec, A. (2006). Flying at the head of the pack: Wnt biology in Drosophila. Oncogene *25*, 7442–7449.

Belliveau, N.M., Barnes, S.L., Ireland, W.T., Jones, D.L., Sweredoski, M.J., Moradian, A., Hess, S., Kinney, J.B., and Phillips, R. (2018). Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. Proc. Natl. Acad. Sci. U. S. A. *115*, E4796–E4805.

Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., et al. (2019). ilastik: interactive machine learning for (bio)image analysis. Nat. Methods *16*, 1226–1232.

Berman, B.P., Nibu, Y., Pfeiffer, B.D., Tomancak, P., Celniker, S.E., Levine, M., Rubin, G.M., and Eisen, M.B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. Proc. Natl. Acad. Sci. U. S. A. *99*, 757–762.

Blomberg, S.P., and Garland, T. (2002). Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. J. Evol. Biol. *15*, 899–910.

Bodmer, R. (1993). The gene tinman is required for specification of the heart and visceral muscles in Drosophila. Development *118*, 719–729.

de Boer, C.G., Vaishnav, E.D., Sadeh, R., Abeyta, E.L., Friedman, N., and Regev, A. (2020). Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. Nat. Biotechnol.

Bolstad, G.H., Cassara, J.A., Márquez, E., Hansen, T.F., Van Der Linde, K., Houle, D., and Pélabon, C. (2015). Complex constraints on allometry revealed by artificial selection on the wing of Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. *112*, 13284–13289.

Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. Elife *4*, 1–14.

Brakefield, P.M., Gates, J., Keys, D., Kesbeke, F., Wijngaarden, P.J., Montelro, A., French, V., and Carroll, S.B. (1996). Development, plasticity and evolution of butterfly eyespot patterns. Nature *384*, 236–242.

Bray, S. (1999). Drosophila development: Scalloped and Vestigial take wing. Curr. Biol. *9*, R245–R247.

van Breugel, F., and Dickinson, M.H. (2017). Superhydrophobic diving flies (Ephydra hians) and the hypersaline waters of Mono Lake. Proc. Natl. Acad. Sci. *114*, 13483–13488.

C., S., N., A., Cui, J., Low, J., and Che, T. (2013). Practical Considerations of Liquid Handling Devices in Drug Discovery. In Drug Discovery, H.A. El-Shemy, ed. (Rijeka: InTech), p.

Campbell, R. (2020). notBoxPlot.

Campbell, S., Inamdar, M., Rodrigues, V., Raghavan, V., Palazzolo, M., and Chovnick, A. (1992). The scalloped gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in Drosophila. Genes Dev. *6*, 367–379.

Campbell, S.D., Duttaroy, A., Katzen, A.L., and Chovnick, A. (1991). Cloning and characterization of the scalloped region of Drosophila melanogaster. Genetics *127*, 367–380.

Cande, J., Goltsev, Y., and Levine, M.S. (2009). Conservation of enhancer location in divergent insects. Proc. Natl. Acad. Sci. U. S. A. *106*, 14414–14419.

Carroll, S.B. (2006). Endless forms most beautiful: the new science of Evo Devo and the making of the animal kingdom (New York: W. W. Norton & Co.).

Chang, M. V., Chang, J.L., Gangopadhyay, A., Shearer, A., and Cadigan, K.M. (2008). Activation of Wingless Targets Requires Bipartite Recognition of DNA by TCF. Curr. Biol. *18*, 1877–1881.

Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F., and Plaza, S. (2006). Shavenbaby Couples Patterning to Epidermal Cell Shape Control. PLoS Biol. *4*, e290.

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.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422–1423.

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

Conrad, C., Wünsche, A., Tan, T.H., Bulkescher, J., Sieckmann, F., Verissimo, F., Edelstein, A., Walter, T., Liebel, U., Pepperkok, R., et al. (2011). Micropilot: automation of fluorescence microscopy–based imaging for systems biology. Nat. Methods *8*, 246–249.

Corradin, O., and Scacheri, P.C. (2014). Enhancer variants: evaluating functions in common disease. Genome Med. *6*, 85.

Crocker, J., and Erives, A. (2008). A Closer Look at the eve Stripe 2 Enhancers of Drosophila and Themira. PLoS Genet. *4*, e1000276.

Crocker, J., and Erives, A. (2013). A Schnurri/Mad/Medea complex attenuates the dorsal–twist

gradient readout at vnd. Dev. Biol. *378*, 64–72.

