University of California, Irvine

Mechanisms of robustness and competition in transcriptional control

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Developmental and Cell Biology

By

Rachel Victoria Waymack

Dissertation Committee:
Assistant Professor Zeba Wunderlich, Chair
Professor Thomas Schilling
Associate Professor Rahul Warrior
Professor German Enciso
Professor Anthony Long

DEDICATION

To Mom and Dad, and the rest of the family.

For constant love and support.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my PhD advisor and chair, Dr. Zeba Wunderlich for being an incredible mentor all throughout my PhD. Thank you for always being willing to talk through any and all of my questions with me, modeling for me what a great scientist and mentor is, and believing in my abilities more than I did sometimes. You also fostered and encouraged an incredibly supportive and fun lab environment that helped make a completely new place and experience that could have been quite lonely feel a lot more like home, so thank you.

I also want to thank my PhD committee members Dr. Rahul Warrior, Dr. Thomas Schilling, Dr. German Enciso, and Dr. Anthony Long. The questions and suggestions they have offered have greatly shaped and improved this work and I am incredibly thankful to them for their time and support.

I must thank every member of the Wunderlich lab for making it a great place to work and learn (and sometimes live) for the past five years. Thank you to Lily Li, Bryan Ramirez, and Kevin Cabrera for your help, support, humor, and the comradery that only comes from people struggling through the same thing as you.

I want to also thank all of the amazing friends I made outside of the lab. From hiking buddies, to fantasy football, to the best softball team someone could hope to captain, thank you for making my time at UCI full of laughter and support. A special thank you to the Party People – Jasper Liedtke, Zane Norman, Andrew Oliver, and (only with a slight delay) Megan Newcomb. Thank you for being the people I know I can turn to in good times or bad. In addition to the great people I've met at UCI, a huge thank you to the friends who have been with me before grad school and have stuck it out with me through grad school, your perspectives on life being much bigger than a PhD have helped pull me through the hard times. In particular, thank you to Rachel Andricosky for being a great friend, role model, and roomie for life.

Lastly, I have to thank the people who have been with me from the beginning - my parents and family back home. Thank you for believing in me all those years ago when I packed up my car and drove to California. I would not be here if it were not for everything you all have done for me and while I can never repay you or thank you enough, know that I am grateful for everything.

CURRICULUM VITAE

Rachel Victoria Waymack

2021 PhD in Developmental and Cell Biology, University of California Irvine

2015 B.S. in Neruoscience, minor in Biochemistry, College of William & Mary

FIELD OF STUDY

Developmental and Cell Biology

PUBLICATIONS

Waymack, R., Elabd, M., & Wunderlich, Z. (2021) Molecular competition shapes enhancer-driven expression in the *Drosophila* embryo.
bioRxiv 2021.05.07.443186; doi: https://doi.org/10.1101/2021.05.07.443186

Li, L., **Waymack, R**., Elabd, M., & Wunderlich, Z. (2021) Two promoters integrate multiple enhancer inputs to drive wild-type knirps expression in the D. melanogaster embryo. *Biorxiv* 2021.03.23.436657; https://doi.org/10.1101/2021.03.23.436657.

Chang, C., Vila, J., Bender, M., Li, R., Mankowski, M., Bassette, M., Borden, J., Golfier, S., Sanchez, P. G., **Waymack, R**., Zhu, X., Diaz-Colunga, J., Estrela, S., Rebolleda-Gomez, M., & Sanchez, A. (2021). Engineering complex communities by directed evolution. *Nature Ecology and Evolution*. https://doi.org/10.1038/s41559-021-01457-5.

Kvon, E.Z., **Waymack, R**., Elabd, M., & Wunderlich, Z. (2021). Enhancer redundancy in development and disease. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-020-00311-x

Waymack, R., Fletcher, A,. Enciso, G., & Wunderlich. Z (2020). Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. *eLife* 2020;9:e59351. DOI: 10.7554/eLife.59351.

Texter, K.B., **Waymack, R**., Kavanagh, P.V., O'Brien, J.E., Talbot, B., Brandt, S.D., and Gardner, E.A. (2018). Identification of Pyrolysis Products of the New Psychoactive Substance 2-Amino-1-(4-Bromo-2,5-Dimethoxyphenyl)Ethanone Hydrochloride (Bk -2C-B) and Its Iodo Analogue Bk -2C-I. *Drug Testing and Analysis.* https://doi.org/10.1002/dta.2200.

ABSTRACT OF THE DISSERTATION

Mechanisms of robustness and competition in transcriptional control

By

Rachel Victoria Waymack

Doctor of Philosophy in Developmental and Cell Biology

University of California, Irvine, 2021

Assistant Professor Zeba Wunderlich, Chair

During normal development, organisms must establish and maintain precise patterns of gene expression that are also robust to internal and external perturbations. The key cell fate decisions made during early development that establish the organism's body plan rely on these precise gene expression patterns and even relatively minor deviations can cause significant defects. This precision and robustness must be achieved in the context of the inherently noisy process of transcription, which is controlled by the stochastic molecular interactions between DNA, transcription factors (TFs), and the other pieces of the transcriptional machinery. While the question of how regulatory regions of DNA, such as enhancers, accurately interact with TFs to control gene expression has been the topic of many studies, there is still much we do not know about how enhancers and TFs combine to regulate gene expression.

Here, I investigate novel mechanisms of gene expression regulation by enhancers and TFs. In Chapter 2, I show that shadow enhancers, groups of seemingly redundant enhancers, are able to buffer fluctuations in upstream regulators by separating TF inputs between individual enhancers in a case study of the shadow enhancers regulating *Kruppel* expression during early embryonic *Drosophila* development. In Chapter 3, I investigate competition observed between transcriptional reporters and find that transgenic reporters compete with one another and an endogenous gene for TF activation, likely due to the non-homogenous nature of the nucleus.

Then, complementing my findings in Chapter 2, I analyze the TF regulation of a large set of *Drosophila* shadow enhancers and show that these groups of enhancers display a wide range of regulatory logic configurations in Chapter 4. In Chapter 5, I discuss a proposed project to investigate the relationship between TF dynamics and enhancer-controlled transcription during NF-kB signaling in the immune response. Taken together, these projects inform our understanding of how gene expression is regulated by enhancers and TFs in early development and novel regulatory mechanisms that likely apply more broadly throughout the genome and life stages.

Chapter 1

Introduction: Transcription factors and enhancers interact to regulate gene expression

Note: Portions of this chapter were published in *Nature Reviews Genetics* in January 2021 as Kvon, E., Waymack, R., Gad, M., & Wunderlich, Z. (2021)

1. Investigation of enhancer-driven transcription

Proper regulation of gene expression, in both time and space, is critical for organismal development and survival. Such spatiotemporal control of gene expression is achieved by *cis*-regulatory regions of DNA called enhancers, which regulate the activity of one or more target genes. Enhancers consist of binding sites for transcription factors (TFs), whose binding determines the activity of the enhancer and consequent expression of the target gene. The interactions between enhancers and TFs to precisely control gene expression is critical for all aspects of life, from early development to proper functioning of adult tissues to immune response.

In this thesis, I investigate the mechanisms by which enhancers and TFs work together to regulate gene expression during early development. In Chapter 2, I study how the configuration of TF regulation of a pair of seemingly redundant enhancers impacts expression noise driven by the enhancers. In Chapter 3, I investigate how competition for TF binding between enhancers regulates gene expression and suggest that this competition depends on the non-uniform environment of the nucleus. In Chapter 4, I perform computational analysis on a large set of *Drosophila* shadow enhancers to assess the different forms of TF regulatory logic used. Lastly, in Chapter 5, I propose a study of TF and enhancer dynamics in the immune system to understand how regulatory rules do or do not differ in a system facing very different demands on gene expression. Taken together, this work furthers our understanding of how enhancers and TFs combine to control gene expression within the reality of a developing organism subject to the inherent perturbations, constraints, and uncertainties of life. This study therefore stands to inform studies of how development is robust and the circumstances under which this robustness breaks down.

1.1 Shadow enhancers in development

To establish a properly formed organism, gene expression during early development must be both precise and robust. One mechanism organisms use to meet these demands are groups of seemingly redundant enhancers, called shadow enhancers. Here I discuss shadow enhancers and their role in regulating gene expression during early development. I connect this brief review of shadow enhancers to the findings presented in Chapter 2 and how these findings help fill holes in our understanding of shadow enhancer function.

1.1.1 Developmental systems establish precise patterns of gene expression

The precise gene expression patterns formed during early development, both in terms of time and space, are critical for establishing cell and tissue fates and ensuring proper development of the overall organism. Examples of this precision are perhaps best studied in the developing *Drosophila* embryo. The concentration of Bicoid (Bcd), a key TF expressed in a decreasing gradient from the anterior to the posterior of the embryo, is reproducible to within the width of one nucleus across a population of embryos [1]. This precision is maintained through development, with expression boundaries of the Bcd-regulated gap genes also differing between embryos by less than the width of a single nucleus [2].

1.1.2 Developmental systems must be robust to internal and external fluctuations

Contrary to what one may expect based on the reproducible gene expression levels and boundaries measured across developing embryos, transcription itself is an inherently noisy process. For the majority of investigated genes across a range of organisms, transcription occurs in series of random bursts of activity interspaced by periods of relative silence [3-7]. How transcriptional bursting is regulated is still an area of active investigation, and different

regulation strategies have different effects on resulting gene expression noise [3,8-12]. Regardless of the mode of modulation, the bursting process of transcription carries inherent noise that has meaningful implications for an organism, as much of the variance in protein levels can be explained by variance in transcript levels [13]. Genes encoding key regulatory proteins, such as TFs or RNA polymerase II, are also transcribed in bursts [6,14] and the resulting fluctuations of these factors can drive noise in the expression of downstream targets [15,16]. In addition to the noise inherent in the process of transcription, developmental systems also must contend with potential environmental and genetic perturbations that can impact gene expression [17-20].

Organisms have evolved multiple mechanisms to meet the developmental demands of precision and robustness in the face of fluctuating internal and external environments. Temporal and spatial averaging of transcripts have been studied as potential mechanisms of coping with variable inputs [21,22]. Redundancy in TF function [23] as well as in the identity of TF binding sites within individual enhancers [24] further act to ensure precise and robust gene expression. Shadow enhancers are increasingly appreciated as another mechanism of redundancy that safeguards against genetic and environmental perturbations.

1.1.3 Shadow enhancers are widespread and often associated with developmental genes

Shadow enhancers are groups of two or more enhancers that drive the same or similar pattern of expression of the same target gene [25]. While such seemingly redundant enhancers had been identified many times prior, these examples were often overlooked until 2008 when Mike Levine and colleagues introduced the term “shadow enhancer”. In that study, redundant enhancers were designated either as ‘primary’ (the enhancers closest to the core promoter) or ‘shadow’ (the enhancers located at a greater distance from the core promoter) [26]. This distinction was later revised owing to a lack of functional differences between primary and shadow enhancers [25,27].

The revised definition of shadow enhancers, as sets of enhancers that regulate a common target gene and drive expression patterns that partially or completely overlap in space and time, has become increasingly accepted in the gene regulation community [24-35].

Prior to the coining of the term "shadow enhancers", enhancers with redundant activity have been described for more than 30 years, with examples from plants [36,37], flies [38–45], zebrafish [46], mice [47-50] and humans [51–54]. These individual gene locus studies showed that shadow enhancers are found in a broad set of multicellular organisms, but within a single genome the prevalence of shadow enhancers was unknown. Since these studies were often focused on enhancers that control important developmental regulators, it was also not clear whether shadow enhancers are associated with other classes of genes. Substantial increases in the throughput of enhancer identification and characterization (reviewed in [55-57]) have allowed researchers to determine the prevalence of shadow enhancers genome-wide.

Genome-wide enhancer predictions based on chromatin features, such as chromatin accessibility, histone modifications, and TF binding, have suggested that shadow enhancers might be common in animal genomes. Using a combination of mesodermal TF chromatin immunoprecipitation (ChIP) data and computational models, Cannavo et al. generated an exhaustive catalogue of muscle development enhancers in *D. melanogaster* [28]. They found that nearly two-thirds of examined muscle developmental genes were controlled by shadow enhancers, and the majority of these genes had three or more predicted shadow enhancers [28]. A genome-wide analysis combining ENCODE transcriptomic and epigenomic data from multiple mouse tissues showed ample enhancer redundancy among developmentally regulated genes [58,59]. Whereas housekeeping genes are typically controlled by one enhancer, developmentally regulated genes can have 10 or more shadow enhancers [60,28,58].

In human cells, ChIP-based profiling of TFs, cofactors, chromatin regulators and enhancer-associated histone modifications revealed that hundreds of key cell identity genes are regulated by large clusters of putative transcriptional enhancers (super-enhancers and stretch enhancers), which could be clusters of shadow enhancers [61–66]. Many mammalian enhancers including human enhancers are actively transcribed, and the presence of enhancer-derived RNAs (eRNAs) was suggested to be predictive of enhancer activity [67-69]. Profiling of eRNAs using cap analysis of gene expression (CAGE) across hundreds of human cell lines and tissues revealed that ~80% of 2,206 examined genes were associated with two or more co-transcribed enhancers [67], suggesting that enhancer redundancy is common in the human genome. Computational approaches have also found widespread evidence for shadow enhancers in the human genome, particularly in association with developmental and disease-causing genes [70].

Most chromatin and TF profiling methods are based on indirect measures of enhancer activity, which is why they have to be followed by functional testing. Large-scale transgenic enhancer-reporter screens have verified that *bona fide* redundant enhancers are common in the *D. melanogaster* and *Caenorhabditis elegans* genomes. An analysis of nearly 8,000 enhancer fragments during *D. melanogaster* embryogenesis revealed that many developmentally regulated genes are controlled by two or more enhancers with overlapping activities [71]. Single-neuron imaging data from hundreds of enhancer-reporter constructs in *C. elegans* demonstrated that shadow enhancers control nearly all 23 studied pan-neuronal genes [72]. Even within a single cell type, massively parallel reporter assays (MPRAs) have shown that hundreds of genes in *D. melanogaster* cell lines are potentially controlled by two or more redundant enhancers [60]. Taken together, these studies indicate that enhancers driving overlapping expression patterns are

common in organisms from worms to insects to mammals and are preferentially, albeit not exclusively, associated with the control of developmental genes.

1.1.4 Shadow enhancers confer robustness

Several early studies in *Drosophila* demonstrated that shadow enhancers are required to drive normal development under conditions of stress, but they may be dispensable in 'ideal' conditions. For example, the TF Snail is required for normal gastrulation, and its expression in mesoderm is controlled by two shadow enhancers. Deletion of either of *snail*'s shadow enhancers caused no apparent gastrulation defect under ideal conditions[73]. However, individual shadow enhancer deletion led to abnormal gastrulation under elevated temperatures or in a sensitized genetic background where the dosage of an upstream regulator, Dorsal, was reduced [73]. Similarly, a deletion of three of six epidermal shadow enhancers of *shavenbaby* had no phenotype under normal conditions but caused a decrease of trichome numbers under temperature or genetic stress conditions [74].

Advances in genome editing have enabled the efficient introduction of multiple mutations in mice[75–77], enabling experiments to test whether shadow enhancers similarly provide developmental robustness in vertebrates. Whereas single shadow enhancer deletions in mice typically show either mild or no observable phenotypes, double enhancer deletions show severe phenotypes, often comparable to complete gene loss-of-function in relevant tissues [78,58,79,64,80]. Together, these observations indicate that both enhancers regulate the gene, and at least one shadow enhancer is required for normal development in ideal conditions (Figure 1.1). Despite driving similar expression patterns, the individual shadow enhancers are not strictly redundant. In a sensitized genetic background with a reduced dosage of the target gene, single enhancer deletions show abnormal phenotypes, indicating that shadow enhancers can confer robustness to

genetic perturbations (Figure 1.1). This pattern has been demonstrated for *Pax6*, a gene required for early eye lens morphogenesis [78], *Shh* in developing teeth [79] and several limb development loci [58].

1.1.5 Potential mechanisms of action of shadow enhancers

Taken together, both fruit fly and mouse studies emphasize that, while ostensibly redundant in the expression patterns they drive, the necessity of shadow enhancers is revealed when enhancer-deficient organisms are placed in stressful conditions. How shadow enhancers provide this robustness remains an area of open investigation and more than one mechanism may be at play. One potential scenario is that each enhancer alone can drive sufficient levels of gene expression for normal development, similar to the haplosufficiency of many developmental genes. By having multiple enhancers the probability that at least one is active increases, improving the chance for normal development [24,25]. Shadow enhancers have also been implicated in establishing or maintaining so-called "transcriptional hubs". These hubs are large nuclear sub-regions formed by a high concentration of TFs, components of the core transcriptional machinery [81,82] and RNA polymerase II [83,84] and may explain why some enhancers activate promoters even in the absence of close enhancer–promoter proximity [85,86]. The hub model suggests that shadow enhancers and their target promoter can simultaneously participate in the same microenvironment, forming a multi-enhancer hub. The observation of transcriptional coactivator condensates on super-enhancer-associated genes provides support for this model [61,81,82] Recent work on the *D. melanogaster shavenbaby* locus showed that deleting one of the shadow enhancers results in decreased local density of the key activating TFs, suggesting that shadow enhancers are critical for maintaining high concentrations of TFs within the transcriptional hub (Figure 1.2) [87]. Through the formation of multi-enhancer transcriptional hubs with high

concentrations of TFs, transcriptional coactivators and RNA polymerase II, shadow enhancers may improve phenotypic resilience to stress by buffering against environmental and genetic perturbations.

In Chapter 2, I investigate another potential mechanism by which shadow enhancers achieve robust gene expression. Due to their overlapping expression patterns, it has largely been assumed that groups of shadow enhancers are regulated by the same sets of TFs. In fact, scanning genomic sequences for clusters of similar TF binding sites has been used to identify novel potential shadow enhancers [26,28]. Breaking with this assumption, I ask whether differential TF regulation between a pair of shadow enhancers enables them to drive stable and robust gene expression. An investigation of over 800 mesodermal enhancers in the *Drosophila* embryo suggests that independent regulation of shadow enhancers may be a widespread, potentially functional, feature [28] (Figure 4.1). By responding to different sets of TFs, but converging on a single output, shadow enhancers could provide a mechanism to buffer against not only mutations in their sequences but, more importantly, perturbations in one of their upstream TFs (Figure 1.3).

1.2 Molecular competition

Much of our understanding of enhancers and their role in controlling gene expression has come from the use of transgenic reporters. While used frequently and extensively, little work has addressed potential effects of transgenic reporters on the systems in which they are placed. In Chapter 3, I investigate the molecular competition that is revealed by transgenic reporters. Here I provide a brief overview of the concept of transgenic reporters and the existing knowledge on their impact on endogenous systems.

1.2.1 Transgenic reporters enable investigation of enhancer activity

Transgenic reporters are DNA sequences in which the expression of measurable reporters, such as fluorescent proteins or LacZ, are placed under the control of *cis*-regulatory elements, such as enhancers. These sequences are incorporated into a living organism, through genomic integration or extrachromosomal arrays, to enable observation of the element's activity in different life stages, tissue types, and experimenter-chosen conditions. Like the endogenous *cis*-regulatory elements, these transgenic reporters bind a specific combination of TFs and related co-factors to drive expression of their target, in this case the reporter as opposed to the endogenous gene product, in a particular time and or space. Transgenic animals have been used extensively in the study of enhancers and have greatly improved our catalog of identified enhancers and our understanding of how these enhancers function [56,88-91]. We now have multiple metrics by which to predict enhancers genome-wide, although no technique is fully predictive and our understanding of the *cis*-regulatory code is far from complete [55,92,93]. Consequently, transgenic reporter measurements have long been the gold standard in the functional validation of potential enhancers [55,94]. While continuously improving genome-editing techniques have enabled investigations of enhancer function in the endogenous locus [95-97], transgenic reporters continue to be widely used to test the sufficiency of genomic regions to drive expression and for their relative simplicity.

1.2.2 Few studies address the impact of transgenes on endogenous systems

Despite the widespread use of transgenic reporters, very few studies have investigated the potential effects of these sequences on the endogenous systems in which they are placed. As discussed above, transgenic reporters work by interacting with the same pool of TFs, co-factors, and other pieces of the transcriptional machinery that bind and regulate the expression of endogenous sequences. This raises the possibility of competition for one or many of these

molecular factors between the transgenic sequence and the surrounding endogenous sequences. Although many TFs and other pieces of the transcriptional machinery are present at high copy number, for example 150,000 Grainyhead TF molecules per nucleus in the *Drosophila* embryo [98,99] (BNID 106848) or over 80,000 RNAP molecules in human cells [99,100] (BNID 112321), competition for these molecules can occur. In mice and flies, repetitive DNA sequences have been shown to sequester TF molecules and consequently reduce expression of genes regulated by those TFs [101,102]. Experimental and computational approaches have demonstrated that both the overall levels as well as noise in expression of a gene regulated by a particular TF are influenced by the number of other targets for that TF in the cell [103,104]. Repetitive sequences of "decoy" TF binding sites in yeast can not only decrease expression of the TF-targeted gene, but also alter the input-output response pattern of the gene to TF levels [106]. Whether or not frequently used transgenic reporters induce such molecular competition has yet to be systematically investigated.

A study by Laboulaye and colleagues investigated the effect of transgenic reporters on endogenous gene expression in mice [107]. Using three mouse lines with three different reporters inserted at three different locations in the genome, the authors measured the expression of the endogenous gene closest to each transgene. In all three cases, endogenous gene expression was decreased in the presence of the transgene, compared to mice lacking the transgene. Additionally, the authors found that this transgene effect depends on the distance between the reporter and the endogenous gene; the genes closer to the respective reporter showed a larger decrease in expression than did the genes further removed from the reporter. Expression of each reporter was also correlated to the expression of the closest endogenous gene, further suggesting interactions between the transgenic and endogenous sequences. As few other studies have mentioned effects of transgenic reporters on endogenous gene expression [107], the question

remains regarding the mechanisms behind the observed effect on endogenous gene expression and if such effects are limited to certain reporters or organisms. In Chapter 3, I investigate molecular competition that underlies transgenic reporter and endogenous gene interactions in the *Drosophila* embryo.

1.3 Conclusions

By integrating the input of multiple TFs, enhancers enable the regulation of gene expression across time and space that is necessary for organismal survival. Enhancers and TFs function together to control gene expression across all stages of life and in biological systems with distinct purposes from early development to immune responses. Understanding how these interactions are integrated across multiple enhancers, in the context of the 3D nucleus, and in various biological processes is needed for a full understanding of gene regulation. Therefore, the overall goal of my work has been to further our understanding of how enhancers interpret and integrate TF inputs in the context of a living organism.

In Chapter 2, I investigate the pair of shadow enhancers controlling *Kruppel* expression in the early *Drosophila* embryo. Using live imaging of transcription, I show that the separation of TF inputs between the two *Kruppel* enhancers enables the combined pair to drive lower expression noise than duplicated enhancers and that this noise buffering ability is robust to a wide range of temperatures. With collaborators, we develop a model of this system, which shows that the separation of TF inputs between the *Kruppel* enhancers is sufficient to explain the lower expression noise seen with the pair of enhancers relative to duplicated enhancers. This work was published in the journal *eLife* in August, 2020.

In Chapter 3, I study the effect of transgenic reporters on the expression of other reporters as well as endogenous genes in the *Drosophila* embryo. I combine live imaging, quantitative gene expression measurements, and thermodynamic modeling to show that our enhancer reporters compete with one another and endogenous genes for TFs. My modeling indicates a role for so-called TF hubs in regulating this competition and consequently gene expression. This work has been posted onto *BioRxiv*.

In Chapter 4, I analyze the TF regulatory patterns of a large dataset of *Drosophila* developmental shadow enhancers. Through computational analysis of TF binding sites, I show that a slight majority of shadow enhancers have less similar TF inputs than expected by chance. My analysis indicates the potential wide use of the separated TF inputs discussed in Chapter 2, as well as the possible existence of different "classes" of shadow enhancers that likely serve different functions.

In Chapter 5, I propose and discuss the beginnings of a study to investigate TF signal integration at enhancers in the immune system. The immune system serves a very different function and consequently faces very different demands than the developmental system. Therefore, comparison of enhancer activity in these two systems will help us determine global as well as context-specific rules regulating enhancer function.

Lastly, in Chapter 6, I discuss some of the overall conclusions of my work and remaining questions. While our understanding of specific enhancers and our ability to edit the genome has come quite far, a full understanding of the non-coding regions of genomes that enables prediction of function from sequence is still a long way off. Future studies that integrate both the contributions of DNA sequence and the surrounding nuclear environment will be needed for a full understanding of gene expression.

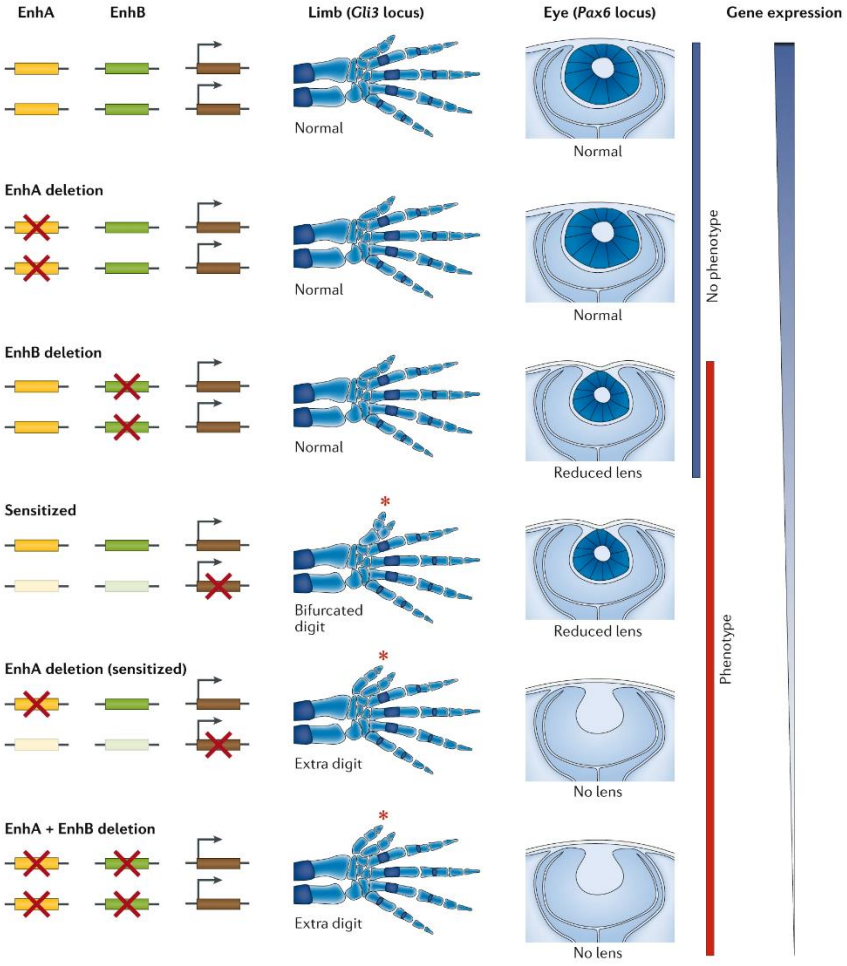


Figure 1.1 - Shadow enhancers confer phenotypic robustness in mammals. In mice, many individual shadow enhancer deletions yield no observable phenotypes. However, either the deletion of individual shadow enhancers in a sensitized background or the deletion of pairs of shadow enhancers leads to observable phenotypes. Schematics of perturbations (left) and resulting phenotypes in mice (right) are shown for two gene loci: *Gli3* [58] and *Pax6* [30,108]. GLI3 is critical for proper limb development, and its knockout causes the formation of extra digits (among other phenotypes) [109]. Skeletal phenotypes in the absence of individual *Gli3* shadow enhancers, pairs of shadow enhancers, or an individual shadow enhancer in a sensitized background are shown (center). Red asterisks indicate the presence of extra digits. *Pax6*-deficient mice have arrested eye development and no lens formation [110,111]. A schematic diagram of an eye section showing a developing lens in the absence of individual PAX6 shadow enhancers, pairs of shadow enhancers, or an individual shadow enhancer in a sensitized background is shown (center-right). A schematic of gene dosage in the mutants is shown on the right. Figure adapted with permission from[58].

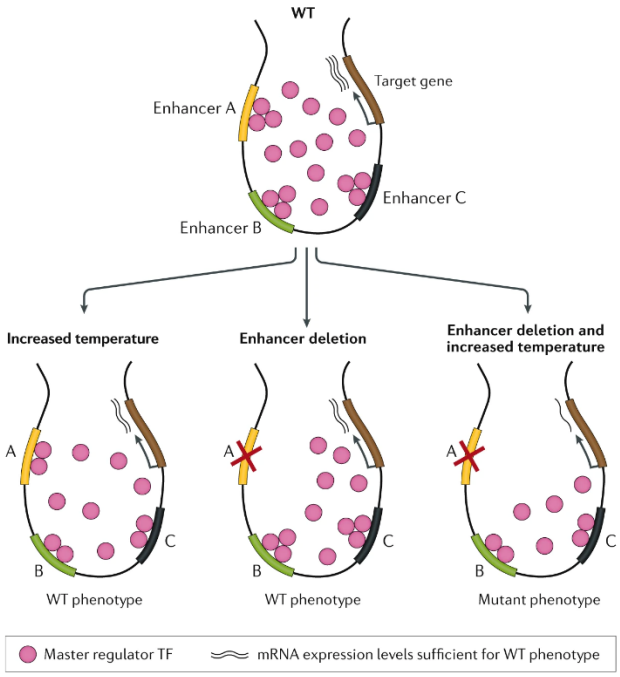


Figure 1.2 – Shadow enhancers function in the formation of transcriptional hubs. Multiple shadow enhancers may aid in the formation of transcriptional hubs by recruiting a high local amount of a master regulator TF (pink). Such transcription hubs can buffer against environmental stress and genetic perturbations [87].

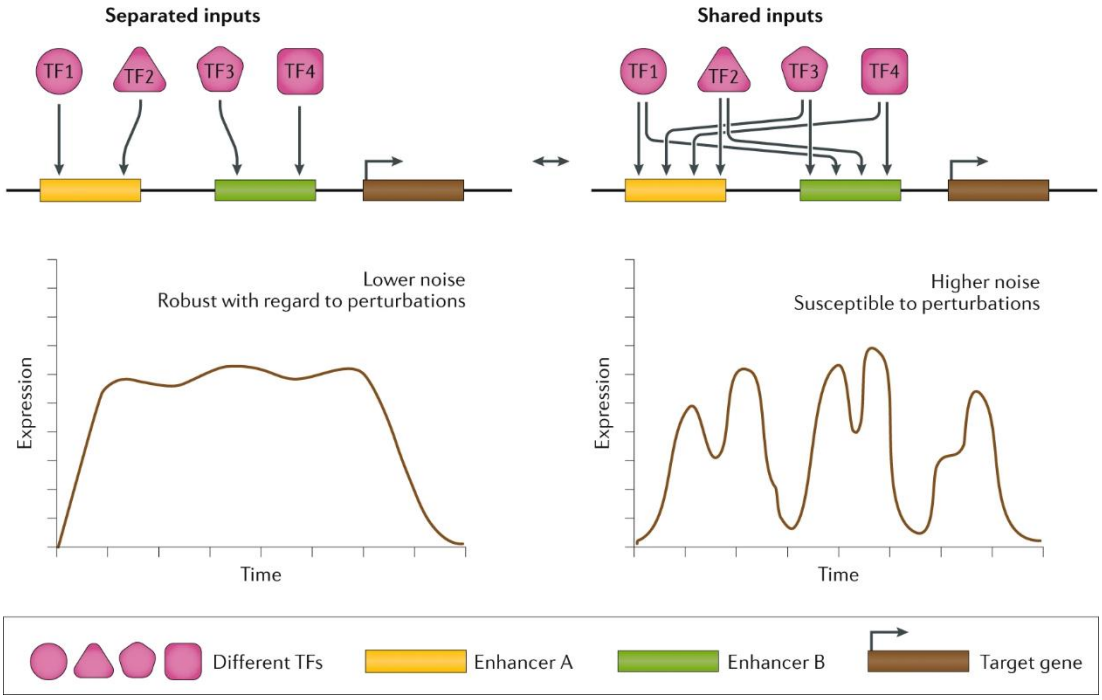


Figure 1.3 - Independent TF inputs to shadow enhancers lead to more robust transcriptional output. Shared and separated TF inputs to the individual shadow enhancers can have different effects on gene expression noise. In case of separated inputs, shadow enhancers regulating the same target gene do not share any of the same TF regulators (top left), while in case of shared inputs, shadow enhancers are regulated by the same set of TFs (top right). Below these two different models, we show the corresponding target gene expression dynamics in single cells as a function of time. Lower expression noise is seen with shadow enhancers with separated TF inputs than with shadow enhancers using shared TF inputs.

1.4 References

1. Gregor, T., Tank, D. W., Wieschaus, E. F., & Bialek, W. (2007). Probing the Limits to Positional Information. *Cell*, *130*(1), 153–164. https://doi.org/10.1016/j.cell.2007.05.025
2. Dubuis, J. O., Samanta, R., & Gregor, T. (2013). Accurate measurements of dynamics and reproducibility in small genetic networks. *Molecular Systems Biology*, *9*, 639. https://doi.org/10.1038/msb.2012.72
3. Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*, *166*(2), 358–368. https://doi.org/10.1016/j.cell.2016.05.025
4. Dar, R. D., Razooky, B. S., Singh, A., Trimeloni, T. V., McCollum, J. M., Cox, C. D., … Weinberger, L. S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences*, *109*(43), 17454–17459. https://doi.org/10.1073/pnas.1213530109
5. Chubb, J. R., Trcek, T., Shenoy, S. M., & Singer, R. H. (2006). Transcriptional Pulsing of a Developmental Gene. *Current Biology*, *16*(10), 1018–1025. https://doi.org/10.1016/j.cub.2006.03.092
6. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biology*, *4*(10), e309. https://doi.org/10.1371/journal.pbio.0040309
7. Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, *123*(6), 1025–1036. https://doi.org/10.1016/j.cell.2005.09.031
8. Zoller, B., Little, S. C., & Gregor, T. (2018). Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting. *Cell*, *175*(3), 835-847.e25. https://doi.org/10.1016/j.cell.2018.09.056
9. Wang, Y., Qi, J., Shao, J., & Tang, X. Q. (2020). Signaling mechanism of transcriptional bursting: A technical resolution-independent study. *Biology*, *9*(10), 1–11. https://doi.org/10.3390/biology9100339
10. Bakker, R., Mani, M., & Carthew, R. W. (2020). The WG and DPP morphogens regulate gene expression by modulating the frequency of transcriptional bursts. *ELife*, *9*, 1–26. https://doi.org/10.7554/eLife.56076
11. Wu, S., Li, K., Li, Y., Zhao, T., Li, T., Yang, Y.-F., & Qian, W. (2017). Independent regulation of gene expression level and noise by histone modifications. *PLOS Computational Biology*, *13*(6), e1005585. https://doi.org/10.1371/journal.pcbi.1005585
12. Carey, L. B., van Dijk, D., Sloot, P. M. A., Kaandorp, J. A., & Segal, E. (2013). Promoter Sequence Determines the Relationship between Expression Level and Noise. *PLoS Biology*, *11*(4), e1001528. https://doi.org/10.1371/journal.pbio.1001528
13. Csárdi, G., Franks, A., Choi, D. S., Airoldi, E. M., & Drummond, D. A. (2015). Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS Genetics*, *11*(5), e1005206. https://doi.org/10.1371/journal.pgen.1005206
14. Fukaya, T. (2021). Dynamic regulation of anterior-posterior patterning genes in living Drosophila embryos. *Current Biology*. https://doi.org/10.1016/j.cub.2021.02.050
15. Pedraza, J. M., & van Oudenaarden, A. (2005). Noise propagation in gene networks. *Science (New York, N.Y.)*, *307*(5717), 1965–1969. https://doi.org/10.1126/science.1109090
16. Blake, W. J., KÆrn, M., Cantor, C. R., & Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature*, *422*(6932), 633–637. https://doi.org/10.1038/nature01546

17. Huang, W., Carbone, M. A., Lyman, R. F., Anholt, R. R. H., & Mackay, T. F. C. (2020). Genotype by environment interaction for gene expression in Drosophila melanogaster. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-19131-y
18. Politis, S. N., Mazurais, D., Servili, A., Zambonino-Infante, J.-L., Miest, J. J., Sørensen, S. R., … Butts, I. A. E. (2017). Temperature effects on gene expression and morphological development of European eel, Anguilla anguilla larvae. *PLOS ONE*, *12*(8), e0182726. https://doi.org/10.1371/journal.pone.0182726
19. Cheung, D., & Ma, J. (2015). Probing the impact of temperature on molecular events in a developmental system. *Scientific Reports*, *5*(1), 13124. https://doi.org/10.1038/srep13124
20. Pastinen, T., Sladek, R., Gurd, S., Sammak, A., Ge, B., Lepage, P., … Hudson, T. J. (2004). A survey of genetic and epigenetic variation affecting human gene expression. *Physiological Genomics*, *16*(2), 184–193. https://doi.org/10.1152/physiolgenomics.00163.2003
21. Stapel, L. C., Zechner, C., & Vastenhouw, N. L. (2017). Uniform gene expression in embryos is achieved by temporal averaging of transcription noise. *Genes & Development*, *31*(16), 1635–1640. https://doi.org/10.1101/gad.302935.117
22. Erdmann, T., Howard, M., & ten Wolde, P. R. (2009). Role of Spatial Averaging in the Precision of Gene Expression Patterns. *Physical Review Letters*, *103*(25), 258101. https://doi.org/10.1103/PhysRevLett.103.258101
23. Raj, A., Rifkin, S. A., Andersen, E., & van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature*, *463*(7283), 913–918. https://doi.org/10.1038/nature08781
24. Lagha, M., Bothma, J. P., & Levine, M. (2012). Mechanisms of transcriptional precision in animal development. *Trends in Genetics : TIG*, *28*(8), 409–416. https://doi.org/10.1016/j.tig.2012.03.006
25. Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *34*(2), 135–141. https://doi.org/10.1002/bies.201100121
26. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
27. Hobert, O. Gene regulation: enhancers stepping out of the shadow. *Curr. Biol.* **20**, R697-9 (2010).
28. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr. Biol.* **26**, 38–51 (2016).
29. Wunderlich, Z. *et al.* Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. *Cell Rep.* **12**, 1740–1747 (2015).
30. Antosova, B. *et al.* The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. *PLoS Genet.* **12**, e1006441 (2016).
31. Letelier, J. *et al.* A conserved Shh cis-regulatory module highlights a common developmental origin of unpaired and paired fins. *Nat. Genet.* **50**, 504–509 (2018).
32. Nolte, C., Jinks, T., Wang, X., Martinez Pastor, M. T. & Krumlauf, R. Shadow enhancers flanking the HoxB cluster direct dynamic Hox expression in early heart and endoderm development. *Dev. Biol.* **383**, 158–173 (2013).
33. Swami, M. Transcription: Shadow enhancers confer robustness. *Nature reviews. Genetics* vol. 11 454 (2010).

34. Ghiasvand, N. M. *et al.* Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. *Nat. Neurosci.* **14**, 578–586 (2011).
35. Watts, J. A. *et al.* Study of FoxA pioneer factor at silent genes reveals Rfx-repressed enhancer at Cdx2 and a potential indicator of esophageal adenocarcinoma development. *PLoS Genet.* **7**, e1002277 (2011).
36. Twell, D., Yamaguchi, J., Wing, R. A., Ushiba, J. & McCormick, S. Promoter analysis of genes that are coordinately expressed during pollen development reveals pollen-specific enhancer sequences and shared regulatory elements. *Genes Dev.* **5**, 496–507 (1991).
37. Poulsen, C. & Chua, N. H. Dissection of 5' upstream sequences for selective expression of the Nicotiana plumbaginifolia rbcS-8B gene. *Mol. Gen. Genet.* **214**, 16–23 (1988).
38. Hoch, M., Schröder, C., Seifert, E. & Jäckle, H. cis-acting control elements for Krüppel expression in the Drosophila embryo. *The EMBO Journal* vol. 9 2587–2595 (1990).
39. Kassis, J. A. Spatial and temporal control elements of the Drosophila engrailed gene. *Genes Dev.* **4**, 433–443 (1990).
40. Zeitlinger, J. *et al.* Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. *Genes Dev.* **21**, 385–390 (2007).
41. Camprodón, F. J. & Castelli-Gair, J. E. Ultrabithorax protein expression in breakpoint mutants: localization of single, co-operative and redundant cis regulatory elements. *Rouxs. Arch. Dev. Biol.* **203**, 411–421 (1994).
42. Bachmann, A. & Knust, E. Dissection of cis-regulatory elements of the Drosophila gene Serrate. *Dev. Genes Evol.* **208**, 346–351 (1998).
43. Schroeder, M. D. *et al.* Transcriptional control in the segmentation gene network of Drosophila. *PLoS Biol.* **2**, E271 (2004).
44. Pappu, K. S. *et al.* Dual regulation and redundant function of two eye-specific enhancers of the Drosophila retinal determination gene dachshund. *Development* **132**, 2895–2905 (2005).
45. Yao, L. C. *et al.* Multiple modular promoter elements drive graded brinker expression in response to the Dpp morphogen gradient. *Development* **135**, 2183–2192 (2008).
46. Ertzer, R. *et al.* Cooperation of sonic hedgehog enhancers in midline expression. *Dev. Biol.* **301**, 578–589 (2007).
47. Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M. & Epstein, D. J. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* **133**, 761–772 (2006).
48. Kurokawa, D. *et al.* Regulation of Otx2 expression and its functions in mouse forebrain and midbrain. *Development* **131**, 3319–3331 (2004).
49. Nakada, Y., Parab, P., Simmons, A., Omer-Abdalla, A. & Johnson, J. E. Separable enhancer sequences regulate the expression of the neural bHLH transcription factor neurogenin 1. *Dev. Biol.* **271**, 479–487 (2004).
50. Menke, D. B., Guenther, C. & Kingsley, D. M. Dual hindlimb control elements in the Tbx4 gene and region-specific control of bone size in vertebrate limbs. *Development* **135**, 2543–2553 (2008).
51. Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077–3086 (2002).
52. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**, 975–985 (1987).

53. Zhou, X. & Sigmund, C. D. Chorionic enhancer is dispensable for regulated expression of the human renin gene. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology* **294**, R279 (2008).
54. Allan, C. M., Walker, D. & Taylor, J. M. Evolutionary duplication of a hepatic control region in the human apolipoprotein E gene locus. Identification of a second region that confers high level and liver-specific expression of the human apolipoprotein E gene in transgenic mice. *J. Biol. Chem.* **270**, 26278–26281 (1995).
55. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
56. Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).
57. Babbitt, C. C., Markstein, M. & Gray, J. M. Recent advances in functional assays of transcriptional enhancers. *Genomics* **106**, 137–139 (2015).
58. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).
59. Gorkin, D. U. *et al.* An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744–751 (2020).
60. Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
61. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
62. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
63. Parker, S. C. J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17921–17926 (2013).
64. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016).
65. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).
66. Dukler, N., Gulko, B., Huang, Y.-F. & Siepel, A. Is a super-enhancer greater than the sum of its parts? *Nature genetics* vol. 49 2–3 (2016).
67. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
68. Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* **17**, 207–223 (2016).
69. Arner, E. *et al.* Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010–1014 (2015).
70. Wang, X. & Goldstein, D. B. Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease. *Am. J. Hum. Genet.* **106**, 215–233 (2020).
71. Kvon, E. Z. *et al.* Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).
72. Stefanakis, N., Carrera, I. & Hobert, O. Regulatory Logic of Pan-Neuronal Gene Expression in C. elegans. *Neuron* **87**, 733–750 (2015).
73. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).

74. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–493 (2010).
75. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
76. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2013).
77. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
78. Antosova, B. *et al.* The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. *PLoS Genet.* **12**, e1006441 (2016).
79. Sagai, T. *et al.* SHH signaling directed by two oral epithelium-specific enhancers controls tooth and oral development. *Sci. Rep.* **7**, 13004 (2017).
80. Dickel, D. E. *et al.* Ultraconserved Enhancers Are Required for Normal Development. *Cell* **172**, 491-499.e15 (2018).
81. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
82. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).
83. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
84. Cisse, I. I. *et al.* Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science* **341**, 664–667 (2013).
85. Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol. Cell* **76**, 473-484.e7 (2019).
86. Alexander, J. M. *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* **8**, e41769 (2019).
87. Tsai, A., Alves, M. R. & Crocker, J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *Elife* **8**, e45325 (2019).
88. Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., … Rubin, E. M. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, *444*(7118), 499–502. (2006). https://doi.org/10.1038/nature05295
89. O'Kane, C. J., & Gehring, W. J. (1987). Detection in situ of genomic regulatory elements in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(24), 9123–9127. https://doi.org/10.1073/pnas.84.24.9123
90. Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., … Grell, E. (1989). Searching for pattern and mutation in the Drosophila genome with a P-lacZ vector. *Genes & Development*, *3*(9), 1273–1287. https://doi.org/10.1101/gad.3.9.1273
91. Swanson, C. I., Evans, N. C., & Barolo, S. (2010). Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye Enhancer. *Developmental Cell*, *18*(3), 359–370. https://doi.org/10.1016/j.devcel.2009.12.026
92. Benton, M. L., Talipineni, S. C., Kostka, D., & Capra, J. A. Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function. *BMC Genomics*, *20*(1), 511. (2019). https://doi.org/10.1186/s12864-019-5779-x
93. Zeitlinger, J. Seven myths of how transcription factors read the cis-regulatory code. *Current Opinion in Systems Biology*. Elsevier Ltd. (2020). https://doi.org/10.1016/j.coisb.2020.08.002

94. Halfon, M. S. Studying Transcriptional Enhancers: The Founder Fallacy, Validation Creep, and Other Biases. *Trends in Genetics*. Elsevier Ltd. (2019). https://doi.org/10.1016/j.tig.2018.11.004
95. Shukla, A., & Huangfu, D. Decoding the noncoding genome via large-scale CRISPR screens. *Current Opinion in Genetics and Development*. Elsevier Ltd. (2018). https://doi.org/10.1016/j.gde.2018.06.001
96. Lim, B., Fukaya, T., Heist, T., & Levine, M. Temporal dynamics of pair-rule stripes in living Drosophila embryos. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(33), 8376–8381. (2018). https://doi.org/10.1073/pnas.1810430115
97. Dunipace, L., Ákos, Z., & Stathopoulos, A. Coacting enhancers can have complementary functions within gene regulatory networks and promote canalization. *PLOS Genetics*, *15*(12), e1008525. (2019). https://doi.org/10.1371/journal.pgen.1008525
98. Biggin, M. D. (2011). Animal Transcription Networks as Highly Connected, Quantitative Continua. *Developmental Cell*, *21*(4), 611–626. https://doi.org/10.1016/J.DEVCEL.2011.09.008
99. Milo, R., Jorgensen, P., Moran, U., Weber, G., & Springer, M. (2009). BioNumbers The database of key numbers in molecular and cell biology. *Nucleic Acids Research*, *38*(SUPPL.1), D750. https://doi.org/10.1093/nar/gkp889
100. Zhao, Z. W., Roy, R., Gebhardt, J. C. M., Suter, D. M., Chapman, A. R., & Xie, X. S. Spatial organization of RNA polymerase II insidea mammalian cell nucleus revealed by reflectedlight-sheet superresolution microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(2), 681–686. (2014). https://doi.org/10.1073/pnas.1318496111
101. Liu, X., Wu, B., Szary, J., Kofoed, E. M., & Schaufele, F. (2007). Functional sequestration of transcription factor activity by repetitive DNA. *Journal of Biological Chemistry*, *282*(29), 20868–20876. https://doi.org/10.1074/jbc.M702547200
102. Janssen, S., Cuvier, O., Müller, M., & Laemmli, U. K. (2000). Specific Gain- and Loss-of-Function Phenotypes Induced by Satellite-Specific DNA-Binding Drugs Fed to Drosophila melanogaster. *Molecular Cell*, *6*(5), 1013–1024. https://doi.org/10.1016/S1097-2765(00)00100-3
103. Brewster, R. C., Weinert, F. M., Garcia, H. G., Song, D., Rydenfelt, M., & Phillips, R. (2014). The transcription factor titration effect dictates level of gene expression. *Cell*, *156*(6), 1312–1323. https://doi.org/10.1016/j.cell.2014.02.022
104. Das, D., Dey, S., Brewster, R. C., & Choubey, S. (2017). Effect of transcription factor resource sharing on gene expression noise. *PLoS Computational Biology*, *13*(4), e1005491. https://doi.org/10.1371/journal.pcbi.1005491
105. Lee, T. H., & Maheshri, N. (2012). A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Molecular Systems Biology*, *8*, 576. https://doi.org/10.1038/msb.2012.7
106. Laboulaye, M. A., Duan, X., Qiao, M., Whitney, I. E., & Sanes, J. R. (2018). Mapping transgene insertion sites reveals complex interactions between mouse transgenes and neighboring endogenous genes. *Frontiers in Molecular Neuroscience*, *11*. https://doi.org/10.3389/fnmol.2018.00385
107. Thompson, A., & Gasson, M. J. (2001). Location effects of a reporter gene on expression levels and on native protein synthesis in Lactococcus lactis and Saccharomyces cerevisiae.

