INSIGHTS INTO NUCLEAR ARCHITECTURE AND STOCHASTIC GENE EXPRESSION THROUGH THE STUDY OF HOMOLOGOUS CHROMOSOME PAIRING AND TRANSVECTION

by

Kayla Viets

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March 2019

Abstract

Chromosomes are organized in a complex manner within the nucleus. Their localization relative to activating and repressing nuclear compartments and to other chromosomes can have major effects on gene expression. One important aspect of nuclear architecture involves the physical colocalization, or “pairing” of different alleles of the same gene. Pairing is involved in mammalian processes including genomic imprinting and X-inactivation, but the most well-studied example occurs in *Drosophila melanogaster*, where homologous chromosomes are paired along their entire lengths throughout interphase. Homologous chromosome pairing in fruit flies facilitates a gene-regulatory phenomenon known as transvection, in which DNA elements on one mutant allele of a gene act between chromosomes on another mutant allele to rescue expression. While homologous chromosome pairing and transvection were first described over 60 years ago, the mechanisms that drive homologous chromosome pairing and transvection across the genome are still unclear. Here, we develop a DNA FISH-based approach to identify button regions across the fly genome that drive pairing. These buttons allow “reconstitution” of genes that are split apart by chromosome rearrangements, suggesting that buttons play a role in maintaining the structural integrity of the genome. Buttons are enriched for topologically associated domains (TADs), indicating that TADs are responsible for homologous chromosome interactions. Using the stochastically expressed *spineless* (*ss*) locus as a paradigm, we gain deeper insight into the mechanisms that control transvection. We find that pairing is necessary but not sufficient for

transvection, and that pairing and transvection are cell-type specific. We also identify the DNA elements required for the separable mechanisms of activation and repression between chromosomes. Furthermore, we find a biological role for transvection in regulating the expression of naturally occurring *ss* alleles to control photoreceptor patterning. Our work suggests a model in which specialized TADs drive homologous chromosome pairing to facilitate cell-type-specific interchromosomal gene regulation. This work has important implications for our understanding of nuclear architecture-linked diseases including Prader-Willi and Angelman syndromes, breast, renal, and pancreatic cancers, and limb malformations.

Primary readers: Xin Chen and Robert J. Johnston, Jr.

Secondary readers: Joseph Gall, Karen Reddy, and Yixian Zheng

Acknowledgements

My heartfelt thanks go to all of the members of the Johnston lab who have provided invaluable scientific ideas, experimental advice, and friendship throughout my PhD. I would especially like to thank my PI, Bob Johnston, for his supportive and enthusiastic mentorship, championing of his students, and copious free chocolate.

I would also like to thank the many colleagues and collaborators who have contributed to the success of this work by generously providing their feedback, reagents, and time, including Eric Joyce, Ting Wu, Jelena Erceg, Jumana AlHaj Abed, Marco Gallio, James Taylor, Mike Sauria, Max Echterling, Peter DeFord, Andrew Gordus, Claude Desplan, Geraldine Seydoux, Jeff Corden, Judy Kassis, and the Johns Hopkins Tri-Fly labs.

Additionally, I would like to thank my thesis committee members, Joe Gall, Karen Reddy, and Yixian Zheng. Their constructive feedback, excellent scientific guidance, and support of my personal and career goals have been essential to my success in graduate school.

Finally, I would like to thank my amazing friends and family. My friends have made the last five and a half years in Baltimore some of the best of my life. I am so grateful to have gone through graduate school with such an amazing group of people who support each other through tough scientific times and celebrate each other's successes. I am also certain that I have the best family around. Without their love, their cheerleading, and their wholehearted support of my career goals throughout my life, I would never have made it as a scientist. I

am also extremely lucky to have found a wonderful second family in my soon-to-be in-laws, and want to thank them for all their support over the past few years, especially these past few months as I've been writing my thesis.

Most importantly of all, I want to thank my wonderful fiancé, Alex, for always being there for me through five and a half years of grad school plus five years of long distance. I can't imagine a better, sweeter, more understanding person to have had by my side for the past five and a half years, and I am so excited to finally start a new adventure together as the Master and the Doctor.

Publications

1) “Probing chromosome pairing and transvection to understand nuclear organization” (Chapter 1) is a review article currently in preparation.
2) “Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates” (Chapter 2) is a review article that was published in *Trends in Genetics* in 2016.
3) “TADs pair homologous chromosomes to promote interchromosomal gene regulation” (Chapter 3) is a research article that is in revision at *Developmental Cell*. It was also posted to *bioRxiv* in 2018.
4) “Activating and repressing stochastic gene expression between chromosomes” (Chapter 4) is a research article that is in preparation for submission to *Developmental Cell*.
5) “The *mir-279/996* cluster represses receptor tyrosine kinase signaling to determine cell fates in the *Drosophila* eye” (Appendix 1) is a research article that was published in *Development* in 2018.
6) “Regulatory logic driving stable levels of *defective proventriculus* expression during terminal photoreceptor specification in flies” (Appendix 2) is a research article that was published in *Development* in 2017.
7) “The insulator protein BEAF-32 is required for Hippo pathway activity in the terminal differentiation of neuronal subtypes” (Appendix 3) is a research article that was published in *Development* in 2016.