Crocker, J., and Ilsley, G.R. (2017). Using synthetic biology to study gene regulatory evolution. Curr. Opin. Genet. Dev. *47*, 91–101.

Crocker, J., and Stern, D.L. (2013). TALE-mediated modulation of transcriptional enhancers in vivo. Nat. Methods *10*, 762–767.

Crocker, J., and Stern, D.L. (2017). Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer. Development dev.149427.

Crocker, J., Potter, N., and Erives, A. (2010). Dynamic evolution of precise regulatory encodings creates the clustered site signature of enhancers. Nat. Commun. *1*, 99.

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., Ilsley, G.R., and Stern, D.L. (2016a). Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors. Nat. Genet. *48*, 292–298.

Crocker, J., Preger-Ben Noon, E., Stern, D.L.D.L., Noon, E.P.B., and Stern, D.L.D.L. (2016b). The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. In Current Topics in Developmental Biology, pp. 455–469.

Crocker, J., Tsai, A., and Stern, D.L. (2017). A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote. Cell Rep. *18*, 287–296.

Davis, J.E., Insigne, K.D., Jones, E.M., Hastings, Q.A., Boldridge, W.C., and Kosuri, S. (2020). Dissection of c-AMP Response Element Architecture by Using Genomic and Episomal Massively Parallel Reporter Assays. Cell Syst. *11*, 75-85.e7.

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.

Delon, I., Chanut-Delalande, H., and Payre, F. (2003). The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in Drosophila. Mech. Dev. *120*, 747–758.

Dessaud, E., McMahon, A.P., and Briscoe, J. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. Development *135*, 2489–2503.

Dey, S.S., Foley, J.E., Limsirichai, P., Schaffer, D. V, and Arkin, A.P. (2015). Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol. Syst. Biol. *11*, 806.

Diel, E.E., Lichtman, J.W., and Richardson, D.S. (2020). Tutorial: avoiding and correcting sample-induced spherical aberration artifacts in 3D fluorescence microscopy. Nat. Protoc. *15*, 2773–2784.

van Dijk, D., Sharon, E., Lotan-Pompan, M., Weinberger, A., Segal, E., and Carey, L.B. (2017). Large-scale mapping of gene regulatory logic reveals context-dependent repression by transcriptional activators. Genome Res. *27*, 87–94.

Dodt, H.U., Leischner, U., Schierloh, A., Jährling, N., Mauch, C.P., Deininger, K., Deussing, J.M., Eder, M., Zieglgänsberger, W., and Becker, K. (2007). Ultramicroscopy: Three-dimensional visualization of neuronal networks in the whole mouse brain. Nat. Methods *4*, 331–336.

Draganescu, A., and Tullius, T.D. (1998). The DNA binding specificity of engrailed homeodomain. J. Mol. Biol. *276*, 529–536.

Driever, W., and Nüsslein-Volhard, C. (1988). A gradient of bicoid protein in Drosophila embryos. Cell *54*, 83–93.

Duveau, F., Yuan, D.C., Metzger, B.P.H., Hodgins-Davis, A., and Wittkopp, P.J. (2017). Effects of mutation and selection on plasticity of a promoter activity in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. *114*, E11218–E11227.

EKER, R. (1929). THE RECESSIVE MUTANT ENGRAILED IN DROSOPHILA MELANOGASTER. Hereditas *12*, 217–222.

Ekker, S.C., Jackson, D.G., von Kessler, D.P., Sun, B.I., Young, K.E., and Beachy, P.A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. EMBO J. *13*, 3551–3560.

Emera, D., Yin, J., Reilly, S.K., Gockley, J., and Noonan, J.P. (2016). Origin and evolution of developmental enhancers in the mammalian neocortex. Proc. Natl. Acad. Sci. *113*, E2617–E2626.

Erives, A., and Levine, M. (2004). Coordinate enhancers share common organizational features in the Drosophila genome. Proc. Natl. Acad. Sci. *101*, 3851–3856.

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.

Fernandes, J., Celniker, S.E., Lewis, E.B., and VijayRaghavan, K. (1994). Muscle development in the four-winged Drosophila and the role of the Ultrabithorax gene. Curr. Biol. *4*, 957–964.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. *9*, 397–405.

Fischer, J.A., Giniger, E., Maniatis, T., and Ptashne, M. (1988). GAL4 activates transcription in Drosophila. Nature *332*, 853–856.