Applied and Environmental Microbiology, *67*(8), 3434–3439. https://doi.org/10.1128/AEM.67.8.3434-3439.2001
108. Dimanlig, P. V, Faber, S. C., Auerbach, W., Makarenkova, H. P., & Lang, R. A. (2001). The upstream ectoderm enhancer in Pax6 has an important role in lens induction. *Development (Cambridge, England)*, *128*(22), 4415–4424.
109. Hui, C. C., & Joyner, A. L. (1993). A mouse model of Greig cephalo–polysyndactyly syndrome: The extra–toesJ mutation contains an intragenic deletion of the Gli3 gene. *Nature Genetics*, *3*(3), 241–246. https://doi.org/10.1038/ng0393-241.
110. Hogan, B. L., Horsburgh, G., Cohen, J., Hetherington, C. M., Fisher, G., & Lyon, M. F. (1986). Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse. *Journal of Embryology and Experimental Morphology*, *97*, 95–110.
111. Hill, R. E., Favor, J., Hogan, B. L. M., Ton, C. C. T., Saunders, G. F., Hanson, I. M., … Van Heyningen, V. (1991). Mouse Small eye results from mutations in a paired-like homeobox-containing gene. *Nature*, *354*(6354), 522–525. https://doi.org/10.1038/354522a0

CHAPTER 2

Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic

Note: This chapter was published in *eLife* in August, 2020 as Waymack, R., Fletcher, A., Enciso, G., & Wunderlich, Z. (2020)

Chapter 2

Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic

2.1 Abstract

Shadow enhancers, groups of seemingly redundant enhancers, are found in a wide range of organisms and are critical for robust developmental patterning. However, their mechanism of action is unknown. We hypothesized that shadow enhancers drive consistent expression levels by buffering upstream noise through a separation of transcription factor (TF) inputs at the individual enhancers. By measuring transcriptional dynamics of several Kruppel shadow enhancer configurations in live Drosophila embryos, we showed individual member enhancers act largely independently. We found that TF fluctuations are an appreciable source of noise that the shadow enhancer pair can better buffer than duplicated enhancers. The shadow enhancer pair is uniquely able to maintain low levels of expression noise across a wide range of temperatures. A stochastic model demonstrated the separation of TF inputs is sufficient to explain these findings. Our results suggest the widespread use of shadow enhancers is partially due to their noise suppressing ability.

2.2 Introduction

The first evidence that transcription occurred in bursts, as opposed to as a smooth, continuous process, was observed in *Drosophila* embryos, where electron micrographs showed that even highly transcribed genes had regions of chromatin lacking associated transcripts between regions of densely associated nascent transcripts[1]. As visualization techniques have improved, it is increasingly clear that transcriptional bursting is the predominant mode of expression across

organisms from bacteria to mammals[2–6]. These bursts of transcriptional activity, separated by periods of relative silence, have important implications for cellular function, as mRNA numbers and fluctuations largely dictate these quantities at the protein level[7,8]. Such fluctuations in regulatory proteins, like TFs and signaling molecules, can propagate down a gene regulatory network, significantly altering the expression levels or noise of downstream target genes[9].

Given that protein levels fluctuate and that these fluctuations can cascade down regulatory networks, this raises the question of how organisms establish and maintain the precise levels of gene expression seen during development, where expression patterns can be reproducible down to half-nuclear distances in *Drosophila* embryos[10,11]. Many mechanisms that buffer against expression noise, either inherent or stemming from genetic or environmental variation, have been observed[12–14]. For example, organisms use temporal and spatial averaging mechanisms and redundancy in genetic circuits to achieve the precision required for proper development[12–15]. Here, we propose that shadow enhancers may be another mechanism by which developmental systems manage noise[16].

Shadow enhancers are groups of two or more enhancers that control the same target gene and drive overlapping spatiotemporal expression patterns[16,17]. Shadow enhancers are found in a wide range of organisms, from insects to plants to mammals, particularly in association with developmental genes[18–21]. While seemingly redundant, the individual enhancers of a shadow enhancer group have been shown to be critical for proper gene expression in the face of both environmental and genetic perturbations[19,22,23]. Such perturbations may exacerbate fluctuations in upstream regulators[24,25]. Although shadow enhancers are shown to be pervasive in developmental systems and necessary for robust gene expression, their precise mechanism of action is still unknown. One proposed mechanism puts forth that having multiple enhancers

controlling the same promoter reduces the effective "failure rate" of the promoter and ensures a critical threshold of gene expression is reached[26,27]. An alternative, but not mutually exclusive, possibility is that shadow enhancers ensure robust expression by buffering noise in upstream regulators. Several studies suggest that individual enhancers of a shadow enhancer group tend to be controlled by different sets of TFs, which we call a "separation of inputs"[18,28,29]. We hypothesize that this separation of inputs allows shadow enhancers to buffer against fluctuations in upstream TF levels.

The *Drosophila* gap gene *Kruppel* (*Kr*) provides a useful system in which to address the mechanisms of action of shadow enhancers. During early embryogenesis, *Kr* is controlled by the activity of two enhancers, proximal and distal[27], that drive overlapping expression in the center of the embryo (Figure S2.1). We call the two individual enhancers together the *shadow enhancer pair*. Previous experiments have shown that each enhancer is activated by different TFs[28] (Figure 2.1A). Here we focus on differences in activation, as the key repressors of *Kr*, Knirps and Giant, are likely to regulate both enhancers. *Kr* expression during this time is critical for thorax formation, and like the other gap genes in the *Drosophila* embryo, has quite low noise[10,30]. By measuring live mRNA dynamics, we can use the *Kr* system in *Drosophila* embryos to assess whether and how shadow enhancers act to buffer noise and identify the sources of noise in the developing embryo.

To test our hypothesis, we measured live mRNA dynamics driven by either single *Kr* enhancer, duplicated enhancers, or the shadow enhancer pair and compared the dynamics and noise associated with each. We showed that the individual *Kr* enhancers can act largely independently in the same nucleus, while identical enhancers display correlated activity. We constructed a simple mathematical model to describe this system and found that TF fluctuations are necessary

to reproduce the correlated activity of identical enhancers in the same nucleus. The shadow enhancer pair drives lower noise than either duplicated enhancer, and using the model, we found that this is a natural consequence of the separation of TF inputs. Additional experiments, including simultaneous measurements of TF levels and expression and a decomposition of noise sources, further demonstrate that the shadow enhancer pair is less sensitive to fluctuations in TF levels than is a single enhancer. Additionally, the shadow enhancer pair is uniquely able to maintain low levels of expression noise across a wide range of temperatures. We suggest that this noise suppression ability is one of the key features that explains the prevalence of shadow enhancers in developmental systems.

2.3 Results

2.3.1 Individual enhancers in the shadow enhancer pair act nearly independently within a nucleus

To test our hypothesis that the separation of inputs between *Kruppel's (Kr)* shadow enhancers provides them with noise-buffering capabilities, we needed to first test the ability of each enhancer to act independently. In previous work, we found that the individual enhancers in the shadow enhancer pair are controlled by different activating TFs[28]. These experiments established that the enhancers responded differently to perturbations in key TFs, indicating that each enhancer uses a distinct regulatory logic. The proximal enhancer is activated by Hunchback (Hb) and Stat92E, and the distal enhancer is activated by Bicoid (Bcd) and Zelda (Zld) (Figure 2.1A). Given this separation of inputs, the shadow enhancer pair could provide a form of noise buffering if variability in gene expression is driven primarily by fluctuations in upstream factors. Conversely, variability in upstream regulators may be low enough in the developing embryo that

these fluctuations are not the primary driver of downstream expression noise. If this were the case, the separation of inputs is unlikely to be a key requirement of shadow enhancer function.

To investigate these possibilities, we measured and compared the correlation of allele activity in homozygous or heterozygous embryos that carry two reporter genes. *Proximal homozygotes* contained the proximal enhancer driving a reporter, inserted in the same location on both homologous chromosomes, and *distal homozygotes* similarly had the distal enhancer driving reporter expression on both homologous chromosomes (Figure 2.1B). We also made heterozygous embryos, called *shadow heterozygotes*, which had one proximal and one distal reporter, again in the same location on both homologous chromosomes. To measure live mRNA dynamics and correlations in allele activity, we used the MS2-MCP reporter system (Figure 2.1C, D). This system allows the visualization of mRNAs that contain the MS2 RNA sequence, which is bound by an MCP-GFP fusion protein[31]. In the developing embryo, only the site of nascent transcription is visible, as single transcripts are too dim, allowing us to measure the rate of transcription[32,33]. In blastoderm-stage embryos with two MS2 reporter genes, we can observe two distinct foci of fluorescence corresponding to the two alleles (Figure 2.1D; Videos 1-6), in line with previous results that suggest there are low levels of transvection at this stage[34,35]. To confirm our ability to distinguish the two alleles, we imaged transcription in embryos hemizygous for our reporter constructs, which only show one spot of fluorescence per nucleus. Our counts of active transcription sites in homozygous embryos correspond well to the expected value calculated from hemizygous embryos (Figure S2.2). Therefore, we are able to measure the correlation of allele activity, though we cannot identify which spot corresponds to which reporter.

We predicted that if variability in gene expression is driven by fluctuations in input TFs, we would observe lower correlations of allele activity in shadow heterozygotes than in either the proximal or distal homozygotes. However, if global factors affecting both enhancers dominate, there would be no difference in allele activity correlations. During the ~1 hour of nuclear cycle 14 (nc14) we found that allele activity is more than twice as correlated in both proximal and distal homozygotes than in shadow heterozygote embryos at 47-57% egg length, which encompasses the central region of *Kr* expression during this time period (Figure 2.1). The difference in our ability to measure allele correlation in the more anterior and posterior regions of the embryo stems from the slightly different expression patterns driven by the proximal and distal enhancers (Figure S2.1 & S2.3). The lower allele correlation in shadow heterozygote embryos indicates not only that the individual member enhancers of the shadow enhancer pair can act largely independently in the same nucleus, but that differential TF inputs are likely the primary determinants of transcriptional bursts in this system. Notably, heterozygotes still show marginal allele correlation, indicating that some correlation is induced by either shared input TFs or factors that affect transcription globally. The independence of individual *Kr* enhancers allows for the possibility that shadow enhancers can act to buffer noise by providing distinct inputs to the same gene expression output.

2.3.2 Transcription factor fluctuations are required for the observed differences in the correlations of enhancer activity

To explore the conditions needed for the two *Kr* enhancers to act nearly independently within the same nucleus, we generated a simple model of enhancer-driven dynamics. We considered an enhancer E_i that interacts with a transcription factor T_i, which together bind to the promoter to form the active promoter-enhancer complex C_i (Figure 2.2A). When the promoter is bound by

the enhancer, it drives the production of mRNA. Since the MS2 system only allows us to observe mRNA at the site of transcription, we modeled the diffusion of mRNA away from the transcription site as decay. The transcription factor T_i is produced in bursts of n_i molecules at a time, and it degrades linearly. For simplicity, the transcription factor T_i is an abstraction of the multiple activating TFs that interact with the enhancer, and T_i corresponds to a different set of TFs for the proximal and distal enhancer. This nonlinear model generalizes the linear model by Bothma et al.[36] by explicitly taking into account the presence of TFs.

We estimated some model parameters directly from experimental data and others by fitting using simulated annealing. The mRNA degradation parameter α and production parameter r were measured directly from fluorescence data without any input from the model (see Methods for details). The remaining parameters were first estimated using mathematical analysis, then fine-tuned using simulated annealing. We found separate parameter sets for the proximal and distal enhancers that, when used to simulate transcription, fit the experimentally measured characteristics of the transcriptional traces, including transcription burst size, frequency, and duration, as well as the total mRNA produced (Figure S2.4).

We hypothesized that a model that lacks fluctuations in the input TFs could not recapitulate the high correlation of transcriptional activity in homozygotes versus the low correlation in heterozygotes. To test this hypothesis, we generated another model of TF production. We call our original model described above *bursting TFs*. The other model is one in which TF numbers are constant over time, which we call *constant TFs* and is equivalent to the model in[36]. If the difference in transcription correlation between homozygotes and heterozygotes is due to fluctuating numbers of TFs, we expected that the bursting TFs model will recapitulate this behavior, while the constant TFs model will not. However, if the constant TFs model is also

able to recapitulate the observed difference in correlations, then the correlations are likely a consequence of the identical enhancers simply being regulated by the same set of TFs.

For each model, we used the 10 best parameter sets to simulate transcriptional activity in homozygote and heterozygote embryos and analyzed the resulting allele correlations. We found that the bursting TFs model always produced results in which both homozygote allele correlations are significantly higher than the heterozygote, which qualitatively mirrors the experimental observations (Figure 2.2B). None of the best fitting parameter sets for the constant TF model were able to produce the experimentally-observed behavior and always resulted in near zero correlations for both the homozygote and heterozygote embryos (Figure 2.2C). Notably, using the bursting TFs model, all the simulated allele correlations were lower than the experimentally observed values, e.g. the simulated heterozygote allele correlation was near zero, while the experimental value was 0.14 at the embryo's midpoint. We hypothesized that this discrepancy was because the model assumes complete independence of the proximal and distal enhancer input TFs, while in reality, there may be some degree of shared inputs, either of known TFs or a general component of the transcriptional machinery. To test this hypothesis, we generated a model that added a common TF to the bursting TFs model and attempted to fit the model parameters. Some of the best parameter sets recapitulated the nonzero correlation of the heterozygote embryos, indicating that a shared factor may play a role in this system, however this behavior was inconsistent from one parameter set to the next (Figure S2.5). Therefore, we concluded that the simpler bursting TFs model, which consistently recapitulated the key features of the allele correlation data, was more suitable for subsequent analysis.

In conclusion, in our minimalist model of enhancer-driven transcription, the presence of TF fluctuations is required for the observed differences in allele correlation. These results also

demonstrate the advantage of using a single generic TF for each enhancer. By abstracting away TF interactions, we reduced the complexity and number of parameters in the model, which allowed us to explore the relationship between TF production and allele correlation.

2.3.4 The shadow pair's activity is less sensitive to fluctuations in Bicoid levels than is the activity of a single enhancer

Both the experimental measurements of allele correlation and the computational model suggest that input TF fluctuations are an appreciable source of noise for enhancer activity. Further, previous experimental work[28] and the low correlation of transcriptional activity in heterozygotes (Figure 2.1E) indicates that each individual *Kr* enhancer receives different TF input signals. This suggests that the shadow enhancer pair will be less sensitive to an input TF fluctuation than a single enhancer, because the shadow enhancer pair's activity is dependent on a broader range of TF inputs. To directly observe the relationship between input TF levels and enhancer output, we simultaneously tracked bcd levels and enhancer activity in individual nuclei (Figure 2.3; Supp. Movies 7-8). We measured this relationship for both the distal enhancer, which is activated by bcd, and the shadow enhancer pair, and predicted that the distal enhancer's transcription dynamics are more strongly influenced by fluctuations in bcd levels.

To allow for tracking of both Bcd levels and enhancer activity, we crossed female flies that express eGFP-tagged Bcd in the place of endogenous Bcd[37] (called Bcd-GFP from here on) and MCP-mCherry with male flies homozygous for either the shadow pair or distal enhancer reporter. As the Bcd-GFP transgene was inserted in a Bcd null background, the resulting embryos should receive roughly WT levels of Bcd. The females flies were heterozygous for the maternally-deposited Bcd-GFP, and therefore, we estimate that roughly half of the Bcd proteins

were labeled. Given previous work demonstrating the normal function and expression levels of tagged Bcd, we expect the Bcd-GFP levels to be a representative sample of total Bcd[37].

Higher activator TF levels increase enhancer activity. We therefore measured the correlation of nuclear Bcd-GFP levels to the slope of MS2 signal. When the enhancer is active, MS2 signal has a positive slope, when the enhancer is inactive, slope is negative. If the shadow enhancer pair is less sensitive than the distal enhancer to fluctuations in Bcd levels, we would predict higher correlation between Bcd-GFP levels and the activity of the distal enhancer than that of the shadow enhancer pair. We find that the transcription dynamics driven by the distal enhancer are indeed significantly more correlated to nuclear Bcd-GFP levels (median r = 0.18) than the dynamics driven by the shadow pair (median r = 0.14; Figure 2.3F; *p*-value = 6.1×10^{-3}), though both correlations are modest (see Discussion). The lower correlation indicates that transcription driven by the shadow pair is less sensitive to Bcd level fluctuations than is the distal enhancer. Our modeling recapitulates this finding, showing that the separated TF inputs of the shadow pair are sufficient to explain the observed decreased sensitivity to TF fluctuations (median r = 0.14 for the distal enhancer; 0.11 for the shadow pair; *p*-value = 2.2×10^{-2}; Figure 2.3G). These findings indicate that the shadow enhancer pair is better able to buffer fluctuations in a single activating TF than a single enhancer, likely due to the shadow enhancer pair's separation of TF inputs.

2.3.5 The shadow enhancer pair drives less noisy expression than enhancer duplications

We wanted to further test whether the shadow enhancer pair drives less noisy gene expression output than a simple enhancer duplication. We compared the noise in expression driven by the shadow enhancer pair to that driven by two copies of either the distal or proximal enhancer

(Figure 2.4). If the shadow enhancer pair drives lower noise, this suggests that having two independently acting enhancers is a critical feature of shadow enhancers' ability to reduce variability and mediate robustness. Alternatively, if duplicated enhancers drive similar levels of expression noise, this suggests that enhancer independence is not critical for shadow enhancer's function and that shadow enhancers mediate robustness through a different mechanism, such as ensuring a critical threshold of expression is met[26,27].

We tracked transcriptional activity in embryos expressing MS2 under the control of the shadow enhancer pair, a duplicated proximal enhancer, or a duplicated distal enhancer (Figure 2.4). To measure noise associated with each enhancer, we used these traces to calculate the coefficient of variation (CV) of transcriptional activity across nc14. CV is the standard deviation divided by the mean and provides a unitless measure of noise to allow comparisons among our enhancer constructs. We then grouped these CV values by the embryo position of the transcriptional spots and found the average CV at each position for each enhancer construct. All of the enhancer constructs display the lowest expression noise at the embryo position of their peak expression (Figure 2.4A), in agreement with previous findings of an inverse relationship between mean expression and noise levels[38] (Figure S2.8). The shadow enhancer pair's expression noise is ~ 30% or 15% lower, respectively, than that of the duplicated proximal or distal enhancers in their positions of maximum expression (Figure .4C).

If the primary function of shadow enhancers is only to ensure a critical threshold of expression is reached, we would not expect to also see the lower expression noise associated with the shadow enhancer pair compared to either duplicated enhancer. Furthermore, this decreased expression noise is not simply a consequence of higher expression levels, as the shadow enhancer pair produces less mRNA than the duplicated distal enhancer during nc14

(Figure 2.4B). The lower expression noise associated with the shadow enhancer pair suggests that it is less susceptible to fluctuations in upstream TFs than multiple identical enhancers.

2.3.6 Modeling indicates the separation of input TFs is sufficient to explain the low noise driven by the shadow enhancer pair

To explore which factors drive the difference in CVs between the duplicated and shadow enhancer constructs, we extended our model to have a single promoter controlled by two enhancers (Figure 2.5A). To do so, we assumed that either or both enhancers can be looped to the promoter and drive mRNA production. The rate of mRNA production when both enhancers are looped is the sum of the rates driven by the individual enhancers. We assumed that some parameters, e.g. the TF production rates and mRNA decay rate, are the same as the single enhancer case. We allowed the parameters describing the promoter-enhancer looping dynamics (the k_{on} and k_{off} values) to differ, depending on the enhancer's position in the construct relative to the promoter and whether another enhancer is present. To fit the k_{on} and k_{off} values, we used the medians of the 10 best single enhancer parameter sets as a starting point and performed simulated annealing to refine them.

This approach allowed us to examine how the model parameters that describe promoter-enhancer looping dynamics change when two enhancers are controlling the same promoter. We compared the k_{off} and k_{on} values for each enhancer in the two enhancer constructs to their values from the single enhancer model. We generally found that k_{off} values increased and k_{on} values decreased (Figure 2.5B). The effect is most pronounced in the duplicated distal enhancer, with large changes to the k_{off} and k_{on} values for the enhancer in the position far from the promoter (position 2). Given that our model assumes that enhancers act additively and only allows for changes in the k_{off} and k_{on} values, these observed effects may indicate that either the presence of a

second enhancer interferes with promoter-enhancer looping or that the promoter can be saturated. Our model cannot distinguish between these two possibilities, but these observations are consistent with our (Figure S2.9) and previous results indicating that the *Kr* enhancers can act sub-additively[39]. Additionally, the dramatic changes in k_{off} and k_{on} values in the duplicated distal enhancer are consistent with a previous assertion that enhancer sub-additivity is most pronounced in cases of strong enhancers[36].

We used these models to simulate transcription and predict the resulting CVs from the duplicated enhancer and shadow pair constructs. In line with experimental data, we found the model predicts that the shadow pair construct drives lower noise than the duplicated distal or duplicated proximal enhancer constructs in the middle of the embryo (Figure 2.5C). This is particularly notable, as we did not explicitly fit our model to reproduce the experimentally observed CVs. There is only one fundamental difference between the shadow pair and duplicated enhancer models, namely the use of separate TF inputs for the shadow pair. Therefore, in our simplified model, we can conclude that the separation of input TFs is sufficient to explain the lower noise driven by the shadow enhancer pair construct.

2.3.7 The shadow enhancer pair buffers against intrinsic and extrinsic sources of noise

To further understand the sources of noise the shadow enhancer pair is able to buffer, we compared the extrinsic and intrinsic noise associated with the shadow enhancer pair to that associated with either single or duplicated enhancers. To do so, we measured the transcriptional dynamics of embryos with two identical reporters in each nucleus and calculated noise sources using the approach of Elowitz, et al.[40]. Intrinsic noise corresponds to sources of noise, such as TF binding and unbinding, that affect each allele separately. It is quantified by the degree to which the activities of the two reporters in a single nucleus differ. Extrinsic noise corresponds to global

sources of noise, such as TF levels, that affect both alleles simultaneously. It is measured by the degree to which the activities of the two reporters change together. Intrinsic and extrinsic noise are defined such that, when squared, their sum is equal to total noise^2, which corresponds to the CV^2 of the two identical alleles in each nucleus in our system (see Methods). Because our data do not meet one key assumption needed to measure extrinsic and intrinsic noise with the two-reporter approach (see Discussion; Figure S2.10), we use the terms inter-allele noise and covariance in place of intrinsic and extrinsic noise.

Based on our separation of inputs hypothesis and CV data, we expected the total noise associated with the shadow enhancer pair to be lower than that associated with the duplicated enhancers. We predicted that the shadow enhancer pair will mediate lower total expression noise through lower covariance, as the two member enhancers are regulated by different TFs. Given the complexity of predicting inter-allele noise from first principles (Methods; Figure S2.11), we predicted that constructs with two enhancers will have lower inter-allele noise than single enhancer constructs but did not have a strong prediction regarding the relative inter-allele noise among the different two-enhancer constructs. Comparisons of noise between the single and duplicated enhancer constructs would further allow us to discern whether reductions in noise are generally associated with two-enhancer constructs or whether this is a particular feature of the shadow enhancer pair.

Neither the duplicated proximal nor distal enhancers drive significantly lower total noise than the corresponding single enhancers, indicating that the addition of an identical enhancer is not sufficient to reduce expression noise in this system (Figure 2.6A). The shadow enhancer pair drives lower total expression noise than either single or duplicated enhancer, consistent with the temporal CV data in Figure 2.4. The median total expression noise associated with the duplicated

distal and duplicated proximal enhancers is 1.4 or 2.4 times higher, respectively, than that associated with the shadow enhancer pair (Figure 2.6A). Note that for measurements of noise, our distal construct places the enhancer at the endogenous spacing from the promoter, as we wanted to control for positional effects on expression and noise[39] (Figure S2.12).

In line with our expectations, the shadow enhancer pair has significantly lower covariance levels than either single or duplicated enhancer (Figure 2.6B). The shadow enhancer pair also has lower inter-allele noise than all of the other constructs, though these differences are only marginally significant ($p = 0.13$) when compared to the duplicated distal enhancer. Covariance makes a larger contribution to the total noise for the duplicated distal enhancer and the shadow enhancer pair, while inter-allele noise is the larger source of noise for the single distal enhancer and the single or duplicated proximal enhancers (Figure 2.6B).

The lower total noise and covariance of the shadow enhancer pair support our hypothesis that, by separating regulation of the member enhancers, the shadow enhancer pair can buffer against upstream fluctuations. The lower inter-allele noise associated with the shadow enhancer pair warrants further investigation. A simple theoretical approach predicts that two enhancer constructs will have lower inter-allele noise (Figure S2.11). Given that this is not universally observed in our data, this suggests that there is still much to discover about how inter-allele noise changes as additional enhancers control a gene's transcription.

2.3.8 The shadow enhancer pair drives low noise at several temperatures

We showed the *Kr* shadow enhancer pair drives expression with lower total noise than either single or duplicated enhancer, yet previous studies have generally found individual member enhancers of a shadow enhancer set are dispensable under ideal conditions[19,22,27]. However, in

the face of environmental or genetic stress, the full shadow enhancer group is necessary for proper development[19,22,27]. We therefore decided to investigate whether temperature stress causes significant increases in expression noise and whether the shadow enhancer pair or duplicated enhancers can buffer these potential increases in noise.

Similar to our findings at ambient temperature (26.5°C), the shadow enhancer pair drives lower total noise than all other tested enhancer constructs at 32°C (Figure 2.7B). At 32°C, the duplicated distal and duplicated proximal enhancers display 35% or 52%, respectively, higher total noise than the shadow enhancer pair. At 17°C, the shadow enhancer pair has approximately 46% lower total noise than either the single or duplicated proximal enhancer, 21% lower total noise than the single distal enhancer, and is not significantly different than the duplicated distal enhancer (Figure 2.7A). As seen by the variety of shapes in the temperature response curves (Figure 2.7C), temperature perturbations have enhancer-specific effects, suggesting input TFs may differ in their response to temperature change. The low noise driven by the shadow enhancer pair across conditions is consistent with previous studies showing shadow enhancers are required for robust gene expression at elevated and lowered temperatures[22,23].

2.4 Discussion

Fluctuations in the levels of transcripts and proteins are an unavoidable challenge to precise developmental patterning[8,41,42]. Given that shadow enhancers are common and necessary for robust gene expression[19,22,23], we proposed that shadow enhancers may function to buffer the effects of fluctuations in the levels of key developmental TFs. To address this, we have, for the first time, extensively characterized the noise associated with shadow enhancers critical for patterning the early *Drosophila* embryo. By either tracking biallelic transcription or simultaneously measuring input TF levels and transcription, we tested the hypothesis that

shadow enhancers buffer noise through a separation of TF inputs to the individual member enhancers. Our results show that TF fluctuations play a significant role in transcriptional noise and that a shadow enhancer pair is better able to buffer both extrinsic and intrinsic sources of noise than duplicated enhancers. Using a simple mathematical model, we found that fluctuations in TF levels are required to reproduce the observed correlations between reporter activity and that the low noise driven by the shadow enhancer pair may be a natural consequence of the separation of TF inputs to the member enhancers. Lastly, we showed that a shadow enhancer pair is uniquely able to buffer expression noise across a wide range of temperatures. Together, these results support the hypothesis that shadow enhancers buffer input TF noise to drive robust gene expression patterns during development.

2.4.1 Temporal fluctuations in transcription factor levels drive expression noise in the embryo

When measured in fixed embryos, the TFs used in *Drosophila* embryonic development show remarkably precise expression patterns, displaying errors smaller than the width of a single nucleus[10,11,43,44]. It therefore was unclear whether fluctuations in these regulators play a significant role in transcriptional noise in the developing embryo. By measuring the temporal dynamics of the individual *Kr* enhancers, each of which is controlled by different transcriptional activators, we show that TF fluctuations do significantly contribute to the noise in transcriptional output of a single enhancer. Within a nucleus, expression controlled by the two different *Kr* enhancers is far less correlated than expression driven by two copies of the same enhancer, indicating that TF inputs, as opposed to more global factors, are the primary regulators of transcriptional bursting in this system. Our current findings leave open the possibility that additional mechanisms, such as differences in 3D nuclear organization between different reporters, may also contribute to the differences in noise that we see.

We also showed that activity driven by the *Kr* shadow enhancer pair is less sensitive to levels of a single TF than is activity driven by an individual *Kr* enhancer. While prior work has shown that changes in TF levels precede changes in target transcription[45], the sensitivity of individual enhancers to changes in TF levels had not been previously quantified. The correlation between Bcd levels and activity of the distal enhancer is modest, and we expect that this reflects both the influence of additional TF inputs and nuclear heterogeneity that causes the local Bcd levels available to the enhancer to differ from total nuclear levels[46]. We suspect that the correlation between the activity of the distal enhancer and Bcd levels in the microenvironment surrounding the enhancer is higher than what we were able to measure here. New and emerging technologies will likely allow for live measurements of multiple TF inputs at higher spatial resolution, enabling further insights into the dynamics of expression regulation.

The finding that the *Kr* shadow enhancer pair is less sensitive to TF levels helps reconcile our finding that the individual *Kr* enhancers are influenced by fluctuations in input TFs with previous studies showing that endogenous *Kr* expression patterns are rather reproducible[43]. Previous work has cited the role of spatial and temporal averaging, which buffers noisy nascent transcriptional dynamics to generate more precise expression levels. Shadow enhancers operate upstream of this averaging, driving less noisy nascent transcription than either single enhancers or enhancer duplications.

2.4.2 A stochastic model underscores importance of transcription factor fluctuations

We developed a stochastic mathematical model of *Kr* enhancer dynamics and mRNA production that recapitulates our main experimental results. This model is based on that by[36], but it is expanded to include the dynamics of a TF that regulates each enhancer. We placed a strong emphasis on the simplicity of this model, e.g. by using a single abstract TF for each enhancer.

This choice both avoids a combinatorial explosion of parameters and makes the model results and parameters easier to interpret. One of the most notable features of the model is that it recreates the differences in noise between shadow and duplicated enhancer constructs without any additional fitting, indicating that these differences in the model system are a direct result of the separation of input TFs to the proximal and distal enhancers.

Future versions of this model can include refinements. For example, in the current model, we do not include the influence of repressive TFs or consider the multiple modes of action used by activating TFs. Future experiments and models can also be designed to identify the mechanism of enhancer non-additivity: changes in promoter-enhancer looping, saturation of the promoter, or other mechanisms.

2.4.3 Noise source decomposition suggests competition between reporters

In our investigation of sources of noise, we decomposed total noise into extrinsic and intrinsic components as in[40]. In that study, the authors showed that the activity of one reporter did not inhibit expression of the other reporter, and therefore their calculations assumed no negative covariance between the reporters' expression output. In our system, we found a small amount of negative covariance between the activity of two alleles in the same nucleus (Figure S2.10). For this reason, we called our measurements covariance and inter-allele noise. The negative covariance we observe indicates that activity at one allele can sometimes interfere with activity at the other allele, suggesting competition for limited amounts of a factor necessary for reporter visualization. The two possible limiting factors are MCP-GFP or an endogenous factor required for transcription. If MCP-GFP were limiting, we would expect to see the highest levels of negative covariance at the center of the embryo, where the highest number of transcripts are produced and bound by MCP-GFP. Since the fraction of nuclei with negative covariance is

highest at the edges of the expression domain (Figure S2.10), the limiting resource is likely not MCP-GFP, but instead a spatially-patterned endogenous factor, like a TF.

Currently, the field largely assumes that adding reporters does not appreciably affect expression of other genes. However, sequestering TFs within repetitive regions of DNA can impact gene expression[47,48], and a few case studies show that reporters can affect endogenous gene expression[49,50]. If TF competition is responsible for the observed negative covariance between reporters, a closer examination of the effects of transgenic reporters on the endogenous system is warranted. In addition, TF competition may be a feature, not a bug, of developmental gene expression control, as modeling has indicated that molecular competition can decrease expression noise and correlate expression of multiple targets[51].

2.4.4 Additional functions of shadow enhancers and outlook

There are likely several features of shadow enhancers selected by evolution outside of their noise-suppression capabilities. Preger-Ben Noon, et al. showed that all shadow enhancers of *shavenbaby*, a developmental TF gene in *Drosophila*, drive expression patterns in tissues and times outside of their previously-characterized domains in the larval cuticle[52]. This suggests that shadow enhancers, while seemingly redundant at one developmental stage, may play separate, non-redundant roles in other stages or tissues. Additionally, a recent study investigating shadow enhancer pairs associated with genes involved in *Drosophila* embryonic development found that CRISPR deletions of the individual enhancers result in different phenotypes, suggesting each plays a slightly different role in regulating gene expression[53]. In several other cases, both members of a shadow enhancer pair are required for the precise expression pattern generated by the endogenous locus[27,54–57]. These sharpened expression patterns achieved by a shadow

enhancer pair may reflect enhancer dominance or other forms of enhancer-enhancer interaction and are likely another important function of shadow enhancers[54].

In the case of *Kr*, the endogenous expression pattern is best recapitulated by the shadow enhancer pair, with the individual enhancers driving slightly more anterior or posterior patterns of expression[54] (Figure S2.1). Additionally, the early embryonic *Kr* enhancers drive observable levels of expression in additional tissues and time points, but these expression patterns overlap those driven by additional, generally stronger, enhancers, suggesting that the primary role of the proximal and distal enhancers is in early embryonic patterning[58]. Therefore, while we cannot rule out the possibility that the proximal and distal enhancers perform separate functions at later stages, it seems that their primary function, and evolutionary substrate, is controlling *Kr* expression pattern and noise levels during early embryonic development.

Here, we have investigated the details of shadow enhancer function for a particular system, and we expect that some key observations may generalize to many sets of shadow enhancers. Shadow enhancers seem to be a general feature of developmental systems[18,19], but the diversity among them has yet to be specifically addressed. While we worked with a pair of shadow enhancers with clearly separated TF activators, shadow enhancers can come in much larger groups and with varying degrees of TF input separation between the individual enhancers[18,19]. To discern how expression dynamics and noise driven by shadow enhancers depend on their degree of TF input separation, we are investigating these characteristics in additional sets of shadow enhancers with varying degrees of differential TF regulation. Our current results combined with data gathered from additional shadow enhancers will inform fuller models of how developmental systems ensure precision and robustness.

2.5 Materials and Methods

2.5.1 Generation of transgenic reporter fly lines

The single, duplicated, or shadow enhancers were each cloned into the pBphi vector, upstream of the *Kruppel* promoter, 24 MS2 repeats, and a *yellow* reporter gene as in (Fukaya, et al., 2016). We defined the proximal enhancer as chromosome 2R:25224832-25226417, the distal enhancer as chromosome 2R:25222618-25223777, and the promoter as chromosome 2R:25226611-25226951, using the *Drosophila melanogaster* dm6 release coordinates. The precise sequences for each reporter construct are given in Supplemental File 4. For the allele correlation experiments, each enhancer was cloned 192 bp upstream of the *Kr* promoter, separated by the endogenous sequence found between the proximal enhancer and the promoter. For transcriptional noise experiments, the distal enhancer was placed at its endogenous spacing, 2835 bp upstream of the promoter, and the proximal enhancer sequence was replaced by a region of the lambda genome that is predicted to have few relevant TF binding sites. In the shadow enhancer pair or duplicated enhancer constructs, the two enhancers were separated by the sequence separating the proximal and distal enhancers in the endogenous locus.

Using phiC31-mediated integration, each reporter construct was integrated into the same site on the second chromosomes by injection into yw; PBac{y[+]-attP-3B}VK00002 (BDRC stock #9723) embryos by BestGene Inc. (Chino Hills, CA). To produce embryos with biallelic expression of the MS2 reporter, female flies expressing RFP-tagged histones and GFP-tagged MCP (yw; His-RFP/Cyo; MCP-GFP/TM3.Sb) were crossed with males containing one of the enhancer-MS2 reporter constructs. Virgin female F1 offspring were then mated with males of the same parental genotype, except in the case of shadow heterozygous flies, which were mated with males containing the other single enhancer-MS2 reporter.

2.5.2 Sample preparation and image acquisition

Live embryos were collected prior to nc14, dechorionated, mounted on a permeable membrane, immersed in Halocarbon 27 oil, and put under a glass coverslip as in (Garcia, et al., 2013). Individual embryos were then imaged on a Nikon A1R point scanning confocal microscope using a 60X/1.4 N.A. oil immersion objective and laser settings of 40uW for 488nm and 35uW for 561nm. To track transcription, 21 slice Z-stacks, at 0.5um steps, were taken throughout the length of nc14 at roughly 30 second intervals. To identify the Z-stack's position in the embryo, the whole embryo was imaged after the end of nc14 at 20x using the same laser power settings. Later in the analysis, each transcriptional spot's location is described as falling into one of 42 anterior-posterior (AP) bins, with the first bin at the anterior of the embryo. Unless otherwise indicated, embryos were imaged at ambient temperature, which was on average 26.5°C. To image at other temperatures, embryos were either heated or cooled using the Bioscience Tools (Highland, CA) heating-cooling stage and accompanying water-cooling unit.

2.5.3 Burst calling and calculation of transcription parameters

Tracking of nuclei and transcriptional spots was done using the image analysis Matlab pipeline described in Garcia, et al., 2011 (Garcia, et al., 2011). For every spot of transcription imaged, background fluorescence at each time point is estimated as the offset of fitting the 2D maximum projection of the Z-stack image centered around the transcriptional spot to a gaussian curve, using Matlab *lsqnonlin*. This background estimate is subtracted from the raw spot fluorescence intensity. The resulting fluorescence traces across the time of nc14 are then subject to smoothing by the LOWESS method with a span of 10%. The smoothed traces were used to measure transcriptional parameters and noise. Traces consisting of fewer than three time frames were removed from calculations. To calculate transcription parameters, we used the smoothed traces

to determine if the promoter was active or inactive at each time point. A promoter was called active if the slope of its trace (change in fluorescence) between one point and the next was greater than or equal to the instantaneous fluorescence value calculated for one mRNA molecule (F_{RNAP} , described below). Once called active, the promoter is considered active until the slope of the fluorescence trace becomes less than or equal to the negative instantaneous fluorescence value of one mRNA molecule, at which point it is called inactive until another active point is reached. The instantaneous fluorescence of a single mRNA was chosen as the threshold because we reasoned that an increase in fluorescence greater than or equal to that of a single transcript is indicative of an actively producing promoter, while a decrease in fluorescence greater than that associated with a single transcript indicates transcripts are primarily dissociating from, not being produced from, this locus. Visual inspection of fluorescence traces agreed well with the burst calling produced by this method (Figure S2.7).

Using these traces and promoter states, we measured burst size, frequency and duration. Burst size is defined as the integrated area under the curve of each transcriptional burst, from one "ON" frame to the next "ON" frame, with the value of 0 set to the floor of the background-subtracted florescence trace (Figure S2.7 C). Duration is defined as the amount of time occurring between the frame a promoter is determined active and the frame it is next determined inactive (Figure S2.7 F). Frequency is defined as the number of bursts occurring in the period of time from the first time the promoter is called active until 50 minutes into nc14 or the movie ends, whichever is first (Figure S2.7 E). The time of first activity was used for frequency calculations because the different enhancer constructs showed different characteristic times to first transcriptional burst during nc14. For these, and all other measurements, we control for the embryo position of the transcription trace by first individually analyzing the trace and then using

all the traces in each AP bin (anterior-posterior; the embryo is divided into 41 bins each containing 2.5% of the embryo's length) to calculate summary statistics of the transcriptional dynamics and noise values at that AP position.

All Matlab codes used for burst calling, noise measurements, and other image processing are available at the Wunderlich Lab GitHub (https://github.com/WunderlichLab/KrShadowEnhancerCode)

2.5.4 Simultaneous tracking of Bcd-GFP and enhancer activity

To compare the sensitivity of the activity of the shadow pair and distal enhancer to Bcd levels, we tracked the fluorescence of Bcd-GFP and MCP-mCherry in individual nuclei across the time of nc14. To obtain embryos for simultaneous tracking, we crossed female flies heterozygous for Bcd-GFP and MCP-mcherry with male flies homozygous for either the shadow pair or distal enhancer reporter. Bcd-GFP and MCP-mCherry are maternally deposited and thereby allow us to measure levels of Bcd and enhancer activity in individual nuclei of the resulting embryos. Embryo collection and preparation was performed as described above. The same microscope, objective, and Z-step profile were used as described above, but laser settings were switched to 40uW for 561nm and 35uW for 488nm. Analysis of transcriptional activity was performed as described above. Time traces of Bcd-GFP levels in individual nuclei were subjected to background correction by subtracting the average fluorescence of the regions of the image not containing a nucleus at each time point from the raw Bcd-GFP fluorescence. The resulting Bcd-GFP time traces were then subjected to smoothing by the MATLAB *smooth* function, using the LOWESS method with a span of 10%. To measure the sensitivity of enhancer activity to Bcd levels, we correlated the slope of MS2 traces to the corresponding Bcd-GFP levels in the same nucleus. Slope was calculated between the MS2 values at consecutive time

points and compared to the Bcd-GFP value at the earlier of the two time points. This process was done for all time points through 50 minutes into nc14.

2.5.5 Conversion of integrated fluorescence to mRNA molecules

To put our results in physiologically relevant units, we calibrated our fluorescence measurements in terms of mRNA molecules. As in (Lammers, et al., 2020), for our microscope, we determined a calibration factor, α, between our MS2 signal integrated over nc13, F_{MS2},and the number of mRNAs generated by a single allele from the same reporter construct in the same time interval, N_{FISH}, using the *hunchback* P2 enhancer reporter construct (Garcia et al., 2013). Using this conversion factor, we can calculate the integrated fluorescence of a single mRNA (F_1) as well as the instantaneous fluorescence of an mRNA molecule (F_{RNAP}). With our microscope, F_{RNAP} is 379 AU/RNAP and F_1 is 1338 AU/RNAP·min. With these values, we are able to convert both integrated and instantaneous fluorescence into total mRNAs produced and number of nascent mRNAs present at a single time point, by dividing by F_1 and F_{RNAP}, respectively.

2.5.6 Calculation of noise metrics

To calculate the temporal CV each transcriptional spot *i*, we used the formula:

$$CV(i) = \text{standard deviation}(m^i(t))/\text{mean}(m^i(t))$$

where $m^i(t)$ is the fluorescence of spot *i* and time *t.*

We also decomposed the total noise experienced in each nucleus to inter-allele noise and co-variance, analogous to the approach of (Elowitz, et al., 2002).

Inter-allele noise is calculated one nucleus at a time. It is the mean square difference between the fluorescence of the two alleles in a single nucleus:

$$\eta_{IA}^2 = \frac{\langle (m_1(t) - m_2(t))^2 \rangle}{2\langle m_1(t) \rangle \langle m_2(t) \rangle}$$

where $m_1(t)$ is the fluorescence of one allele in the nucleus at time t, and $m_2(t)$ is the fluorescence of the other allele in the same nucleus and the angled brackets indicate the mean across the time of nc14.

Covariance is the covariance of the activity of the two alleles in the same nucleus across the time of nc14:

$$\eta_{CV}^2 = \frac{\langle m_1(t) m_2(t) \rangle - \langle m_1(t) \rangle \langle m_2(t) \rangle}{\langle m_1(t) \rangle \langle m_2(t) \rangle}$$

The inter-allele and covariance values are defined such that they sum to give the total transcriptional noise displayed by the two alleles in a single nucleus.

$$\eta_{tot}^2 = \frac{\langle m_1(t)^2 + m_2(t)^2 \rangle - 2\langle m_1(t) \rangle \langle m_2(t) \rangle}{2\langle m_1(t) \rangle \langle m_2(t) \rangle}$$

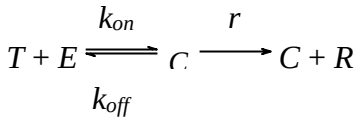
This total noise value is equal to the coefficient of variation of the expression of the two alleles in a single nucleus across the time of nc14.

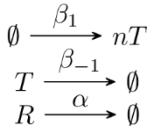
2.5.7 Statistical methods

To determine any significant differences in total noise, covariance, or inter-allele noise values between the different enhancer constructs, we performed Kruskal-Wallis tests with the Bonferroni multiple comparison correction.

2.5.8 Description of the single enhancer model and associated parameters

We constructed a model of enhancer-driven transcription based on the following chemical reaction network,





where *E* is an enhancer that interacts with a transcription factor *T*, which together bind to the promoter at a rate k_{on} to form the active promoter-enhancer complex *C*. When the promoter is in this active form, it leads to the production of mRNA denoted by *R*, which degrades by diffusion from the gene locus at a rate α. Transcription is interrupted whenever the complex *C* disassociates spontaneously at a rate k_{off}. In the bursting TFs model, the transcription factor *T* appears at a rate β_1 and degrades at a rate β_{-1}. To recapitulate *Kruppel* expression patterns, the value of β_1 was assumed to be given by

1. $$f(x) = c\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where *x* is the percentage along the length of the egg and *c* is a scaling constant. Since *Kruppel* activity peaks near the center of the egg, we chose $\mu = 50$, while *c* and σ were fitted along with the other parameters. Lastly, *n* was assumed to be fixed across the length of the egg.