Table of Contents

List of Tables

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Chapter 1: Probing chromosome pairing and transvection to understand nuclear organization

1.1: Abstract

The organization of chromosomes within the nucleus is a major factor regulating gene expression. At the genome-wide level, chromosomes localize to specific domains known as chromosome territories. At the individual chromosome level, chromosomes fold into regions known as topologically associated domains (TADs), regions of close self-association that are hypothesized to isolate genes into regulatory domains. While a handful of studies have indicated that disruption of TADs and localization of genes to improper regions of the nucleus can have major effects on gene expression, it is unclear how nuclear architecture in general is regulated to maintain transcriptional states. One important aspect of nuclear organization involves long-distance interactions between different regions of the genome, which is observed in processes including compartmentalization, X-inactivation, and genomic imprinting. One of the most well-studied examples of interchromosomal interactions is the phenomenon of homologous chromosome pairing in *Drosophila melanogaster*. Homologous chromosome pairing facilitates an interchromosomal gene-regulatory process known as transvection, in which two mutant alleles of a gene interact between chromosomes to rescue gene expression. While homologous chromosome pairing and transvection have been studied for decades, much remains unknown about the mechanisms facilitating these processes. Here, we discuss recent developments in our understanding of homologous chromosome

pairing and transvection and their relevance to similar phenomena in mammalian systems. These classical *Drosophila* processes offer an excellent paradigm for understanding the role of nuclear architecture in regulating gene expression across organisms.

1.2: Introduction

Chromosomes are organized non-randomly within the 3D space of the nucleus. In metazoans, chromosomes occupy distinct regions of the nucleus known as chromosome territories (**Fig. 1A**)(*1*). While the localization of chromosome territories varies widely from cell to cell, larger chromosomes and gene-poor chromosomes tend to localize near the nuclear periphery, whereas smaller chromosomes and gene-rich chromosomes tend to localize to the nuclear interior (**Fig. 1A**)(*1-4*). Regions along chromosomes interact to form compartments, which are segregated into A (active) and B (repressive) categories based on gene expression states (**Fig. 1B**)(*5-8*). Individual chromosomes are organized further into topologically associated domains (TADs), regions of close self-association that are hypothesized to isolate genes into regulatory domains and ensure their activation by the appropriate *cis*-regulatory elements (**Fig. 1C**)(*6, 9-13*).

More locally, targeting of individual genes to different nuclear subcompartments can have major effects on gene expression. For example, a number of yeast and *Drosophila* genes localize to the nuclear pore complex (NPC) upon activation, and their retention at the NPC is essential in establishing

an epigenetic “memory” of their active state (*14-16*). In *Drosophila* neuroblast development, the *hunchback* locus moves from the activating nuclear interior to the repressive nuclear periphery, ending the competence window for the generation of early-born neurons (*17*). Similarly, in human lymphocyte development, the IgH and Igκ loci localize to the nuclear periphery in hematopoietic progenitor cells and pro-T cells, where they are not expressed (*18*). In pro-B cells, these loci localize to the nuclear interior, likely allowing their activation and recombination (*18*).

Disruptions of TADs and targeting of loci to improper regions of the nucleus can have major effects on gene expression and is associated with limb malformations, cancer, Rett syndrome, Prader-Willi syndrome, Angelman syndrome, autism, and gliosis (*13, 19-26*). However, we do not have a clear understanding of how the precise folding and localization of chromosomes and genes within the nucleus works to maintain proper gene expression. Chromosomes are segregated non-randomly into territories, but we understand little about how interchromosomal interactions between chromosome territories affect gene expression. While there are many examples of individual loci whose expression is disrupted by changes in nuclear localization and chromosome conformation, dissolution of mammalian TADs has minimal effects on transcription genome-wide (*27-29*). TAD organization is largely conserved between cell types and even between organisms (*10, 30*), but gene expression varies widely between cells to ensure the development of distinct cell fates. What is the function of TADs? What drives interactions between chromosomes, and

how do interchromosomal interactions affect gene expression? In this review, we discuss how the *Drosophila* phenomena of homologous chromosome pairing and transvection can help us to answer these fundamental questions.