Fisher, R.A. (1919). The Correlation between Relatives on the Supposition of Mendelian Inheritance. Trans. R. Soc. Edinburgh *52*, 399–433.

Flores, G. V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial Signaling in the Specification of Unique Cell Fates. Cell *103*, 75–85.

Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2019). JASPAR 2020: update of

the open-access database of transcription factor binding profiles. Nucleic Acids Res. *48*, D87–D92.

Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807.

Fowlkes, C.C., Hendriks, C.L.L., Keränen, S.V.E., Weber, G.H., Rübel, O., Huang, M.-Y., Chatoor, S., DePace, A.H., Simirenko, L., Henriquez, C., et al. (2008). A Quantitative Spatiotemporal Atlas of Gene Expression in the Drosophila Blastoderm. Cell *133*, 364–374.

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., van Breugel, M.E., Halavatyi, A., Tischer, C., Polidoro, P., Abe, N., Tsai, A., Mann, R.S., Stern, D.L., et al. (2020). Dense and pleiotropic regulatory information in a developmental enhancer. Nature *587*, 235–239.

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. Rep. *11*, 10314.

Garner, M.M., and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res. *9*, 3047–3060.

Gaudet, J. (2002). Regulation of Organogenesis by the Caenorhabditis elegans FoxA Protein PHA-4. Science (80-.). *295*, 821–825.

Gehring, W.J. (1996). The master control gene for morphogenesis and evolution of the eye. Genes to Cells *1*, 11–15.

Gerber, L.C., Calasanz-Kaiser, A., Hyman, L., Voitiuk, K., Patil, U., and Riedel-Kruse, I.H. (2017). Liquid-handling Lego robots and experiments for STEM education and research. PLOS Biol. *15*, e2001413.

Gilbert, S.F. (2006). Developmental biology (Sunderland, Mass.: Sinauer Associates, Inc. Publishers).

Giniger, E., and Ptashne, M. (1988). Cooperative DNA binding of the yeast transcriptional activator GAL4. Proc. Natl. Acad. Sci. *85*, 382–386.

Glassford, W.J., and Rebeiz, M. (2013). Assessing constraints on the path of regulatory sequence evolution. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20130026.

Goldwater, M.B., Markman, A.B., and Stilwell, C.H. (2010). The empirical case for role-governed categories. Dev. Cell *18*, 359–376.

Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., and Carroll, S.B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature *433*, 481–487.

Goto, T., Macdonald, P., and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell *57*, 413–422.

Gould, S.J., and Vrba, E.S. (1982). Exaptation—a Missing Term in the Science of Form. Paleobiology *8*, 4–15.

Halder, G., Polaczyk, P., Kraus, M.E., Hudson, A., Kim, J., Laughon, A., and Carroll, S. (1998). The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in Drosophila. Genes Dev. *12*, 3900–3909.

Harding, K., Hoey, T., Warrior, R., and Levine, M. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. EMBO J. *8*, 1205–1212.

Al Hayek, S., Alsawadi, A., Kambris, Z., Boquete, J., Bohère, J., Immarigeon, C., Ronsin, B., Plaza, S., Lemaitre, B., Payre, F., et al. (2021). Steroid-dependent switch of OvoL/Shavenbaby controls self-renewal versus differentiation of intestinal stem cells. EMBO J. *40*, 1–23.

HELL, S., REINER, G., CREMER, C., and STELZER, E.H.K. (1993). Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index. J. Microsc. *169*, 391–405.

Hill, R.E., and Lettice, L.A. (2013). Alterations to the remote control of Shh gene expression cause congenital abnormalities. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20120357.

Hong, J.-W., Hendrix, D.A., and Levine, M.S. (2008). Shadow Enhancers as a Source of Evolutionary Novelty. Science (80-.). *321*, 1314–1314.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ramia, M., Tarone, A.M., Turlapati, L., Zichner, T., Zhu, D., Lyman, R.F., et al. (2014). Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. *24*, 1193–1208.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng.

Inestrosa, N.C., Sunkel, C.E., Arriagada, J., Garrido, J., and Godoy-Herrera, R. (1996). Abnormal development of the locomotor activity in yellow larvae of Drosophila: a cuticular defect? Genetica *97*, 205–210.

Jacob, F. (1977). Evolution and tinkering. Science (80-.). *196*, 1161–1166.

Jaw, T.J., You, L.-R., Knoepfler, P.S., Yao, L.-C., Pai, C.-Y., Tang, C.-Y., Chang, L.-P., Berthelsen, J., Blasi, F., Kamps, M.P., et al. (2000). Direct interaction of two homeoproteins, Homothorax and Extradenticle, is essential for EXD nuclear localization and function. Mech. Dev. *91*, 279–291.