We also generated a constant TF model, which is an adaptation of the model in (Bothma et al., 2015). This model implicitly assumes that TF numbers are constant and, therefore, are incorporated into the value of k_{on} as described by the reactions

$$E \underset{k_{\text{off}}}{\overset{Tk_{\text{on}}}{\rightleftharpoons}} C \xrightarrow{r} C + R$$

$$R \xrightarrow{\alpha} \emptyset$$

In this case, the value for *T* was fitted for each bin in a similar way to β_1, i.e. the constant number of TFs was assumed to be described by equation (1) (values were rounded to the nearest integer).

To simulate the transcriptional traces, we implemented a stochastic approach. Individual chemical events such as enhancer-promoter looping take place at random times and are influenced by transcription factor numbers. Individual trajectories of chemical species over time were calculated using the Gillespie algorithm (Gillespie, 1976), and these trajectories are comparable to the experimentally measured transcriptional traces. Since the enhancer is either bound or not bound to the promoter, we imposed the constraint that $C + E = 1$ when simulating model dynamics.

2.5.9 Estimation of model parameters from experimental data

To yield a starting estimate for the k_{on} and k_{off} parameters, we defined the start and end of a burst as the time when the reactions $E\ k_{on} \rightarrow\ C$ and $C\ k_{off} \rightarrow\ E$ occur, respectively. The length of the *i*th burst was defined as the range of $[b_i,p_i]$ where b_i corresponds to the time of the i^{th} instance of the reaction $E\ k_{on} \rightarrow\ C$ and p_i to the time of the i^{th} instance of the reaction $C\ k_{off} \rightarrow\ E$. The time between the i^{th} burst and the $i + 1^{\text{th}}$ burst is $[p_i,b_{i+1}]$. The Gillespie algorithm dictates that the time spent in any given state is determined by an exponentially distributed random variable with a rate parameter equal to the product of two parts: the sum of rate constants of the outgoing reactions, and the number of possible reactions. If the enhancer is either bound or unbound, we

have that $C = 1$ or $E = 1$, respectively. Therefore, by letting t_b be the average time between bursts and t_d be the average duration of a burst, we can write

$$t_b = \lim_{M\to\infty} \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{N-1}\sum_{i=1}^{N-1}(b_{i+1_j} - p_{i_j})\right) = \frac{1}{k_{\text{on}}ET} \approx k$$

and

$$\lim_{M\to\infty} \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{N}\sum_{i=1}^{N}(p_{i_j} - b_{i_j})\right) = \frac{1}{k_{\text{off}}C} = k$$

where N is the number of bursts for spot j, b_{ij} and p_{ij} denote the beginning and end of burst i in spot j respectively, and M denotes the total number of spots in the egg. The right-hand sides are given by the expected value of the exponential distribution and the assumption that, on average, T is close to 1. While this may not be the case for T, the assumption provides a convenient upper bound for the average time between bursts, which is likely not to have a much smaller value for a lower bound. (A low enough value of t_b would imply nearly constant fluorescence intensity instead of bursts.) Finally, the average duration of a burst t_d can be calculated directly from the data and used to obtain k_{off} by calculating $1/t_d$. Similarly, the average time between bursts t_b is readily available from the data giving us $k_{\text{on}} \approx 1/t_b$.

We were able to directly estimate mRNA production and degradation rates from the experimental data. To estimate α, we focused on periods of mRNA decay; i.e. periods where no active transcription is taking place and are thus described by

$$R' = -\alpha R,$$

which in turn can be solved to be

2. $R = ce^{-t\alpha}$,

where c is a constant of integration. Taking the derivative of equation 2 yields

3. $R'(t) = -\alpha ce^{-t\alpha}$,

which corresponds to the slope of the decaying burst. We define the interval of decay of the i^{th} burst as $[p_i,b_{i+1}]$. For some point $t_0 \in (p_i,b_{i+1})$, let $R_0 = R(t_0) = ce^{-t0\alpha}$. Solving this expression for c gives that $c = R_0e^{t0\alpha}$. Substituting for c in equation 3 evaluated at t_0 results in $R'(t_0) = -\alpha R_0e^{t0\alpha}e^{-t0\alpha} = -\alpha R_0$. Then, it follows that

(4) $$\alpha = -\frac{R'(t_0)}{R_0}.$$

In other words, the rate of decay of mRNA fluorescence can be calculated from any trace by taking the ratio of the slope during burst decay and its intensity at a given time $t_0 \in (p_i,b_{i+1})$.

Adjacent measurements of fluorescence intensity from the single enhancer systems were used to approximate the slope at each point in the traces. Then, equation 4 was applied to each point. A histogram of all calculated values was generated (Figure S2.6). In this figure, there was a clear peak, which provided us with an estimate of $\alpha \approx 1.95$.

The estimation of r was done for periods of active transcription, which are also accompanied by simultaneous mRNA decay. By noting that $C = 1$ during mRNA transcription, we can approximate these periods as the zeroth order process

$$\emptyset \underset{\alpha}{\overset{r}{\rightleftharpoons}} R$$

The differential equation associated with this system is given by

5. $R' = r - \alpha R,$

and has steady state $R^* = r/\alpha$. Equation 5 can be solved explicitly for R by choosing

$$R(t) = \frac{r}{\alpha} + ce^{-t\alpha},$$

where c is a constant of integration. For two adjacent measurements at times t_1 and t_2 we can write their respective measured amounts of mRNA as

6. $$R_1 = \frac{r}{\alpha} + c_1 e^{-t_1\alpha},$$

and

7. $$R_2 = \frac{r}{\alpha} + c_2 e^{-t_2\alpha}.$$

Solving for c_1 and c_2 gives

$$c_1 = (R_1 - \frac{r}{\alpha})e^{t_1\alpha}$$

$$c_2 = (R_2 - \frac{r}{\alpha})e^{t_2\alpha}.$$

The short-term fluctuations of mRNA from R_1 to R_2 between two adjacent discrete time points in the stochastic system can be approximated by equations 6 and 7. This implies that

$$(R_1 - \frac{r}{\alpha})e^{t_1\alpha} = (R_2 - \frac{r}{\alpha})e^{t_2\alpha},$$

which in turn gives

$$r = \alpha \frac{R_1 - R_2 e^{\alpha \Delta t}}{1 - e^{\alpha \Delta t}}.$$

Therefore, the estimation of *r* can be computed given two adjacent measurements of fluorescence and the time between them. Finally, we used a similar approach as done with α to calculate values of *r* from fluorescence data. However, unlike α, *r* was calculated for each bin to account for differences in transcriptional efficiency across the length of the embryo.

2.5.10 Parameter fitting with simulated annealing

Simulations and parameter fitting were done with MATLAB®. Optimization in fitting was done by minimizing the sum of squared errors (SSE) between the normalized vectors of burst properties and allele correlations of the experimental and simulated data. In particular, a vector *y* of experimental data was created by concatenating the following vectors: burst size, integrated fluorescence, frequency, duration, and allele correlation across the length of the embryo. The vector *y* was subsequently normalized by dividing each burst property by the largest element in their respective vectors (except correlation which by definition is unitless between -1 and 1). A vector *x* was created in an analogous fashion to *y* but using simulated instead of experimental data. However, *x* was normalized using the same elements that were used to normalize *y*. Then, the discrepancy between the experimental and simulated data was measured by

$$SSE = \sum_{i=1}^{n} (y_i - x_i)^2 .$$

We used a high-performance computing cluster to compute 200 independent runs of parameter fitting with simulated annealing for each model variant. The algorithm requires an initial guess of the parameter set P_0, an initial temperature Γ_0, a final temperature Γ', the number of iterations per temperature *N*, and a cooling factor μ. Then, each iteration is as follows:

1. If the current iteration i is such that $i > N$, then update the current temperature $\Gamma_k = \mu^k\Gamma_0$ to $\mu^{k+1}\Gamma_0$ and set $i = 0$. Otherwise, set i to $i + 1$.
2. Check if $\Gamma_k < \Gamma'$. If so, return the current parameter set P_j and terminate.
3. Choose a parameter randomly from P_j and multiply it by a value sampled from a normal distribution with a mean equal to 1. The standard deviation of such distribution should be continuously updated to be Γ_k. The result of this step is the newly generated parameter set P_{j+1}.
4. Calculate ΔE as the difference in SSE between the data generated by P_j and that generated by P_{j+1}. Update P_j to P_{j+1} if $\Delta E < 0$ or with probability $p < e^{\Delta E/\Gamma k}$ where p is a uniformly distributed random number.
5. Repeat all steps until termination.

To generate our results, we chose $\Gamma_0 = 1$, $\Gamma' = \Gamma_0/10$, $N = 30$, and $\mu = 0.8$. We observed an improvement in the quality of the fittings by using analysis-derived parameter values as initial guesses instead of values given through random sampling. The sampled space ranged from 10^{-3} to 10^3 for all parameters, except n, which was sampled from 10^0 to 10^2, and σ, which was randomly chosen to be an integer between 1 and 20. Equal numbers of parameter values were sampled at each order of magnitude. The analysis in the section above was used to estimate the parameters in P_0. Parameters that were not estimated in the previous section were given the following initial guesses: $n = 10$, $\beta_{-1} = 1$, $\sigma = 6$, and $c = 40$. Initial guesses for c and σ were based on the experimental observation that there is little transcription outside of 20-80% egg length. Based on this observation, simulations were limited to this egg length range, as well. For the constant TFs model, both analysis-derived and random initial parameter values were used to

maximize the likelihood of finding any parameter set capable of recapitulating the observed allele correlation.

2.5.11 Generation of simulated experimental data

Parameter sets resulting from fitting were sorted in ascending order based on their sum of squared errors, and the 10 lowest error parameter sets are what we called the 10 best parameter sets. For all figures, we simulated 80 spots per bin and simulated each bin 5 times to generate error bars. Data for the distal enhancer at the proximal location was used to reproduce simulated allele correlations in all cases.

Gillespie simulations update the counts of each chemical species at random time intervals. However, for ease of parameter fitting and to better recapitulate the experiments, we generated data in two distinct timescales: one consisting of 30 second intervals after which mRNA counts were recorded, and another consisting of random time intervals generated by the algorithm after which chemical counts were updated. The former one was used for all parameter fitting rounds and generation of figures.

2.5.12 Description of two enhancer model, parameter estimation, and fitting

To explore two enhancer systems, we expanded our previous model to include an additional enhancer. First, we considered duplicated enhancer systems, which consist of either two proximal or two distal enhancers. Enhancers were denoted by E_1 and E_2, which correspond to two identical enhancers that exist in different locations relative to the promoter. They are activated by the same transcription factors as described by the reactions

$$T + E_1 \underset{k_{\mathrm{off}_1}}{\overset{k_{\mathrm{on}_1}}{\rightleftharpoons}} C_1 \xrightarrow{r_1} C_1 + R$$
$$T + E_2 \underset{k_{\mathrm{off}_2}}{\overset{k_{\mathrm{on}_2}}{\rightleftharpoons}} C_2 \xrightarrow{r_2} C_2 + R$$
$$\emptyset \xrightarrow{\beta_1} nT$$
$$T \xrightarrow{\beta_{-1}} \emptyset$$
$$R \xrightarrow{\alpha} \emptyset$$

Without loss of generality, we used E_1 to denote the enhancer at the proximal location and E_2 to denote the enhancer at the distal location. This model describes independent enhancer dynamics; i.e. the behavior of one enhancer does not affect the behavior of the other, and, as such, both enhancers can be simultaneously looped to the promoter. Consequently, to account for potential enhancer interference or competition for the promoter, we assumed distinct k_{on} and k_{off} values for each enhancer in the duplicated enhancer constructs. We also used distinct values of r for each distal enhancer in the duplicated distal construct since fluorescence data was available for this enhancer at the proximal and endogenous location. For proximal enhancers, we assume $r_1 = r_2$.

To describe the dynamics of the shadow enhancer pair, we denoted the activators for E_1 (the proximal enhancer) and E_2 (the distal enhancer) by T_1 and T_2, respectively:

$$T_1 + E_1 \underset{k_{\mathrm{off}_1}}{\overset{k_{\mathrm{on}_1}}{\rightleftharpoons}} C_1 \xrightarrow{r_1} C_1 + R$$
$$T_2 + E_2 \underset{k_{\mathrm{off}_2}}{\overset{k_{\mathrm{on}_2}}{\rightleftharpoons}} C_2 \xrightarrow{r_2} C_2 + R$$
$$\emptyset \xrightarrow{\beta_1} n_1 T_1$$
$$\emptyset \xrightarrow{\gamma_1} n_2 T_2$$
$$T_1 \xrightarrow{\beta_{-1}} \emptyset$$
$$T_2 \xrightarrow{\gamma_{-1}} \emptyset$$
$$R \xrightarrow{\alpha} \emptyset$$

The production rate of T_2, γ_1, was calculated in the same way as production rate of T_1, β_1, but differed in the values of c and σ. The two enhancer models were also used to calculate allele correlation between homozygotes and heterozygotes because a distinction between the mRNA

produced by C_1 and C_2 was made. This approach works because, e.g., when considering the homozygote embryos, each single enhancer resides in the same nucleus and is therefore affected by the same fluctuating TF numbers. In the duplicated enhancer model, each enhancer E_1 or E_2 is affected by the same fluctuations in the number of transcription factor T. An analogous logic applies to the heterozygotes.

To fit the two enhancer models to experimental data, we retained several parameters from the single enhancer models. Parameters r and α were directly calculated from the data, and, as such, did not vary across models. We assume that parameters concerning transcription factors (β_1, β_{-1}, γ_1, γ_{-1}, n_1, and n_2) are not affected by the presence of an additional enhancer. Therefore, in our model, only k_{on} and k_{off} are free to change. To fit the values of k_{on1}, k_{on2}, k_{off1}, and $k_{off2,}$ we set the other model parameters to the median values of the 10 best parameter sets in the respective single enhancer model. We then used a similar simulating annealing approach to fit the k_{on} and k_{off} values. We used the resulting values to simulate transcriptional traces and to calculate the predicted CV values shown in Figure 2.5.

2.5.13 Theoretical modeling of inter-allele noise

To make a prediction about the expected change in inter-allele noise between single and two enhancer reporter constructs, we used the theory put forth in (Sánchez and Kondev, 2008; Sanchez et al., 2011). This formalism can be used to calculate the expected mean and variance of the transcriptional output of a promoter, given the possible states of the promoter, transition rates between states, and the rate of transcription resulting from each state. In these papers, the authors apply their formalism to different promoter architectures. Here, we generate a simpler model, in which we abstract away the individual transcription factor (TF) binding configurations, which would be numerous and poorly parametrized, and simply define states by whether an enhancer is

looped to the promoter and activating transcription. Since these models do not account for fluctuations that would contribute to extrinsic noise, e.g. fluctuations in TF or RNA polymerase levels, they can predict the dependence of intrinsic noise on enhancer arrangement.

To apply this model to our system, we used these parameters: γ, degradation rate of mRNA; *p*, production rate of mRNA; *k*, on rate for enhancer-promoter looping; *l*, off rate for enhancer-promoter looping. We generated 5 models that represent different configurations of either one or two enhancers controlling a single promoter. Key assumptions are that the parameters describing this system are independent of both the position of the enhancer relative to the promoter and the presence of a second enhancer controlling the same promoter. In Model 1, there is a single enhancer controlling one promoter. There are two states, when the enhancer and promoter are not looped (mRNA production rate of 0), and when the enhancer and promoter are looped (mRNA production rate of *p*). In Model 2 (OR model) there are two enhancers controlling one promoter, transcription is activated if either enhancer is looped, and both enhancers can't be bound at the same time. In Model 3 (additive model), there are two enhancers controlling one promoter, transcription is activated if either enhancer is looped, and, if both enhancers are bound, transcription occurs at twice the rate of single enhancer looping states. In Model 4 (synergistic model), there are two enhancers controlling one promoter, transcription is activated if either enhancer is looped, and, if both enhancers are bound, transcription occurs at three times the rate of single enhancer looping states. In Model 5 (XOR model), there are two enhancers controlling one promoter, transcription is activated if either enhancer is looped, and, if both enhancers are bound, no transcription occurs. Results of these models are shown in Figure S2.11.

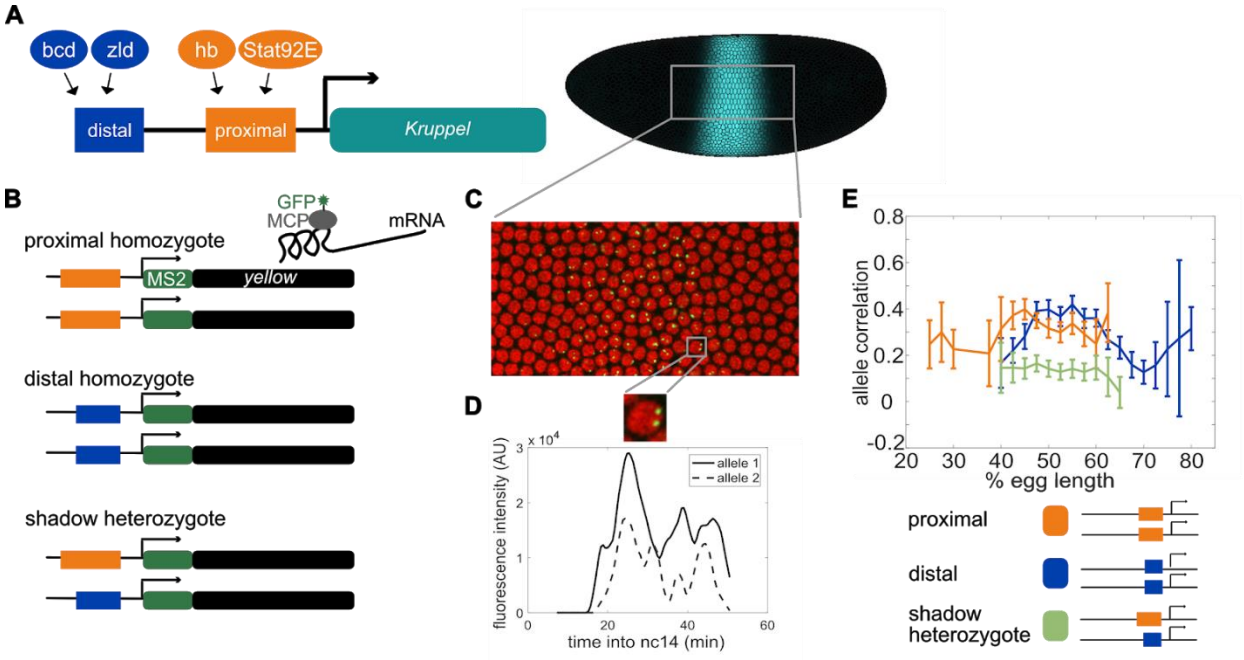


Figure 2.1: Dual allele imaging shows the individual *Kruppel* enhancers drive largely independent transcriptional dynamics. A. Schematic of the endogenous *Kruppel* locus with distal (blue) and proximal (orange) shadow enhancers driving *Kr* (teal) expression in the central region of the embryo. Known transcriptional activators of the two enhancers are shown. **B**. Schematics of single enhancer reporter constructs driving expression of MS2 sequence and a *yellow* reporter. When transcribed, the MS2 sequence forms stem loops that are bound by GFP-tagged MCP expressed in the embryos. Proximal embryos have expression on each allele controlled by the 1.5kb proximal enhancer at its endogenous spacing from the *Kr* promoter, while distal embryos have expression on each allele controlled by the 1.1kb distal enhancer at the same spacing from the *Kr* promoter. Shadow heterozygote embryos have expression on one allele controlled by the proximal enhancer and expression on the other allele controlled by the distal enhancer. **C**. Still frame from live imaging experiment where nuclei are red circles and active sites of transcription are green spots. MCP-GFP is visible as spots above background at sites of nascent transcription (Garcia, et al., 2013). **D.** The fluorescence of each allele in individual nuclei can be tracked across time as a measure of transcriptional activity. Graph shows a representative trace of transcriptional activity of the two alleles in a single nucleus across the time of nc14. These traces are used to calculate the Pearson correlation coefficient between the transcriptional activity of the two alleles in a nucleus across the time of nc14. Correlation values are grouped by position of the nucleus in the embryo and averaged across all imaged nuclei in all embryos of each construct. **E**. Graph of average correlation between the two alleles in each nucleus as a function of egg length. 0% egg length corresponds to the anterior end. Error bars indicate 95% confidence intervals. The shadow heterozygotes have much lower allele correlation than either homozygote, demonstrating that the individual shadow enhancers drive nearly independent transcriptional activity and that upstream fluctuations in regulators are a significant driver of transcriptional bursts. The total number of nuclei used in calculations for each construct by anterior-posterior (AP) bin are given in Supplemental file 1.

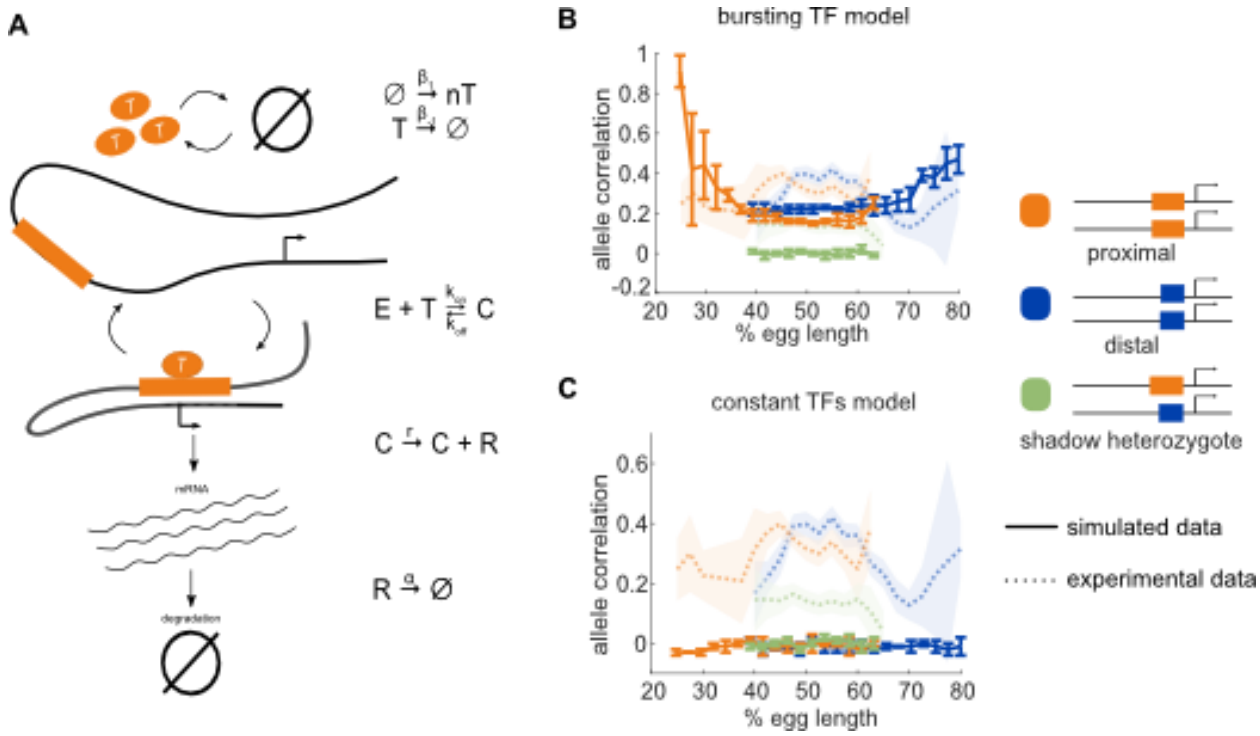
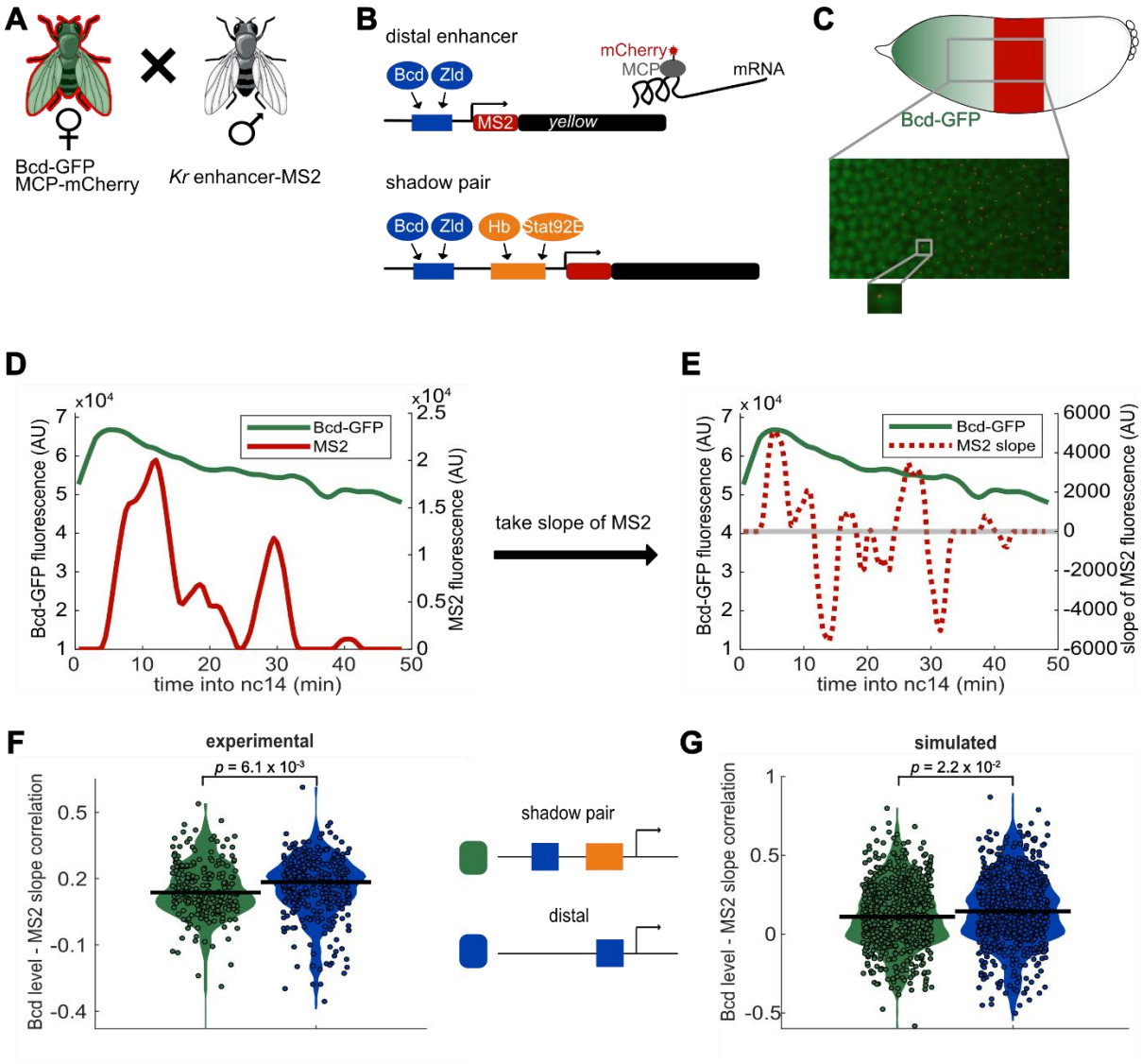


Figure 2.2: Model of enhancer-driven dynamics demonstrates TF fluctuations are required for correlated reporter activity. To investigate the factors required for the observed correlated behavior of identical enhancers and largely independent behavior of the individual enhancers, we developed a simple stochastic model of enhancer-driven transcription. **A**. Schematic of model of transcription driven by a single enhancer (the *bursting TFs* model). For each enhancer, we assume there is a single activating TF, T_i, that appears in bursts of size n_i molecules at a rate β_i, which varies by the position in the embryo. TFs degrade linearly at rate β_{-1}. When present, T_i can bind the enhancer, E_i, to form a transcriptionally active complex, C_i, at a rate k_{on} and dissociates at rate k_{off}. This complex then produces mRNA at an experimentally determined rate r that degrades at an experimentally determined rate, α. **B**. The bursting TFs model is able to recapitulate the experimentally observed pattern of allele correlation. We plot the correlation between the two alleles in a nucleus as a function of egg length. Simulated data is created using the lowest energy parameter set for each enhancer. The data shown is the average of five simulated embryos that have 80 transcriptional spots per AP bin. In B and C simulated data are shown by solid lines, experimental data are shown by dotted lines. **C**. The constant TFs model fails to recapitulate the experimentally observed pattern of allele correlation. Without TF fluctuations, both heterozygous and homozygous embryos display independent allele activity. Error bars and shaded regions in B and C represent 95% confidence intervals.

Figure 2.3: Activity of *Kr* shadow pair is less correlated with Bcd levels than is activity of single distal enhancer.

To assess whether fluctuations in enhancer activity across time are associated with fluctuations in TF levels, we simultaneously measured Bcd levels and enhancer-driven transcription in individual nuclei. **A**. To track Bcd levels and enhancer activity in the same nuclei, we crossed flies expressing a *Kr* enhancer-MS2 transgene to flies expressing Bcd-GFP and MCP-mCherry. In the resulting embryos, Bcd levels can be measured by GFP fluorescence and enhancer reporter activity can be measured by mCherry fluorescence. **B**. Schematic of the enhancer reporters used for simultaneous tracking of TF levels and enhancer activity. As in Figure 2.1, the transcribed MS2 sequence forms stem loops that are bound by MCP, which is here tagged with mCherry. **C**. Bcd-GFP expression forms a gradient from the anterior to posterior of the embryo, whereas the *Kr* enhancer reporters drive expression in the center region of the embryo. The magnified section of the embryo shows a still frame from live imaging indicating nuclei (green) and active transcription spots (red). **D**. Bcd levels and enhancer activity can be simultaneously tracked in individual nuclei. Graph shows a representative trace of Bcd-GFP levels (in green) and distal enhancer transcriptional activity (in red) in a nucleus across the time of nc14. **E.** Activator TF levels regulate enhancer activity, so to assess the sensitivity of our enhancer constructs to input TF fluctuations, we compare the levels of nuclear Bcd-GFP to the slope of MS2 fluorescence across the time of nc14. Positive slope values indicate an increase in enhancer activity while negative values indicate a decrease in enhancer activity. The graph shows nuclear Bcd-GFP levels (as in D), in solid green line, and MS2 slope values (of the MS2 trace shown in D), in dashed red line, across the time of nc14. Horizontal grey line indicates a slope value of 0. **F**. Changes in the shadow pair's activity are significantly less correlated with Bcd-GFP levels than are changes in the distal enhancer's activity. Shown are violin plots of the distribution of correlation values between Bcd-GFP levels and MS2 slopes in individual nuclei for either the shadow pair or distal enhancer. Circles correspond to the correlation values of individual nuclei and the horizontal lines indicate the median. This correlation is significantly higher for the distal enhancer than it is for the shadow pair (median r values are 0.18 and 0.14, respectively. *p*-value = 6.1×10^{-3} from Kruskal-Wallis pairwise comparison.) The total number of nuclei used in calculations for each construct by AP bin are given in Supplemental file 2. **G**. Our enhancer model recapitulates the lower correlation between Bcd-GFP levels and enhancer activity seen with the shadow pair than with the distal enhancer. Graph is as in F, but showing the distribution of correlation values in simulated nuclei, using 100 nuclei per AP bin. Median r values for simulation are 0.14 for the distal enhancer and 0.11 for the shadow pair. *p*-value = 2.2×10^{-2} from Kruskal-Wallis pairwise comparison of correlations.


A
Bcd-GFP
MCP-mCherry
Kr enhancer-MS2
B
distal enhancer
Bcd
Zld
MS2
yellow
mCherry
MCP
mRNA
shadow pair
Bcd
Zld
Hb
Stat92E
C
Bcd-GFP
D
Bcd-GFP
MS2
Bcd-GFP fluorescence (AU)
MS2 fluorescence (AU)
time into nc14 (min)
take slope of MS2
E
Bcd-GFP
MS2 slope
Bcd-GFP fluorescence (AU)
slope of MS2 fluorescence (AU)
time into nc14 (min)
F
experimental
p = 6.1 x 10-3
Bcd level - MS2 slope correlation
shadow pair
distal
G
simulated
p = 2.2 x 10-2
Bcd level - MS2 slope correlation

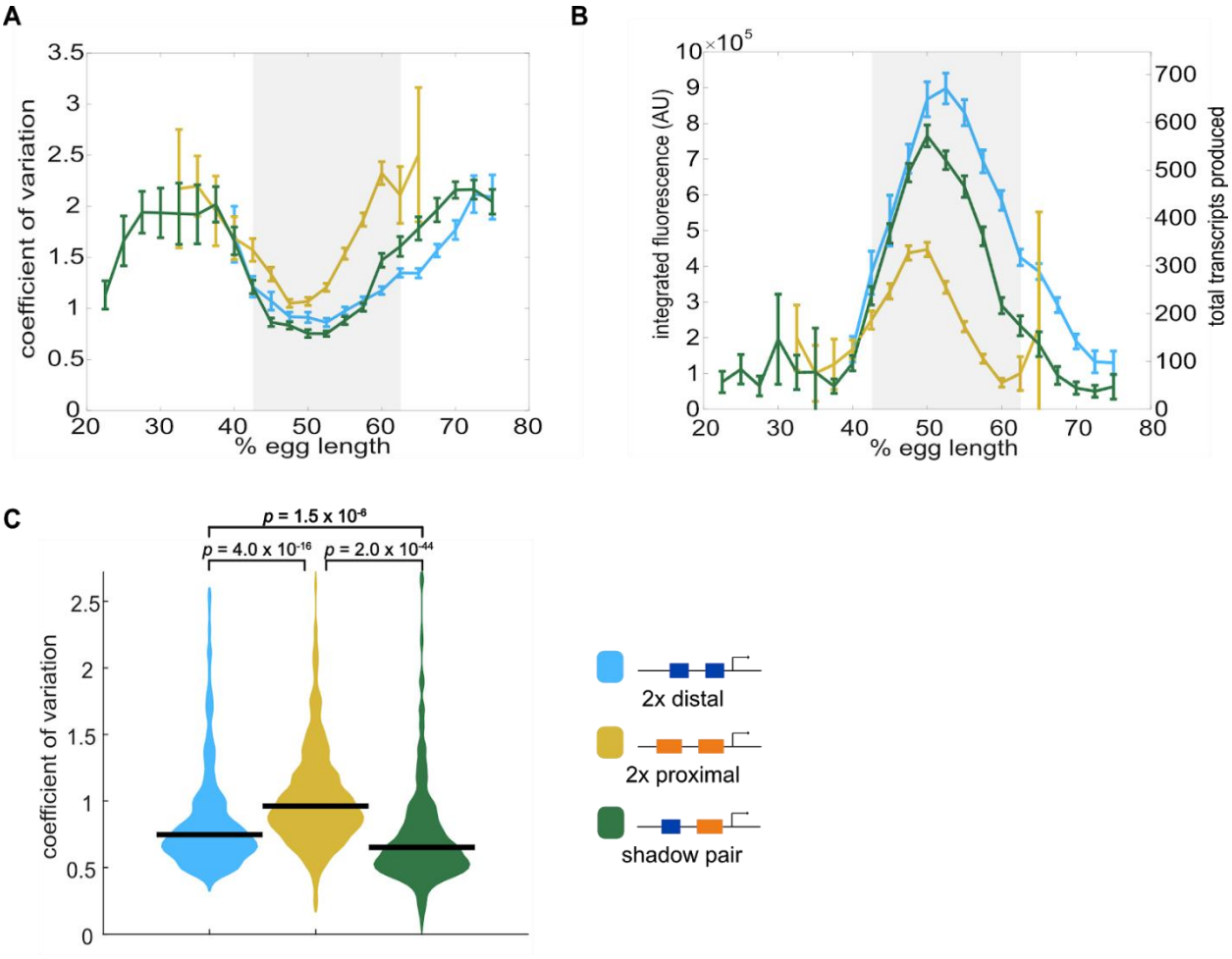


Figure 2.4: Shadow enhancer pair produces lower expression noise than duplicated enhancers. To investigate whether the shadow enhancer pair drives less noisy expression, we calculate the coefficient of variation (CV) associated with the shadow enhancer pair or either duplicated enhancer across time of nc14. **A**. The shadow enhancer pair displays lower temporal expression noise than either duplicated enhancer. Graph is mean coefficient of variation of fluorescence traces across time as a function of embryo position. The grey rectangle in A and B highlights the region of endogenous *Kr* expression (boundaries where 33% maximal expression occurs). **B**. The shadow enhancer pair shows the lowest expression noise, but not the highest expression levels, indicating that the lower noise is not simply a function of higher expression. Graph is average total expression during nc14 as a function of embryo position. Error bars in A and B represent 95% confidence intervals. Total number of transcriptional spots used for graphs are given in Supplemental file 3 by construct and AP bin. **C**. Violin plot of distribution of CV values at AP bin of peak expression for each enhancer construct (corresponding to 50% egg length for shadow pair and duplicated proximal, 52.5% egg length for duplicated distal), horizontal bar indicates median. Y-axis limited to 99th percentile of the construct with highest expression noise (duplicated proximal). The shadow pair drives significantly lower expression noise than either duplicated enhancer (*p*-value = 1.5 x10^{-6} for duplicated distal and shadow pair. *p*-value = 2.0 x10^{-44} for duplicated proximal and shadow pair). *p*-values calculated using Kruskal Wallis pairwise comparison with Bonferroni multiple comparison correction.

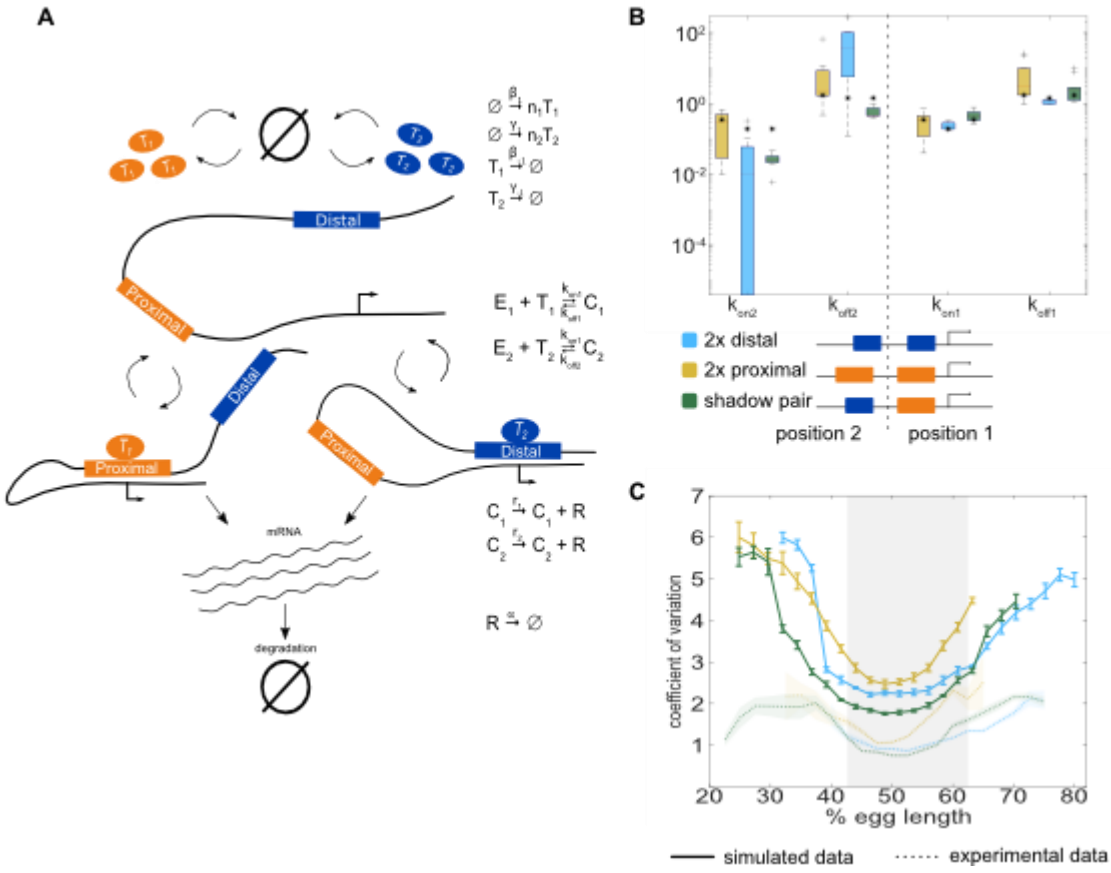
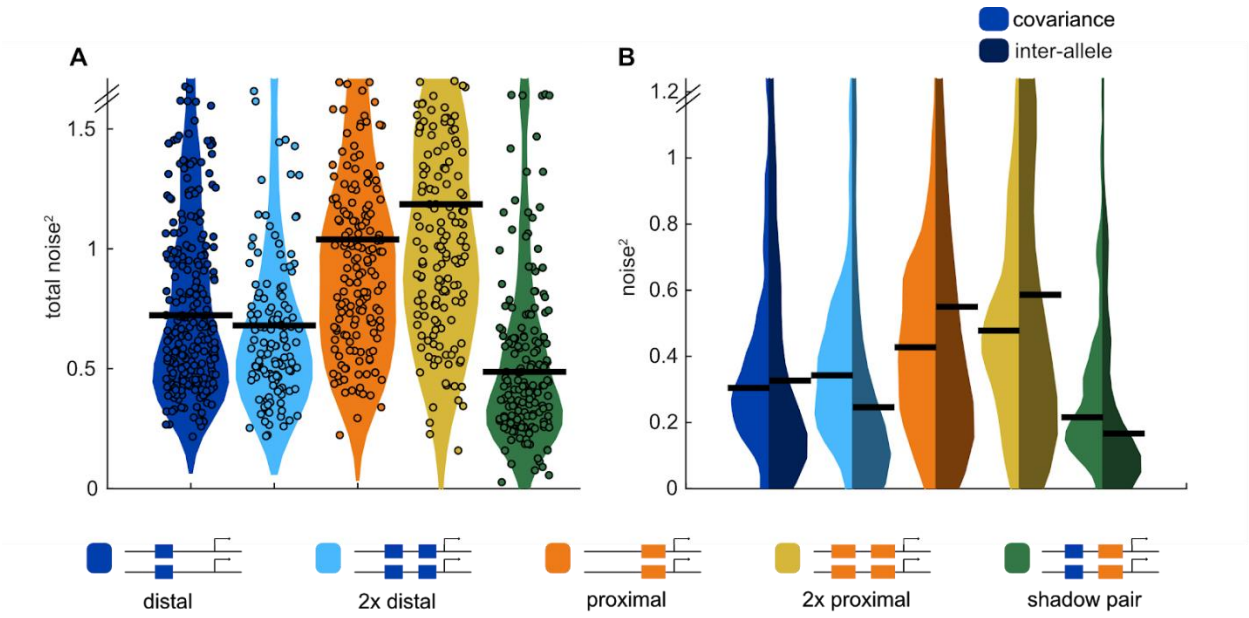
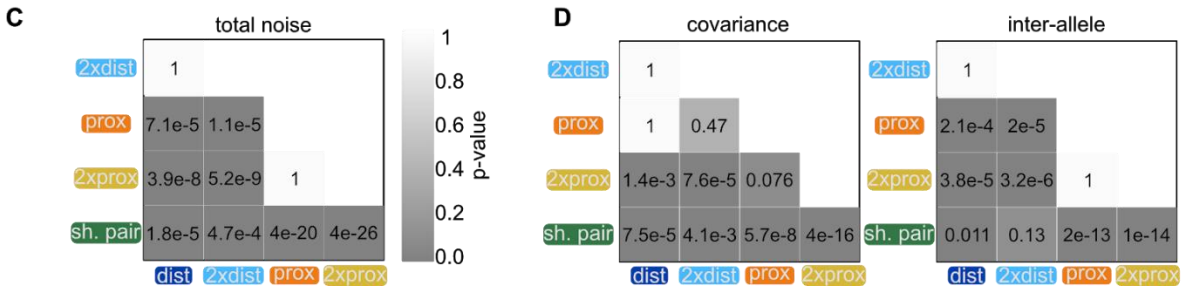


Figure 2.5: The two enhancer model recapitulates low expression noise associated with the shadow enhancer pair. To assess whether the separation of input TFs mediates the lower expression noise driven by the shadow enhancer pair, we expanded our model to incorporate two enhancers driving transcription. **A**. Schematic of the two enhancer model. We assume that when two enhancers control a single promoter, either or both can loop to the promoter and drive transcription. We defined model parameters as in Figure 2.2, and only allowed the k_{on} and k_{off} values to vary from the single enhancer model. **B**. To understand the effect of adding a second enhancer, we examined how the k_{on} and k_{off} values vary from those in the single enhancer model. We plotted the distribution of the values for k_{on} and k_{off} for each enhancer in the three different constructs measured. The distribution shows the values derived from the 10 best-fitting parameter sets, and the black star in each column indicates the k_{on} or k_{off} value from the corresponding single enhancer model. In general, the k_{off} values increased relative to the single enhancer model, and the k_{on} values decreased, indicating that the presence of a second enhancer inhibits the activity of the first. **C.** Graph of average coefficient of variation of simulated (solid lines) or experimental (dotted lines) transcriptional traces as a function of egg length. The model is able to recapitulate the lower expression noise seen with the shadow enhancer pair with no additional fitting, indicating that the separation of TF inputs to the two enhancers is sufficient to explain this observation. Error bars of simulated data and shaded region of experimental data indicate 95% confidence intervals.

Figure 2.6: Shadow enhancer pair achieves lower total noise by buffering global and allele-specific sources of noise. To determine how the shadow enhancer pair produces lower expression noise, we calculated the total noise associated with each enhancer construct and decomposed this into the contributions of covariance and inter-allele noise. Covariance is a measure of how the activities of the two alleles in a nucleus change together and is indicative of global sources of noise. Inter-allele noise is a measure of how the activities of the two alleles differ and is indicative of allele-specific sources of noise. **A.** The shadow enhancer pair has lower total noise than single or duplicated enhancers. Circles are total noise values for individual nuclei in AP bin of peak expression for the given enhancer construct. Horizontal line represents the median. The y-axis is limited to 75th percentile of the proximal enhancer, which has the largest noise values. The shadow enhancer pair has significantly lower total noise than all other constructs. **B.** The shadow enhancer pair displays significantly lower covariance than either single or duplicated enhancer and significantly lower inter-allele noise than both single enhancers and the duplicated proximal enhancer. The left half of each violin plot shows the distribution of covariance values of nuclei in the AP bin of peak expression, while the right half shows the distribution of inter-allele noise values. Horizontal lines represent median. The y-axis is again limited to the 75th percentile of enhancer with the largest noise values, which is duplicated proximal. The lower covariance and inter-allele noise associated with the shadow enhancer pair indicates it is better able to buffer both global and allele-specific sources of noise. **C.** *p*-value table of Kruskal-Wallis pairwise comparison of the total noise values of each enhancer construct. *p*-value gradient legend applies to C and D. **D.** *p*-value table of Kruskal-Wallis pairwise comparison of covariance (on left) and inter-allele noise (on right) values for each enhancer construct. Bonferroni multiple comparison corrections were used for *p*-values in C and D. Total number of nuclei used in noise calculations are given in Supplemental file 1.