1.3: Interchromosomal interactions and their role in gene expression

Chromosomes are organized into discrete territories within the nucleus, but what effect do interactions between territories have on gene expression? Interchromosomal associations are involved in a variety of developmental and gene-regulatory processes, including co-regulation of genes with similar expression profiles, imprinting, and X-inactivation. These associations can be divided into two categories: interactions between heterologous chromosomes, and interactions between homologous chromosomes.

Heterologous chromosome interactions

Frequently, interchromosomal interactions between non-homologous loci provide a means of regulating multiple genes simultaneously. One way this occurs is through the interaction of shared regulatory elements with multiple loci. For example, in murine $CD4^+$ T cells, immune response genes on chromosomes 10 and 11 form a complex with a shared locus control region, poising them for activation (*31, 32*). In mouse olfactory neurons, olfactory receptor enhancers drive interactions between olfactory receptor genes located on 18 different chromosomes, forming a super enhancer that drives olfactory receptor gene expression (*33*).

The regulation of multiple genes at once also occurs through gene clustering. In mouse olfactory neurons, silent olfactory receptor genes from multiple chromosomes interact in heterochromatic foci to ensure exclusive expression of a single olfactory gene per neuron (*34*). More generally, gene clustering allows the concentration of transcription factors or chromatin proteins at a single site, allowing more efficient gene activation or repression. Active genes colocalize at subnuclear regions called transcription factories, where high concentrations of RNA polymerase allow coordination of gene expression within a single structure (*35-37*). Transcription factories also contain high concentrations of the insulator protein CTCF, supporting a role for this protein in driving interchromosomal interactions (*38*). Conversely, repressed genes containing Polycomb Response Elements (PREs) come together in Polycomb bodies, which contain a high concentration of the repressive Polycomb Group complex (*35, 39, 40*). Thus, interchromosomal interactions between non-homologous loci facilitate the proper expression of multiple genes simultaneously.

Homologous chromosome interactions

Interactions between identical sequences located on homologous chromosomes represent another important aspect of nuclear organization. Unlike interactions between heterologous chromosomes, which simply require targeting to a similar area of the nucleus, homologous chromosome interactions require a sequence to find a single unique partner within the entire nuclear volume. The *cis*

and *trans* factors required for the co-localization, or pairing, of homologous sequences are less well understood than for interchromosomal heterologous clustering.

Homologous pairing is important for a number of cellular processes (reviewed in (*41*)), including mammalian X-inactivation and imprinting. Through X-inactivation, one copy of the X chromosome is repressed in females to ensure a similar gene dosage between males and females (reviewed in (*42*)). A region known as the X inactivation center (Xic) counts the number of X chromosomes present in a cell and determines which copy of the X chromosome to inactivate (*43, 44*). At the onset of X-inactivation, the two copies of the Xic transiently pair (*45, 46*). Deletion of Xic disrupts pairing and abolishes X-inactivation, suggesting that Xic-driven pairing is required for X-inactivation (*45, 46*). Small fragments of a 15-kb region within the Xic are sufficient to drive pairing (*47, 48*). These fragments bind to CTCF and the transcription factor Oct4, both of which are required for pairing (*47, 49*).

Genomic imprinting is an epigenetic phenomenon in which either the maternal or paternal allele of a gene is heritably silenced (*50*)(reviewed in (*51*)). In certain cases, imprinting appears to require transient homologous pairing between the two copies of the imprinted locus. For example, the imprinted human 15q11-q13 locus is paired in 58% of wild type lymphocyte nuclei during late S-phase (*52*). This association is lost in cells taken from patients with imprinting-related Prader-Willi, Rett, and Angelman syndromes (*26, 52*). Knockdown of the gene *MECP2* causes decreased pairing, suggesting that

MECP2-mediated pairing allows proper imprinting at this region (*26*). Homologous pairing of imprinted sites has also been observed at the human *H19* locus and the distal region of mouse chromosome 7 (*52, 53*), suggesting that pairing may be a more general mechanism regulating genomic imprinting.

While interchromosomal interactions are essential for a number of mammalian processes, these interactions typically occur transiently or in a low frequency of cells (*45, 52*). As we describe below, *Drosophila* homologous chromosome pairing occurs stably and at a high frequency, providing an ideal paradigm for the study of interactions between chromosomes.

1.4: Homologous chromosome pairing in *Drosophila* provides insight into nuclear organization

One of the most well-studied examples of pairing occurs in *Drosophila*, where homologous chromosome pairing occurs throughout interphase in nearly all somatic cells (*54*). This stable pairing provides an excellent paradigm to study interchromosomal interactions. *Drosophila* pairing was first described over a century ago by Nettie Stevens (*54*), but the mechanisms driving the colocalization of homologous chromosomes have remained unclear. Recently, a number of studies have shed light on this long-standing mystery in the field by identifying chromosome structures and *trans* factors involved in bringing homologous chromosomes together.