Jindal, G.A., and Farley, E.K. (2021). Enhancer grammar in development, evolution, and disease: dependencies and interplay. Dev. Cell *56*, 575–587.

Jonas (2020). Violin Plots for plotting multiple distributions (distributionPlot.m).

Jones, K.A., and Tjian, R. (1985). Sp1 binds to promoter sequences and activates herpes simplex virus 'immediate-early' gene transcription in vitro. Nature *317*, 179–182.

Karin, M. (1990). Too many transcription factors: positive and negative interactions. New Biol.

2, 126–131.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, a. D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T.S., and Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. Genome Res. *23*, 800–811.

Kim, A.-R., Martinez, C., Ionides, J., Ramos, A.F., Ludwig, M.Z., Ogawa, N., Sharp, D.H., and Reinitz, J. (2013). Rearrangements of 2.5 Kilobases of Noncoding DNA from the Drosophila even-skipped Locus Define Predictive Rules of Genomic cis-Regulatory Logic. PLoS Genet. *9*, e1003243.

Kim, J., Zwieb, C., Wu, C., and Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene *85*, 15–23.

Kircher, M., Xiong, C., Martin, B., Schubach, M., Inoue, F., Bell, R.J.A., Costello, J.F., Shendure, J., and Ahituv, N. (2019). Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. Nat. Commun. *10*, 3583.

Kittelmann, S., Buffry, A.D., Franke, F.A., Almudi, I., Yoth, M., Sabaris, G., Couso, J.P., Nunes, M.D.S., Frankel, N., Gómez-Skarmeta, J.L., et al. (2018). Gene regulatory network architecture in different developmental contexts influences the genetic basis of morphological evolution. PLOS Genet. *14*, e1007375.

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.

KLUG, A. (1968). Rosalind Franklin and the Discovery of the Structure of DNA. Nature *219*, 808–810.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191–201.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature *512*, 91–95.

Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissières, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive Loss of Function in a Limb Enhancer during Snake Evolution. Cell *167*, 633-642.e11.

Kvon, E.Z., Zhu, Y., Kelman, G., Novak, C.S., Plajzer-Frick, I., Kato, M., Garvin, T.H., Pham, Q., Harrington, A.N., Hunter, R.D., et al. (2020). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. Cell *180*, 1262-1271.e15.

Kwasnieski, J.C., Mogno, I., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2012). Complex effects of nucleotide variants in a mammalian cis-regulatory element. Proc. Natl. Acad. Sci. U. S. A. *109*, 19498–19503.

Lack, J.B., Lange, J.D., Tang, A.D., Corbett-Detig, R.B., and Pool, J.E. (2016). A Thousand Fly

Genomes: An Expanded Drosophila Genome Nexus. Mol. Biol. Evol. *33*, 3308–3313.

Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J., et al. (2013). XPaused Pol II coordinates tissue morphogenesis in the drosophila embryo. Cell *153*, 976.

Lamb, A.M., Walker, E.A., and Wittkopp, P.J. (2017). Tools and strategies for scarless allele replacement in Drosophila using CRISPR/Cas9. Fly (Austin). *11*, 53–64.

Leal, F., and Cohn, M.J. (2016). Loss and Re-emergence of Legs in Snakes by Modular Evolution of Sonic hedgehog and HOXD Enhancers. Curr. Biol. *26*, 2966–2973.

Lee, T.I., and Young, R.A. (2000). Transcription of Eukaryotic Protein-Coding Genes. Annu. Rev. Genet. *34*, 77–137.

Levario, T.J., Zhan, M., Lim, B., Shvartsman, S.Y., and Lu, H. (2013). Microfluidic trap array for massively parallel imaging of Drosophila embryos. Nat. Protoc. *8*, 721–736.

Li, Z., Schulz, M.H., Look, T., Begemann, M., Zenke, M., and Costa, I.G. (2019). Identification of transcription factor binding sites using ATAC-seq. Genome Biol. *20*, 45.

Lifanov, A.P. (2003). Homotypic Regulatory Clusters in Drosophila. Genome Res. *13*, 579–588.

Liu-Johnson, H.-N., Gartenberg, M.R., and Crothers, D.M. (1986). The DNA binding domain and bending angle of E. coli CAP protein. Cell *47*, 995–1005.