A
B
covariance
inter-allele
total noise²
noise²
distal
2x distal
proximal
2x proximal
shadow pair


C
total noise
2xdist
prox
2xprox
sh. pair
dist
p-value
D
covariance
inter-allele

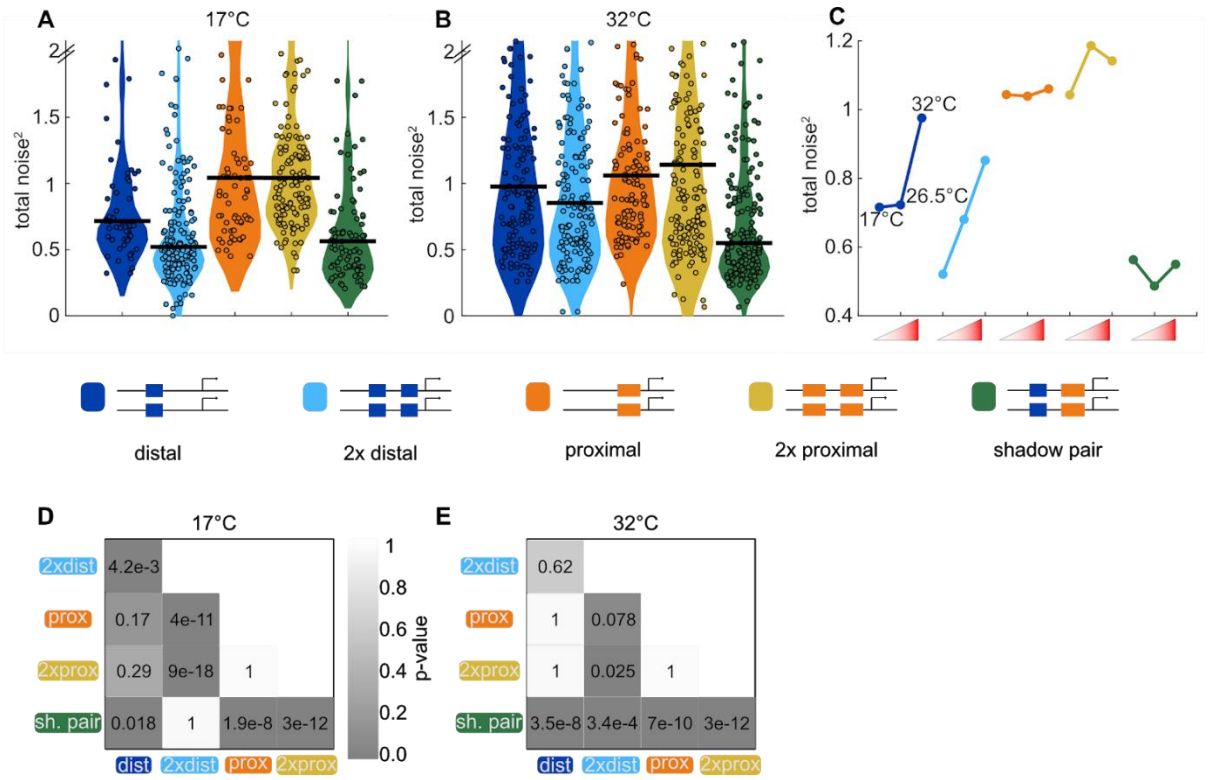


Figure 2.7: Shadow enhancer pair maintains lower total noise across temperature perturbations. To test the ability of each enhancer construct to buffer temperature perturbations, we measured the total expression noise associated with each for embryos imaged at 17°C or 32°C. **A**. The shadow enhancer pair displays significantly lower total noise than the single or duplicated proximal enhancer and the single distal enhancer at 17°C. Circles are total noise values for individual nuclei in AP bin of peak expression for the given enhancer construct and horizontal bars represent medians. The y-axis is limited to 75th percentile of construct with highest total noise at 17°C (single proximal). **B**. The shadow enhancer pair has significantly lower total noise than all other constructs at 32°C. The y-axis is limited to 75th percentile of the enhancer construct with highest total noise at 32°C (duplicated proximal). **C.** Temperature changes have different effects on the total noise associated with the different enhancers. The median total noise value at the AP bin of peak expression at the three measured temperatures is shown for each enhancer construct. Within each enhancer, the median total noise values are shown left to right for 17°C, 26.5°C, and 32°C. **D**. *p*-value table of Kruskal-Wallis pairwise comparison of the total noise values of each enhancer construct at 17°C. *p*-value gradient legend applies to D and E. **E**. *p*-value table of Kruskal-Wallis pairwise comparison of the total noise values of each enhancer construct at 32°C. Bonferroni multiple comparison corrections were used for *p*-values in D and E.

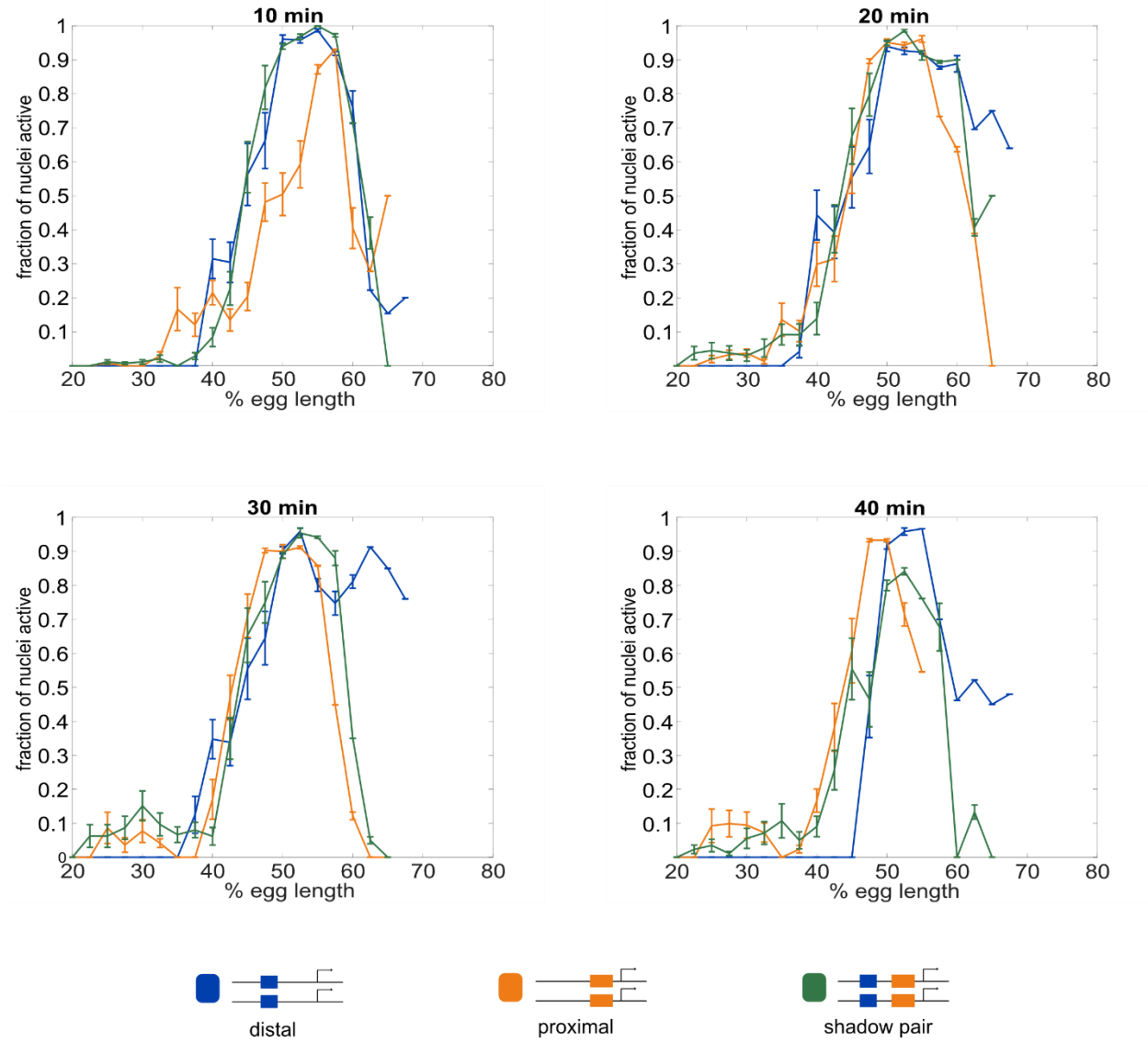


Figure S2.1: Fraction of nuclei transcribing as a function of embryo position. The different enhancer constructs display different spatial and temporal patterns of activity. Shown in all graphs are the fraction of nuclei actively transcribing as a function of embryo position at each indicated time point into nc14. **A**. 10 minutes into nc14. **B**. 20 minutes into nc14. **C**. 30 minutes into nc14. **D**. 40 minutes into nc14. Error bars are 95% confidence intervals. We note that differences in the individual *Kr* enhancers become more pronounced throughout progression of nc14. The more anterior pattern driven by the proximal enhancer in the second half of nc14 mimics the anterior shift previously observed for the *Kr* expression domain (Jaeger, et al., 2004; El-Sherif & Levine, 2016).

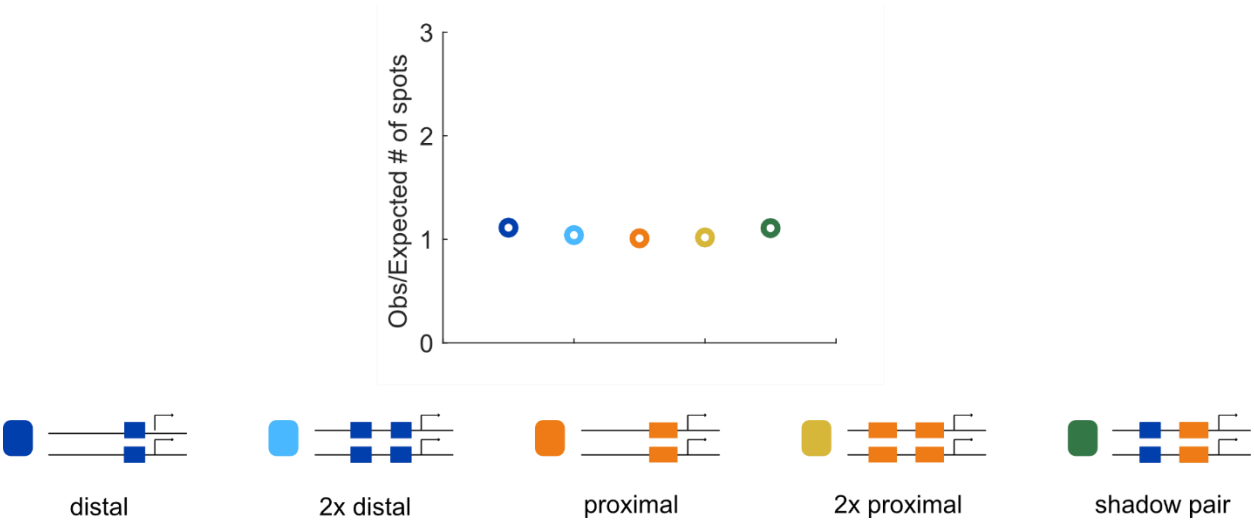


Figure S2.2: **Correspondence of observed and expected number of spots.** To ensure that we can accurately measure two spots of expression in the embryo, we compared the number of transcriptional spots seen in embryos hemizygous or homozygous for each construct. Our rationale was that in the absence of transvection, the number of transcriptional spots in homozygous embryos should be twice the number in embryos expressing the reporter on only one allele. The number of transcriptional spots tracked during nc14 in the AP bin of maximum expression was counted for all embryos imaged for each homozygous and hemizygous construct. The graph shows the average of this value for homozygous embryos, divided by double the value observed in the corresponding hemizygous construct. Assuming no transvection occurs, this value should be close to 1. The ratio of observed to expected number of spots is close to 1 for all of our enhancer constructs, indicating we are reliably able to track the two individual spots of transcription in single nuclei.

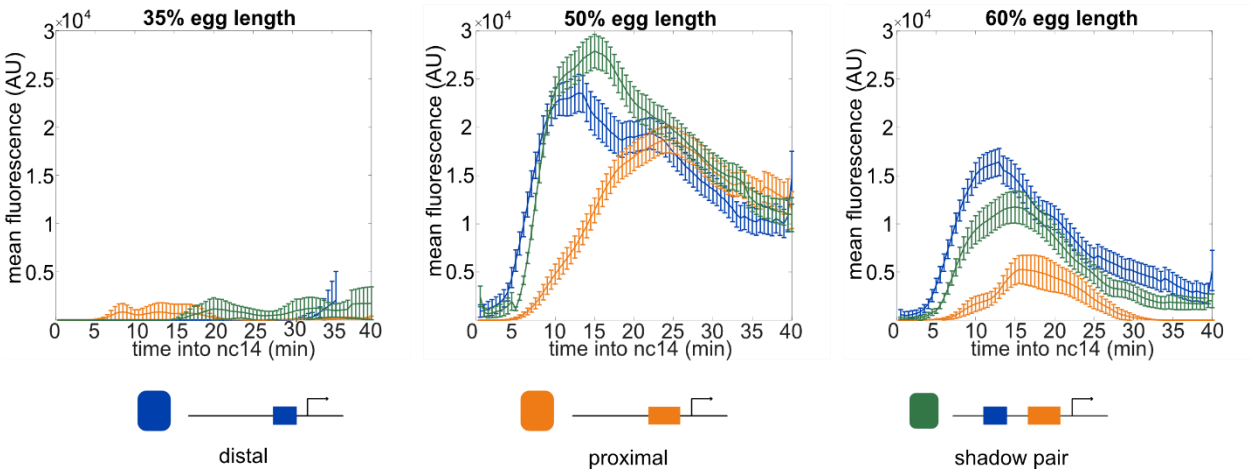
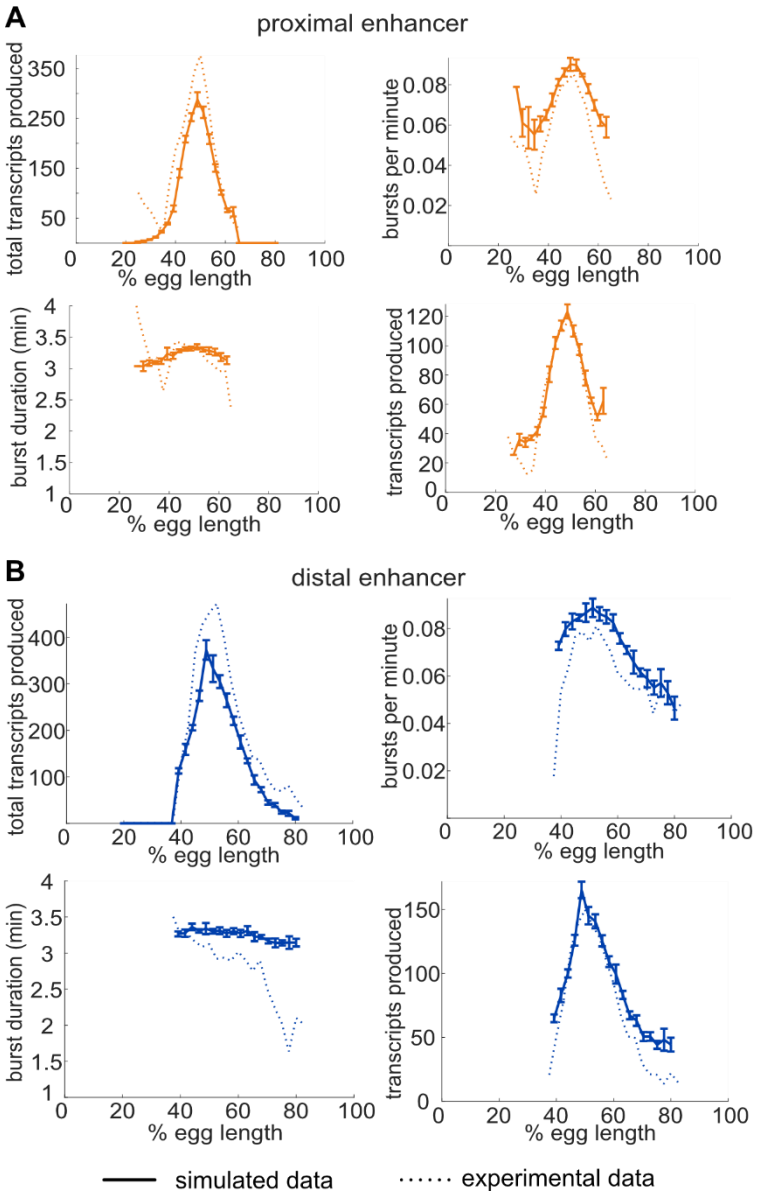


Figure S2.3: Expression across time at different embryo positions. The activity levels of the different enhancer constructs vary both across time and space. Shown in each graph is the mean fluorescence of all transcriptional traces of the indicated enhancer construct as a function of time into nc14 at the indicated position in the embryo. The earlier activation of the shadow pair and distal enhancer compared to the proximal enhancer at 50% and 60% egg length may stem from the input of the pioneer TF Zelda (Zld) to the distal and shadow pair. Error bars are 95% confidence intervals.

Figure S2.4: Single enhancer models recreate observed transcriptional bursting properties. To investigate whether our model is accurately simulating our experimental system, we compared the transcriptional burst properties produced by model simulations of transcription to those observed experimentally (see Figure S2.7 for description of burst properties). **A**. Graphs of average values of transcriptional burst properties, total mRNA produced during nc14, burst frequency, burst duration, and burst size associated with the proximal enhancer as a function of egg length. In A and B, simulated data are represented with solid lines and experimental data are shown with dotted lines. **B**. Graphs of average values of transcriptional burst properties as in A, associated with the distal enhancer. For both the proximal and distal enhancers, our model is largely able to recapitulate the experimentally observed transcriptional burst properties associated with each enhancer. **C**. The median and CV values of the model parameters for the proximal enhancer in the top 10 performing parameter sets. **D**. The median and CV values of the model parameters for the distal enhancer in the top 10 performing parameter sets. Explanations of model parameters are given in the Methods. Error bars represent 95% confidence intervals.



C

	median	CV
β_{-1}	2.7	2.4
k_{off}	1.8	0.40
k_{on}	0.36	0.90
n	4	2.0
c	6.8	0.75
σ	8.1	0.08

D

	median	CV
β_{-1}	3.9	1.2
k_{off}	1.5	0.13
k_{on}	0.19	1.3
n	12.5	1.3
c	9.2	0.34
σ	12.5	0.13

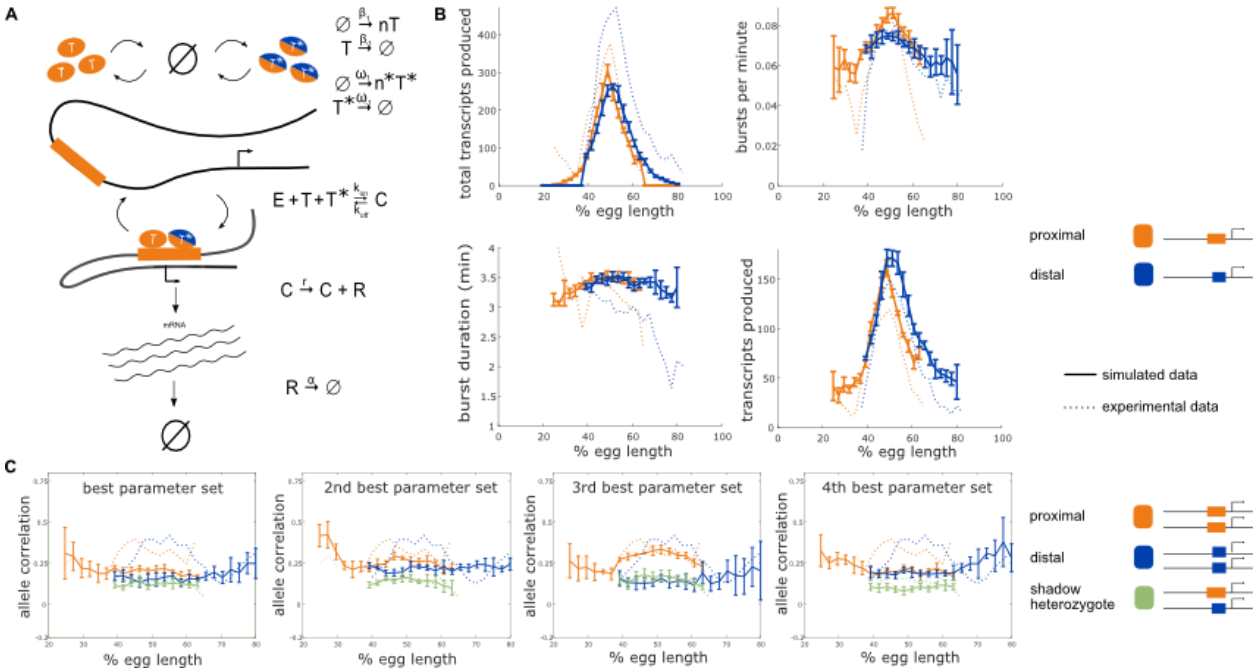


Figure S2.5: Incorporating a common TF into the model yields nonzero heterozygote allele correlations. To determine whether the observed nonzero heterozygote correlation can be explained by common TF activity, we incorporated into our model a TF that can bind to both the proximal and distal enhancers. **A**. Schematic of a model that includes an additional TF denoted T^* which can bind to both the proximal and distal enhancers. The production of T^* occurs at a rate ω_1 which varies across the embryo in a similar manner to β_1. T^* degrades linearly at a rate ω_{-1} and appears in bursts of size n^*. The presence of both the enhancer-specific TF T_i and the common TF T^* are necessary to initiate transcription. **B**. The addition of a common TF does not hinder the model from recapitulating the experimentally observed burst properties of single enhancer constructs. Simulated data is created using the second-best parameter set for each enhancer. The data shown is the average of five simulated embryos that have 80 transcriptional spots per AP bin. In B, C, and D simulated data are shown by solid lines, experimental data are shown by dotted lines. **C**. The addition of the common TF T^* consistently produces nonzero heterozygote allele correlations. However, some of the best parameter sets do not conserve the experimental relationship between homozygote and heterozygote correlations. Other parameter sets do not match the experimental data well suggesting that the model accepts a narrower range of parameter combinations than the bursting TF model. Error bars in B and C represent 95% confidence intervals.

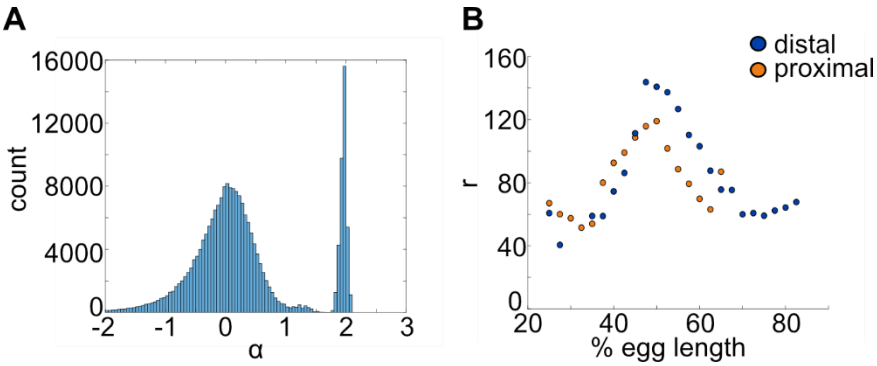
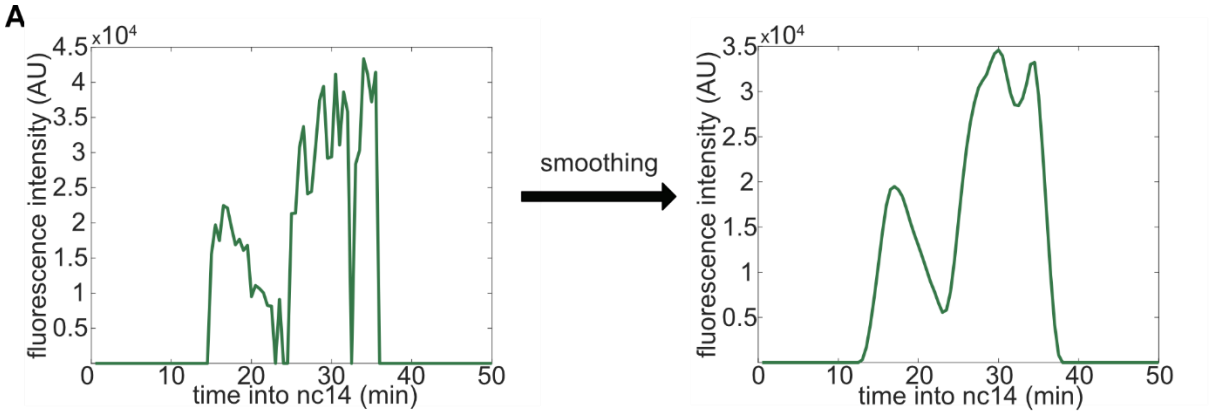
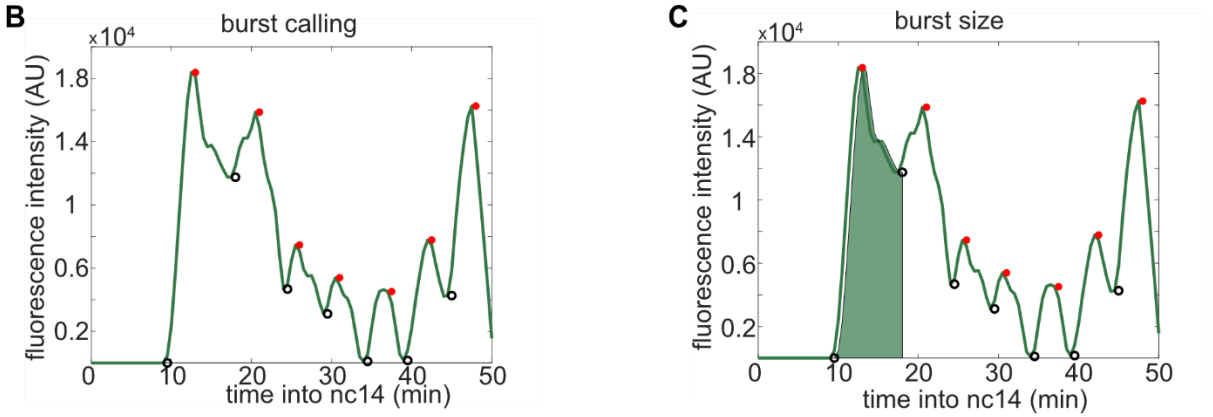
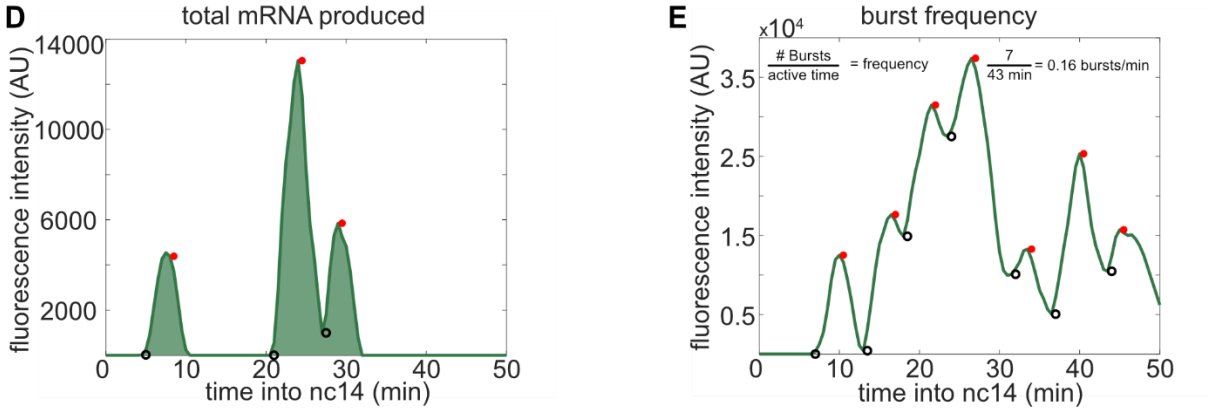


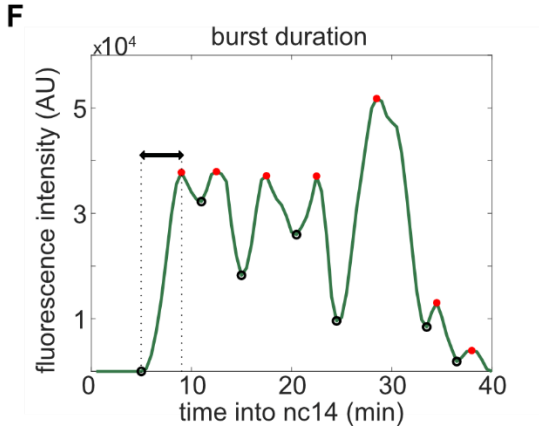
Figure S2.6: mRNA production and decay rates can be directly estimated from experimental data. The mRNA degradation parameter α and production parameter r were measured directly from fluorescence data without any input from the model. **A**. To estimate α, we used adjacent measurements of fluorescence intensity to approximate the slope at each point in the fluorescence traces. These values are compared with an exponential rate of mRNA decay (see Methods) and the resulting predicted values are shown in the histogram. Periods of mRNA production have negative α values and periods of decay have positive values. The histogram shows a distinct peak for $\alpha > 0$, which provided us with an estimate of $\alpha \approx 1.95$. **B**. A similar computational approach was used to calculate values of r from fluorescence data (see Methods). We calculated different values of r for each bin to account for differences in transcriptional efficiency across the length of the embryo due to factors that are not explicitly included in the model. For example, different combinations of TF bound to the enhancer may give rise to different mRNA production rates. Different values of r were found for the proximal and distal enhancers. Notice that distal r values shown correspond to the distal enhancer at the proximal location.

Figure S2.7: Visual inspection of burst calling algorithm. To extract the bursting parameters examined (burst size, frequency, and duration), individual fluorescence traces were first smoothed using the LOWESS method with a span of 0.1. Our burst calling algorithm then determined the periods of promoter activity or inactivity based on the slope of the fluorescence trace. **A**. Representative example of smoothing of transcriptional traces. **B**. Representative fluorescence trace of a single spot across the time of nc14. Black open circles indicate time points where the promoter is switched to being called "on", red filled circles indicate time points where the promoter is switched to being called "off". **C**. A representative transcriptional trace with shading representing the area under the curve used to calculate the size of the first burst. This area is calculated using the trapz function in MATLAB and is done for each burst, from the time point the promoter is called "on" until the next time it is called "on". D-F show additional representative fluorescence traces of single transcriptional spots across the time of nc14. **D**. A trace with shading representing the area under the entire curve during nc14 used to calculate the total amount of mRNA produced. This area is calculated using the trapz function in MATLAB and is done from the time the promoter is first called active until 50 minutes into nc14 or the movie ends, whichever comes first. **E**. Burst frequency is calculated by dividing the number of bursts that occur from the time the promoter is first called active until 50 minutes into nc14 or the movie ends, whichever comes first. **F**. Burst duration is defined as the amount of time between when the promoter is called active and it is next called inactive.


A
x10^4
fluorescence intensity (AU)
time into nc14 (min)
smoothing


B
burst calling
C
burst size
fluorescence intensity (AU)
time into nc14 (min)


D
total mRNA produced
E
burst frequency
Bursts / active time = frequency
7 / 43 min = 0.16 bursts/min
fluorescence intensity (AU)
time into nc14 (min)


F
burst duration
fluorescence intensity (AU)
time into nc14 (min)

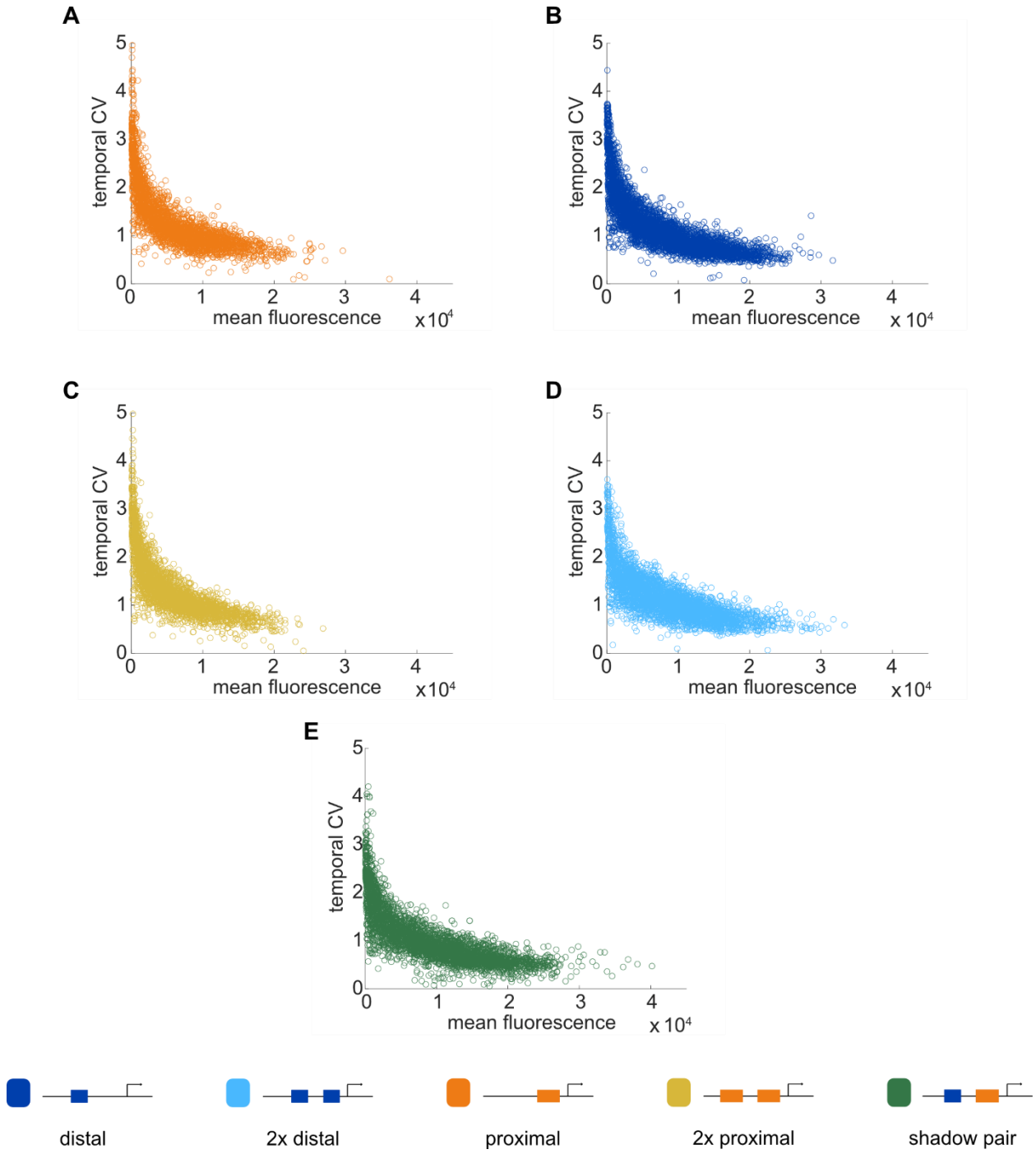


Figure S2.8: Temporal CV as a function of mean fluorescence. To investigate the relationship between our noise measurement of temporal CV and the mean activity of each construct, we plotted the temporal CV of each transcription spot as a function of its mean fluorescence. **A**. Distal; **B**. Proximal; **C**. 2x Proximal; **D**. 2x Distal; E. Shadow pair. With all constructs, we find the general trend that CV decreases with increasing average expression, flattening out at a baseline noise level specific to each enhancer construct.

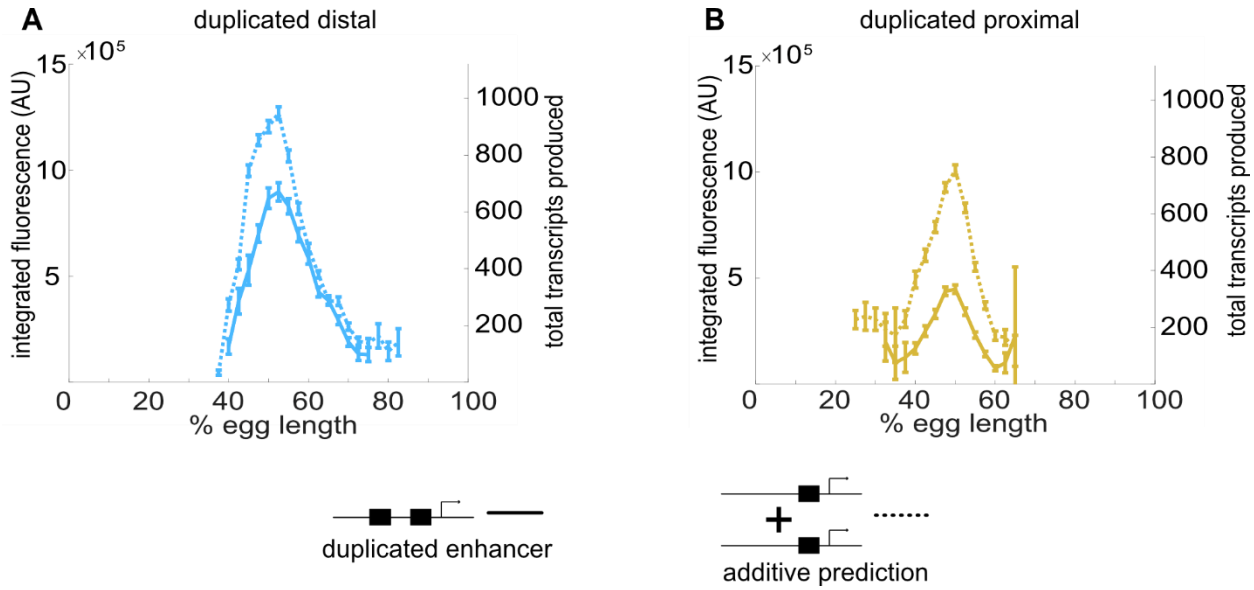
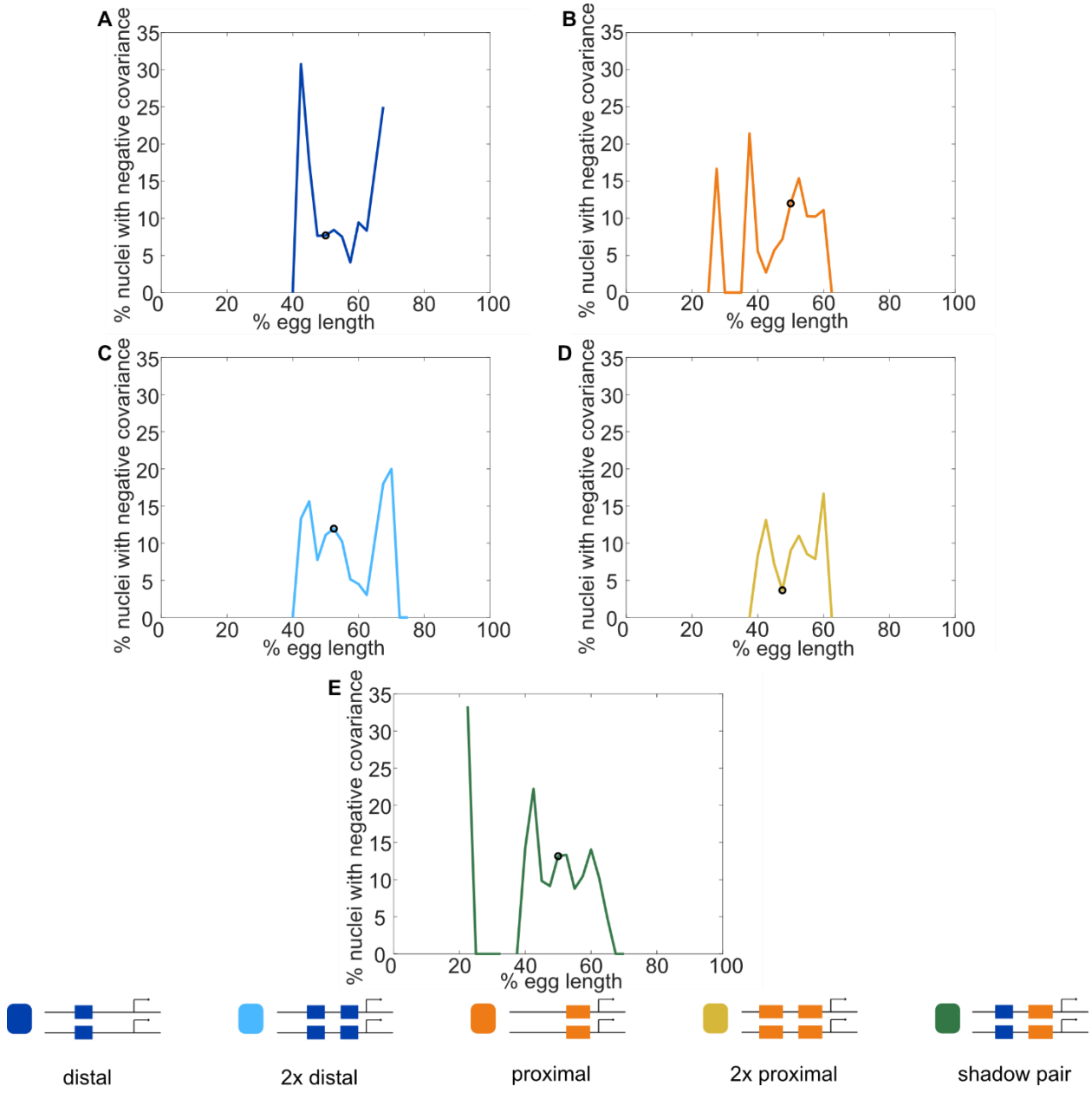


Figure S2.9: Individual *Kr* enhancers display sub-additive behavior. To assess the way input from two enhancers is integrated at the *Kr* promoter, we compared the experimentally observed mRNA production of duplicated enhancers to that predicted from additive behavior of the single enhancers. **A**. The duplicated distal enhancer displays sub-additive behavior. The solid line is the experimentally observed total mRNA produced by the duplicated distal enhancer during nc14 as a function of egg length and the dotted line is that expected by doubling the total mRNA produced by the single distal enhancer. **B**. The duplicated proximal enhancer also acts sub-additively. The solid line is the experimentally observed total mRNA produced by the proximal enhancer during nc14 as a function of egg length and the dotted line is that expected by doubling the total mRNA produced by the single proximal enhancer. These results, along with the observation that k_{off} values increased and k_{on} values decreased in our model with the addition of a second enhancer, suggests that the *Kr* enhancers compete with each other for interactions with the promoter. Error bars represent 95% confidence intervals.

Figure S2.10: Fraction of nuclei with negative covariance of allele activity. To identify the likely cause of the observed negative covariance between allele activity in some nuclei, we calculated the fraction of nuclei displaying negative covariance out of all nuclei that had active reporter transcription. Graphs show the fraction of transcribing nuclei with negative covariance as a function of egg length for each reporter construct, with a black circle indicating the position along the embryo of maximal expression for that construct. **A**. Distal; **B**. Proximal; **C**. 2x Proximal; **D**. 2x Distal; **E**. Shadow pair. Note that for all constructs, the highest rates of negative covariance are outside of the region of maximal reporter expression. MCP-GFP is expressed uniformly along the length of the embryo and we would therefore expect if MCP-GFP were the limiting factor, we would see the highest rates of negative covariance in the center of the expression pattern, where the highest number of transcripts are produced. Instead, the highest rates of negative covariance are seen at the edges of the *Kr* expression pattern, suggesting a spatially patterned factor, such as a TF, may be what is limiting.


A
B
C
D
E
% nuclei with negative covariance
% egg length
0
5
10
15
20
25
30
35
0
20
40
60
80
100
distal
2x distal
proximal
2x proximal
shadow pair

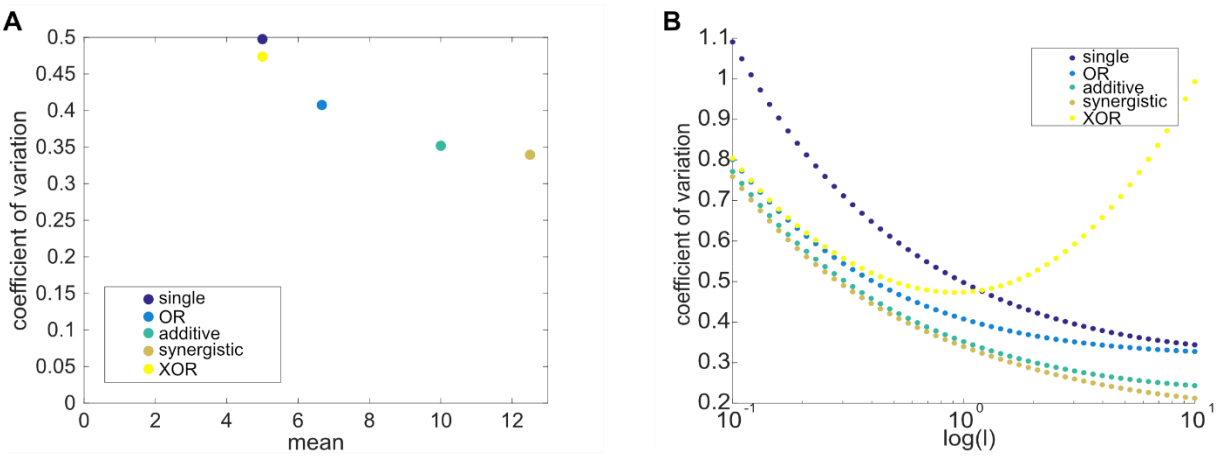
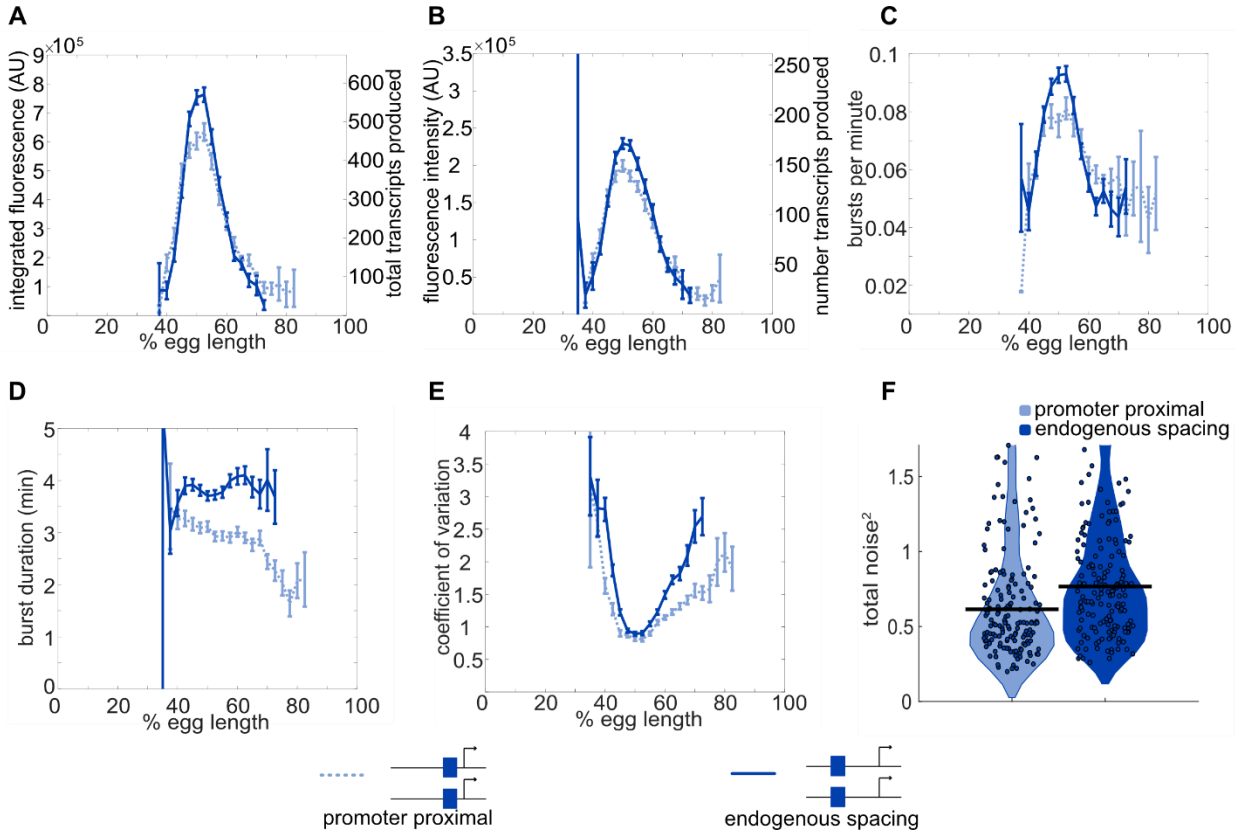


Figure S2.11: In most cases, two enhancer models drive lower noise than the single enhancer model. To explore the behavior of CV in these different models, we use several approaches. **A**. We plot the mean expression level versus CV for the five models above and one set of parameters, k = l = 1, p = 1, γ = 0.1. The single enhancer model (dark purple) drives the highest CV, indicating that, under the assumptions of our models, adding an additional enhancer generally lowers intrinsic noise. Except for XOR model (yellow), all other models produce more mRNA than the single enhancer model. The other colors are: blue, OR model; green, additive model; brown, synergistic model. **B**. Here we plot the CV as a function of l, the rate of promoter-enhancer dissociation, for the five models above and vary l from 0.1 to 10 on a logarithmic scale with k = 1, p = 1, γ = 0.1. With the exception of the XOR model with low l, the single enhancer model drives a higher CV than the models with two enhancers for the same value of l. These results show that, under the simplifying assumptions that the production rates and on-off rates of enhancers are independent of the position and number of enhancers, the addition of a second enhancer generally lowers the predicted intrinsic noise. In our experimental data (Figure 2.6), we only observe a significant decrease in interallele noise for the shadow enhancer pair compared to the single distal or single proximal enhancer. Duplications of either the proximal or distal enhancer do not have significantly lower noise than their respective single enhancer constructs. Therefore, we expect that the simple addition of an identical enhancer likely does not fulfill the simplifying parameter assumptions used here and suggests that further investigation is needed to understand the complexity of the relationship between interallele noise and the numbers of enhancers controlling a promoter.

Figure S2.12: Position-dependent effects on distal enhancer. To best mimic the endogenous system, we looked at expression driven by the distal enhancer at its endogenous spacing from the promoter for our noise calculations. In this construct we replaced the sequence of the proximal enhancer with sequence of the same length from the lambda phage genome predicted to have low number of Drosophila TF binding sites. This increased distance from the promoter had observable effects on the transcriptional dynamics and noise associated with the distal enhancer. **A**. Comparison of total transcriptional expression mediated by the distal enhancer at its endogenous spacing or proximal to the promoter. The distal enhancer at its endogenous spacing, shown as the solid line, produces significantly more total mRNA in the center region of expression than the distal enhancer proximal to the promoter, shown as the dotted line. **B**. Comparison of the average number of transcripts produced per transcriptional burst by each distal enhancer configuration as a function of egg length. **C**. Average burst frequency associated with either distal enhancer configuration as a function of egg length. **D**. Average burst duration associated with either distal enhancer configuration as a function of egg length. **E**. Coefficient of variation of transcriptional activity across nc14 for each distal enhancer configuration as a function of egg length. **F**. Total expression noise associated with either distal enhancer configuration at the AP bin of that construct's peak expression. The total noise distribution for the distal enhancer proximal to the promoter is on the left and that for the distal enhancer at its endogenous spacing from the promoter is on the right. The distal enhancer at its endogenous spacing displays significantly higher total noise ($p = 0.018$) than the distal enhancer proximal to the promoter. Each circle represents the total noise of an individual nucleus and the horizontal bar marks the median total noise value. Y-axis limited to the 75th percentile of the construct with the highest total noise values (distal promoter at endogenous spacing). Error bars in A-E represent 95% confidence intervals.


A
integrated fluorescence (AU)
×10^5
total transcripts produced
% egg length
B
fluorescence intensity (AU)
×10^5
number transcripts produced
% egg length
C
bursts per minute
% egg length
D
burst duration (min)
% egg length
E
coefficient of variation
% egg length
F
promoter proximal
endogenous spacing
total noise²
promoter proximal
endogenous spacing

2.6 References

1. Miller Jr. OL, McKnight SL. Post-replicative nonribosomal transcription units in D. melanogaster embryos. *Cell*. 1979;17(3):551-563. doi:10.1016/0092-8674(79)90263-0
2. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional Pulsing of a Developmental Gene. *Curr Biol*. 2006;16(10):1018-1025. doi:10.1016/j.cub.2006.03.092
3. Dar RD, Razooky BS, Singh A, et al. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci*. 2012;109(43):17454-17459. doi:10.1073/pnas.1213530109
4. Sanchez A, Golding I. Genetic determinants and cellular constraints in noisy gene expression. *Science*. 2013;342(6163):1188-1193. doi:10.1126/science.1242975
5. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol*. 2008;15(12):1263-1271. doi:10.1038/nsmb.1514
6. Bothma JP, Garcia HG, Esposito E, Schlissel G, Gregor T, Levine M. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. *Proc Natl Acad Sci U S A*. 2014;111(29):10598-10603. doi:10.1073/pnas.1410022111
7. Csárdi G, Franks A, Choi DS, Airoldi EM, Drummond DA. Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS Genet*. 2015;11(5):e1005206. doi:10.1371/journal.pgen.1005206
8. Hansen MMK, Desai R V., Simpson ML, Weinberger LS. Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. *Cell Syst*. 2018;7(4):384-397.e6. doi:10.1016/j.cels.2018.08.002
9. Blake WJ, KÆrn M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. *Nature*. 2003;422(6932):633-637. doi:10.1038/nature01546
10. Dubuis JO, Samanta R, Gregor T. Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol Syst Biol*. 2013;9:639. doi:10.1038/msb.2012.72
11. Gregor T, Tank DW, Wieschaus EF, Bialek W. Probing the limits to positional information. *Cell*. 2007;130(1):153-164. doi:10.1016/j.cell.2007.05.025
12. Lagha M, Bothma JP, Levine M. Mechanisms of transcriptional precision in animal development. *Trends Genet*. 2012;28(8):409-416. doi:10.1016/j.tig.2012.03.006
13. Stapel LC, Zechner C, Vastenhouw NL. Uniform gene expression in embryos is achieved by temporal averaging of transcription noise. *Genes Dev*. 2017;31(16):1635-1640. doi:10.1101/gad.302935.117
14. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. *Nature*. 2010;463(7283):913-918. doi:10.1038/nature08781
15. Erdmann T, Howard M, ten Wolde PR. Role of Spatial Averaging in the Precision of Gene Expression Patterns. *Phys Rev Lett*. 2009;103(25):258101. doi:10.1103/PhysRevLett.103.258101
16. Barolo S. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays*. 2012;34(2):135-141. doi:10.1002/bies.201100121
17. Hong JW, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty. *Science (80-)*. 2008;321(5894):1314. doi:10.1126/science.1160631

18. Cannavò E, Khoueiry P, Garfield DA, et al. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol*. 2016;26(1):38-51. doi:10.1016/j.cub.2015.11.034
19. Osterwalder M, Barozzi I, Tissières V, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*. 2018;554(7691):239-243. doi:10.1038/nature25461
20. Garnett AT, Square TA, Medeiros DM. BMP, Wnt and FGF signals are integrated through evolutionarily conserved enhancers to achieve robust expression of Pax3 and Zic genes at the zebrafish neural plate border. *Development*. 2012;139(22):4220-4231. doi:10.1242/dev.081497
21. Bomblies K, Dagenais N, Weigel D. Redundant Enhancers Mediate Transcriptional Repression of AGAMOUS by APETALA2. *Dev Biol*. 1999;216(1):260-264. doi:10.1006/DBIO.1999.9504
22. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*. 2010;466(7305):490-493. doi:10.1038/nature09158
23. Perry MW, Boettiger AN, Bothma JP, Levine M. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr Biol*. 2010;20(17):1562-1567. doi:10.1016/j.cub.2010.07.043
24. Cheung D, Ma J. Probing the impact of temperature on molecular events in a developmental system. *Sci Rep*. 2015;5. doi:10.1038/srep13124
25. Chen J, Nolte V, Schlötterer C. Temperature stress mediates decanalization and dominance of gene expression in Drosophila melanogaster. *PLoS Genet*. 2015;11(2):e1004883. doi:10.1371/journal.pgen.1004883
26. Lam DD, de Souza FSJ, Nasif S, et al. Partially redundant enhancers cooperatively maintain Mammalian pomc expression above a critical functional threshold. *PLoS Genet*. 2015;11(2):e1004935. doi:10.1371/journal.pgen.1004935
27. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci*. 2011;108(33):13570-13575. doi:10.1073/PNAS.1109873108
28. Wunderlich Z, Bragdon MDJ, Vincent BJ, White JA, Estrada J, DePace AH. Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. *Cell Rep*. 2015;12(11):1740-1747. doi:10.1016/j.celrep.2015.08.021
29. Ghiasvand NM, Rudolph DD, Mashayekhi M, et al. Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. *Nat Neurosci*. 2011;14(5):578-586. doi:10.1038/nn.2798
30. Preiss A, Rosenberg UB, Kienlin A, Seifert E, Jäckle H. Molecular genetics of Krüppel, a gene required for segmentation of the Drosophila embryo. *Nature*. 313(5997):27-32. http://www.ncbi.nlm.nih.gov/pubmed/3917552. Accessed January 11, 2019.
31. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA Particles in Living Yeast. *Mol Cell*. 1998;2(4):437-445. doi:10.1016/S1097-2765(00)80143-4
32. Garcia HG, Tikhonov M, Lin A, Gregor T. Quantitative live imaging of transcription in Drosophila embryos links polymerase activity to macroscopic patterns. doi:10.1016/j.cub.2013.08.054
33. Lucas T, Ferraro T, Roelens B, et al. Live Imaging of Bicoid-Dependent Transcription in

Drosophila Embryos. *Curr Biol*. 2013;23(21):2135-2139. doi:10.1016/J.CUB.2013.08.053
34. Lim B, Heist T, Levine M, Fukaya T. Visualization of Transvection in Living Drosophila Embryos. *Mol Cell*. 2018;70(2):287-296.e6. doi:10.1016/J.MOLCEL.2018.02.029
35. Fukaya T, Levine M. Transvection. *Curr Biol*. 2017;27(19):R1047-R1049. doi:10.1016/J.CUB.2017.08.001
36. Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife*. 2015;4. doi:10.7554/eLife.07956
37. Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW. Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient. *Cell*. 2007;130(1):141-152. doi:10.1016/j.cell.2007.05.026
38. Dar RD, Shaffer SM, Singh A, et al. Transcriptional Bursting Explains the Noise–Versus–Mean Relationship in mRNA and Protein Levels. Chauhan A, ed. *PLoS One*. 2016;11(7):e0158298. doi:10.1371/journal.pone.0158298
39. Scholes C, Biette KM, Harden TT, DePace AH. Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the Drosophila Embryo. *Cell Rep*. 2019;26(9):2407-2418.e5. doi:10.1016/j.celrep.2019.01.115
40. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic Gene Expression in a Single Cell. *Science (80-)*. 2002;297(5584). http://science.sciencemag.org/content/297/5584/1183.full. Accessed July 11, 2017.
41. Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. *Science*. 2004;304(5678):1811-1814. doi:10.1126/science.1098641
42. Arias AM, Hayward P. Filtering transcriptional noise during development: concepts and mechanisms. *Nat Rev Genet*. 2006;7(1):34-44. doi:10.1038/nrg1750
43. Little SC, Tikhonov M, Gregor T. Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell*. 2013;154(4):789-800. doi:10.1016/j.cell.2013.07.025
44. He F, Wen Y, Deng J, et al. Probing Intrinsic Properties of a Robust Morphogen Gradient in Drosophila. *Dev Cell*. 2008;15(4):558-567. doi:10.1016/J.DEVCEL.2008.09.004
45. Bothma JP, Norstad MR, Alamos S, Garcia HG. LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. *Cell*. 2018;173(7):1810-1822.e16. doi:10.1016/j.cell.2018.03.069
46. Mir M, Stadler MR, Ortiz SA, et al. Dynamic multifactor hubs interact transiently with sites of active transcription in drosophila embryos. *Elife*. 2018;7. doi:10.7554/eLife.40497
47. Liu X, Wu B, Szary J, Kofoed EM, Schaufele F. Functional sequestration of transcription factor activity by repetitive DNA. *J Biol Chem*. 2007;282(29):20868-20876. doi:10.1074/jbc.M702547200
48. Janssen S, Cuvier O, Müller M, Laemmli UK. Specific Gain- and Loss-of-Function Phenotypes Induced by Satellite-Specific DNA-Binding Drugs Fed to Drosophila melanogaster. *Mol Cell*. 2000;6(5):1013-1024. doi:10.1016/S1097-2765(00)00100-3
49. Laboulaye MA, Duan X, Qiao M, Whitney IE, Sanes JR. Mapping Transgene Insertion Sites Reveals Complex Interactions Between Mouse Transgenes and Neighboring Endogenous Genes. *Front Mol Neurosci*. 2018;11:385. doi:10.3389/fnmol.2018.00385
50. Thompson A, Gasson MJ. Location effects of a reporter gene on expression levels and on native protein synthesis in Lactococcus lactis and Saccharomyces cerevisiae. *Appl Environ Microbiol*. 2001;67(8):3434-3439. doi:10.1128/AEM.67.8.3434-3439.2001

51. Yuan Y, Wei L, Hu T, et al. Quantitative understanding of molecular competition as a hidden layer of gene regulatory network. *bioRxiv*. February 2018:258129. doi:10.1101/258129
52. Ella Preger-Ben Noon A, Sabarís G, Ortiz DM, et al. Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer Function. *Cell Rep*. 2018;22:3021-3031. doi:10.1016/j.celrep.2018.02.073
53. Dunipace L, Ákos Z, Stathopoulos A. Coacting enhancers can have complementary functions within gene regulatory networks and promote canalization. Desplan C, ed. *PLOS Genet*. 2019;15(12):e1008525. doi:10.1371/journal.pgen.1008525
54. El-Sherif E, Levine M. Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo. *Curr Biol*. 2016;26(9):1164-1169. doi:10.1016/j.cub.2016.02.054
55. Perry MW, Bothma JP, Luu RD, Levine M. Precision of Hunchback Expression in the Drosophila Embryo. *Curr Biol*. 2012;22(23):2247-2252. doi:10.1016/J.CUB.2012.09.051
56. Dunipace L, Ozdemir A, Stathopoulos A. Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. *Development*. 2011;138(18):4075-4084. doi:10.1242/dev.069146
57. Yan J, Anderson C, Viets K, et al. Regulatory logic driving stable levels of defective proventriculus expression during terminal photoreceptor specification in flies. *Development*. 2017;144(5):844-855. doi:10.1242/dev.144030
58. Hoch M, Schroder1 C, Seifert2 E, Jackle H. Cis-acting control elements for Kruppel expression in the Drosophila embryo. *EMBO J*. 1990;9(8):2587-2595. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC552291/pdf/emboj00235-0229.pdf.

CHAPTER 3

Molecular competition can shape enhancer activity in the Drosophila embryo

CHAPTER 3

Molecular competition can shape enhancer activity in the Drosophila embryo

3.1 Summary

Transgenic reporters allow the measurement of regulatory DNA activity *in vivo* and consequently have long been useful tools in the study of enhancers. Despite the utility of transgenic reporters, few studies have investigated the potential effects these reporters have on the expression of other transgenic reporters or endogenous genes. A full understanding of the impacts transgenic reporters have on expression is required for both the accurate interpretation of transgenic reporter data and the characterization of gene regulatory mechanisms. Here, we investigate the impact transgenic reporters have on the expression of other transgenic reporters and endogenous genes. By measuring the expression of *Kruppel* (*Kr*) enhancer reporters in live *Drosophila* embryos that contain either one or two copies of identical reporters, we find reporters have an inhibitory effect on one another's expression. Further, expression of a nearby endogenous gene is decreased in the presence of a *Kr* enhancer reporter. Through the use of competitor transcription factor (TF) binding site arrays, we present evidence that reporters, and potentially endogenous genes, are competing for TFs. Increasing numbers of competitor binding sites decrease both peak levels and the spatial extent of expression. To understand how small numbers of added TF binding sites could significantly impact gene expression, we develop a simple thermodynamic model of our system. Our model predicts competition of the measured magnitude specifically if TF binding is restricted to distinct nuclear subregions, underlining an unexpected role of the non-homogenous nature of the nucleus in regulating gene expression.

3.2 Introduction

An organism's ability to precisely control gene expression is dependent on the activity of enhancers. Through the binding of specific combinations of transcription factors (TFs), which can be activating or repressive, enhancers are able to control the expression of their target genes in time and space. Enhancers control gene expression across all aspects of organismal functioning, from the immune system to the nervous system, and play a particularly important and well-studied role in the process of embryonic development[1,2]. During this period, enhancers regulate the expression of genes that determine critical cell fate decisions underlying patterning and organogenesis.

A significant amount of our understanding of enhancers and other *cis*-regulatory elements has come from the use of transgenic reporter lines. These transgenic animals have measurable reporters, such as fluorescent proteins or LacZ, under the control of *cis*-regulatory elements to enable observation of that element's activity in living organisms in different life stages, tissue types, or conditions[3,4]. Studies of transgenic animals have enabled the discoveries of previously unknown enhancers, the modularity of enhancers, and the importance of the arrangement of TF binding sites or enhancer "grammar", among others[5–9].

Despite the remarkable utility of transgenic reporters, or perhaps in part because of it, little work has been done to look at the effect of these reporters on expression of other reporters or endogenous genes. Although reporters are exogenous regions of DNA that can originate from completely different species than the host animal, once integrated into the genome, these transgenes rely on the same pool of transcription factors, polymerases, and other molecular factors required for gene expression as endogenous genes. Given that most of these factors are present at relatively high copy numbers in the cell, for example 250,000 Zld TF molecules per nucleus in the *Drosophila* embryo[10,11] or over 80,000 RNAP molecules per nucleus in human

cells[11,12], it is commonly assumed that adding an additional enhancer would have little impact on the availability of key expression machinery. However, a couple of examples suggest that there may be competition between transgenic reporters and endogenous genes. A study by Laboulaye, et al. measured the effect of three different transgenic reporters on endogenous gene expression in mice[13]. The authors found that the transgenic reporters all decreased the expression of the closest endogenous gene. Thompson & Gasson noted that endogenous protein levels may be slightly decreased in *Saccharomyces cerevisiae* and *Lactococcus lactis* expressing transgenic reporters, but the results were inconclusive[14]. These examples suggest that transgenic reporters may decrease endogenous gene expression but leave open the questions of the mechanisms behind these decreases and whether such an effect is limited to certain organisms or reporters.

Like much of the field, we often used transgenic reporters under the assumption that they had no effect on the expression of other genes until we saw evidence to the contrary in our own data. In a study investigating gene expression noise in *Drosophila* embryos, we observed evidence of competition between identical copies of transcriptional reporters integrated on homologous chromosomes[15]. We were surprised to find that homozygous reporter embryos produced less mRNA per reporter allele than hemizygous embryos, with a reporter present on only one of the two homologous chromosomes (Figure 3.1). We suspected this could have important implications not only for the use of transgenic reporters, but also for our understanding of the balance between the supply of and demand for transcriptional machinery within the nucleus.

Here we track the activity of multiple configurations of transgenic reporters in *Drosophila* embryos to assess the impact of these reporters on one another and endogenous genes. We measured live mRNA dynamics driven by the embryonic enhancers of the gap gene

Kruppel (*Kr*) in the presence or absence of a second transcriptional reporter or a competitor TF binding array. We find that enhancer reporter expression is lower not only in the presence of a second reporter, but also in the presence of non-transcribing TF binding arrays, suggesting that there is competition for locally limited levels of certain TFs. This effect is not restricted to reporters; expression of a nearby endogenous gene is also decreased in transgenic embryos. To understand how the addition of the relatively small number of TF binding sites present in our constructs can measurably decrease reporter expression, we developed a thermodynamic model of our system. We predict reduced expression of the magnitude observed in transgenic embryos if we assume TF binding is restricted to so-called "hub" regions, but not if we assume TFs have access to the whole genome. This work reconciles the question of how tens of TF binding sites in a transgenic reporter construct can impact the available supply of tens of thousands of TF molecules. We suggest that the TF supply relevant to a particular enhancer is limited to a smaller pool of the TFs in a nucleus.

3.3 Results

3.3.1 Homozygous reporters display evidence of competition

To test whether transgenic reporters affect the expression of other alleles, we compared the expression output in embryos either homozygous or hemizygous for different reporter constructs. In the absence of reporter interactions, we will see the same levels of mRNA production per allele in hemizygous embryos and homozygous embryos. Conversely, if the reporters do affect one another's expression, then expression levels per allele will differ in hemizygous vs homozygous embryos, depending on the nature of this interaction. A synergistic interaction,

perhaps through a mechanism such as increasing the local concentration of a key TF, would lead to higher levels of transcription in homozygous embryos than hemizygous embryos (Figure 3.1B upper half). An antagonistic interaction, such as competition for a limited shared resource, would lead to lower levels of transcription in homozygous embryos than hemizygous embryos (Figure 3.1B lower half).

To assess the nature of potential reporter interactions, we measured transcriptional output of different enhancers in living embryos using the MS2 reporter system. When transcribed, the MS2 sequence forms stem loops that are then bound by an MCP-GFP fusion protein expressed in the embryo, enabling us to visualize sites of nascent transcription[16] (Figure 3.1A). We can track these individual transcriptional spots across the time of nuclear cycle 14 (nc14), when these enhancers are most active, to measure total transcriptional output and dynamics. As a test case, we used different combinations of the two *Kruppel* (*Kr*) embryonic shadow enhancers. The *Kr* shadow enhancer pair, together or individually, drives a stripe of expression in the central 20% of the embryo (Figure 3.1A). We generated transgenic flies with each individual enhancer, the shadow enhancer pair, and each enhancer duplicated in tandem driving an MS2 reporter (Figure 3.1B). Despite the similar pattern of expression driven by the two individual enhancers, the distal and proximal enhancers are each activated by different sets of TFs[17]. We previously showed that this separation of TF inputs plays an important role in suppressing gene expression noise[15]. Here, this separation of TF inputs allows us to investigate whether the reporter interactions we observe are influenced by specific regulatory factors or are more general consequences of having two reporters present.

In the majority of cases, hemizygous embryos produce more mRNA per allele than do homozygous embryos (Figure 3.1B). To calculate the mRNA produced by each reporter, we

integrate the area under the fluorescence traces of activity measured during nc14 at the anterior-posterior position in the embryo of peak expression (Figure S3.1). The single and duplicated distal constructs produce 62% and 40%, respectively, more mRNA per allele in hemizygous embryos than in homozygous embryos. The shadow pair and proximal enhancer reporters produce 27% and 22% more mRNA per allele at their respective regions of peak expression in hemizygous embryos than in homozygous embryos. The duplicated proximal construct drives the same level of expression in hemizygous and homozygous embryos. By comparing the competition exhibited by duplicated and single enhancers, we do not find evidence that longer reporter sequences drive stronger reporter competition (Figure 3.1B inset). We suspect this trend may arise because duplicated enhancers with a large array of similar binding sites can recruit a larger pool of TFs[18] or because there can be synergy between the enhancers in promoter activation[19]. In sum, when two reporters are present in the same nucleus, neither typically transcribes to its full potential, suggesting that there is some form of competition between the two reporters. We hypothesized that the reporters are competing for one or more molecular factors required for reporter transcription or visualization.

3.3.2 Reporter competition is not an artifact of imaging system

To assess whether reporter competition is the result of a biological phenomena, such as limiting levels of a TF, or an artifact of our reporter system, such as limiting levels of MCP-GFP, we measured reporter output in the presence of a second non-transcribing transgenic construct. We produced a version of our distal enhancer construct that lacks both a promoter and the MS2 cassette. This construct therefore can bind the same regulatory TFs as the original distal construct but will not drive transcription. Therefore, it should not interact with promoter-bound

factors, such as RNA polymerase, or the MCP-GFP coat protein. If the observed competition is for MCP-GFP or is dependent on transcription, we expect to see no effect on reporter expression when the enhancer-only construct is present on the homologous chromosome. Conversely, if one or more regulatory factors binding the enhancer is limiting, we expect to see a decrease in reporter expression, similar to the lower expression of homozygous versus hemizygous embryos (Figure 3.1B).

In the presence of the distal enhancer-only construct, the distal enhancer reporter drives 11% lower levels of expression at its region of peak expression than in the hemizygous configuration (Figure 3.1C). While significant (t-test *p*-value = 0.02), this decrease is not as large as the one we see when a second transcribing distal enhancer reporter is present on the homologous chromosome. We suspect the smaller effect of the non-transcribing distal enhancer construct is due to differences in the exact composition and levels of factors that are recruited to transcriptionally active versus inactive enhancers[20–22]. To further rule out that the observed reporter competition is a result of limiting levels of the MCP-GFP reporter, we looked at the pattern of reporter competition across the length of the embryo. The MCP-GFP coat protein is expressed ubiquitously across the length of the embryo, while many of the TFs regulating the *Kr* enhancers are spatially patterned. If reporters are competing for limited levels of MCP-GFP, we would expect to see the highest rates of competition in the center of the embryo where expression driven by our reporters is highest (Figure S3.2B). Instead, with all of our reporter constructs we find that rates of competition are highest outside of the region of peak expression (Figure S3.22C-E), strongly suggesting that reporters are not competing for limited levels of MCP-GFP. Instead, this pattern of competition rates combined with the finding that a non-

transcribing enhancer construct can reduce reporter activity suggest that our reporters are competing for an endogenous factor required for transcription.

3.3.3 Reporters are competing for transcription factors

While reporter competition seems to be independent of the MS2 system or weakly dependent on transcription itself, it does depend on the identity of the enhancer driving reporter expression (Figure 3.1B). The presence of a second identical reporter has a large effect on expression driven by the shadow construct and an even larger effect on the duplicated distal construct, while it has no significant effect on the duplicated proximal construct (Figure 3.1B). Since the *Kr* distal and proximal enhancers are regulated by separate sets of TFs[17], we hypothesized that reporters may be competing for one or more of these TFs and that this may underlie the difference in competition levels between constructs. To test this hypothesis, we measured the effect of TF binding site arrays on the activity of the reporters. As the level of competition is not significantly different between the single and duplicated enhancer constructs, we focused on the two duplicated enhancers and the shadow pair, which are similar lengths and therefore have similar numbers of TF binding sites. We created DNA sequences consisting of six strong TF binding sites for each of the key activating TFs of the *Kr* enhancers and inserted them into the identical site on the homologous chromosome opposite one of the enhancer-MS2 reporters (Figure 3.2A). Critically, these TF binding site arrays lack promoter and MS2 sequences. We reasoned that these TF binding site arrays would function to sequester TF molecules without affecting factors specifically involved in transcript production (such as RNAP) or reporter visualization (i.e. MCP-GFP). Therefore, any changes in transcriptional output by the enhancer-MS2 reporter observed in the presence of a TF binding site array should stem from decreased levels of available TF, not higher demand for basal transcriptional machinery or MCP-GFP.

Specifically, we created four binding site arrays corresponding to the four key TF activators of the shadow pair (Bicoid (Bcd), Hunchback (Hb), Stat92E, and Zelda (Zld); Figure 3.2A), which each contain six binding sites for the respective TF across 236bp. As the shadow pair is the only construct known to be regulated by all four TFs, we first assessed the impact of these binding site arrays on the activity of the shadow pair reporter. We find that the binding site arrays for the Zld and Stat92E each reduce the activity of the shadow pair down to the levels seen in hemizygous embryos, while the Bcd and Hb binding site arrays do not have a significant effect on the shadow pair's activity (Figure 3.2B). We suspect that the Stat92E and Zld binding arrays may have the largest effect on the shadow pair's activity due to their essential roles in early gene activation[23,24].

As we observe a stark difference in the levels of competition in the duplicated distal versus duplicated proximal constructs, we asked whether the Bcd binding site array affects expression of either construct. Bcd is a key activator of the distal enhancer, but not the proximal enhancer (Figure 3.2A). In line with this, the duplicated distal reporter's activity is reduced 37% compared to hemizygous levels at their regions of peak activity in the presence of the Bcd binding site array (Figure 3.2C), while the activity of the duplicated proximal enhancer is not significantly changed (Figure 3.2D). The large effect the Bcd array (from here on called 1xBcd) on the duplicated distal reporter is striking, as the TF binding site array is less than one-fifth the size of either *Kr* enhancer and contains only six, albeit strong, binding sites for Bcd. The specificity of the 1xBcd array in reducing expression only of the Bcd-activated duplicated distal reporter, but not of the duplicated proximal reporter, suggests that the effect we observe is specific to sequestering Bcd molecules, and not a general effect of inserted DNA sequences.

3.3.4 Reporters show dosage-dependent response to increasing number of Bcd competitor sites

As a whole, these experiments suggest that limiting levels of TFs play an important role in reporter competition. When comparing the effects of the 1xBcd array across constructs, the expression of the duplicated distal reporter is dramatically reduced in the presence of the array, while the expression of the duplicated proximal and shadow pair constructs are unaffected. Given that the shadow pair is regulated by more TF inputs beyond Bcd than is the duplicated distal reporter, we hypothesized that the shadow pair may be less sensitive to Bcd competition. To test this hypothesis, we attempted to sequester larger amounts of Bcd and measure the effect on the shadow pair's activity. We measured the activity of the shadow pair reporter in the presence of larger binding site arrays consisting of three (3xBcd; 3 x 6 copies = 18 Bcd binding sites) or six (6xBcd; 6 x 6 = 36 Bcd binding sites) copies of the original Bcd binding site array.

In line with our hypothesis, we find that the shadow pair reporter activity decreases with increasing number of Bcd binding sites in the competitor array. Peak expression is reduced 1% in the presence of the 1xBcd array, 21% with the 3xBcd array, and 38% with the 6xBcd array relative to expression in hemizygotes (Figure 3.3A). We also measured expression of the duplicated distal enhancer with the larger Bcd binding arrays and find a non-linear effect of increasing the number of competitor Bcd binding sites (Figure S3.3A; Discussion). To assess whether the reduction in mRNA output in the presence of the Bcd array is specific to the *Kr* enhancers or a general phenomenon, we measured the expression driven by the *hunchback* (*hb*) P2 Bcd-responsive enhancer in the presence or absence of the 1xBcd and 6xBcd arrays. Similar to our findings with the *Kr* enhancers, the Bcd binding arrays decrease the expression of the *hb* P2 enhancer (Figure 3.3B). Relative to hemizygous levels, peak expression of the *hb* P2 enhancer is decreased 44% with the 1xBcd array and 49% with the 6xBcd array. We suspect that the relatively modest effect of larger numbers of Bcd competitor sites on reporter activity stems

from an upper limit to the amount of Bcd molecules that can be effectively sequestered away from the enhancers at our binding site arrays along with the activating function of other TFs .

Since TFs can control both the level and pattern of enhancer activity, we measured how the expression boundaries of our reporters changed in response to the Bcd binding arrays. Bcd is expressed in a gradient from the anterior to the posterior of the embryo. Even though Bcd is present at high levels in the anterior of the embryo, *Kr* enhancers do not drive expression there because of repression by Giant (Gt), Knirps (Kni), and Hb, which can act as both an activator and a repressor[25–27]. Therefore, we expected little effect on the anterior boundary by the Bcd binding site arrays. In contrast, since the posterior boundary is partially set by the decreasing levels of Bcd, if our Bcd arrays functionally reduce Bcd levels available for enhancer activation, we would expect to see a larger effect at the posterior boundary. We find that the posterior boundary of the shadow pair's expression domain moves towards the anterior in response to increasing number of competitor Bcd binding sites (Figure 3.3C). Relative to the homozygous configuration, the posterior border of shadow pair expression shifts anteriorly 2.5% of embryo length in the presence of the 3xBcd array and 5% in the presence of the 6xBcd array. Similar to peak expression levels, the 1xBcd array does not change the expression boundaries of the shadow pair reporter. We see similar anterior shifts of the duplicated distal expression pattern with the Bcd binding arrays that qualitatively match the decrease in peak expression seen with each array (Figure S3.3B). We note that the anterior boundary shifts towards the posterior in the presence of the 3xBcd and 6x Bcd arrays, which we suspect stems from the balance of activity between Bcd and the repressive TFs in this region[27–30].

3.3.5 Competition occurs at another genomic site and with an endogenous gene

Based on our findings thus far suggesting that reporter competition stems from competition for Bcd and other TFs, we reasoned that this competition should occur at other genomic insertion sites and with endogenous genes reliant on the same TFs. To first assess whether the observed reporter competition occurs at other genomic insertion sites, we measured the expression of the reporters in homozygous versus hemizygous configurations when inserted into a different chromosome (chromosome 3). Similar to our findings at the chromosome 2 insertion site (Figure 3.1), expression levels driven by the duplicated distal and shadow pair reporters are significantly lower in the presence of a second identical reporter (Figure 3.4A, B). On chromosome 3, expression in homozygous embryos is 82% and 75% of expression in hemizygous embryos for the duplicated distal and shadow pair reporters, respectively. With both of these reporters, the degree of competition is consistent between the two insertion sites, indicating that the observed competition occurs at different genomic locations (Figure 3.4 A and B insets).

Based on previous work in the mouse, we suspected that the reporter-induced competition would be limited to endogenous genes that are a short linear distance from the reporter insertion site[13]. To assess whether this is true, we measured the expression of three genes likely to be regulated by Bcd at varying linear distances from the chromosome 2 insertion site. We measured the expression of *Piezo* (22kb from insertion site), *Mcr* (58kb from insertion site), and *Btk29A* (160kb from insertion site) via qPCR in embryos with or without two copies of the duplicated distal transgene. All three of these genes are predicted to be regulated by Bcd, based on both previously measured expression patterns and Bcd binding near these genes in the early embryo[31,32] (Figure S3.4). In transgenic embryos, expression of the gene closest to the insertion site, *Piezo*, is significantly reduced to 60% of the levels seen in embryos of the same genetic background but lacking the transgene (Figure 3.4B). The expression levels of *Mcr* and

Btk29A, which are further removed from the transgenic insertion site, are not significantly changed in transgenic embryos (Figure 3.4B). This potential distance-dependent effect of our transgene on endogenous gene expression is consistent with our finding that, in homozygous reporter embryos, there is more competition in nuclei in which the MS2 spots are physically closer together (Figure S3.5B).

3.3.6 A hub-based model of TF-enhancer interactions predicts TF competition

We were initially surprised to find that a reporter construct, with a length less than 0.001% of the genome, can have measurable effects on the expression of both other reporters and a nearby endogenous gene. Even more surprising is our finding that Bcd binding site arrays, which do not themselves drive any expression and have as few as six binding sites, also significantly reduce the expression of our Bcd-regulated enhancer reporters (Figure 3.2 & 3.3). This suggests that competition for Bcd can be induced by the addition of a relatively small number of binding sites, despite the fact that Bcd copy numbers vary between approximately 1500 and 3000 molecules per nucleus in the region of *Kr* expression[10,33]. To better understand how the addition of a small sequence could induce competition for TFs, we developed a simple thermodynamic model of our system. The goal of our modeling effort is not to fit parameters such that the model precisely recapitulates our experimental data, but rather to see if our experimental observations are sensible by generating ballpark estimates of molecular competition using models that only rely on measured biophysical parameters.

Our model predicts the probability of a TF being bound to a target site, such as one of the binding sites that exist in an enhancer (Figure 3.5). For simplicity, we assume that TF binding at the target site is proportional to enhancer activity[34,35]. In reality, enhancer activity depends on the combined occupancy of many TF binding sites[36]. The simplifying assumption that enhancer

activity is proportional to binding site occupancy allows us to avoid the need to test multiple models with different components, such as cooperative TF binding or activation behavior. In addition to the target site, TF molecules can bind to specific or non-specific competitor sites. Since most TFs have sequence-independent affinity for DNA[37], the number of non-specific binding sites, *N*, is set to $1x10^6$, roughly the size of the *Drosophila melanogaster* genome. The number of specific competitor sites, *C*, is varied. To maintain the simplicity of the model, the binding energy of specific competitor sites, E_s, is equal to the binding energy of the target site, while the binding energy of all non-specific sites is represented as E_{ns}. Since the specific binding energy is representing that of multiple binding sites, which may differ in their affinities, we vary the difference between specific and non-specific binding energies. Lastly, to allow for comparisons with our experimental data and to measure the effect of TF levels on binding, we vary the levels of our input TF *T* as a function of embryo length, *l*, in accordance with the measured Bcd gradient[33]. In this way, we can look at how the probability of TF binding to a single target site, *p*(*bound; T(l)*), changes as a function of number of specific competitor sites, binding strength relative to non-specific binding, and TF abundance.

$$p(bound;T(l)) = \frac{\sum_{x=0}^{c} \frac{N^{T-1-x}}{(T-1-x)!} \times \frac{C!}{x!\,(C-x)!} \times e^{-[(T-x)\times E_{ns}+(1+xE_s)]}}{\sum_{x=0}^{c} \frac{N^{T-1-x}}{(T-1-x)!} \times \frac{C!}{x!\,(C-x)!} \times e^{-[(T-x)\times E_{ns}+(1+xE_s)]} + \sum_{x=0}^{c} \frac{N^{T-x}}{(T-x)!} \times \frac{C!}{x!\,(C-x)!} \times e^{-[(T-x)\times E_{ns}+xE_s} + \frac{N^T}{T!} \times e^{-TE_{ns}}} \quad (1)$$

As we vary the parameters, we find that *p(bound; T(l))* changes in a qualitatively intuitive way. *p(bound; T(l))* decreases as a function of increasing competitor sites, decreasing difference in specific and non-specific binding strength, and decreasing TF levels (Figure S3.6). To test the accuracy of our model, we compared our experimental measurements of expression changes as a

function of additional Bcd binding sites to predicted changes in *p(bound; T(l))* as a function of added competitor sites. In our model, we assume there are 2000 specific competitor sites, based on experimental measurements of genome-wide Bcd binding in nc14 embryos[32], and then add additional competitor sites to mimic the addition of a reporter of Bcd binding site array. Our findings are similar even if we do not assume these "background" sites exist (Figure S3.6). We recognize that the relationship between TF binding at an enhancer and gene transcription is complex[38–40] and do not expect our predicted *p(bound; T(l))* values to exactly predict gene expression levels. Still, gene expression is dependent on TF binding[41,42] and so our *p(bound; T(l))* values provide useful ballpark estimates of how gene expression is expected to change as new competitor sites are introduced.

We compared our model predictions to the experimentally measured changes in activity of the *hb*P2 reporter. The *hb*P2 enhancer is a well-studied, Bcd-responsive enhancer and therefore makes a useful point of comparison for our model of Bcd binding[43–45]. For simplicity, we compared our experimental data and model predictions at one position in the embryo, 27% egg length, where the *hb*P2 enhancer drives peak levels of expression in homozygous embryos. This means we hold *l*, and consequently *T*, constant and therefore refer to our model output as *p(bound)* from here on. To observe the effect specifically of introducing new specific competitor binding sites, we also used experimental measurements to estimate the difference between *Es* and *Ens* and held this constant (see Methods). In our experimental data, we see a 28% reduction in activity driven by the *hb*P2 reporter when a second reporter is present on the homologous chromosome. In contrast, our model predicts a 0.0003% decrease in *p(bound)* from the addition of 6 specific competitor sites, which is the number of known Bcd sites within the *hb*P2 reporter[46]. With this model, over 6,000 competitor sites are needed to get a 28% reduction in

p(bound) (Figure 3.5C). Similarly, while we observe a 49% decrease in *hb*P2 expression in the presence of the 6xBcd array, which contains 36 Bcd binding sites, our model predicts only a 0.002% decrease in *p(bound)* with this number of added competitor sites. Based on model predictions, 14,100 specific competitor sites need to be added to achieve a 49% reduction in *p(bound)*. Thus, a simple thermodynamic model of molecular competition produces estimates a couple of orders of magnitude different from experimental measurements.

We suspected that the large discrepancy between our measured decreases in reporter activity and our model's predictions of decrease in *p(bound)* are partially due to the model's assumption that any Bcd molecule in the nucleus can bind the target site. Growing evidence indicates that TFs and other pieces of the transcriptional machinery are not distributed evenly throughout the nucleus, but instead tend to cluster in regions of high density, called hubs, separated by low density regions[18,41,47–49]. This non-homogenous distribution seems functional, as transcription itself is also associated with these hubs[18,47,50]. Compared to the whole nucleus, hubs have a higher concentration of TFs and a lower number of specific and non-specific binding sites. We predicted that the addition of a small number of binding sites, similar to the numbers found in our reporter constructs, may have a sizable impact on *p(bound)* in the context of individual TF hubs.

To test this, we modified our previous model (*genome model*) to look at the probability of TF binding at the same target site, assuming all TF binding happens within hubs (*hub model*). In our hub model, we divide the nucleus into 1000 hub-sized regions, based on the size of *Drosophila* embryonic nuclei and previous estimates of the distance between enhancers associated with the same TF hub[18] (see Methods). Based on the measured distribution of distances between transcriptional spots in homozygous embryos, it is likely that reporters and TF

binding site arrays transiently co-localize to the same hub-sized region (Figure S3.7). Within each region, we assume there are 100,000 non-specific binding sites, which is the number of non-specific sites in the genome model ($1x10^8$) divided by 1000. The number of specific competitor sites is varied from 0 to 100. Based on previous measurements ,the number of Bcd TFs present in a hub, T_{hub}, is held constant at 20 molecules per hub, but the number of total Bcd molecules per nucleus, *T(l)*, follows the Bcd gradient along the embryo[41]. Regions that are not a hub are assumed to have 0 Bcd molecules. Consequently, the *p(bound)* value in our hub model is found by multiplying the *p(bound)* value calculated using the same formula as the genome model (equation 1) by the probability that a given region is a TF hub (*p(hub; T(l));* equation 2). As in our genome model, we varied the difference between specific and non-specific binding energies.

$$p(hub;\, T(l)) = \frac{\frac{T(l)}{T_{hub}}}{1000} \qquad (2)$$

In comparison to the genome model, the hub model shows far better agreement with our experimental data. As with the genome model, we assume that some specific competitor sites already exist and ask how *p(bound)* changes as additional specific competitor sites are added. In the hub model, we assume the 2000 specific competitor sites of the genome model are evenly distributed throughout the genome and consequently two specific competitor sites are present in each sub-region of the nucleus. We again focus on one position in the embryo, 27% egg length, and therefore hold *T(l)* constant. Experimentally, we see a 28% decrease in the activity of the *hb*P2 reporter with the addition of a second *hb*P2 reporter. The hub model predicts a 5% decrease in *p(bound)* from the addition of the 6 Bcd binding sites in the *hb*P2 reporter (Figure 3.5D). Unlike the genome model, which requires over 6000 competitor sites for a 28% reduction in

p(bound), this magnitude reduction is achieved by 22 competitor sites in the hub model. With 36 competitor sites, the number of Bcd binding sites in the 6xBcd array, the hub model predicts a 46% decrease in *p(bound)* compared to the 49% decrease in *hb*P2 reporter expression we measure in the presence of the 6xBcd array.

It is notable the hub model better predicts the effect of a 6xBcd array than the second *hb*P2 reporter. While there are many simplifying assumptions in the model, for example, assuming *p(bound)* is proportional to expression output, there also is a key difference between the two experimental measurements. The 6xBcd array lacks a promoter, while the *hb*P2 reporter actively drives transcription. The model assumes that the only effect of adding the *hb*P2 reporter is the addition of competitor Bcd binding sites, but this reporter may also siphon away other key pieces of transcriptional machinery, which may explain why the measured effect of adding the *hb*P2 reporter is larger than predicted by either model.

3.4 Discussion

Since the discovery of enhancers 40 years ago[51,52], transgenic reporters have been invaluable tools to study the principles governing *cis*-regulatory regions. With a few exceptions, it has largely been assumed that transgenic reporters do not meaningfully affect the expression of other genes. Here we challenge this assumption and investigate the observed competition between transgenic transcriptional reporters in developing *Drosophila* embryos. Using reporters controlled by different configurations of the *Kruppel* shadow enhancers, we show that expression of a single reporter is decreased in the presence of a second identical reporter. We further show that this effect is not limited to transgenic reporters, but that the expression of a nearby endogenous gene is also decreased in transgenic embryos. Using non-transcribing arrays of TF binding sites, we find evidence that decreased reporter expression is due in part to decreased

availability of key activating TFs of the *Kr* enhancers. Focusing on enhancer competition for the TF Bcd, we show that competitor Bcd binding arrays specifically affect the expression of Bcd-regulated enhancers, have a dosage-dependent effect on these enhancers, and shrink the width of the expression pattern of the enhancers. By developing a simple thermodynamic model, we predicted that the introduction of tens of additional Bcd binding sites can appreciably decrease gene expression, but only when TF binding is assumed to be limited to nuclear subregions.

3.4.1 Transgenic reporters can affect the expression of other reporters and an endogenous gene

Due to the widespread use of transgenic reporters, we were surprised to find that our small reporters reduce the expression of not only other reporters, but a nearby endogenous gene. A deeper search of literature revealed that Laboulaye, et al., also found a distance-dependent decrease in endogenous gene expression in mice that is similar to our own results[13]. While further systematic investigation is needed, these similar results in these distantly related organisms suggest that decreased endogenous gene expression may be a common consequence of transgenic reporters.