Liu, F., and Posakony, J.W. (2012). Role of Architecture in the Function and Specificity of Two Notch-Regulated Transcriptional Enhancer Modules. PLoS Genet. *8*, e1002796.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187.

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

Ludwig, M.Z., Bergman, C., Patel, N.H., and KreLtman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. Nature *403*, 564–567.

Ludwig, M.Z., Palsson, A., Alekseeva, E., Bergman, C.M., Nathan, J., and Kreitman, M. (2005). Functional Evolution of a cis-Regulatory Module. PLoS Biol. *3*, e93.

Ludwig, M.Z., Manu, Kittler, R., White, K.P., and Kreitman, M. (2011). Consequences of Eukaryotic Enhancer Architecture for Gene Expression Dynamics, Development, and Fitness. PLoS Genet. *7*, e1002364.

M. Woltering, J. (2012). From Lizard to Snake; Behind the Evolution of an Extreme Body Plan. Curr. Genomics *13*, 289–299.

Macdonald, P.M., Ingham, P., and Struhl, G. (1986). Isolation, structure, and expression of even-skipped: A second pair-rule gene of Drosophila containing a homeo box. Cell *47*, 721–734.

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., et al. (2012). The Drosophila melanogaster Genetic

Reference Panel. Nature *482*, 173–178.

Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. Curr. Top. Dev. Biol. *88*, 63–101.

Maricque, B.B., Chaudhari, H.G., and Cohen, B.A. (2019). A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. Nat. Biotechnol. *37*, 90–95.

Martinez, C., Rest, J.S., Kim, A.-R., Ludwig, M., Kreitman, M., White, K., and Reinitz, J. (2014). Ancestral Resurrection of the Drosophila S2E Enhancer Reveals Accessible Evolutionary Paths through Compensatory Change. Mol. Biol. Evol. *31*, 903–916.

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. Nature *308*, 428–433.

McGregor, A.P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D.G., Payre, F., and Stern, D.L. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature *448*, 587–590.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proc. 9th Python Sci. Conf.

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Kinney, J.B., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. *30*, 271–277.

Menoret, D., Santolini, M., Fernandes, I., Spokony, R., Zanet, J., Gonzalez, I., Latapie, Y., Ferrer, P., Rouault, H., White, K.P., et al. (2013). Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. Genome Biol. *14*, R86.

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.

Mighell, T.L., Evans-Dutson, S., and O'Roak, B.J. (2018). A Saturation Mutagenesis Approach to Understanding PTEN Lipid Phosphatase Activity and Genotype-Phenotype Relationships. Am. J. Hum. Genet. *102*, 943–955.

Mogno, I., Kwasnieski, J.C., and Cohen, B.A. (2013). Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. Genome Res. *23*, 1908–1915.

Müller, B., and Basler, K. (2000). The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. Development *127*, 2999–3007.

Nadeau, N.J., Pardo-Diaz, C., Whibley, A., Supple, M.A., Saenko, S. V., Wallbank, R.W.R.,

Wu, G.C., Maroja, L., Ferguson, L., Hanly, J.J., et al. (2016). The gene cortex controls mimicry and crypsis in butterflies and moths. Nature *534*, 106–110.

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.

Nüsslein-volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in drosophila. Nature *287*, 795–801.

O'Neill, E.M., Rebay, I., Tjian, R., and Rubin, G.M. (1994). The activities of two Ets-related transcription factors required for drosophila eye development are modulated by the Ras/MAPK pathway. Cell *78*, 137–147.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243.

Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Sato, H., Hori, M., Nakamura, Y., et al. (2002). Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. Nat. Genet. *32*, 650–654.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An Atomic Model of the Interferon-β Enhanceosome. Cell *129*, 1111–1123.

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

Parker, D.S., White, M.A., Ramos, A.I., Cohen, B.A., and Barolo, S. (2011). The cis-Regulatory Logic of Hedgehog Gradient Responses: Key Roles for Gli Binding Affinity, Competition, and Cooperativity. Sci. Signal. *4*, ra38–ra38.

Passamaneck, Y.J., Katikala, L., Perrone, L., Dunn, M.P., Oda-Ishii, I., and Di Gregorio, A. (2009). Direct activation of a notochord cis-regulatory module by Brachyury and FoxA in the ascidian Ciona intestinalis. Development *136*, 3679–3689.

Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'Er, D., and Shendure, J. (2009). High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat. Biotechnol. *27*, 1173–1175.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265–270.

Payne, J.L., and Wagner, A. (2014). The Robustness and Evolvability of Transcription Factor Binding Sites. Science (80-.). *343*, 875–877.

Payne, J.L., and Wagner, A. (2015). Mechanisms of mutational robustness in transcriptional regulation. Front. Genet. *6*, 1–10.

Payne, J.L., and Wagner, A. (2019). The causes of evolvability and their evolution. Nat. Rev. Genet. *20*, 24–38.

Payre, F. (2004). Genetic control of epidermis differentiation in Drosophila. Int. J. Dev. Biol. *48*, 207–215.

Payre, F., Vincent, A., and Carreno, S. (1999). ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. Nature *400*, 271–275.

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.

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.

Plessy, C., Dickmeis, T., Chalmel, F., and Strähle, U. (2005). Enhancer sequence conservation between vertebrates is favoured in developmental regulator genes. Trends Genet. *21*, 207–210.

Politi, A.Z., Cai, Y., Walther, N., Hossain, M.J., Koch, B., Wachsmuth, M., and Ellenberg, J. (2018). Quantitative mapping of fluorescently tagged cellular proteins using FCS-calibrated four-dimensional imaging. Nat. Protoc. *13*, 1445–1464.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.

Le Poul, Y., Xin, Y., Ling, L., Mühling, B., Jaenichen, R., Hörl, D., Bunk, D., Harz, H., Leonhardt, H., Wang, Y., et al. (2020). Regulatory encoding of quantitative variation in spatial activity of a Drosophila enhancer. Sci. Adv. *6*, eabe2955.

Preger-Ben Noon, E., Davis, F.P., and Stern, D.L. (2016). Evolved Repression Overcomes Enhancer Robustness. Dev. Cell *39*, 572–584.

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

Prince, V.E. (2002). The Hox Paradox: More Complex(es) Than Imagined. Dev. Biol. *249*, 1–15.

Prud'homme, B., Gompel, N., and Carroll, S.B. (2007). In the Light of Evolution (Washington, D.C.: National Academies Press).

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. Proc. Natl. Acad. Sci. *115*, E3692–E3701.

Reetz, M.T., and Carballeira, J.D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat. Protoc. *2*, 891–903.

Riley, T.R., Slattery, M., Abe, N., Rastogi, C., Liu, D., Mann, R.S., and Bussemaker, H.J. (2014). SELEX-seq: A Method for Characterizing the Complete Repertoire of Binding Site Preferences for Transcription Factor Complexes. In Methods in Molecular Biology, pp. 255–278.

Rizzo, N.P., and Bejsovec, A. (2017). SoxNeuro and shavenbaby act cooperatively to shape denticles in the embryonic epidermis of Drosophila. Development *144 (12):*, 2248–2258.

de Rosa, R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte, A., Akam, M., Carroll, S.B., and Balavoine, G. (1999). Hox genes in brachiopods and priapulids and protostome evolution. Nature *399*, 772–776.

Rothwell, W.F., and Sullivan, W. (2007). Fixation of Drosophila embryos. CSH Protoc. *2007*, pdb.prot4827.

Rudnicki, S., and Johnston, S. (2009). Overview of Liquid Handling Instrumentation for High-Throughput Screening Applications. Curr. Protoc. Chem. Biol. *1*, 43–54.

Sabarís, G., Laiker, I., Preger-Ben Noon, E., and Frankel, N. (2019). Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. Trends Genet. *35*, 423–433.

Santangelo, A.M., de Souza, F.S.J., Franchini, L.F., Bumaschny, V.F., Low, M.J., and Rubinstein, M. (2007). Ancient Exaptation of a CORE-SINE Retroposon into a Highly Conserved Mammalian Neuronal Enhancer of the Proopiomelanocortin Gene. PLoS Genet. *3*, e166.

Schaerli, Y., Jiménez, A., Duarte, J.M., Mihajlovic, L., Renggli, J., Isalan, M., Sharpe, J., and Wagner, A. (2018). Synthetic circuits reveal how mechanisms of gene regulatory networks constrain evolution. Mol. Syst. Biol. *14*, 1–18.

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.

Scholes, C., Biette, K.M., Harden, T.T., and DePace, A.H. (2019). Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the Drosophila Embryo. Cell Rep. *26*, 2407-2418.e5.

Sheng, G., Thouvenot, E., Schmucker, D., Wilson, D.S., and Desplan, C. (1997). Direct regulation of rhodopsin 1 by Pax-6/eyeless in Drosophila: evidence for a conserved function in photoreceptors. Genes Dev. *11*, 1122–1131.