At first glance, our findings are also reminiscent of the transgene silencing previously reported in *Drosophila*[53–55]. Pioneering studies found that flies containing multiple copies of a transgene showed reduced expression of the transgene as well as the corresponding endogenous gene[56,57]. This silencing was shown to depend on Polycomb-mediated repression in the case of transgenes, and post-transcriptional RNAi mechanisms in the case of the endogenous gene[55]. While our findings share some similarities with transgene silencing, and may well rely on related mechanisms, numerous differences lead us to believe we are observing a distinct phenomenon. Unlike these previous studies, our transgenic flies, which contain a mini-*white* marker, do not show lighter eye color in homozygotes compared to hemizygotes (Figure S3.8). This suggests

that overall expression from our transgenic insertion sites is not being ubiquitously repressed. If our observations were only the consequence of silencing mechanisms, triggered by increased amounts of transgenic DNA, we would expect to see larger reporter competition effects with our duplicated enhancer constructs compared to the single enhancer versions. Instead, for both the distal and proximal enhancers, we see a trend of larger decreases in homozygous expression levels compared to hemizygous expression levels with the single enhancer constructs (Figure 3.1B). Further, unlike the findings of Pal-Bhadra, our transgenic reporter decreases expression of an endogenous gene with which it does not share sequence homology[55].

3.4.2 Reporters and non-transcribed DNA sequences can induce competition for TFs

In addition to the studies described above, which describe how a transgene can alter the expression of genes, there are several studies that describe how the presence of non-transcribing pieces of DNA can alter expression. Work in flies and mouse cells showed that highly repetitive genomic sequences can alter gene expression, likely by binding and sequestering TF molecules away from their target genes[58,59]. In yeast, repetitive sequences of "decoy" tetO TF binding sites can change the relationship between tetO levels and the expression of a gene regulated by tetO[60], and individual competitor binding sites in bacteria also have a similar effect[61]. These studies underscore the regulatory importance of repetitive non-coding DNA sequences, which make up the majority of many genomes, by titrating available TF levels.

The repetitive sequences investigated in these studies are much longer (6Mb of major satellite DNA in mice, 7Mb of satellite V DNA in flies) than our transgenic constructs, which all contain less than 5kb of regulatory sequence and 10s of TF binding sites[59,62]. The large effect our transgenic constructs have on gene expression levels are therefore initially surprising. It is easier to imagine how very large, repetitive DNA sequences could sequester meaningful amounts of

TFs than small sequences containing as few as six TF binding sites. In particular, this competition is surprising because we observe it even in the peak regions of the reporter expression patterns, where we expect activating TF levels to be high. For example, there are 250,000 molecules of the TF Zld per nucleus in the embryo[11,63], yet we see evidence of competition by introducing only six new strong Zld binding sites.

We suspect that the effect of our reporters and binding site arrays on expression levels, as well as the effect of large repetitive sequences, partially stems from the non-uniform distribution of TFs in the nucleus. Although heterogeneity in the nucleus has been long observed with DNA[64,65], recent studies revealed that TFs and other pieces of the transcriptional machinery are also distributed unevenly throughout the nucleus[18,41,47–49]. There are several potential consequences of the organization of TFs into hubs. First, if our competitive binding arrays end up outside of a so-called TF "hub" with the enhancer reporter (or nearby endogenous genes), TF levels functionally available to the enhancer may be low enough to disrupt reporter activity. Second, if our binding arrays and reporters are found in the same hub, they may be competing for a fairly small pool of TFs. Previous measurements suggest there are roughly 20 Bcd molecules per hub[66]. Lastly, the presence of a binding array may affect the properties of the hubs themselves. Another study showed that the deletion of TF-recruiting enhancers can decrease TF hub size and therefore lower gene expression[18]. In addition, Zld plays a key role in the formation of Bcd hubs[41], suggesting that our Zld binding site arrays may sequester both Zld and Bcd molecules.

Several aspects of our data support the hypothesis of local competition. First, reporters that spend more time in close physical proximity in the nucleus compete more than reporters that are further apart (Figure S3.5). Similarly, the endogenous gene *Piezo*, whose expression is

decreased in the presence of the duplicated distal reporter, is within the same topologically associating domain (TAD) as the insertion site of the transgene during nc14[67] (Figure S3.9). This suggests that *Piezo* and the reporter likely inhabit the same nuclear subregion and have access to the same local pool of TFs.

3.4.3 Thermodynamic model of TF binding implicates TF hubs in competition

In line with our experimental data, our modeling results suggest that local competition for TFs is consistent with the observed decrease in expression levels. To rationalize how the addition of a small number of competitor TF binding sites could meaningfully decrease expression levels, we developed two simple thermodynamic models of TF binding. Our hub model, which assumes all TF binding is restricted to nuclear subregions matches our experimental data more closely than the genome model, which assumes that all TF molecules have access to the whole genome. Our findings suggest an unexplored consequence of TF hubs. Previous studies have shown that TF hubs help to increase local concentration of TFs to increase gene expression[18,50]. Here, we show the flip side of this coin -- the non-uniform distribution of TFs can also induce competition among binding sites. We note that we have used only strong binding sites in our competitor arrays and plan to test the effect of binding arrays consisting of non-optimal TF binding sites, as enhancers containing sub-optimal binding sites have been shown to be important for establishing TF hubs[18]. There may be a balance between sequestering TFs, as we see here, and recruiting TFs to a local region that could depend on the affinity of the binding sites present.

Our goal in developing a simple model of our system was to generate ballpark predictions about the behavior of the system, using experimentally measured parameters and minimal assumptions. While our hub model better matches our experimental findings than the genome model, we recognize that it is a simplification of reality and as such cannot fully describe our

system. For example, we assume that existing specific competitor sites are evenly distributed throughout the genome, but in reality, chromatin accessibility and the clustering of TF binding sites in cis-regulatory regions alters the distribution of available binding sites[63,68]. The "true" number of specific competitor sites in any given sub-nuclear region will vary and consequently *p(bound)* of a given target site will depend on the surrounding sequences in the same region (Figure S3.10). With these simplifying assumptions, come an incomplete ability to explain some experimental data. We find that expression levels driven by the duplicated distal reporter significantly decrease in the presence of the smallest and largest of our Bcd binding site arrays, but, unexpectedly, are not affected by the presence of the intermediate sized array (Figure S3.3). We do not fully understand this observation, but suspect that it has to do with the exact recruitment of TFs and other molecular factors mediated by this combination of DNA sequences.

3.4.4 Implications for transgenic reporters and underlying biology

Our work adds to the evidence that transgenic reporters can have measurable effects on endogenous gene expression[13,56] and also builds on our understanding of the mechanisms behind this phenomenon. We note that our transgenic fly lines develop without any gross phenotypic defects in ideal laboratory conditions, making it tempting to assume that any effects of transgenic reporters are negligible. While much about the mechanisms and effects of reporters on endogenous gene expression remains to be discovered, our findings provide some practical lessons for using transgenic reporters. First, investigators should use caution in interpreting changes in expression levels or patterns when comparing assays using one reporter to those using multiple reporters simultaneously. We find clear evidence that our reporters compete with one another when present in the same nucleus and as this seems to be mediated by competition for

TFs, we suspect this finding is true beyond our specific reporters and system. Additionally, potential effects of reporters on nearby endogenous gene expression should be considered in study design and data interpretation.

Beyond the implications for the use of transgenic reporters, our findings suggest that the distribution of TF binding sites, both in the genome and in 3D space, is a potential tuning mechanism for dose-response relationships between TF levels and target genes. Previous studies in bacteria and yeast have shown that competitor TF binding sites can modulate the dose-response relationship of TF levels and gene expression, and that this modulation depends on the relative affinity of competitor versus gene-regulating TF binding sites[60,61]. This effect may generate an unappreciated selection pressure to either retain or eliminate TF binding sites that are not directly regulating a specific target gene. The observations of TF sequestration across a wide range of organisms suggest that this phenomenon is conserved and likely plays a functional role in regulating gene expression.

Acknowledgements:

We thank Hernan Garcia for flies containing MCP-GFP and His-RFP transgenes, as well as for useful discussion of our observations. We thank Lily Li for helpful suggestions on the model and all members of the Wunderlich lab for feedback on the project. We also thank German Enciso, Anthony Long, Thomas Schilling, and Rahul Warrior for helpful feedback and suggestions on the project.

Author Contributions:

RW: conceptualization, software, formal analysis, investigation, writing - original draft, writing - review & editing, visualization. **MG:** Investigation, resources, writing - review & editing. **ZW:** conceptualization, resources, writing - review & editing, supervision, funding acquisition.

Declaration of Interests:

The authors declare no competing interests.

3.5 Methods

3.5.1 Generation of transgenic fly lines

Transgenic fly lines containing an enhancer-MS2 reporter were generated by phiC31-mediated insertion into the second or third chromosome, as described in Waymack, et al., 2020[15]. Unless otherwise indicated, all reporter constructs and TF binding site arrays were integrated into the same site on the second chromosome via phiC31-mediated integration. These constructs were injected into y[1] w[1118]; PBac{y[+]-attP-3B}VK00002 (BDRC stock #9723) embryos by BestGene Inc (Chino Hills, CA). For the reporter constructs inserted into chromosome 3, plasmids were injected into y[1] w[1118]; PBac{y[+]-attP-3B}VK00033 (BDRC stock #9750) embryos by BestGene Inc (Chino Hills, CA). The *Kruppel* enhancer reporters contained a single, duplicated, or shadow enhancer pair and the *Kruppel* promoter upstream of 24 MS2 repeats and a *yellow* reporter gene cloned into the pBphi vector[16]. These are the same enhancer-MS2 reporters as used and described in Waymack, et al., 2020. The *hunchback* P2 enhancer reporter is that used in Garcia, et al., 2013 and consists of the *hunchback* P2 enhancer and P2 promoter upstream of 24 MS2 repeats and a *lacZ* reporter[16]. Exact genomic sequences used in each reporter construct are given in Supplementary file 1.

Hemizygous embryos were generated by crossing male flies homozygous for an enhancer-MS2 reporter to females expressing RFP-tagged histones and MCP-GFP[16]. Homozygous embryos were generated by crossing virgin females of the F1 hemizygous offspring just described with males homozygous for the same enhancer-MS2 reporter. Embryos with one copy of an enhancer reporter and one copy of a TF binding site array were generated by crossing the virgin female hemizygous offspring (i.e. containing one enhancer-MS2 reporter allele) with males homozygous for the corresponding TF binding site array.

3.5.2 Generation of TF binding site arrays

To generate our competitor binding site arrays for the four different TFs investigated, we started with the sequence of the *hb* P2 enhancer, which is well known to be Bcd responsive and contains six Bcd binding sites[43–45]. This 236bp sequence was our 1xBcd array and the starting point for our other binding site arrays. To generate the Hb, Zld, and Stat92E arrays used we modified the six Bcd binding sites of the *hb* P2 enhancer to be the consensus motif for the corresponding TF (while retaining the same 236bp total length of the array) and had these sequences synthesized by Integrated DNA Technologies Inc (San Diego, CA). Previously defined consensus motifs were used for Zld[69], Hb[30], and Stat92E[70]. To generate the 3x and 6xBcd arrays we performed Golden Gate assembly to ligate three or six copies of the 1xBcd array together with 10bp random sequences between each repeat, to avoid potential repeat removal during transformations. All of the described TF binding site arrays were inserted into the same plasmid backbone, which was a modified version of the pBphi vector used for our enhancer-MS2 reporters, which lacks any enhancers, promoters, or MS2 sequence. We generated this vector by removing the *Kr* distal enhancer, *Kr* promoter, MS2 cassette, *yellow* sequence, and termination sequence from our distal MS2 reporter through digestion with NotI and XbaI. We then used Gibson assembly to ligate in

the appropriate TF binding site array to this backbone. Sequences for the TF binding site arrays are provided in Supplemental File 1.

3.5.3 Embryo preparation and image acquisition

Living embryos were collected and dechorinated before being mounted onto a permeable membrane in halocarbon 27 oil and placed under a glass coverslip as in Garcia, et al., 2013. Individual embryos were then imaged as described in Waymack, et al., on a Nikon A1R point scanning confocal microscope using a 60X/1.4 N.A. oil immersion objective and laser settings of 40uW for 488nm and 35uW for 561nm[15]. To track transcription, 21 slice Z-stacks, at 0.5um steps, were taken throughout the length of nc14 at roughly 30s intervals. To identify the imaged position in the embryo, the whole embryo was imaged after nc14 prior to gastrulation at 20X using the same laser power settings. This whole embryo image was used to assign each transcription spot into one of 42 bins across the anterior-posterior (AP) axis of the embryo. The first bin corresponds to the anterior end of the embryo.

3.5.4 Measurement of transcriptional reporter activity

Tracking of nuclei and MCP-GFP bound MS2 transcriptional spots was done using the image analysis Matlab pipeline described in Garcia et al., 2013, which can be accessed at the Garcia lab Github (https://github.com/GarciaLab/mRNADynamics). Calling of transcriptional bursts to use for analysis was done as in Waymack et al., 2020. In short, transcriptional traces captured during nc14 consisting of at least three points were used for analysis. To measure total mRNA produced by all of our reporter configurations, we integrated the area under the curve of the transcriptional spot's fluorescence across the time of nc14 (Figure S3.1).

For every tracked spot of transcription, background fluorescence at each time point is estimated as the offset of fitting the 2D maximum projection of the Z-stack image centered around the transcriptional spot to a gaussian curve, using Matlab *lsqnonlin*. This background estimate is subtracted from the raw spot fluorescence intensity. The resulting fluorescence traces across the time of nc14 are then subject to smoothing by the LOWESS method with a span of 10%. The smoothed traces were used to measure transcriptional parameters and noise. Traces consisting of fewer than three time frames were removed from calculations. The area under each smoothed transcriptional trace is integrated using the Matlab *trapz* function, which gives the total integrated fluorescence value for that transcriptional spot. This integrated fluorescence is proportional to the number of transcripts produced by an enhancer reporter[16,71]. We group all transcriptional spots of a given reporter configuration by AP bin (position in the embryo) and calculate the average total integrated fluorescence value in each AP bin. For each reporter configuration we identify the AP bin with the highest average integrated fluorescence value as the region of peak expression. In the text, unless otherwise indicated, the integrated fluorescence or peak expression values correspond to the average integrated fluorescence value at this AP bin (Figure S3.1).

3.5.5 qRT-PCR to measure expression of endogenous genes in varying genetic backgrounds

Flies were allowed to lay eggs on molasses plates for 2.5 hours, so that most embryos collected were in nc14. Flies were either homozygous for the duplicated distal reporter on chromosome 2 or of the same genetic background (BDSC #9723) but did not have the transgene. The embryos collected from each plate were pooled and total RNA was extracted and purified using TRIzol (Thermo Fisher Scientific) and the Direct-zol RNA Miniprep kit (Zymo Research). cDNA was generated using SuperScript III RT Kit (Thermo Fisher Scientific). qPCR amplification was then

done using the TaqMan Gene Expression Master Mix (Thermo Fisher Scientific). The data for each group, transgenic or non-transgenic, is from three separate biological replicates (i.e. the colored circles in Figure 3.4C are biological replicates), each done in technical triplicates. Relative RNA levels of each measured gene was calculated using the $2^{-ddC(t)}$ method, using *RpII140* as the reference gene. The TaqMan FAM probes used for each gene were DM01803576_g1 for *Piezo*, Dm01825813_g1 for *Mcr*, Dm01803642_g1 for *Btk29A*, and DM02134593_g1 for *RpII140*

3.5.6 Description of the genome model of TF binding

We developed a simple thermodynamic model that looks at the probability of a TF molecule being bound at a single target site. For simplicity, we assume all TF molecules are bound either specifically or non-specifically. The probability of TF being bound at the target site, *p(bound)*, is then determined by the number of TF molecules, *T(l)*, the number of non-specific competitor binding sites, *N*, the number of specific competitor sites, *C*, and the difference in specific versus non-specific binding energies, E_s and E_{ns} respectively. The number of TF molecules, *T(l)*, follows the Bcd gradient[33] and is determined by position in the embryo, *l*, with a maximum value of 20,000 at the anterior tip of the modeled embryo[10]. For ease of comparison with our experimental data, we only consider binding probability at one position (*l* = 27% egg length) and thereby hold *T* constant, unless otherwise indicated. We hold the number of non-specific binding sites, *N*, constant at $1x10^8$. Using statistical mechanics, we first enumerated the possible states of our system and their associated Boltzman weights[34]. In these states, x indicates the number of TFs that are bound at specific competitor sites.

TF binding configuration	**State**	**Statistical weight**
Non-specific binding only	$\frac{N^T}{T!}$	$e^{-TE_{ns}}$
Competitor sites and non-specific binding	$\sum_{x=0}^{C} \frac{N^{T-x}}{(T-x)!} \times \frac{C!}{x!\,(C-x)!}$	$e^{-[(T-x)E_{ns}+xE_S]}$
Target site, competitor sites and non-specific binding	$\sum_{x=0}^{C} \frac{N^{T-1-x}}{(T-1-x)!} \times \frac{C!}{x!(C-x)!}$	$e^{-[(T-1-x)E_{ns}+xE_S+E_S]}$

With all of the possible states of the system and the associated statistical weights, we can calculate *p(bound)* by dividing the statistical weight of the state with TF bound at the target site by the combined statistical weights of all possible states:

$$p(bound) = \frac{\sum_{x=0}^{C} \frac{N^{T-1-x}}{(T-1-x)!} \times \frac{C!}{x!(C-x)!} \times e^{-[(T-x)\times E_{ns}+(1+xE_S)]}}{\sum_{x=0}^{C} \frac{N^{T-1-x}}{(T-1-x)!} \times \frac{C!}{x!(C-x)!} \times e^{-[(T-x)\times E_{ns}+(1+xE_S)]} + \sum_{x=0}^{C} \frac{N^{T-x}}{(T-x)!} \times \frac{C!}{x!(C-x)!} \times e^{-[(T-x)\times E_{ns}+xE_S+\frac{N^T}{T!}\times e^{-TE_{ns}}}} \quad (1)$$

With this equation we can calculate the probability of a TF molecule being bound at the target site for a given number of TF molecules, specific competitor sites, and difference in binding affinity at specific vs non-specific sites. In the main text (Figure 3.5), we assume 2,000 background competitor sites already exist and span a range of 0 to 100,000 added specific competitor sites. To facilitate comparison with our experimental data, we looked at binding probability at one position in the embryo by holding *l* and consequently *T* constant. We focus on

the effect of adding specific competitor binding sites and as such hold the difference between E_{ns} and E_s , and *T(l)* constant. E_{ns} is held at $0k_BT$ and E_s is held at $-10k_BT$, based on previously measured differences in specific vs non-specific DNA binding[34,72]. Specifically, the formula $\Delta E = k_B T ln(\frac{k_S}{k_{ns}})$ was applied to the measured binding affinities, from Jung, et al., of TFs to their consensus sequence and highly mutated consensus sequences[72]. *T(l)* is held at 5468, corresponding to 27% egg length in the modeled embryo. In Supplemental Figure 3.6, we explore how *p(bound)* changes as a function of our other parameters (i.e. *T(l),* the difference between E_{ns} and E_s, or the number of background specific competitor sites).

3.5.7 Description of the hub model of TF binding

Calculation of *p(bound)* in the hub model is similar to the genome model but assumes all TF molecules are restricted to "hub" regions (sub-regions of the nucleus containing a high concentration of TFs) and therefore takes into account the probability that the nuclear sub-region containing the target site is a TF hub. In the hub model, each nucleus is divided into 1000 equal-sized regions with a radius of 256nm. This estimate for the size of nuclear regions was based on the average distance between interacting loci found by Tsai, et al.[18], of approximately 360nm, the approximate volume of *Drosophila* embryonic nuclei of $70um^3$, and an estimation of the amount of DNA contained within a TF hub of the size seen in Tsai, et al. The nuclear volume of 70um3 was reached by estimating the nucleus to be a sphere and using the formula $V = 43\pi r^3$ with r = 2.5um (estimated from imaging data).

As the nucleus is divided into 1000 hub regions, we set the number of non-specific binding sites, N, to 100,000, which is 1/1000th of the value in the genome model (10^8). For simplicity, we assume that DNA is distributed uniformly in the nucleus and as such the amount of DNA in each

region is the same, which also allows us to have the same number of total non-specific binding sites per nucleus as in the genome model (10^5 x 1,000 hubs = 10^8). To maintain the same number of total specific competitor sites as the genome model, we assume 2 background specific competitor sites per region and vary the number of added specific competitor sites per region from 0 to 100. These values were reached by dividing the number of background or added specific competitor sites from the genome model by 1,000 (2,000 / 1,000 = 2 and 100,000 / 1,000 = 100). Based on the observation of Mir, et al.,[66] that the number of Bcd molecules per hub did not change along the Bcd gradient, we hold *T* constant at 20 if a region is a TF hub or 0 if it is not a TF hub. To account for this additional condition of whether the target site is within a TF hub or not, we calculate the probability that the region containing the target site is a TF hub:

$$p(hub) = \frac{\frac{T(l)}{T_{hub}}}{1000} \qquad (2)$$

Where T_{hub} is 20 TF molecules found in a hub and *T(l)* is the total number of TFs in the nucleus, as determined by the Bcd gradient[33]. To obtain the final *p(bound)* value from the hub model the *p(hub)* value of equation 2 is multiplied by the value obtained using the above parameters in equation 1.

3.5.8 Plotting p(bound) as a function of added specific competitor sites

To best simulate our experimental system, where we add transgenic constructs containing Bcd binding sites into a genome that already contains Bcd binding sites, we focused on how *p(bound)* changes as a function of added specific competitor sites. We therefore hold *l*, and consequently nuclear TF levels *T*, constant. Similarly, we hold constant the difference in binding energy

between specific and non-specific sites by setting E_s to -10 and E_{ns} to 0. To account for specific Bcd binding sites that already exist in the *Drosophila* genome, we set the *p(bound)* value when a set number of "background" competitor sites exist as our reference maximum *p(bound)* value. We estimated the total number of true Bcd binding sites in nc14 embryos to be 2,000 based on the number of genome-wide Bcd ChIP-seq peaks reported by Hannon, et al[32]. Therefore for the genome model, the graph shown in Figure 3.5B depicts how *p(bound)* changes as a function of additional specific competitor sites beyond 2,000. For the hub model, we assume that these 2,000 Bcd binding sites are equally distributed throughout the genome and consequently there are two specific Bcd binding sites per nuclear region (2,000 / 1,000 = 2) . The graph in Figure 3.5D shows how *p(bound)* changes as a function of additional specific competitor sites per nuclear region beyond the baseline two.

3.5.9 Statistical Methods

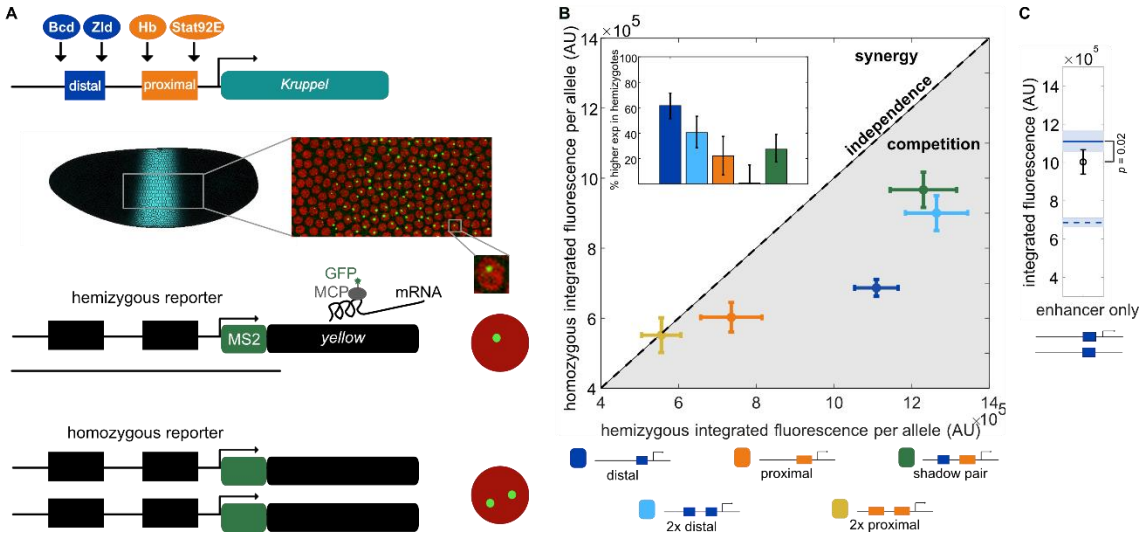
To determine statistical differences in levels of competition (Figures 3.1 and 3.4) and expression boundaries (Figure 3.3 and Supplemental Figure 3) between reporters we performed bootstrapping to estimate 95% confidence intervals. To do so, we randomly sampled with replacement the integrated fluorescence values of all of the transcriptional spots tracked in the AP bin of peak expression for both the hemizygous and homozygous configurations of the respective enhancer reporters. We averaged this value for the hemizygous configuration and for the homozygous configuration and then divided this average homozygous expression by the hemizygous expression to get our competition value (i.e. % hemizygous expression). This was done 1,000 times and each time the difference between the competition value found using the original real data set and that found using the randomly resampled data was calculated. We then

took the top and bottom 2.5 percentiles of these differences as our upper and lower error bounds, respectively.

We estimated the error in expression boundaries in a similar fashion. We again perform 1,000 rounds of bootstrapping by randomly sampling, with replacement, the integrated fluorescence values from rows of transcriptional spots along the AP embryo axis for a given enhancer construct. Each column corresponds to a single AP bin in the embryo. We randomly sampled rows equal to the total number of rows in the original data set and using these found the anterior-most and posterior-most AP bins that produce greater than or equal to 50% of the maximum expression measured in the hemizygous configuration of that reporter. Empirical 95% confidence intervals were calculated as above by finding the 2.5th and 97.5th percentiles of the distribution of differences between the expression boundaries found using the original data and those found using each iteration of resampled data.

To determine whether the distal enhancer reporter produces significantly lower expression levels in the presence of a non-transcribing distal enhancer (Figure 3.1C) we performed a t-test comparing the integrated fluorescence values recorded in the region of peak expression of the two configurations.

Figure 3.1 – Differences in mRNA production in homozygous and hemizygous embryos suggest competition between reporters. A. This panel is adapted from Waymack, et al.[15]. (Top) *Kr* expression in the early embryo is controlled by the activity of a pair of shadow enhancers, termed proximal and distal based on their location relative to the *Kr* promoter, that are each activated by different transcription factors (TFs). (Middle) The expression pattern driven by this pair of shadow enhancers is a stripe in the center 20% of the embryo. We use the MS2 system to image active transcription driven by enhancer reporters in living embryos. The cut out from embryo shows a still frame of a movie where red circles are nuclei and green spots are sites of active transcription. To test whether transgenic reporters affect each other's expression, we generated embryos that are either homozygous or hemizygous for a particular reporter construct. (Bottom) Hemizygous embryos have the enhancer-MS2 reporter inserted on only one homologous chromosome and therefore display one transcriptional spot per nucleus. Homozygous embryos have the same enhancer-MS2 reporter inserted at the same location on both homologous chromosomes and therefore display two transcriptional spots per nucleus. **B**. mRNA production from homozygous reporter constructs compared to production from hemizygous constructs suggests competition between reporters. The graph shows total mRNA produced per allele in homozygous embryos as a function of total mRNA produced per allele in hemizygous embryos for the reporter construct indicated. The dashed diagonal line represents expected expression assuming independent activity of the two reporters in homozygous embryos. Points falling above this line display synergy in the activity of the two reporters in homozygotes and points falling below display competition. Error bars indicate 95% confidence intervals. Inset shows the percent higher expression in hemizygous versus homozygous embryos for each reporter construct. Error bars represent 95% confidence intervals from 1000 rounds of bootstrapping. **C**. To rule out reporter competition being an artifact of our imaging system, we measured expression of the distal enhancer MS2 reporter in the presence of a non-transcribing distal enhancer on the homologous chromosome. The second distal enhancer is identical to the reporter, but lacks both a promoter and MS2 sequence. The graph shows expression driven by the distal enhancer reporter is significantly reduced in the presence of the non-transcribing distal enhancer (*p*-value = 0.02, t-test). The top (solid) horizontal line indicates expression driven by the distal enhancer reporter in the hemizygous configuration and the bottom (dashed) horizontal line indicates peak expression per allele driven in the homozygous configuration. Error bars and shading represent 95% confidence intervals.


A
Bcd
Zld
Hb
Stat92E
distal
proximal
Kruppel
hemizygous reporter
GFP
MCP
mRNA
MS2
yellow
homozygous reporter
B
synergy
independence
competition
% higher exp in hemizygotes
homozygous integrated fluorescence per allele (AU)
hemizygous integrated fluorescence per allele (AU)
distal
proximal
shadow pair
2x distal
2x proximal
C
integrated fluorescence (AU)
p = 0.02
enhancer only

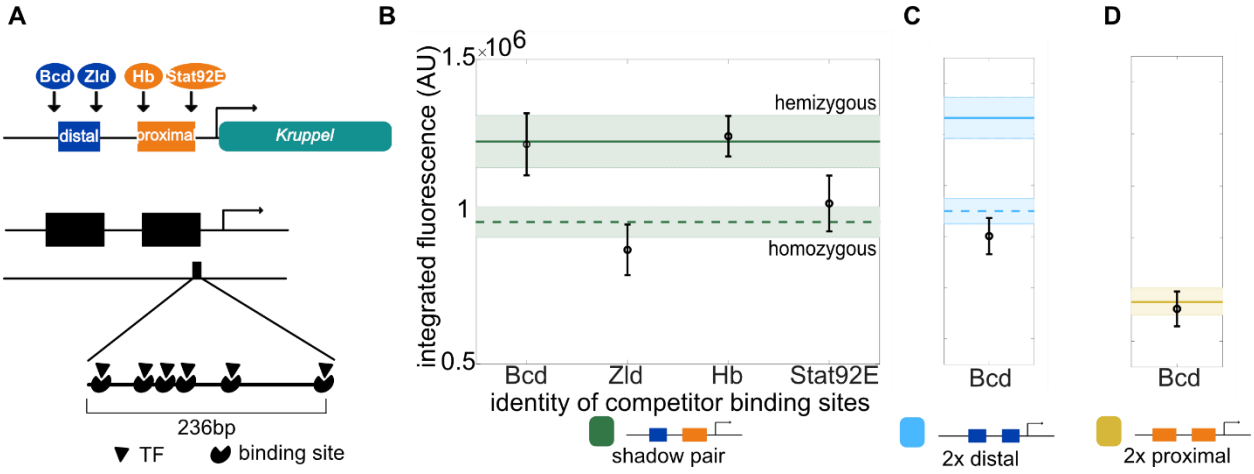
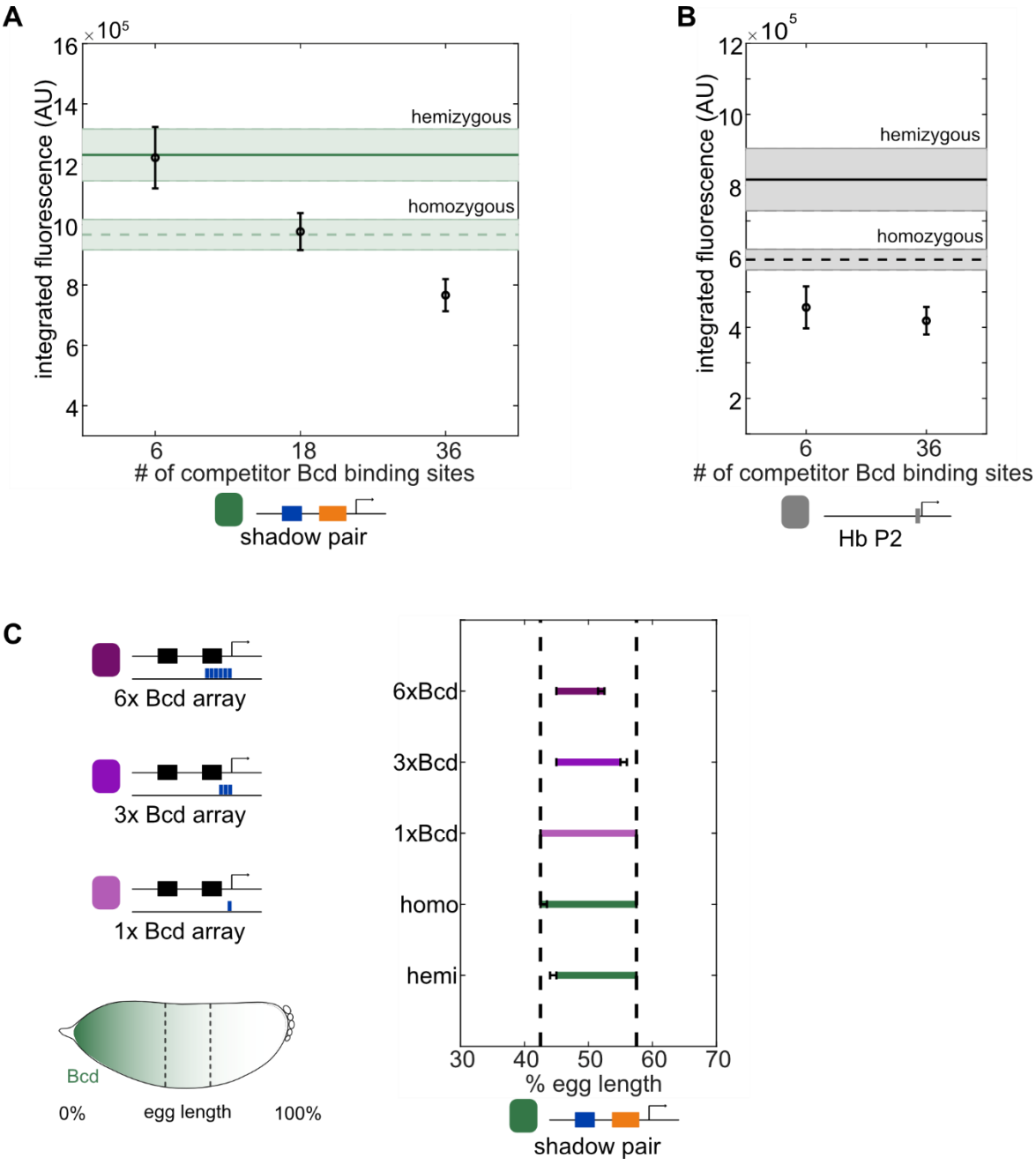


Figure 3.2 - Competitor TF binding sites on homologous chromosome decrease reporter activity. To test whether limiting levels of one or more activating TFs contribute to the reporter competition we observe, we measured the activity of our reporters in the presence of TF binding site arrays. **A**. (Top) The *Kr* shadow enhancers are activated by different sets of TFs. (Bottom) A schematic of TF binding site arrays that are intended to act as sinks for TF molecules. The arrays are each 236bp long, contain six binding sites for the indicated TF, and are inserted at the same genomic site as enhancer-MS2 reporters on the homologous chromosome. The binding site arrays do not contain a promoter or MS2 sequence. **B**. The activity of the shadow pair reporter is reduced in the presence of some TF binding site arrays. Graph shows the peak expression of the shadow pair in the presence of the indicated TF binding site array on the homologous chromosome. In B-D, the horizontal solid line indicates the peak expression level in hemizygous embryos of the indicated reporter construct and the horizontal dashed line indicates the peak expression level per allele in homozygous embryos. **C**. The activity of the duplicated distal reporter is reduced to homozygous levels when the Bcd binding array is present on the homologous chromosome. **D**. Activity of the duplicated proximal reporter, which is not activated by Bcd, is not reduced when the Bcd binding array is present on the homologous chromosome. Note that the homozygous and hemizygous peak expression levels (dashed and solid horizontal lines) overlap for the duplicated proximal reporter. Error bars and shading in B-D indicate 95% confidence intervals.

Figure 3.3 – **Competitor Bicoid binding sites decrease and shift the activity of shadow pair reporter**. To assess whether limiting levels of the activating TF Bicoid (Bcd) cause the apparent competition between reporters observed, we measured the transcriptional output of the shadow pair construct in the presence of Bcd binding site arrays of increasing length on the homologous chromosome. The 1xBcd binding site array consists of six Bcd binding sites but lacks a promoter or MS2 cassette. The 3xBcd and 6xBcd binding site arrays are three and six repeats, respectively, of the 1xBcd array and therefore contain a total of 18 and 36 Bcd sites, while also both lacking a promoter or MS2 cassette. These binding site arrays were inserted into the same location on Chromosome 2 as the enhancer reporters. **A**. Peak expression per allele driven by the shadow pair reporter decreases as the number of competitor Bcd binding sites increases. The horizontal lines mark the peak total expression per allele driven by the shadow pair reporter as hemizygotes (top solid line) or homozygotes (bottom dashed line). Shading and error bars indicate 95% confidence intervals. **B**. Competitor Bcd binding site arrays decrease the expression of an unrelated Bcd-responsive enhancer. To test if the effect of the Bcd binding arrays is specific to the *Kr* enhancers, we measured expression driven by the *hunchback* P2 (*hb*P2) enhancer, which is also activated by Bcd, in the presence of the 1x and 6xBcd binding site arrays. The graph shows the peak expression driven by the *hb*P2 reporter in the presence of the indicated Bcd binding site arrays. Shading and error bars in A and B indicate 95% confidence intervals. **C**. (Left) Bcd is expressed in a gradient from the anterior of the embryo (0% egg length) to the posterior (100% egg length). The *Kr* expression domain is indicated by dashed vertical lines. Schematics above the embryo diagram show the 1x, 3x, or 6xBcd arrays used with the enhancer reporters. (Right) Expression patterns driven by the shadow pair reporter in the presence of increasing numbers of competitor Bcd binding sites. Graph shows the range of the expression pattern of each configuration to 50% of peak expression levels of the homozygous configuration, whose boundaries are indicated with dashed vertical lines. Error bars represent 95% confidence intervals found from 1000 rounds of bootstrapping.


A
×10^5
integrated fluorescence (AU)
hemizygous
homozygous
# of competitor Bcd binding sites
shadow pair
B
×10^5
integrated fluorescence (AU)
hemizygous
homozygous
# of competitor Bcd binding sites
Hb P2
C
6x Bcd array
3x Bcd array
1x Bcd array
Bcd
0%
egg length
100%
6xBcd
3xBcd
1xBcd
homo
hemi
% egg length
shadow pair

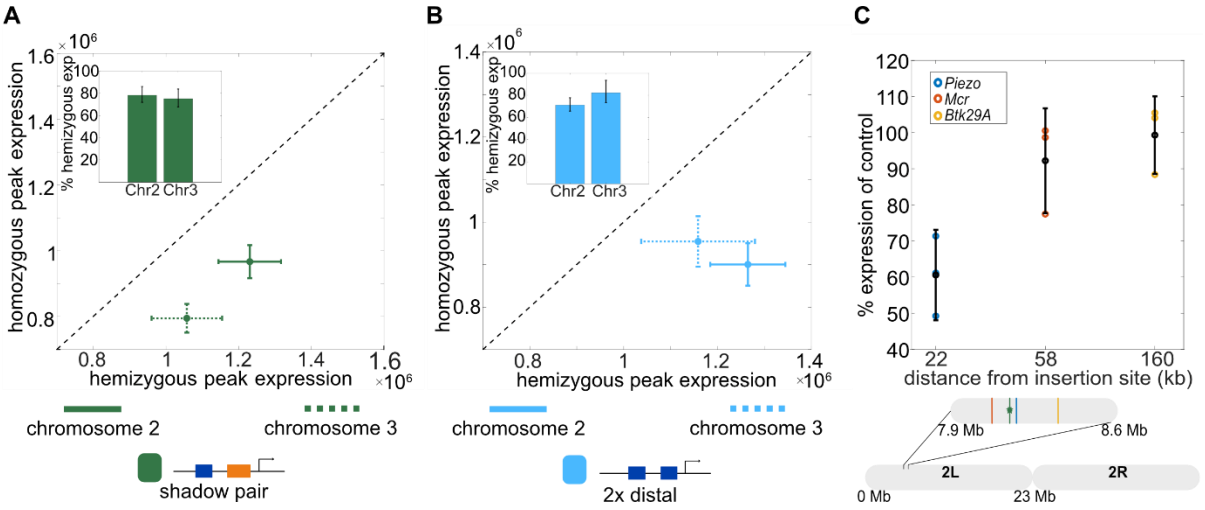


Figure 3.4 – Competition occurs at an additional location and gene. Based on our data suggesting that reporters are competing for limited levels of TFs, we suspected this competition would also occur at other transgenic insertion sites and with endogenous genes. **A**. Reporter competition occurs at multiple genomic insertion sites. Graph shows the peak expression levels per allele in homozygous embryos as a function of the peak expression levels in hemizygous embryos for the shadow pair reporter inserted in either chromosome 2L or 3L. The data for chromosome 2 are the same as in Figure 1B. Diagonal line marks expected values for homozygous expression if reporters do not interact and instead display independent expression. Error bars represent 95% confidence intervals. Inset shows the peak expression levels in homozygous embryos relative to hemizygous embryos with the shadow pair reporter inserted on either chromosome 2 or chromosome 3. Error bars in inset represent 95% confidence intervals from 1000 rounds of bootstrapping. **B**. The graph is as in A with the duplicated distal reporter inserted on chromosome 2 or chromosome 3. **C**. To determine the effect, if any, of transgene's use of resources on endogenous genes' expression, we compared the expression levels of three endogenous genes likely to be Bcd-regulated at increasing genetic distances from the transgenic insertion site in embryos with or without the duplicated distal transgene. Graph shows the fold change in expression of *Piezo*, *Mcr*, and *Bkt29A* in embryos homozygous for the duplicated distal transgene compared to WT embryos as measured by qPCR. Error bars represent 95% confidence intervals and black circles indicates the mean. Schematic below graph shows the genetic distance of the three measured genes (indicated with a blue, red, or yellow vertical line) from the attP site (VK000002) on chromosome 2L (marked with green line and star) where all transgenic constructs, unless otherwise specified, were inserted.

Figure 3.5 - Modeling the impact of competitor binding sites on TF-enhancer binding. To understand how small transgenic sequences could induce the competition for TFs we observe, we created a thermodynamic model of TF binding at a single site as a function of TF levels, competitor sites, and binding strengths. **A**. Schematic of the parameters of the genome model where the whole genome is considered for TF binding. The probability of a TF molecule being bound at the target site, *p(bound)*, is determined by the parameters shown. The number of available TFs, *T*, varies as a function of embryo position *l* to match the measured Bcd gradient[33]. We assume that all TF molecules are bound and can be bound to either the target site or competitor sites, which are divided into specific and non-specific sites. The number of non-specific sites, *N*, is held constant at $1x10^8$ while the number of specific competitor sites, *C*, is varied. TF molecules bind the target site and specific competitor sites with binding energy E_s and bind non-specific sites with binding energy E_{ns}. E_{ns} is held constant at zero and E_s is varied. With each set of parameters, *p(bound)* is calculated using equation 1 from the text. **B**. The fraction of maximum *p(bound)* as a function of number of added competitor sites using the genome model. E_s is held constant at 10 and *l* is held constant at 27% embryo length. Model predictions are in black. Experimental data of the fraction of maximum hemizygous *hb*P2 reporter expression as a function of the number of Bcd binding sites in the transgene on the homologous chromosome is shown in red. Data points indicate the fraction of maximum *hb*P2 expression with a second *hb*P2 reporter, which contains 6 Bcd binding sites, or the 6xBcd array on the homologous chromosome, measured at 27% egg length. Dashed lines indicate the number of additional competitor sites predicted by the genome model to be required to produce the experimentally observed decrease in expression. The inset shows the same data on a linear x-axis. **C**. Schematic of the parameters of the hub model where TF binding is assumed to only occur within nuclear subregions. Each nucleus is divided into 1000 equally-sized regions, one of which contains the target site. As in the genome model, the output of the model is the probability of a TF molecule being bound at the target site, *p(bound)*. Based on previous measurements, the number of available TFs, *T*, is held constant at 20 for hub regions and 0 for non-hub regions[66]. Instead, the probability that a region in a nucleus is a hub is a function of embryo position *l* to match the Bcd gradient and we call this probability *p(hub; T(l))* (equation 2 in the text). As in the genome model, we assume all TFs are bound at the target site, competitor sites, or non-specific sites. In each region, the number of non-specific competitor sites, *N*, is 100,000 while the number of specific competitor sites, *C*, varies. The binding strength parameters E_s and E_{ns} are the same as those used in the genome model. *p(bound)* is calculated as in the genome model using equation 1 from the text and multiplying the resulting value by *p(hub;T(l))*. This product is the final *p(bound)* value. **D**. Results of the hub model. The graph is as in B, with the fraction of maximum *p(bound)* as a function of the number of added competitor sites where the black line is the prediction of the hub model. As in B, E_s is held at 10 and *l* is held constant at 27% egg length. Red points are the same experimental data as in B. Dashed lines indicate the number of additional competitor sites predicted by the hub model to be needed to produce the experimentally observed decrease in expression. **E**. Model results can be compared to experimentally measured decreases in reporter expression. The schematic shows how our models relate to our experimental system. The target site in the models is analogous to the enhancer-MS2 reporters in our experimental system. The added specific competitor sites of the models

represent the TF binding site arrays or second reporter introduced on the homologous chromosome opposite the enhancer reporter. Although the exact relationship is not known, TF binding at enhancers is related to enhancer activity so the *p(bound)* output of our models is related to our measured enhancer reporter activity.