Shively, C.A., Liu, J., Chen, X., Loell, K., and Mitra, R.D. (2019). Homotypic cooperativity and collective binding are determinants of bHLH specificity and function. Proc. Natl. Acad. Sci. *116*, 16143–16152.

Shorr, A.Z., Sönmez, U.M., Minden, J.S., and Leduc, P.R. (2019). High-throughput mechanotransduction in: Drosophila embryos with mesofluidics. Lab Chip *19*, 1141–1152.

Small, S., and Arnosti, D.N. (2020). Transcriptional enhancers in Drosophila. Genetics *216*, 1–26.

Smith, J.M., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., and Wolpert, L. (1985). Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution. Q. Rev. Biol. *60*, 265–287.

Snetkova, V., Ypsilanti, A.R., Akiyama, J.A., Mannion, B.J., Plajzer-Frick, I., Novak, C.S., Harrington, A.N., Pham, Q.T., Kato, M., Zhu, Y., et al. (2021). Ultraconserved enhancer function does not require perfect sequence conservation. Nat. Genet. *53*, 521–528.

Sönmezer, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., and Krebs, A.R. (2021). Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. Mol. Cell *81*, 255-267.e6.

Soutourina, J. (2018). Transcription regulation by the Mediator complex. Nat. Rev. Mol. Cell Biol. *19*, 262–274.

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

Stanojevic, D., Small, S., and Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science (80-.). *254*, 1385–1387.

Starr, B.D., Hoopes, B.C., and Hawley, D.K. (1995). DNA Bending is an Important Component of Site-specific Recognition by the TATA Binding Protein. J. Mol. Biol. *250*, 434–446.

Stathopoulos, A., and Levine, M. (2002). Dorsal Gradient Networks in the Drosophila Embryo. Dev. Biol. *246*, 57–67.

Stern, D.L., and Frankel, N. (2013). The structure and evolution of cis -regulatory regions: the shavenbaby story. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20130028.

Stern, D.L., and Orgogozo, V. (2008). THE LOCI OF EVOLUTION: HOW PREDICTABLE IS GENETIC EVOLUTION? Evolution (N. Y). *62*, 2155–2177.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. *100*, 9440–9445.

Stormo, G.D., Schneider, T.D., Gold, L., and Ehrenfeucht, A. (1982). Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli. Nucleic Acids Res. *10*, 2997–3011.

Su, M., Han, D., Boyd-Kirkup, J., Yu, X., and Han, J.-D.J. (2014). Evolution of Alu Elements toward Enhancers. Cell Rep. *7*, 376–385.

Sucena, E., Delon, I., Jones, I., Payre, F., and Stern, D.L. (2003). Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. Nature *424*, 935–938.

Sundararajan, V., Pang, Q.Y., Choolani, M., and Huang, R.Y.-J. (2020). Spotlight on the Granules (Grainyhead-Like Proteins) – From an Evolutionary Conserved Controller of Epithelial Trait to Pioneering the Chromatin Landscape. Front. Mol. Biosci. *7*, 7–213.

Swanson, C.I., Schwimmer, D.B., and Barolo, S. (2011). Rapid Evolutionary Rewiring of a Structurally Constrained Eye Enhancer. Curr. Biol. *21*, 1186–1196.

Tepass, U., and Knust, E. (1993). crumbs and stardust Act in a Genetic Pathway That Controls the Organization of Epithelia in Drosophila melanogaster. Dev. Biol. *159*, 311–326.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. Cell *83*, 1091–1100.

Thompson, A.C., Capellini, T.D., Guenther, C.A., Chan, Y.F., Infante, C.R., Menke, D.B., and

Kingsley, D.M. (2018). A novel enhancer near the Pitx1 gene influences development and evolution of pelvic appendages in vertebrates. Elife *7*, e38555.

Tischer, C., Hilsenstein, V., Hanson, K., and Pepperkok, R. (2014). Adaptive fluorescence microscopy by online feedback image analysis. In Methods in Cell Biology, pp. 489–503.

Tito, A.J., Cheema, S., Jiang, M., and Zhang, S. (2016). A Simple One-step Dissection Protocol for Whole-mount Preparation of Adult Drosophila Brains. J. Vis. Exp. e55128.

Todd, C.D., Deniz, Ö., Taylor, D., and Branco, M.R. (2019). Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. Elife *8*, e44344.

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*, 1–18.

Tsai, A., Alves, M.R.P., and Crocker, J. (2019). Multi-enhancer transcriptional hubs confer phenotypic robustness. Elife *8*, e45325.

Turing, A.M. (1952). The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *237*, 37–72.

Uller, T., Moczek, A.P., Watson, R.A., Brakefield, P.M., and Laland, K.N. (2018). Developmental Bias and Evolution: A Regulatory Network Perspective. Genetics *209*, 949–966.

Vincent, B.J., Estrada, J., and DePace, A.H. (2016). The appeasement of Doug: a synthetic approach to enhancer biology. Integr. Biol. *8*, 475–484.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2019). SciPy 1.0--Fundamental Algorithms for Scientific Computing in Python. 1–22.

VISSER, T., OUD, J., and BRAKENHOFF, G. (1992). Refractive index and axial distance measurements in 3-D microscopy. Opt. *90*, 17–19.

Wagner, A. (2013). Robustness and evolvability in living systems.

Wagner, G.P. (2014). Homology, genes, and evolutionary innovation.

Wagner, P., and Lynch, V.J. (2008). The gene regulatory logic of transcription factor evolution. 377–385.

Walther, C., and Gruss, P. (1991). Pax-6, a murine paired box gene, is expressed in the developing CNS. Development *113*, 1435–1449.

Wang, B., Fallon, J.F., and Beachy, P.A. (2000). Hedgehog-Regulated Processing of Gli3 Produces an Anterior/Posterior Repressor Gradient in the Developing Vertebrate Limb. Cell *100*, 423–434.

Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021.

WATSON, J.D., and CRICK, F.H.C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature *171*, 737–738.

Waymack, R., Fletcher, A., Enciso, G., and Wunderlich, Z. (2020). Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. Elife *9*, e59351.

Weingarten-Gabbay, S., Nir, R., Lubliner, S., Sharon, E., Kalma, Y., Weinberger, A., and Segal, E. (2019). Systematic interrogation of human promoters. Genome Res. *29*, 171–183.

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

Wieschaus, E., and Nüsslein-Volhard, C. (2016). The Heidelberg Screen for Pattern Mutants of Drosophila : A Personal Account. Annu. Rev. Cell Dev. Biol. *32*, 1–46.

Wilk, R., Hu, J., Blotsky, D., and Krause, H.M. (2016). Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. Genes Dev. *30*, 594–609.

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. *13*, 59–69.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. *25*, 1–47.

Wong, B.G., Mancuso, C.P., Kiriakov, S., Bashor, C.J., and Khalil, A.S. (2018). Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nat. Biotechnol. *36*, 614–623.

Wunderlich, Z., Bragdon, M.D.J., Vincent, B.J., White, J.A., Estrada, J., and DePace, A.H. (2015). Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. Cell Rep. *12*, 1740–1747.

Xu, C., Kauffmann, R.C., Zhang, J., Kladny, S., and Carthew, R.W. (2000). Overlapping Activators and Repressors Delimit Transcriptional Response to Receptor Tyrosine Kinase Signals in the Drosophila Eye. Cell *103*, 87–97.

Yin, Z., and Frasch, M. (1998). Regulation and function oftinman during dorsal mesoderm induction and heart specification inDrosophila. Dev. Genet. *22*, 187–200.

Yuh, C.H., and Davidson, E.H. (1996). Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development *122*, 1069–1082.

Yuh, C.-H., Ransick, A., Martinez, P., Britten, R.J., and Davidson, E.H. (1994). Complexity and organization of DNA-protein interactions in the 5′-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. Mech. Dev. *47*, 165–186.

Yuh, C.H., Bolouri, H., and Davidson, E.H. (2001). Cis-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control. Development *128*, 617–629.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556–559.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A.C., Ghane, T., Di Felice, R., and Rohs, R. (2013). DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. Nucleic Acids Res. *41*, W56–W62.

Zhou, T., Shen, N., Yang, L., Abe, N., Horton, J., Mann, R.S., Bussemaker, H.J., Gordân, R., and Rohs, R. (2015). Quantitative modeling of transcription factor binding specificities using DNA shape. Proc. Natl. Acad. Sci. *112*, 4654–4659.

Zinkel, S.S., and Crothers, D.M. (1987). DNA bend direction by phase sensitive detection. Nature *328*, 178–181.

Herewith I declare that I, Timothy Fuqua, prepared this Ph.D. thesis:

The Evolutionary Landscape of a Developmental Enhancer

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