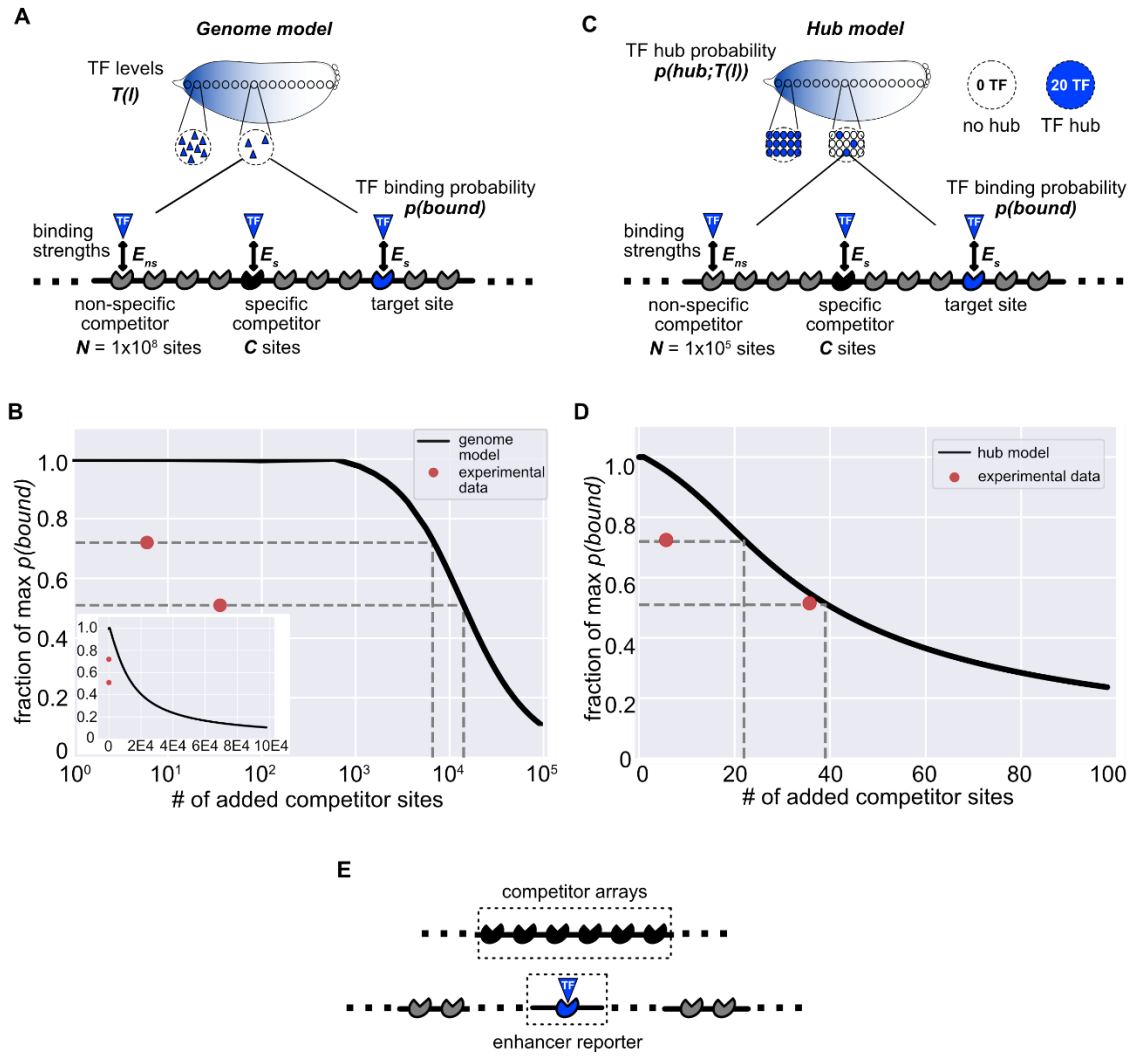


Figure S3.1 – Total enhancer activity is measured by integrating the area under fluorescence time trace curves at the region of peak expression. To measure the activity of our enhancer reporters, we track individual spots of transcription across the time of nc14. Our enhancer reporters drive expression of 24 repeats of MS2 sequence that when transcribed forms stem loops that are bound by the fluorescently tagged coat protein, MCP-GFP. This enables us to visualize and track sites of nascent transcripts as spots of fluorescence above background[16]. **A**. Transcriptional traces are first smoothed using the Lowess method (see Methods for more details). Integrated fluorescence, which is proportional to total mRNA produced and hence indicative of enhancer activity, is measured by calculating the area under the fluorescence curve. This area is calculated with the trapz function in Matlab using the fluorescence points recorded during the first 50 minutes of nc14. **B**. Total expression for each enhancer reporter is calculated at the position of that reporter's peak expression levels along the anterior-posterior axis of the embryo. We divide the embryo along the anterior-posterior axis into 41 equally sized bins that each encompass 2.5% of the embryo and calculate the average total reporter expression per bin. Schematic shows the *Kr* expression pattern during nc14 as the green stripe at the center of the embryo. Within this stripe, the different enhancer reporters have slightly different regions of peak expression as indicated by the darker green rectangles. The enhancer reporter or reporters that drive peak expression in each region is indicated below the corresponding bin. The duplicated proximal reporter shows peak expression at 47.5% egg length (0% corresponding to the anterior tip of the embryo), the proximal and shadow pair reporters at 50% egg length, and the single and duplicated distal reporters at 52.5% egg length.

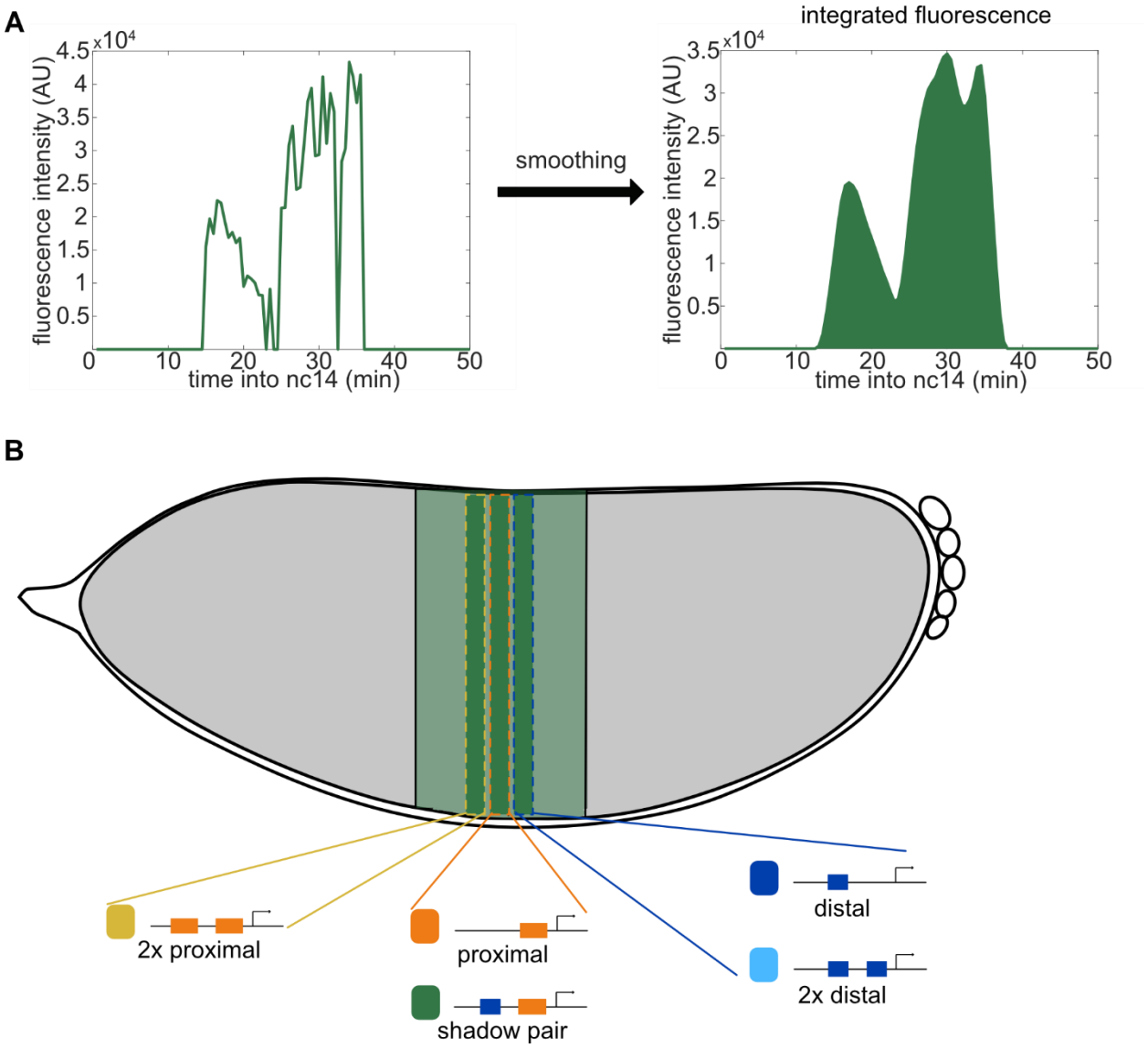
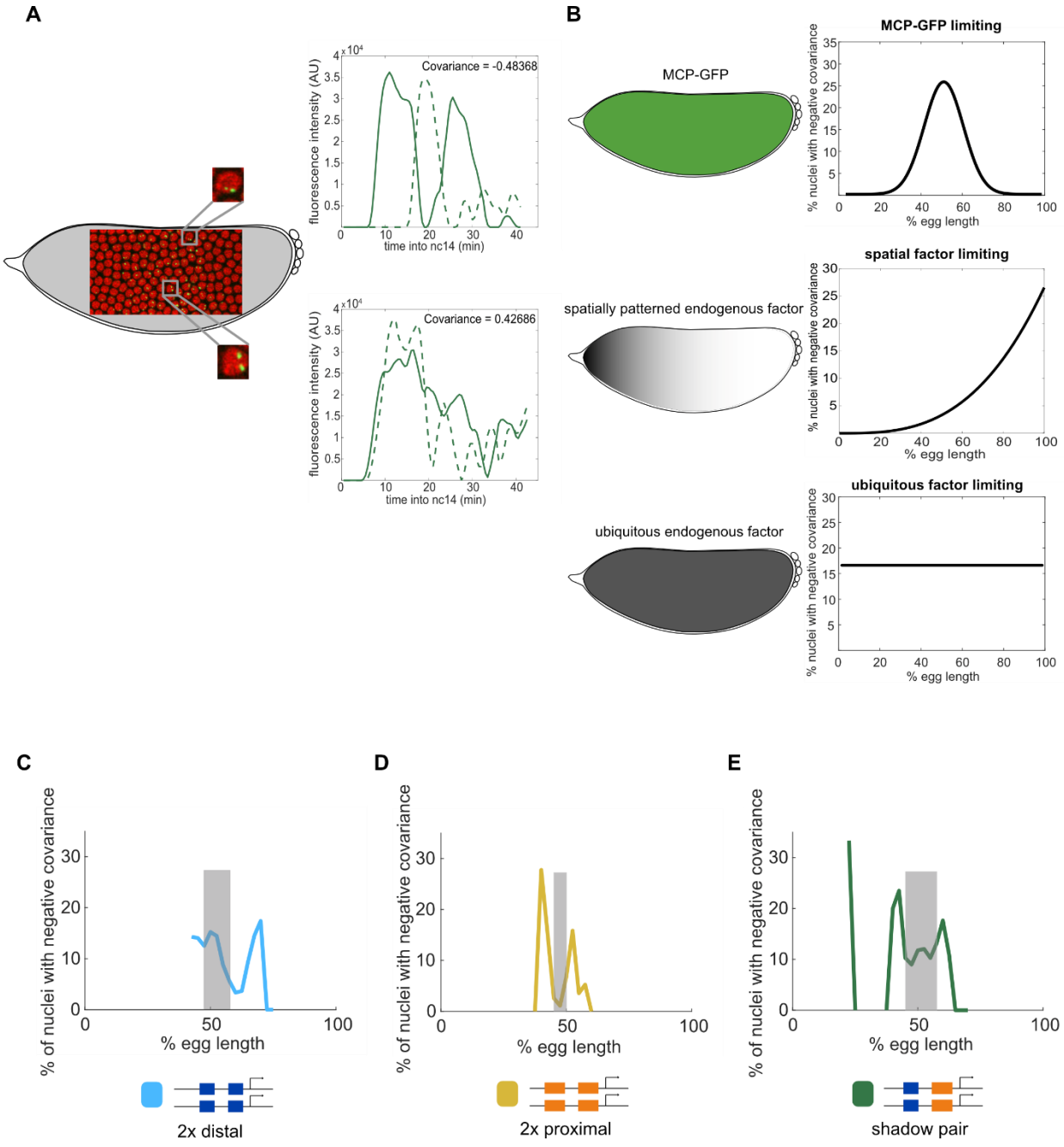

A
integrated fluorescence
smoothing
fluorescence intensity (AU)
x10^4
time into nc14 (min)
B
2x proximal
proximal
shadow pair
distal
2x distal

Figure S3.2 – Patterns of negative covariance suggest competition for spatially-patterned factor. To assess whether the differences in reporter expression levels observed between homozygous and hemizygous embryos stem from competition for factors required for reporter visualization, we measured the covariance of the activity of identical reporters in individual nuclei. A fraction of nuclei display negative covariance, which is indicative of an antagonistic relationship between the activities of the two reporters. **A**. Still image from a live imaging movie where nuclei are colored in red and sites of active transcription are green spots. Insets show zoomed-in example nuclei that display negative covariance (top) or do not (bottom). The graphs to the right show the transcriptional activity of the two reporters in each nucleus across the time of nc14. **B**. Schematics of expression patterns of different possible limiting factors and the corresponding expected patterns of negative covariance. The MCP-GFP reporter is expressed ubiquitously across the length of the embryo, while an endogenous factor may be expressed ubiquitously or in a spatial pattern. Graphs to the right show expected spatial pattern of negative covariance rates if reporters are competing for limiting levels of MCP-GFP (top), a spatially patterned endogenous factor (middle), or a ubiquitously expressed endogenous factor (bottom). **C-E**. The fraction of nuclei that display negative covariance as a function of egg length. Grey highlight indicates the region of 75% max expression for that reporter construct. **C.** Duplicated distal. **D**. Duplicated proximal. **E**. Shadow pair.


A
Covariance = -0.48368
fluorescence intensity (AU)
time into nc14 (min)
Covariance = 0.42686
fluorescence intensity (AU)
time into nc14 (min)
B
MCP-GFP
MCP-GFP limiting
% nuclei with negative covariance
% egg length
spatially patterned endogenous factor
spatial factor limiting
% nuclei with negative covariance
% egg length
ubiquitous endogenous factor
ubiquitous factor limiting
% nuclei with negative covariance
% egg length
C
% of nuclei with negative covariance
% egg length
2x distal
D
% of nuclei with negative covariance
% egg length
2x proximal
E
% of nuclei with negative covariance
% egg length
shadow pair

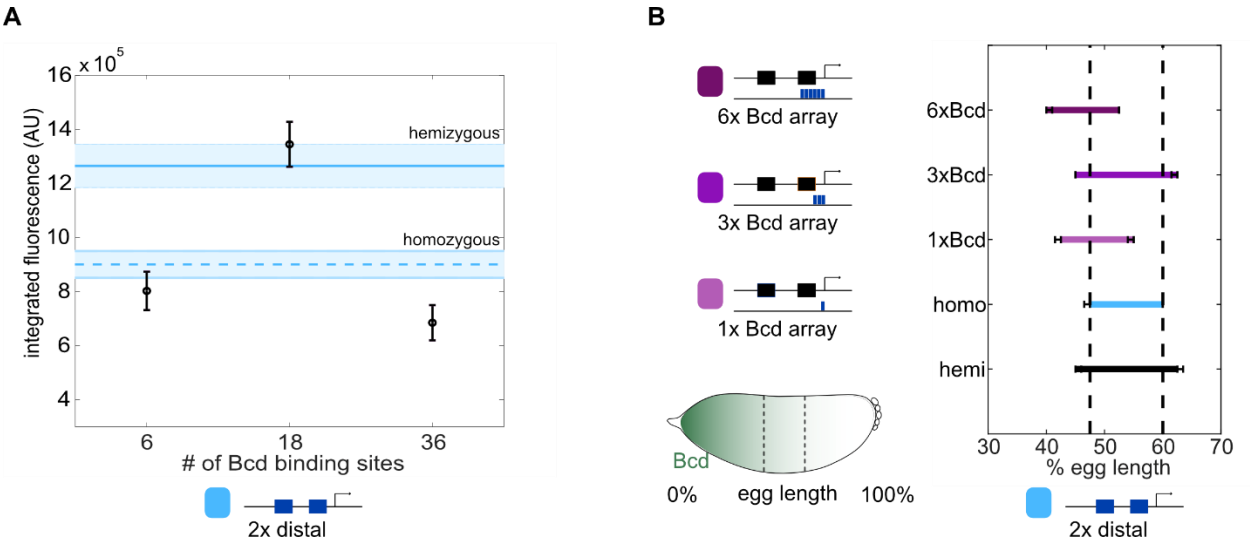


Figure S3.3 - Competitor Bicoid binding sites decrease and shift the activity of the duplicated distal reporter. To assess the effect of increasing number of competitor Bcd sites on the activity driven by the duplicated distal enhancer construct, we measured its expression in the presence of larger Bcd binding site arrays (the same as used in Figure 3). **A**. Duplicated distal reporter expression changes non-linearly with increasing number of Bcd binding sites. Graph shows the total expression driven by the duplicated distal reporter at its region of peak expression in the presence of Bcd binding arrays with the indicated number of Bcd binding sites. Horizontal lines indicate peak expression levels per allele in hemizygous (top) and homozygous (bottom) embryos. Error bars and shading indicate 95% confidence intervals. In the presence of the 1xBcd array, peak expression of the duplicated distal reporter is decreased 37% relative to hemizygous levels and is further reduced 46% relative to hemizygous levels in the presence of the 6xBcd array. Interestingly, with the 3xBcd array the duplicated distal reporter's activity is not significantly affected, which we do not fully understand but propose may be due to the molecular composition of the microenvironment created around the duplicated distal reporter with the 3xBcd array. **B**. The expression domain of the duplicated distal reporter shifts in the presence of competitor Bcd binding arrays and these shifts qualitatively match changes in peak expression levels. The graph shows the range of the expression pattern of the indicated construct to 50% of peak expression levels of the homozygous configuration, whose boundaries are indicated with dashed vertical lines. Error bars represent 95% confidence intervals from 1000 rounds of bootstrapping.

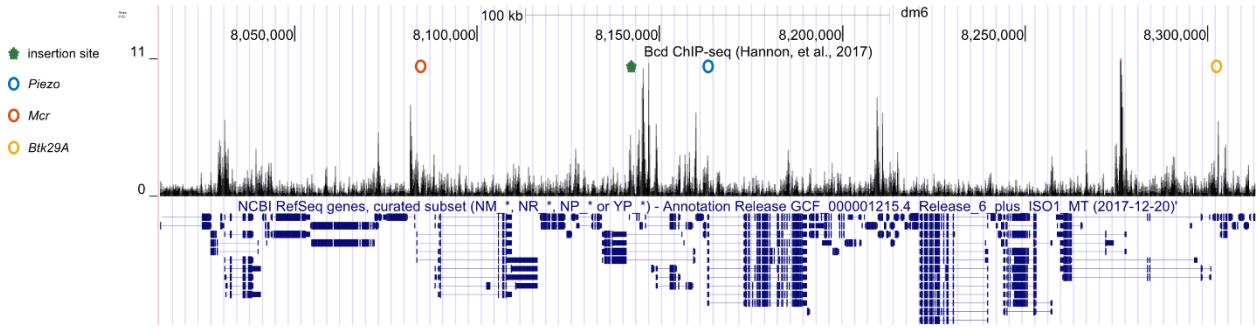


Figure S3.4 – Bcd binding surrounding genes near chromosome 2 transgenic insertion site. To assess the potential effect of our enhancer-MS2 transgenes on endogenous gene expression, we measured the expression of *Piezo*, *Mcr*, and *Btk29A* in embryos with or without the duplicated distal transgene inserted on chromosome 2. We looked at these three genes due to their proximity to the transgenic integration site (Figure 4) and the observation that they are all likely regulated by Bcd. All three genes have expression patterns in the early embryo that are consistent with activation by Bcd and previous Chip-seq in nc14 embryos indicates Bcd binding surrounding these genes[32]. Figure shows UCSC genome browser window of 300kb centered on the *Piezo* transcription start site. Top line shows genomic coordinates, black peaks indicate Bcd binding as measured in Hannon, et al., 2017, and bottom in blue indicates gene annotations. Green star marks site of chromosome 2 attP insertion site. Circles mark the transcription start sites of the three endogenous genes where *Piezo* is blue, *Mcr* is red, and *Btk29A* is yellow.

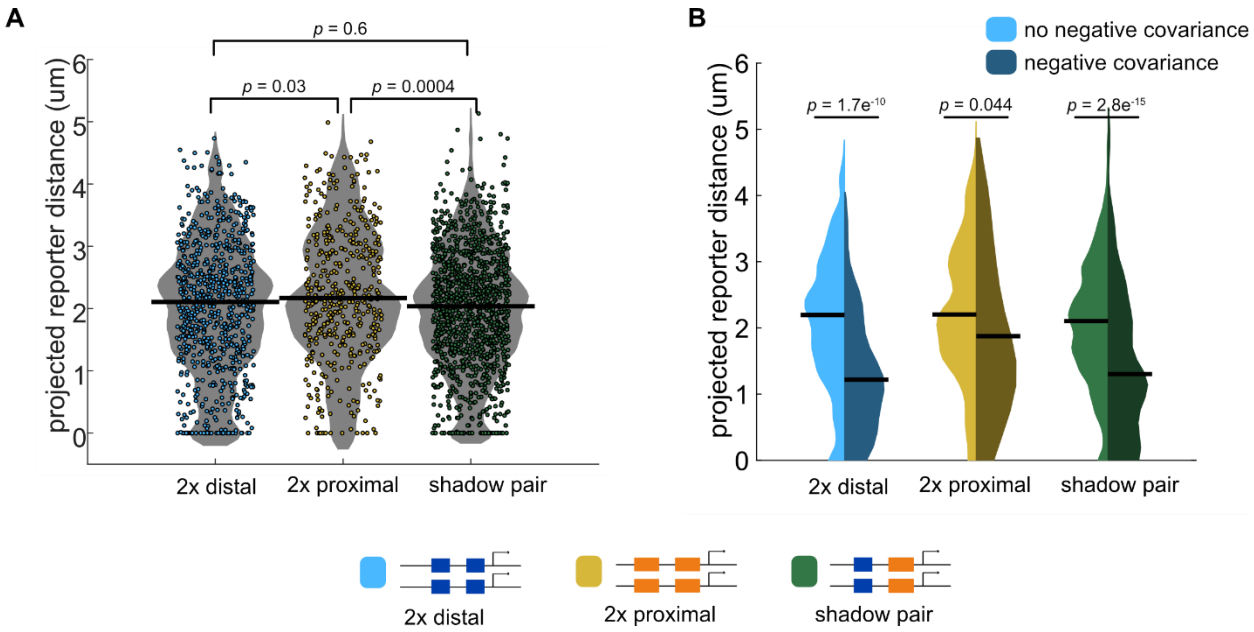
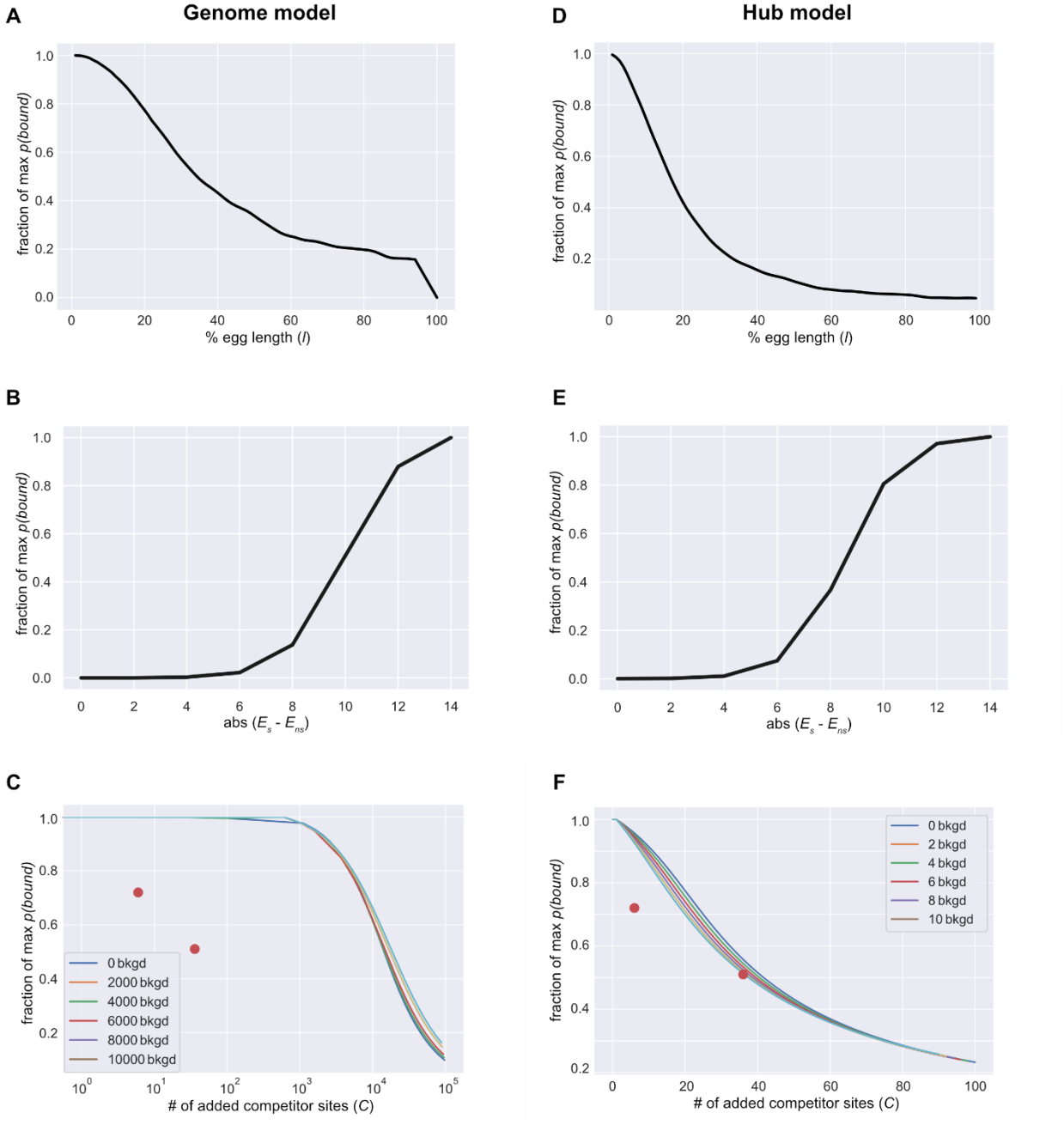


Figure S3.5 – Average distance between reporters is negatively correlated with competition. To test the hypothesis that reporters are competing for locally limited TFs, we compared the average distance between our reporters in a nucleus with the degree of competition that reporter shows. **A**. Graph shows the average projected distance between identical reporters in the same nucleus for the indicated reporters. Each colored circle represents the average distance between the two reporters in a single nucleus across the time of nc14. The duplicated distal and shadow pair reporters, which show significant competition (Figure 1A) on average are much closer together in the nucleus than are the duplicated proximal reporters, which do not show significant competition. Horizontal lines indicate medians. Significance determined by Kruskal-Wallis test with Bonferroni correction. **B**. Graph shows the distributions of projected distances between identical reporters in a nucleus broken down into nuclei with or without negative covariance, which is indicative of reporter competition. The left, lighter colored half of each violin plot shows the distribution of average reporter distance during nc14 in nuclei that do not show negative covariance. Right half of each violin plot shows this distribution in nuclei with negative covariance. Horizontal lines indicate medians. In all three reporter constructs, nuclei whose reporters display negative covariance, which is indicative of competition, have reporters that are on average significantly closer together than are reporters that do not display negative covariance. Significance determined by Kolmogorov-Smirnov test.

Figure S3.6 – How *p(bound)* changes with varying parameters. To get a sense for how our model behaves, we looked at how *p(bound)* changes as a function of our three parameters (TF levels, binding energy difference, and number of specific competitor sites) for both our genome and hub models. **A**. In the genome model, *p(bound*) decreases with decreasing levels of TF. Graph shows model prediction of *p(bound)* as a function of embryo position, *l*, which determines TF levels following the Bcd gradient. We set the maximum Bcd level to 20,000 at the anterior of the embryo[10] and use this value with the measured Bcd gradient[33] to estimate TF levels at each point along the embryo. 0% egg length corresponds to the anterior of the embryo. Number of specific competitor sites is held constant at 2,000 and the difference in specific vs non-specific binding energies is held constant at 10. **B**. In the genome model, binding probability increases as the difference between specific and non-specific binding energies increases. Graph shows genome model prediction of the fraction of maximum binding probability as a function of the difference between specific and non-specific binding energies. Non-specific binding energy is held constant at 0 while specific binding energy is decreased. Here the number of TFs is held constant at 5,468 (corresponding to l = 27% egg length) and the number of specific competitor sites is held constant at 2,000. **C**. The sensitivity of binding probability to added competitor sites in the genome model is dependent on the number of pre-existing or "baseline" specific competitor sites. Graph shows the fraction of maximum binding probability as a function of added specific competitor sites. The different colored lines show model predictions depending on the number of baseline specific competitor sites. Red points show experimental data of the fraction of peak hemizygous expression for the *hb*P2 reporter as homozygotes (top point) or with the 6xBcd array (bottom point). The number of TFs is held constant at 5,468 (corresponding to l = 27% egg length) and the difference in specific versus non-specific binding energies is held constant at 10. **D**. Binding probability predicted by the hub model decreases along the length of the embryo. The graph shows the fraction of maximum binding probability of the hub model as a function of embryo position, which in the hub model determines the probability of a nuclear subregion being a TF hub. The number of specific competitor sites is held constant at 2 and the difference in binding energies between specific and non-specific sites is held constant at 10. **E**. Graph as in B, showing the predictions of the hub model. Like with the genome model, binding probability increases with increasing difference between specific and non-specific binding energies. Non-specific binding energy is held constant at 0 while specific binding energy is decreased. Here embryo position *l*, and consequently *p(hub;T(l))*, is held constant at 27% egg length and the number of specific competitor sites per region is held constant at 2. **G**. As with the genome model, the sensitivity of binding probability to added competitor sites in the hub model is dependent on the number of pre-existing specific competitor sites. Graph is as in C for predictions of the hub model depending on the indicated number of baseline specific competitor sites. Red points are the same experimental data as in C. Here, embryo position *l*, and consequently *p(hub;T(l))*, is held constant at 27% egg and the difference in binding energies between specific and non-specific competitor sites is held constant at 10.


A
Genome model
fraction of max p(bound)
% egg length (l)
B
abs (E_s - E_ns)
C
# of added competitor sites (C)
0 bkgd
2000 bkgd
4000 bkgd
6000 bkgd
8000 bkgd
10000 bkgd
D
Hub model
E
F
0 bkgd
2 bkgd
4 bkgd
6 bkgd
8 bkgd
10 bkgd

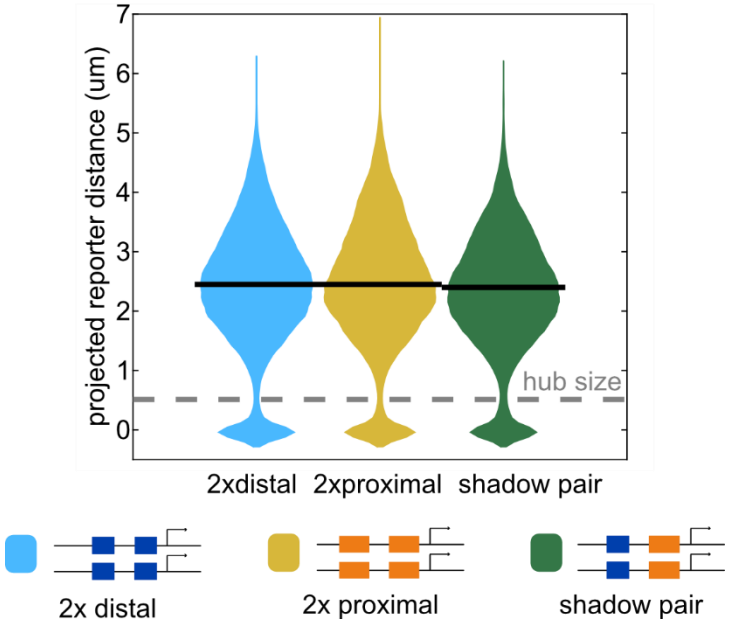


Figure S3.7 – Distribution of all measured distances between transcriptional reporters. To assess whether enhancer reporters were close enough to be accessing the same "hub" region within the nucleus, we looked at the distribution of reporter distances in homozygous embryos of the three indicated reporter constructs. Violin plots show the distribution of projected distances between transcriptional spots for every time point in nc14 where two transcriptional spots were tracked in a nucleus, across all measured nuclei in each of the indicated constructs. Gray dashed line indicates the 512nm diameter of nuclear regions in our hub model. For all three constructs, the two copies of the reporter are within the same hub-sized region in between 7% and 8% of all recorded time points, based on estimates of TF hub size[18] (see Methods for more details).

Figure S3.8 – Comparison of homozygous and hemizygous fly eye color. Unlike previous observations of transgene silencing[56], flies homozygous for our enhancer reporters, which contain a mini-*white* marker, do not show lighter eye color than hemizygous flies. The three flies on the left are hemizygous for the shadow pair reporter on chromosome 2 and the three flies on the right are homozygous for this same reporter. On both sides, the top two flies are female and the bottom-most fly is male.

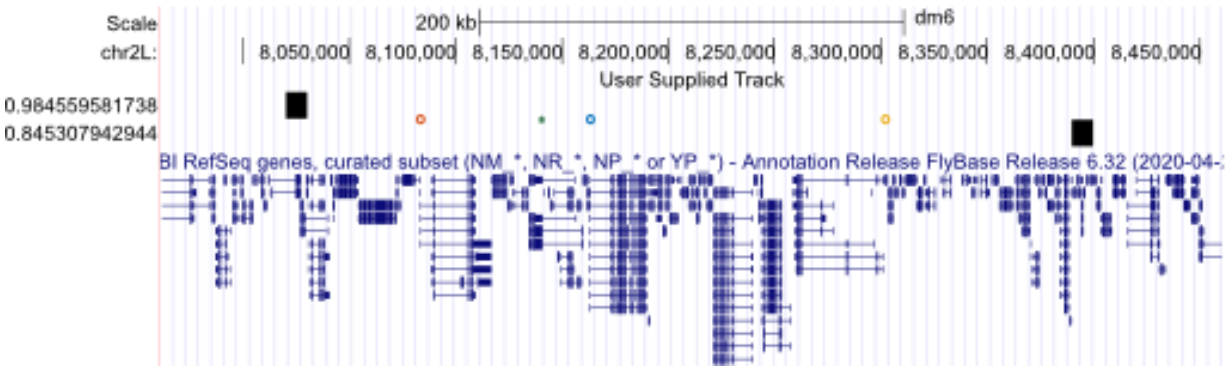


Figure S3.9 – Transgenic insertion site and affected endogenous gene fall within the same TAD. Ours and previous data suggest that transgenes have a larger effect on the expression of endogenous genes closer to the transgene on a linear piece of DNA[13]. As the genome is organized three dimensionally, we suspect that the distance between transgenes and endogenous genes in 3D space is important in determining the effect of a transgenic reporter on endogenous gene expression. As such, we used previously published Hi-C data from nc14 embryos to ask whether our transgene is likely contained within the same TAD as the measured endogenous genes. The figure shows a UCSC genome browser window with our transgenic insertion site marked as a green star and the three measured endogenous genes, *Piezo*, *Mcr*, and *Btk29A*, indicated as the blue, red, and yellow circles, respectively. The closest TAD boundaries as measured in Hug, et al., 2017 are indicated as black rectangles[67]. The transgenic insertion site and all three endogenous genes are contained within the same TAD during nc14.

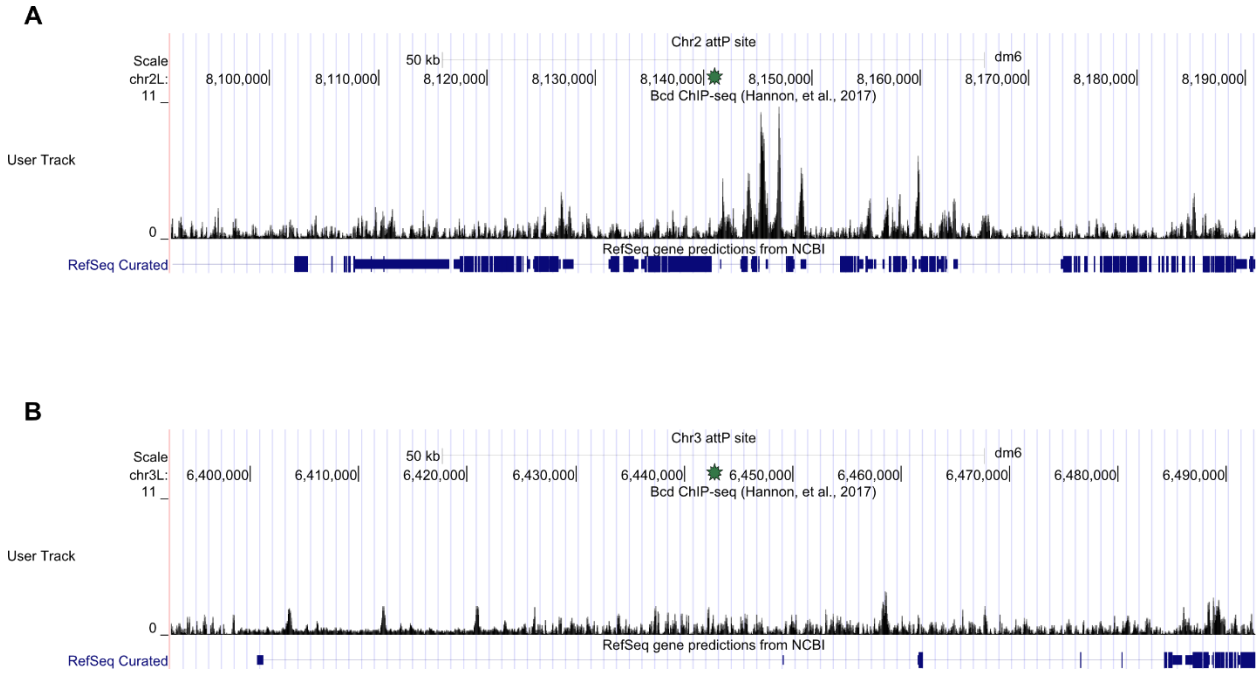


Figure S3.10 – Bcd binding surrounding the two transgenic insertion sites. Our hub model predicts that TF competition is in part dependent on the number of existing TF binding sites within the same nuclear subregion. In our hub model we assume that all existing TF binding sites are distributed uniformly throughout the genome, but in reality, this is not the case. This can be seen with the Bcd binding surrounding our two transgenic insertion sites. **A**. UCSC genome browser window of 100kb centered on the attP insertion site used on chromosome 2. Top line shows genomic coordinates and black peaks indicate Bcd ChIP-seq reads from Hannon, et al., 2017[32]. **B**. UCSC genome browser as in A, but centered on the attP insertion site used on chromosome 3. Green star in A and B indicates the genomic coordinates of the respective insertion sites.

3.6 References:

1. Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol*. 2010;20(17):R754. doi:10.1016/j.cub.2010.06.070
2. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. 2014;15(4):272-286. doi:10.1038/nrg3682
3. Kvon EZ. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics*. 2015;106(3):185-192. doi:10.1016/j.ygeno.2015.06.007
4. Wood K V. Marker proteins for gene expression. *Curr Opin Biotechnol*. 1995;6(1):50-58. doi:10.1016/0958-1669(95)80009-3
5. Pennacchio LA, Ahituv N, Moses AM, et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. 2006;444(7118):499-502. doi:10.1038/nature05295
6. O'Kane CJ, Gehring WJ. Detection in situ of genomic regulatory elements in Drosophila. *Proc Natl Acad Sci U S A*. 1987;84(24):9123-9127. doi:10.1073/pnas.84.24.9123
7. Bier E, Vaessin H, Shepherd S, et al. Searching for pattern and mutation in the Drosophila genome with a P-lacZ vector. *Genes Dev*. 1989;3(9):1273-1287. doi:10.1101/gad.3.9.1273
8. Visel A, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA. Functional autonomy of distant-acting human enhancers. *Genomics*. 2009;93(6):509-513. doi:10.1016/j.ygeno.2009.02.002
9. Swanson CI, Evans NC, Barolo S. Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye Enhancer. *Dev Cell*. 2010;18(3):359-370. doi:10.1016/j.devcel.2009.12.026
10. Biggin MD. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Dev Cell*. 2011;21(4):611-626. doi:10.1016/J.DEVCEL.2011.09.008
11. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers The database of key numbers in molecular and cell biology. *Nucleic Acids Res*. 2009;38(SUPPL.1):D750. doi:10.1093/nar/gkp889
12. Zhao ZW, Roy R, Gebhardt JCM, Suter DM, Chapman AR, Xie XS. Spatial organization of RNA polymerase II insidea mammalian cell nucleus revealed by reflectedlight-sheet superresolution microscopy. *Proc Natl Acad Sci U S A*. 2014;111(2):681-686. doi:10.1073/pnas.1318496111
13. Laboulaye MA, Duan X, Qiao M, Whitney IE, Sanes JR. Mapping transgene insertion sites reveals complex interactions between mouse transgenes and neighboring endogenous genes. *Front Mol Neurosci*. 2018;11. doi:10.3389/fnmol.2018.00385
14. Thompson A, Gasson MJ. Location effects of a reporter gene on expression levels and on native protein synthesis in Lactococcus lactis and Saccharomyces cerevisiae. *Appl Environ Microbiol*. 2001;67(8):3434-3439. doi:10.1128/AEM.67.8.3434-3439.2001
15. Waymack R, Fletcher A, Enciso G, Wunderlich Z. Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. *Elife*. 2020;9:1-57. doi:10.7554/ELIFE.59351
16. Garcia HG, Tikhonov M, Lin A, Gregor T. Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. *Curr Biol*. 2013;23(21):2140-2145. doi:10.1016/j.cub.2013.08.054
17. Wunderlich Z, Bragdon MDJ, Vincent BJ, White JA, Estrada J, DePace AH. Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. *Cell Rep*. 2015;12(11):1740-1747. doi:10.1016/j.celrep.2015.08.021

18. Tsai A, Alves MRP, Crocker J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *Elife*. 2019;8. doi:10.7554/eLife.45325
19. Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife*. 2015;4. doi:10.7554/eLife.07956
20. Savic D, Roberts BS, Carleton JB, et al. Promoter-distal RNA polymerase II binding discriminates active from inactive CCAAT/ enhancer-binding protein beta binding sites. *Genome Res*. 2015;25(12):1791-1800. doi:10.1101/gr.191593.115
21. Bozek M, Gompel N. Developmental Transcriptional Enhancers: A Subtle Interplay between Accessibility and Activity. *BioEssays*. 2020;42(4):1900188. doi:10.1002/bies.201900188
22. Li X, MacArthur S, Bourgon R, et al. Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm. Kadonaga J, ed. *PLoS Biol*. 2008;6(2):e27. doi:10.1371/journal.pbio.0060027
23. Harrison MM, Li X-Y, Kaplan T, Botchan MR, Eisen MB. Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. Copenhaver GP, ed. *PLoS Genet*. 2011;7(10):e1002266. doi:10.1371/journal.pgen.1002266
24. Tsurumi A, Xia F, Li J, Larson K, LaFrance R, Li WX. STAT Is an Essential Activator of the Zygotic Genome in the Early Drosophila Embryo. Rulifson E, ed. *PLoS Genet*. 2011;7(5):e1002086. doi:10.1371/journal.pgen.1002086
25. Jaeger J. The gap gene network. *Cell Mol Life Sci*. 2011;68(2):243-274. doi:10.1007/s00018-010-0536-y
26. Vincent BJ, Staller M V., Lopez-Rivera F, et al. Hunchback is counter-repressed to regulate even-skipped stripe 2 expression in Drosophila embryos. Levine M, ed. *PLOS Genet*. 2018;14(9):e1007644. doi:10.1371/journal.pgen.1007644
27. Papatsenko D, Levine MS. Dual regulation by the Hunchback gradient in the Drosophila embryo. *Proc Natl Acad Sci U S A*. 2008;105(8):2901-2906. doi:10.1073/pnas.0711941105
28. Kraut R, Levine M. *Mutually Repressive Interactions between the Gap Genes Giant and Kruppel Define Middle Body Regions of the Drosophila Embryo*. Vol 111.; 1991.
29. Stanojevic D, Small S, Levine M. Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science (80-)*. 1991;254(5036):1385-1387. doi:10.1126/science.1683715
30. Štanojević D, Hoey T, Levine M. Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Krüppel in Drosophila. *Nature*. 1989;341(6240):331-335. doi:10.1038/341331a0
31. Fisher, B., Weiszmann, R., Frise, E., Hammonds, A., Tomancak, P., Beaton, A., Berman, B., Quan, E., Shu, S., Lewis, S., Rubin, G., Barale, C., Laguertas, E., Quinn, J., Ghosh, A., Hartenstein, V., Ashburner, M., Celniker S. BDGP insitu homepage. https://insitu.fruitfly.org/cgi-bin/ex/insitu.pl. Published 2012.
32. Hannon CE, Blythe SA, Wieschaus EF. Concentration dependent chromatin states induced by the bicoid morphogen gradient. *Elife*. 2017;6. doi:10.7554/eLife.28275
33. Fowlkes CC, Luengo Hendriks CL, Keränen SVE, et al. A Quantitative Spatiotemporal Atlas of Gene Expression in the Drosophila Blastoderm. *Cell*. 2008;133(2):364-374. doi:10.1016/j.cell.2008.01.053

34. Bintu L, Buchler NE, Garcia HG, et al. Transcriptional regulation by the numbers: Models. *Curr Opin Genet Dev*. 2005;15(2):116-124. doi:10.1016/j.gde.2005.02.007
35. Phillips R, Belliveau NM, Chure G, Garcia HG, Razo-Mejia M, Scholes C. Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. *Annu Rev Biophys*. 2019;48:121-163. doi:10.1146/annurev-biophys-052118
36. Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol*. 2010;20(17). doi:10.1016/j.cub.2010.06.070
37. Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordân R, Rohs R. Absence of a simple code: How transcription factors read the genome. *Trends Biochem Sci*. 2014;39(9):381-399. doi:10.1016/j.tibs.2014.07.002
38. Grossman SR, Zhang X, Wang L, et al. Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc Natl Acad Sci U S A*. 2017;114(7):E1291-E1300. doi:10.1073/pnas.1621150114
39. Chen CH, Zheng R, Tokheim C, et al. Determinants of transcription factor regulatory range. *Nat Commun*. 2020;11(1):1-15. doi:10.1038/s41467-020-16106-x
40. Liu Z, Tjian R. Visualizing transcription factor dynamics in living cells. *J Cell Biol*. 2018;217(4):1181-1191. doi:10.1083/jcb.201710038
41. Mir M, Stadler MR, Ortiz SA, et al. Dynamic multifactor hubs interact transiently with sites of active transcription in drosophila embryos. *Elife*. 2018;7. doi:10.7554/eLife.40497
42. Shariati SA, Dominguez A, Xie S, Wernig M, Qi LS, Skotheim JM. Reversible Disruption of Specific Transcription Factor-DNA Interactions Using CRISPR/Cas9. *Mol Cell*. 2019;74(3):622-633.e4. doi:10.1016/j.molcel.2019.04.011
43. Driever W, Nüsslein-Volhard C. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature*. 1989;337(6203):138-143. doi:10.1038/337138a0
44. Struhl G, Struhl K, Macdonald PM. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*. 1989;57(7):1259-1273. doi:10.1016/0092-8674(89)90062-7
45. Chen H, Xu Z, Mei C, Yu D, Small S. A system of repressor gradients spatially organizes the boundaries of bicoid-dependent target genes. *Cell*. 2012;149(3):618-629. doi:10.1016/j.cell.2012.03.018
46. Driever W, Thoma G, Nüsslein-Volhard C. Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. *Nature*. 1989;340(6232):363-367. doi:10.1038/340363a0
47. Cho W-K, Spille J-H, Hecht M, et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*. 2018;361(6400):412. doi:10.1126/SCIENCE.AAR4199
48. Boehning M, Dugast-Darzacq C, Rankovic M, et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol*. 2018;25(9):833-840. doi:10.1038/s41594-018-0112-y
49. Tsai A, Muthusamy AK, Alves MRP, et al. Nuclear microenvironments modulate transcription from low-affinity enhancers. *Elife*. 2017;6. doi:10.7554/eLife.28975
50. Chong S, Dugast-Darzacq C, Liu Z, et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science (80-)*. 2018;361(6400). doi:10.1126/science.aar2555
51. Banerji J, Rusconi S, Schaffner W. Expression of a β-globin gene is enhanced by remote

SV40 DNA sequences. *Cell*. 1981;27(2 PART 1):299-308. doi:10.1016/0092-8674(81)90413-X
52. Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res*. 1981;9(22):6047-6068. doi:10.1093/nar/9.22.6047
53. Kassis JA, VanSickle EP, Sensabaugh SM. A fragment of engrailed regulatory DNA can mediate transvection of the white gene in Drosophila. *Genetics*. 1991;128(4).
54. Kassis JA. Unusual properties of regulatory DNA from the Drosophila engrailed gene: three "pairing-sensitive" sites within a 1.6-kb region. *Genetics*. 1994;136(3).
55. Pal-Bhadra M, Bhadra U, Birchler JA. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. *Mol Cell*. 2002;9(2):315-327. doi:10.1016/S1097-2765(02)00440-9
56. Pal-Bhadra M, Bhadra U, Birchler JA. Cosuppression in Drosophila: Gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. *Cell*. 1997;90(3):479-490. doi:10.1016/S0092-8674(00)80508-5
57. Pal-Bhadra M, Bhadra U, Birchler JA. Cosuppression of nonhomologous transgenes in Drosophila involves mutually related endogenous sequences. *Cell*. 1999;99(1):35-46. doi:10.1016/S0092-8674(00)80060-4
58. Liu X, Wu B, Szary J, Kofoed EM, Schaufele F. Functional sequestration of transcription factor activity by repetitive DNA. *J Biol Chem*. 2007;282(29):20868-20876. doi:10.1074/jbc.M702547200
59. Janssen S, Durussel T, Laemmli UK. Chromatin opening of DNA satellites by targeted sequence-specific drugs. *Mol Cell*. 2000;6(5):999-1011. doi:10.1016/S1097-2765(00)00099-X
60. Lee TH, Maheshri N. A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol Syst Biol*. 2012;8:576. doi:10.1038/msb.2012.7
61. Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R. The Transcription Factor Titration Effect Dictates Level of Gene Expression. *Cell*. 2014;156(6):1312-1323. doi:10.1016/J.CELL.2014.02.022
62. Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol*. 2004;166(4):493-505. doi:10.1083/jcb.200403109
63. Li X-Y, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, Biggin MD. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biol*. 2011;12(4):R34. doi:10.1186/gb-2011-12-4-r34
64. Nagele RG, Freeman T, McMorrow L, Thomson Z, Kitson-Wind K, Lee H y. Chromosomes exhibit preferential positioning in nuclei of quiescent human cells. *J Cell Sci*. 1999;112 (Pt 4:525-535.
65. Manuelidis L. Different central nervous system cell types display distinct and nonrandom arrangements of satellite DNA sequences. *Proc Natl Acad Sci U S A*. 1984;81(10 I):3123-3127. doi:10.1073/pnas.81.10.3123
66. Mir M, Reimer A, Haines JE, et al. Dense bicoid hubs accentuate binding along the morphogen gradient. *Genes Dev*. 2017;31(17):1784-1794. doi:10.1101/gad.305078.117
67. Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*. 2017;169(2):216-228.e19. doi:10.1016/j.cell.2017.03.024

68. Berman BP, Nibu Y, Pfeiffer BD, et al. 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*. 2002;99(2):757-762. doi:10.1073/pnas.231608898
69. Xu Z, Chen H, Ling J, Yu D, Struffi P, Small S. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. *Genes Dev*. 2014;28(6):608-621. doi:10.1101/gad.234534.113
70. Yan R, Small S, Desplan C, Dearolf CR, Darnell JE. Identification of a Stat gene that functions in Drosophila development. *Cell*. 1996;84(3):421-430. doi:10.1016/S0092-8674(00)81287-8
71. Lammers NC, Galstyan V, Reimer A, Medin SA, Wiggins CH, Garcia HG. Multimodal transcriptional control of pattern formation in embryonic development. *Proc Natl Acad Sci U S A*. 2020;117(2):836-847. doi:10.1073/pnas.1912500117
72. Jung C, Bandilla P, Von Reutern M, et al. True equilibrium measurement of transcription factor-DNA binding affinities using automated polarization microscopy. *Nat Commun*. 2018;9(1):1-11. doi:10.1038/s41467-018-03977-4

Chapter 4

Large-scale investigation of shadow enhancer transcription factor input patterns

Chapter 4

Large-scale investigation of shadow enhancer transcription factor input patterns

Enhancers are critical for proper regulation of gene expression in a variety of systems facing very different demands. As the general mechanism by which organisms regulate gene expression in both time and space, enhancers in different systems will likely be under selection for different features. In Chapter 2, I discuss how a separation of transcription factor (TF) inputs is critical for the stable gene expression driven by a pair of developmental shadow enhancers. In this shorter chapter, I provide a brief investigation into the regulatory logic of developmental enhancers more broadly to assess if the findings of Chapter 2 are a general feature of shadow enhancers.

4.1 Transcription factor regulation of developmental shadow enhancers

As discussed in Chapter 2, shadow enhancers are groups of two or more enhancers that drive overlapping expression patterns of the same target gene [1–3]. Shadow enhancers are largely associated with developmental genes and have been shown to be critical for robust gene expression[4–8]. In Chapter 2, I detail our findings that the pair of shadow enhancers regulating *Kruppel* (Kr) expression in the early *Drosophila* embryo, which are each activated by different sets of TFs[9], drive more stable gene expression across time than do single or duplicated enhancers. With this pair of shadow enhancers we showed that a separation of transcription factor (TF) inputs between the enhancers makes the pair uniquely able to buffer against multiple sources of noise across a wide range of temperatures[10]. From these findings, we hypothesized that a separation of TF inputs among the individual enhancers of a shadow enhancer group may be a general mechanism by which shadow enhancers mediate robust gene expression.

To assess how widespread the use of separate TFs is among shadow enhancers, we sought out a large dataset of shadow enhancers. Cannavo and colleagues identified a set of over 1,000 predicted shadow enhancers involved in *Drosophila* mesoderm development[11]. These groups of shadow enhancers, many of which consisted of well over two individual enhancers, allowed us to examine the regulatory logic of shadow enhancers more broadly beyond the *Kr* shadow enhancers.

To determine whether separation of TF inputs among shadow enhancers is a common feature, we compared the predicted TF regulators of enhancers within a shadow enhancer group to that of randomly grouped enhancers. TF regulators of each individual enhancer were predicted via motif analysis of the enhancer sequences defined in the Cannavo dataset[11]. We focused on the key muscle TFs, Bap, Bin, Tin, Twi, Sna, and Dl. For each enhancer, we summed the predicted strengths of all binding sites for a given TF to have a total predicted regulatory score for that TF and did this for all 6 investigated TFs to have a total regulatory profile for each enhancer. Then the pairwise correlation of regulatory profiles between all enhancers of a shadow enhancer group was calculated. A higher correlation score indicates the individual enhancers of the shadow enhancer group are predicted to be more similarly regulated by TFs while a lower correlation score indicates the individual enhancers are predicted to be more differently regulated by TFs. To assess whether shadow enhancers tend to have more similar or dissimilar TF inputs, we performed the same correlation analysis on randomly grouped mesodermal enhancers. For each real set of shadow enhancers, we randomly selected (without replacement) the same number of individual enhancers from the total list of enhancers and then performed TF input correlation analysis as done with the real groups of shadow enhancers. We did this random selection and analysis 1,000 times.

We find that TF inputs tend to be more dissimilar among the individual enhancers of a shadow enhancer group than among randomly grouped mesoderm enhancers (Figure 4.1A). Specifically, shadow enhancers are more likely to be negatively correlated than are randomly grouped enhancers, but this difference is not significant (Figure 4.1B). After each of the 1,000 instances of randomly grouping enhancers and calculating the pairwise regulatory correlations of each group, we found the overall median correlation score to compare to the overall median correlation score of the true shadow enhancer groups. Out of the 1,000 trials of randomly grouping enhancers, 51% percent have a higher median regulatory correlation scores than the true shadow enhancer group (median regulatory correlation score of 0.67; Figure 4.1C). This indicates that while a slight majority of shadow enhancers have less similar TF inputs than expected by chance, many other shadow enhancers do have similar TF inputs. While the general trend of shadow enhancer groups displaying differing degrees of TF input separation likely holds, the fact that these shadow enhancers were predicted in part from similar TF binding motifs[11] suggests that additional shadow enhancers, with less similar TF inputs, may have been missed from this analysis.

Our findings suggest that a separation of TF inputs, as discussed in Chapter 2, is a common feature of many, but not all, shadow enhancers. The degree to which a group of shadow enhancers is regulated by similar or dissimilar TFs is likely determined by the exact role of individual enhancers, which can be different for even seemingly redundant enhancers[12] and can change across developmental time[13].

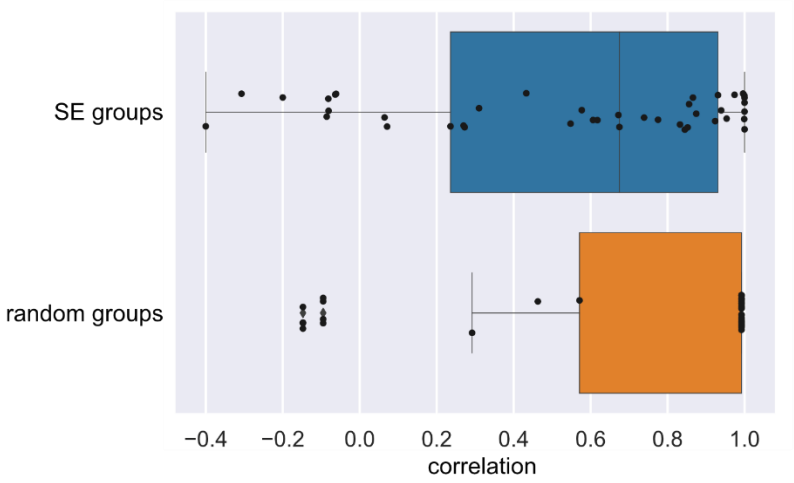
4.2 Methods

Analysis of TF regulation similarity among enhancers

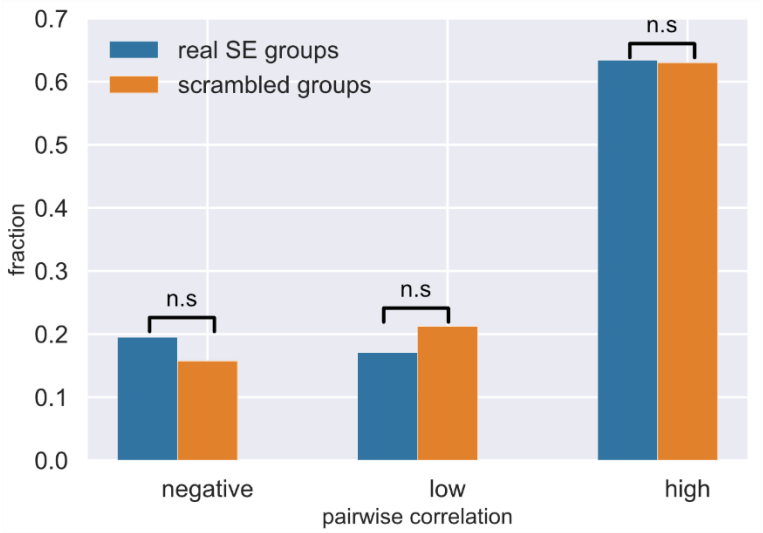
To assess whether individual enhancers of a shadow enhancer group are more likely than chance to be regulated by different TFs, we compared the correlation of TF inputs between enhancers of a shadow enhancer group to that between randomly grouped enhancers. We did this on a large dataset (> 1,000) of predicted shadow enhancers in *Drosophila* mesoderm development[14] . First, binding sites of six key mesoderm TFs; Twi, Bap, Bin, Tin, Sna, and Dl, were predicted in each individual enhancer using Patser[15] and position weight matrices (PWM) from FlyFactor Survey[16]. The cutoffs for p-values of binding sites are determined as in[17] where Patser scores the aligned sequences that informed the PWMs and the top scoring 75% of the aligned sequences are considered "true" binding sites. For each TF, the predicted strength of each binding site was summed to get a total regulatory score by that TF for each enhancer. This means each individual enhancer has a six value (corresponding to the six TFs) regulatory score. To assess the similarity of TF regulation between individual shadow enhancers, we calculated the pairwise Pearson correlation of regulatory scores between all individual enhancers in a shadow enhancer group. For our randomly grouped comparison, we maintained the same number of individual enhancers per group as in the real shadow enhancer groups (i.e. if there were five groups of three enhancers in the real data, there will be five groups of three enhancers in the randomly grouped data), but randomly shuffle the individual enhancers. We then calculated the pairwise correlation of regulatory scores between all individual enhancers in a random enhancer group, as done with the true shadow enhancer groups. We did this random grouping 1000 times. The Python code used to calculate and compare the pairwise correlations of TF regulatory scores between shadow enhancers or randomly grouped enhancers is available at https://github.com/rvwaymack/ShadowEnhancer_TFRegulation_Analysis.

Figure 4.1 – Shadow enhancers have less correlated TF inputs than randomly grouped enhancers. To assess whether the separation of TF inputs among shadow enhancers studied in Chapter 2 (Waymack, et al., 2020) is a general feature of shadow enhancers, we assessed the predicted TF regulators of *Drosophila* mesoderm shadow enhancers. Using the groups of shadow enhancers predicted by Cannavo and colleagues we calculated the predicted regulatory strength of 6 mesoderm TFs for each individual enhancer. Within each shadow enhancer group, we then calculate the pairwise correlation of TF regulatory strengths between all enhancers. As a comparison, we perform the same correlation calculation for randomly grouped enhancers. The randomly grouped enhancers are chosen from the same list of mesoderm enhancers and the number of enhancers per group is the same as in the real shadow enhancer groups. **A**. Comparison of the distribution of pairwise correlation scores of shadow enhancers (top, blue) and randomly grouped enhancers (bottom, orange). The randomly grouped enhancers are a representative result of one instance of randomly grouping the over 1,000 mesoderm enhancers. **B**. To ease comparison, pairwise enhancer correlation scores were divided into three categories: negative (correlation less than or equal to 0), low (correlation between 0 and 0.5), and high (correlation greater than or equal to 0.5). Graph shows the fraction of pairwise comparisons in each group for the real shadow enhancer groups (blue) and the randomly grouped enhancers (orange) from all 1,000 instances of randomly grouping enhancers. The ratio of enhancer groups in each correlation category was not significantly different between real shadow enhancers and randomly grouped enhancers (p-value = 0.16 for negative correlation, 1 for low correlation, and 1 for high correlation; p-value from t-test with Bonferroni multiple testing correction). **C**. Distribution of median correlations across all groups (i.e. the median of all pairwise correlation scores) for all 1,000 instances of randomly grouping enhancers, where the black box indicates the first through third quartiles and the white circles indicates the median. The dashed vertical line indicates the median correlation of the true shadow enhancer groups.

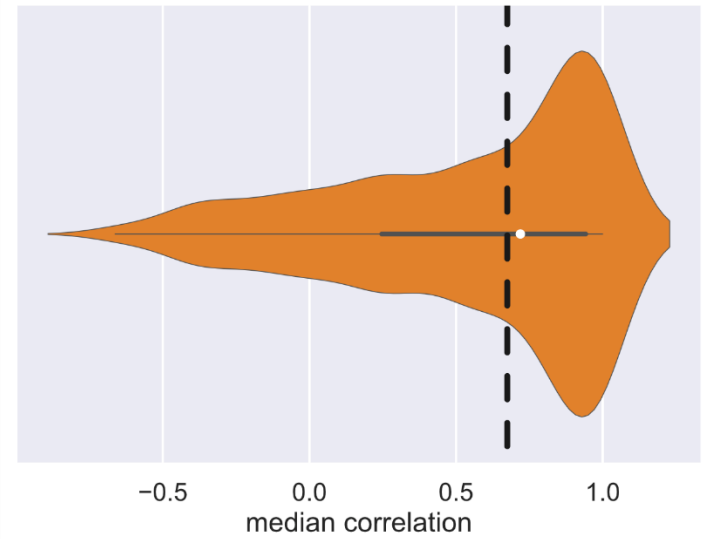
A


SE groups
random groups
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
correlation

B


real SE groups
scrambled groups
fraction
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
n.s
n.s
n.s
negative
low
high
pairwise correlation

C


−0.5
0.0
0.5
1.0
median correlation

4.3 References

1. Barolo S. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays*. 2012;34(2):135-141. doi:10.1002/bies.201100121
2. Hong JW, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty. *Science (80-)*. 2008;321(5894):1314. doi:10.1126/science.1160631
3. Hobert O. Gene regulation: Enhancers stepping out of the shadow. *Curr Biol*. 2010;20(17). doi:10.1016/j.cub.2010.07.035
4. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*. 2010;466(7305):490-493. doi:10.1038/nature09158
5. Osterwalder M, Barozzi I, Tissières V, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*. 2018;554(7691):239-243. doi:10.1038/nature25461
6. Perry MW, Boettiger AN, Bothma JP, Levine M. Shadow Enhancers Foster Robustness of Drosophila Gastrulation. *Curr Biol*. 2010;20(17):1562-1567. doi:10.1016/J.CUB.2010.07.043
7. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci*. 2011;108(33):13570-13575. doi:10.1073/PNAS.1109873108
8. Antosova B, Smolikova J, Klimova L, et al. The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. *PLoS Genet*. 2016;12(12). doi:10.1371/journal.pgen.1006441
9. Wunderlich Z, Bragdon MDJ, Vincent BJ, White JA, Estrada J, DePace AH. Krüppel Expression Levels Are Maintained through Compensatory Evolution of Shadow Enhancers. *Cell Rep*. 2015;12(11):1740-1747. doi:10.1016/j.celrep.2015.08.021
10. Waymack R, Fletcher A, Enciso G, Wunderlich Z. Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. *Elife*. 2020;9:1-57. doi:10.7554/ELIFE.59351
11. Cannavò E, Khoueiry P, Garfield DA, et al. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol*. 2016;26(1):38-51. doi:10.1016/j.cub.2015.11.034
12. Dunipace L, Ákos Z, Stathopoulos A. Coacting enhancers can have complementary functions within gene regulatory networks and promote canalization. Desplan C, ed. *PLOS Genet*. 2019;15(12):e1008525. doi:10.1371/journal.pgen.1008525
13. Preger-Ben Noon E, Sabarís G, Ortiz DM, et al. Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer Function. *Cell Rep*. 2018;22(11):3021-3031. doi:10.1016/j.celrep.2018.02.073
14. Cannavò E, Khoueiry P, Garfield DA, et al. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol*. 2016;26(1):38-51. doi:10.1016/j.cub.2015.11.034
15. Hertz GZ, Stormo GD. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. In: *Bioinformatics*. Vol 15. Oxford University Press; 1999:563-577. doi:10.1093/bioinformatics/15.7.563
16. Zhu LJ, Christensen RG, Kazemian M, et al. FlyFactorSurvey: A database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system.

Nucleic Acids Res. 2011;39(SUPPL. 1):D111. doi:10.1093/nar/gkq858
17. Li L, Wunderlich Z. An enhancer's length and composition are shaped by its regulatory task. *Front Genet*. 2017;8(MAY):63. doi:10.3389/fgene.2017.00063

Chapter 5

Assessment of transcription factor and enhancer dynamics in a non-developmental system

Chapter 5

Assessment of transcription factor and enhancer dynamics in a non-developmental system

As mentioned in the previous chapter, enhancers are critical for gene expression regulation in a variety of systems each facing different demands. As the primary determinants of gene expression, differences in enhancer activities likely underlie these differences, but this has yet to be thoroughly studied. In this short chapter, I provide a brief introduction to a project to uncover the regulatory logic of enhancers controlling the immune response.

5.1 Enhancers in the immune system

The bulk of this work has focused on enhancers regulating transcription in early development, but enhancers are used in systems well beyond development. The demands placed on enhancers for proper gene expression control may be quite different outside of development, for example in the immune system. Unlike development, which requires precise gene expression to reliably produce viable organisms, the immune system needs to be able to respond to the wide range of potential pathogens an organism may encounter. In line with these different needs, a large degree of expression variation is seen with gene targets of the immune system[1,2]. To investigate the role of enhancers in a system where greater gene expression variability is tolerated, I propose studies of how differences in NF-kB nuclear translocation dynamics translate into different transcriptional responses of immune target genes. These proposed experiments will shed light on if and how NF-kB dynamics are interpreted by enhancers to regulate gene expression during immune response. Further, by comparing these data with TF-enhancer input-output relationships measured during development we can gain an understanding of to what degree the principles governing enhancer function are universal or system-specific. Historically,

the study of enhancers has been biased towards those regulating developmental genes[3] so our understanding of how enhancers function may similarly be biased towards development-specific mechanisms.

5.1.1 The transcription factor NF-kB regulates the expression of immune-responsive genes

One specific pathway that provides a good model system in which to investigate the function and dynamics of enhancers in the immune system is that of NF-kB and its downstream target genes in response to immune stimulation[4,5]. NF-kB is the generic term for members of a family of TFs that prior to stimulation are sequestered in the cytoplasm, but upon immune stimulation translocate to the nucleus to regulate expression of a large number of immune-responsive genes[6,7]. This regulation is mediated by NF-kB, and other immune-responsive TFs, binding to enhancer and promoter regions of target genes[8–10]. In both the mammalian and *Drosophila* systems, these targets include genes encoding inhibitors of NF-kB[6,11]. This negative feedback results in the well-documented oscillations in mammalian cells of NF-kB in and out of the nucleus in response to immunological stimuli[12,13]. Significant variation is seen in these oscillations; when a population of cells is treated with the same continuous stimulus, these oscillations quickly become desynchronized[14]. Recent work has shown that differences in NF-kB dynamics across a population treated with the same stimulus are associated with different transcriptional responses[15], but a clear mechanism for this relationship has not been established.

5.1.2 Drosophila NF-kB signaling pathways provide a model system of innate immunity

There is a high degree of conservation between the NF-kB pathways in mammals and *Drosophila* and consequently *Drosophila* have been established as a highly informative model system for investigating pathways and functions of innate immunity[7,16]. *Drosophila* have two

NF-kB signaling pathways, the Toll and Imd pathways, that each are activated by different pathogens and use different NF-kB factors, Dorsal/Dif and Relish, to regulate activity of a distinct but overlapping set of immune-responsive genes[17]. While NF-kB oscillations in response to immune stimulation have not been explicitly studied in *Drosophila*, the NF-kB factor Dorsal also functions in early embryonic patterning and has been shown to move in and out of the nucleus during this process[18]. These pathways provide simpler, yet comparable, systems in which to investigate mechanisms of gene expression regulation that may well be generalizable to higher organisms, including humans[19].

I propose using the *Drosophila* innate immune response as a system in which to study the dynamics of NF-kB signaling and how these are translated into different transcriptional immune responses. I hypothesize that differential activation of immune-responsive enhancers by different NF-kB signals is involved in this transfer of information. An improved understanding of how NF-kB dynamics are translated into transcriptional control of immune-responsive genes will potentially shed light on how the innate immune system mounts pathogen-specific responses as well as what kind of mutations or perturbations may cause NF-kB signaling to go awry. Additionally, findings that different NF-kB signal dynamics stimulate different enhancer activity will spur further study into the molecular mechanisms underlying these differences.

5.1.3 Study of the relationship between NF-kB signaling and transcriptional response of immune genes

As the number of identified immune-responsive enhancers in *Drosophila* is currently lacking (although ongoing work in the lab aims to help address this), a pilot study should first be done using anti-microbial peptide (AMP) genes with known enhancer regions that are activated by NF-kB. Drosophila S2* cells, which are an immune-responsive cell line, combined with

fluorescent microscopy will allow the measurement of how NF-kB dynamics affect enhancer activity. These measurements will provide a useful starting point to compare the dynamics of enhancer-driven transcription in the immune system, which may need to respond only transiently to a wide range of signals, to those in the developmental system, which needs to reliably produce very accurate patterns of gene expression.

The first step in this project will be to determine the best way to stimulate NF-kB signaling in S2* cells. As *Drosophila* have two NF-kB pathways each activated by different types of pathogens, there are different possible ways to stimulate immune response in S2* cells[4]. To assess the response of cells to different kinds of immune stimuli, the endogenous NF-kB's (Dif or Relish) corresponding to the different NF-kB pathways (Toll and Imd) will be fluorescently labeled with GFP through CRISPR-based editing. With this system, one can compare the response of immune stimulation through both qPCR of known target genes of the two NF-kB pathways[20] and imaging of NF-kB translocations into and out of the nucleus. The Toll pathway can be activated by treating cells with cleaved Spatzle protein or using S2* cells that possess a Toll receptor that is constitutively active upon induction with copper[21]. The Imd pathway can be activated by treating cells with heat-killed bacteria or DAP-type peptidoglycan and ecdysone[22]. The optimal NF-kB pathway and mode of induction will be that which stimulates both a robust induction of AMP target genes and variable NF-kB dynamics across the cell population.

To measure the impact of NF-kB dynamics on enhancer activity, target enhancers must be chosen and tagged with aptamers. RNA aptamers, such as mango, work by integrating a sequence encoding a small RNA stem loop into the region of interest. The aptamer is then bound by a fluorescent dye that can be applied to the cells, which has negligible intrinsic fluorescence

but high intensity fluorescence upon binding the aptamer sequence[23]. Depending on the NF-kB pathway being used, the activity of different enhancers, which are only or preferentially activated by the respective NF-kB pathway, will be measured. For this small case study, comparing enhancers associated with genes preferentially activated by one pathway or the other will provide insight into whether differential enhancer activation plays a role in establishing the specificity or synergy of the Toll and Imd responses [16,24,25]. With the Toll pathway, the known enhancer region of *Drosomycin*, an AMP gene preferentially regulated by the this pathway makes a good readout of enhancer activity[10,20]. For the Imd pathway, the known enhancer region of the Imd AMP target *Diptericin* can be tracked[9,10]. With either pathway, the activity of the known enhancer region of *Attacin A*, an AMP gene induced by both Toll and Imd signaling, will also be used[10,24]. This enhancer that is activated by both signaling pathways will provide a comparison to distinguish between TF-enhancer dynamics that are involved in generating a pathway-specific response and those that are involved in generating a general immune response.

Once the above cell lines are established for tracking both NF-kB dynamics and enhancer activity in the same cell, an imaging procedure and analysis pipeline needs to be established. The goal of this imaging will be to simultaneously track NF-kB nuclear translocations (via GFP) and enhancer-driven transcription (via RNA aptamers) in single cells across a population of cells. Imaging populations of cells on 96 well plates should be straightforward and existing image processing codes[12,26,27] can be modified to separate and track individual cells and their corresponding fluorescence across time.

While this project is still in very early stages, much stands to be learned from these studies. Currently, I have stably transformed an S2* cell line to express a copper-inducible GFP-tagged Relish (NF-kB). With stable integration of the RNA aptamer reporters described above to

measure activity of *Diptericin* and *Attacin A*, a simple imaging system will enable analysis of the relationship between NF-kB dynamics and enhancer activity in response to any range of immune stimuli. With this system one can begin to ask questions such as does the same enhancer respond differently to different NF-kB signals, or do different enhancers respond differently to the same NF-kB signal? Further, these data can be compared to my own and others' of enhancer-driven transcription in the context of development to gain an understanding of universal or system-specific principles of enhancer activity.

5.2 Methods

5.2.1 Establishment of stable Rel-GFP cell line

To enable visualization of NF-kB dynamics in response to immune stimulation and in relation to enhancer dynamics, a cell line stably expressing GFP-tagged Relish (Rel) was generated. *Drosophila* S2* cells were transfected with Rel-GFP under the control of a copper inducible promoter (kindly provided by Dr. Edan Foley)[28] via electroporation. 4 million cells were transfected with 8.3ug of plasmid DNA and 1.7ug of a blasticidin resistance plasmid (pCoBlast; Invitrogen, Carlsbad CA) using the SF Cell Line 4D-Nucleofector X kit (Lonza, Basel, Switzerland) and the DG-137 transfection setting on a 4D-Nucleofector (Lonza, Basel, Switzerland). Following transfection, cells were returned to complete Schneiders media and allowed to grow for two days. Cells were then collected and resuspended in complete Schneiders media containing 5ug/mL of blasticidin every 3-4 days for 2 weeks to select for stably transfected cells.

5.2.2 Generation of aptamer-tagged enhancer expression plasmids

To enable visualization of immune-responsive enhancer activity in relation to NF-kB signaling, plasmids were generated containing Mango aptamer sequences[23] and a Luciferase reporter under the control of the *Attacin A* enhancer[29]. DNA sequence of the Mango IV, III, II, and I[23,30] sequences each separated by 30bp of lambda phage DNA, to minimize repeat removal, was synthesized by IDT (San Diego, CA). The *Attacin A* enhancer, a minimal *eve* promoter, and the Mango sequences were integrated into the pGL3 basic plasmid (Promega, Madison, WI), which contains an optimized firefly luciferase sequence. To ensure proper translation of the luciferase protein, a second identical enhancer sequence and minimal *eve* promoter drive luciferase expression. Similar plasmids can be constructed using the enhancer sequences of other immune-responsive genes (i.e. *Drosomycin*, *Diptericin*)[29]. These plasmids can be stably integrated into S2* cells stably expressing GFP-Rel (such as those described above), or with GFP tagging of endogenous Relish, using the same co-transfection with the pCoBlast plasmid as described above.

5.3 References

1. Sackton TB, Lazzaro BP, Clark AG. Genotype and Gene Expression Associations with Immune Function in Drosophila. Begun DJ, ed. *PLoS Genet*. 2010;6(1):e1000797. doi:10.1371/journal.pgen.1000797
2. Nédélec Y, Sanz J, Baharian G, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. 2016;167(3):657-669.e21. doi:10.1016/J.CELL.2016.09.025
3. Sakabe NJ, Savic D, Nobrega MA. Transcriptional enhancers in development and disease. *Genome Biol*. 2012;13(1):1-11. doi:10.1186/gb-2012-13-1-238
4. Ganesan S, Aggarwal K, Paquette N, Silverman N. NF-κB/Rel proteins and the humoral immune responses of Drosophila melanogaster. *Curr Top Microbiol Immunol*. 2011;349:25-60. doi:10.1007/82_2010_107
5. Hoffmann A, Natoli G, Ghosh G. Transcriptional regulation via the NF-κB signaling module. *Oncogene*. 2006;25(51):6706-6716. doi:10.1038/sj.onc.1209933
6. Hoffmann A, Baltimore D. Circuitry of nuclear factor kappaB signaling. *Immunol Rev*. 2006;210(1):171-186. doi:10.1111/j.0105-2896.2006.00375.x
7. Silverman N, Maniatis T. NF-kappaB signaling pathways in mammalian and insect innate immunity. *Genes Dev*. 2001;15(18):2321-2342. doi:10.1101/gad.909001
8. Lindsay SA, Wasserman SA. Conventional and non-conventional Drosophila Toll signaling. *Dev Comp Immunol*. 2014;42(1):16-24. doi:10.1016/j.dci.2013.04.011
9. Meister M, Braun A, Kappler C, Reichhart JM, Hoffmann JA. Insect immunity. A transgenic analysis in Drosophila defines several functional domains in the diptericin promoter. *EMBO J*. 1994;13(24):5958-5966. http://www.ncbi.nlm.nih.gov/pubmed/7813433. Accessed October 1, 2018.
10. Senger K, Armstrong GW, Rowell WJ, Kwan JM, Markstein M, Levine M. Immunity regulatory DNAs share common organizational features in Drosophila. *Mol Cell*. 2004;13(1):19-32. doi:10.1016/S1097-2765(03)00500-8
11. Nicolas E, Reichhart JM, Hoffmann JA, Lemaitre B. In vivo regulation of the IkappaB homologue cactus during the immune response of Drosophila. *J Biol Chem*. 1998;273(17):10463-10469. http://www.ncbi.nlm.nih.gov/pubmed/9553105. Accessed September 30, 2018.
12. Ashall L, Horton CA, Nelson DE, et al. Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription. *Science*. 2009;324(5924):242-246. doi:10.1126/science.1164860
13. Nelson DE, Ihekwaba AEC, Elliott M, et al. Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science*. 2004;306(5696):704-708. doi:10.1126/science.1099962
14. Kellogg RA, Tay S. Noise Facilitates Transcriptional Control under Dynamic Inputs. *Cell*. 2015;160(3):381-392. doi:10.1016/J.CELL.2015.01.013
15. Lane K, Van Valen D, DeFelice MM, et al. Measuring Signaling and RNA-Seq in the Same Cell Links Gene Expression to Dynamic Patterns of NF-κB Activation. *Cell Syst*. 2017;4(4):458-469.e5. doi:10.1016/j.cels.2017.03.010
16. Hoffmann JA, Reichhart J-M. Drosophila innate immunity: an evolutionary perspective. *Nat Immunol*. 2002;3(2):121-126. doi:10.1038/ni0202-121
17. Hetru C, Hoffmann JA. NF-kappaB in the immune response of Drosophila. *Cold Spring*

Harb Perspect Biol. 2009;1(6):a000232. doi:10.1101/cshperspect.a000232
18. DeLotto R, DeLotto Y, Steward R, Lippincott-Schwartz J. Nucleocytoplasmic shuttling mediates the dynamic maintenance of nuclear Dorsal levels during Drosophila embryogenesis. *Development*. 2007;134(23):4233-4241. doi:10.1242/dev.010934
19. Govind S. Innate immunity in Drosophila: Pathogens and pathways. *Insect Sci*. 2008;15(1):29-43. doi:10.1111/j.1744-7917.2008.00185.x
20. Hoffmann JA. The immune response of Drosophila. *Nature*. 2003;426(6962):33-38. doi:10.1038/nature02021
21. Silverman N, Zhou R, Stöven S, Pandey N, Hultmark D, Maniatis T. A Drosophila IκB kinase complex required for relish cleavage and antibacterial immunity. *Genes Dev*. 2000;14(19):2461-2471. doi:10.1101/gad.817800
22. Myllymäki H, Valanne S, Rämet M. The Drosophila imd signaling pathway. *J Immunol*. 2014;192(8):3455-3462. doi:10.4049/jimmunol.1303309
23. Dolgosheina E V., Jeng SCY, Panchapakesan SSS, et al. RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chem Biol*. 2014;9(10):2412-2420. doi:10.1021/cb500499x
24. De Gregorio E, Spellman PT, Tzou P, Rubin GM, Lemaitre B. The Toll and Imd pathways are the major regulators of the immune response in Drosophila. *EMBO J*. 2002;21(11):2568-2579. doi:10.1093/emboj/21.11.2568
25. Tanji T, Hu X, Weber ANR, Ip YT. Toll and IMD Pathways Synergistically Activate an Innate Immune Response in Drosophila melanogaster. *Mol Cell Biol*. 2007;27(12):4578-4588. doi:10.1128/mcb.01814-06
26. RABUT G, ELLENBERG J. Automatic real-time three-dimensional cell tracking by fluorescence microscopy. *J Microsc*. 2004;216(2):131-137. doi:10.1111/j.0022-2720.2004.01404.x
27. Zambrano S, De Toma I, Piffer A, Bianchi ME, Agresti A. NF-κB oscillations translate into functionally related patterns of gene expression. *Elife*. 2016;5:e09100. doi:10.7554/eLife.09100
28. Foley E, O'Farrell PH. Functional dissection of an innate immune response by a genome-wide RNAi screen. *PLoS Biol*. 2004;2(8):e203. doi:10.1371/journal.pbio.0020203
29. Senger K, Armstrong GW, Rowell WJ, Kwan JM, Markstein M, Levine M. Immunity Regulatory DNAs Share Common Organizational Features in Drosophila. *Mol Cell*. 2004;13(1):19-32. doi:10.1016/S1097-2765(03)00500-8
30. Autour A, Jeng SCY, Cawte AD, et al. Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nat Commun*. 2018;9(1):1-12. doi:10.1038/s41467-018-02993-8

Chapter 6

Future Directions

Chapter 6

Future Directions

In the preceding chapters, I discussed my work investigating the details of how transcription factors (TFs) and enhancers work together to regulate gene expression. In Chapter 2, I first showed that a separation of TF inputs enables the *Kruppel* shadow enhancers to buffer noise and drive stable levels of gene expression. I then in Chapter 3 examined how local competition for TFs can regulate enhancer activity and the implications this has for our understanding of nuclear organization. In Chapter 4, I analyzed a large dataset of Drosophila developmental shadow enhancers and found that these enhancers display a range of TF regulatory logic arrangements. Lastly, in Chapter 5 I discuss a proposed study of NF-kB signaling and enhancer dynamics in the immune system to improve our understanding of the rules regulating the relationship between TFs and enhancers in a very different system than development. While my work has provided insight into the complicated ways TFs and enhancers can interact to control gene expression, much still remains to be discovered. In this final chapter I discuss some remaining questions and areas the field should explore.

6.1 Deciphering the cis-regulatory code

Although we have long known about enhancers and other regulatory regions of DNA, a continued but still unreached goal of the field is the ability to predict the function or activity of a regulatory region from its sequence, much as we do with protein-coding sequences. This goal may not be fully achievable, due in part to factors such as the 3D organization of the nucleus as discussed later, but is still a meaningful one to work towards. Numerous studies have implicated mutations in or loss of enhancers in developmental defects and disease[1–4]. The importance of

enhancers and other regulatory DNA elements in human health and disease is made clear by the ever-growing number of GWAS studies implicating non-coding sequences in diseases[5]. In theory, an ability to predict enhancer function from sequence would greatly improve our ability to predict, diagnose, and potentially treat (via genome-editing based methods) such diseases. This is complicated by the fact, as discussed in Chapter 1, that many genes that are associated with human diseases are regulated by multiple enhancers. Further, my findings in Chapter 2 indicate that predicting the combined activity of multiple enhancers from the activity of the individual enhancers is not always straightforward. Although the goal of accurately predicting gene expression from enhancer DNA sequence alone may or may not ever be feasible, it is still a worthwhile pursuit. Our ability to simply identify enhancers within genomes, which is currently still a challenge[6,7], would be vastly improved with an improved understanding of a sequence-function relationship of enhancers.

In Chapters 2 and 4, I discuss the TF regulatory logic controlling shadow enhancers and the implications this has for shadow enhancer functions. I first showed that the pair of *Kruppel* (*Kr*) embryonic shadow enhancers are able to buffer against both global and stochastic sources of noise through a separation of TF inputs between the two enhancers. This separation of TF inputs enables the shadow enhancer pair to drive more stable gene expression across a range of environmental conditions than can a simple enhancer duplication. Such a finding suggests that separated TF inputs may be a general mechanism by which shadow enhancers mediate robust gene expression. My findings in Chapter 4 in part support this idea, as many enhancers in a large dataset of shadow enhancers display greater TF separation than expected by chance, but they also indicate that shadow enhancers vary widely in their degree of TF separation. This suggests (perhaps intuitively) that shadow enhancers serve a wide range of functions and consequently do

not all face the same selective pressure to evolve certain features, such as a separation of TF inputs. In addition to driving robust gene expression, many shadow enhancers have been shown to be important for refining gene expression patterns[8–10], and a recent study found deletions of individual shadow enhancers result in different phenotypes[11]. This suggests that the individual enhancers play distinct roles in regulating gene expression. It is likely that different TF regulation strategies help enable different groups of shadow enhancers to serve these different functions.

An obvious next step for this work is to explore the extent of TF separation among shadow enhancers in different organisms, such as mice[12], and measure the expression noise associated with shadow enhancers with varying degrees of TF separation. These studies will give us insight into how widespread differential TF regulation of shadow enhancers is, whether this regulation is always a functional characteristic that enables buffering of upstream noise, and how adjustable this characteristic is. Further, in combination with experimental validation (i.e. measurements of expression noise associated with shadow enhancers), machine learning approaches seem potentially well suited to help elucidate the patterns of TF regulation of shadow enhancers and how these relate to shadow enhancer activity. With this understanding, we can begin to better understand shadow enhancers function from the individual enhancer sequences, as well as refine our ability to identify shadow enhancers as current methods have often included looking for clusters of similar TF binding sites[13,14].

6.2 Synthesizing studies of cis and trans regulators to understand how gene expression is controlled in the 3D nucleus

An exciting topic the field should continue to explore is the idea of the nucleus as a 3D space in which DNA and the transcriptional machinery are not uniformly distributed, but must

organize and find one another and the impact this has on gene expression. In Chapter 3, I uncover competition for TFs that can modulate gene expression levels and patterns. Our thermodynamic modeling suggests that such competition is dependent on the non-uniform distribution of TFs in the nucleus. In that chapter, I discuss how our findings have implications not only for the use of transgenic reporters, but for our understanding of transcriptional regulation more broadly. It has been increasingly shown that TFs and other pieces of the transcriptional machinery cluster in regions of high local concentration, or hubs[15–18]. How exactly the formation and dynamics of these hubs are regulated will continue to be an active area of interest, as my work and that of others indicates these hubs play a critical role in controlling gene expression[16,17,19]. A recent study showed that shadow enhancers of the gene *shavenbaby* (*svb*) in *Drosophila* are critical for establishing high local TF concentrations that seem to be required for normal levels of gene expression[16]. This raises the question of whether the establishment of high concentration TF hubs is a general feature of shadow enhancers that underlie their ability to drive robust gene expression. One could imagine this working in concert with the noise buffering ability explored in Chapter 2, where shadow enhancers facilitate the formation of TF hubs but also can have built-in features to buffer against fluctuations in the exact size and composition of these hubs as they have been shown to be quite dynamic[17,19,20].

My findings in Chapter 3 in the context of other studies of TF hubs indicate a complex relationship between these hubs and transcriptional regulation. While most previous studies focused on the association between TF hubs and high levels of transcription[19–22], our findings indicate that the non-uniform distribution of TFs in the nucleus can also produce competition among genomic targets for TF activation. Studies simultaneously tracking TF concentrations and the transcriptional activity of two or more target enhancers will help further elucidate the

relationship between TF hubs and enhancer activity. Ideally, one would want to simultaneously track the position of large regions of DNA (i.e. whole TADs or shadow enhancers), TF levels, and transcriptional output of TF targets. This would enable us to address many currently open questions. How much DNA is typically interacting within a hub and how important is the identity of this DNA sequence in hub formation or duration? We know that the hubs of TFs and other proteins are dynamic and relatively short-lived. Does the associated DNA follow similar dynamics, perhaps moving with (or causing!) the shifting hubs? Lastly and most importantly, how do the composition and dynamics of these hubs affect gene expression? Have genomes evolved to ensure that the most critical, or sensitive, genes are preferentially found in the most stable hubs? While serious technological advances in our ability to visualize and track multiple molecular species in living cells will be needed to answer these questions, recent studies provide glimpses of the promise such advances would bring[16].

6.3 References

1. Sakabe NJ, Savic D, Nobrega MA. Transcriptional enhancers in development and disease. *Genome Biol*. 2012;13(1):1-11. doi:10.1186/gb-2012-13-1-238
2. Herz HM, Hu D, Shilatifard A. Enhancer malfunction in cancer. *Mol Cell*. 2014;53(6):859-866. doi:10.1016/j.molcel.2014.02.033
3. Carullo NVN, Day JJ. Genomic enhancers in brain health and disease. *Genes (Basel)*. 2019;10(1). doi:10.3390/genes10010043
4. Chatterjee S, Ahituv N. Gene Regulatory Elements, Major Drivers of Human Disease. *Annu Rev Genomics Hum Genet*. 2017;18:45-63. doi:10.1146/annurev-genom-091416-035537
5. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet*. 2015;24(R1):R102-R110. doi:10.1093/hmg/ddv259
6. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. 2014;15(4):272-286. doi:10.1038/nrg3682
7. Benton ML, Talipineni SC, Kostka D, Capra JA. Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function. *BMC Genomics*. 2019;20(1):511. doi:10.1186/s12864-019-5779-x
8. El-Sherif E, Levine M. Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo. *Curr Biol*. 2016;26(9):1164-1169. doi:10.1016/j.cub.2016.02.054
9. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci*. 2011;108(33):13570-13575. doi:10.1073/PNAS.1109873108
10. Perry MW, Bothma JP, Luu RD, Levine M. Precision of Hunchback Expression in the Drosophila Embryo. *Curr Biol*. 2012;22(23):2247-2252. doi:10.1016/J.CUB.2012.09.051
11. Dunipace L, Ákos Z, Stathopoulos A. Coacting enhancers can have complementary functions within gene regulatory networks and promote canalization. Desplan C, ed. *PLOS Genet*. 2019;15(12):e1008525. doi:10.1371/journal.pgen.1008525
12. Osterwalder M, Barozzi I, Tissières V, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*. 2018;554(7691):239-243. doi:10.1038/nature25461
13. Hong JW, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty. *Science (80-)*. 2008;321(5894):1314. doi:10.1126/science.1160631
14. Cannavò E, Khoueiry P, Garfield DA, et al. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol*. 2016;26(1):38-51. doi:10.1016/j.cub.2015.11.034
15. Mir M, Reimer A, Haines JE, et al. Dense bicoid hubs accentuate binding along the morphogen gradient. *Genes Dev*. 2017;31(17):1784-1794. doi:10.1101/gad.305078.117
16. Tsai A, Alves MRP, Crocker J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *Elife*. 2019;8. doi:10.7554/eLife.45325
17. Cho W-K, Spille J-H, Hecht M, et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*. 2018;361(6400):412.

doi:10.1126/SCIENCE.AAR4199
18. Boehning M, Dugast-Darzacq C, Rankovic M, et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol*. 2018;25(9):833-840. doi:10.1038/s41594-018-0112-y
19. Chong S, Dugast-Darzacq C, Liu Z, et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science (80-)*. 2018;361(6400). doi:10.1126/science.aar2555
20. Mir M, Stadler MR, Ortiz SA, et al. Dynamic multifactor hubs interact transiently with sites of active transcription in drosophila embryos. *Elife*. 2018;7. doi:10.7554/eLife.40497
21. Boija A, Klein IA, Sabari BR, et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*. 2018;175(7):1842-1855.e16. doi:10.1016/j.cell.2018.10.042
22. Sabari BR, Dall'Agnese A, Boija A, et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science (80-)*. 2018;361(6400):eaar3958. doi:10.1126/science.aar3958