Models for pairing initiation

Two main models have been proposed to describe the initiation of homologous chromosome pairing. Classical studies of pairing hypothesized a "zipper" model (**Fig. 2A**), in which chromosomes pair sequentially along their entire lengths, beginning at the centromere and proceeding distally to the telomere (*55, 56*). Supporting this model, chromosome rearrangements with breakpoints proximal to a gene disrupt pairing-dependent gene regulation between chromosomes (see below), whereas rearrangements distal to a specific gene do not (*55, 56*). This model predicts that pairing is based solely on DNA sequence homology, so all regions of the genome have an equal ability to initiate homologous chromosome pairing.

More recent studies of pairing initiation during embryonic development suggest an alternative to the zipper model (*57-59*). Fung and colleagues used DNA fluorescence *in situ* hybridization (FISH) to examine the pairing of 11 sites along chromosome 2L and found that pairing initiated independently at discrete sites, rather than progressing sequentially from the centromere to the telomere as the zipper model predicted (*58*). This work and other related studies (*57, 59*) proposed a "button" model, in which specific regions interspersed along chromosome arms drive pairing at a higher affinity than surrounding regions (**Fig. 2A**).

Trans factors and chromosome structural features involved in homologous chromosome pairing

Chromosome pairing initiates at specific buttons, but what unique features of buttons allow them to find their homologous partners in the nucleus? RNAi knockdown of *topoisomerase II, slmb, zelda*, and certain mitotic genes (*60-64*) decreases homologous chromosome pairing. The binding of insulator proteins including CTCF and Cp190 has also been linked to homologous pairing (*65*). Additionally, button regions are correlated with active chromatin, A compartments, and high levels of transcription (*62*). It is unlikely that transcriptional status or the binding of a single protein alone could provide the specificity required for a button to identify its unique homologous partner within the nucleus. However, a gene's transcriptional status might target it to a specific nuclear compartment, decreasing the search space for a button to find its partner (**Fig. 2B**). Within this decreased search space, binding of unique combinations of these pairing-promoting proteins to each button might provide the specificity required for a button to find its unique homologue (**Fig. 2C**).

The condensin II complex is an "anti-pairer": overexpression of condensin II subunits decreases pairing, while knockout of condensin II subunits increases pairing (*60, 66*). A low level of condensin II binding at buttons may also give them a higher pairing affinity than surrounding regions.

A number of small DNA sequence elements have also been linked to homologous chromosome pairing. The gypsy retrotransposon and a handful of specific insulator and Polycomb Response Elements (PREs), including the Fab-

7, Mcp, and TMR regions of the *Abd-B* locus, drive pairing between identical homologous sequences on different chromosomes (*65, 67-71*). However, these short sequences are unique to specific loci and cannot account for the pairing that is observed genome-wide.

It is thus likely that a larger chromosome structural feature is required to provide the specificity necessary for buttons to drive pairing genome-wide. Recent work supports a model in which a subset of TADs form buttons and drive pairing initiation (*72*). Transgenes containing TADs selected from chromosomes X, 2L, 2R, and 3R drive pairing with their endogenous sites from multiple locations in the genome (*72*). This TAD-driven pairing occurs independently of A and B compartments, transcriptional activity, and Polycomb Group Complex binding (*72*). However, clusters of insulator protein binding sites are associated with pairing (*72*). Together, this evidence suggests that buttons are defined by specialized TADs. Supporting this conclusion, visualization of TADs using super resolution microscopy indicates that homologous TADs are often closely associated (*73*).

The mechanism by which TADs initiate homologous chromosome pairing is still unclear. Transgenes that only cover a portion of a TAD do not drive pairing, suggesting that an entire intact TAD, rather than individual elements contained within a TAD, is required for pairing (*72*). Evidence from haplotype-resolved HiC studies suggests that TAD boundaries are more tightly paired than TAD interiors (*62, 63*). Furthermore, insulator proteins, which have been linked to homologous chromosome pairing, are enriched at TAD boundaries (*11, 65*).

Therefore, TAD boundaries may drive pairing, after which the interiors of TADs associate with each other (**Fig. 3A**). Alternatively, each TAD may have a unique "code" of insulator proteins bound across the TAD region that allows it to find its homologue within the nucleus and initiate chromosome pairing (**Fig. 3B**). It is also possible that TADs take on unique chromatin conformations to create nuclear microcompartments that enable homologous TAD association and pairing (**Fig. 3C**).

Homologous chromosome pairing varies over developmental time and between cell types

The frequency of chromosome pairing varies widely between developmental stages and between cell types. Notably, only 15-30% of nuclei in the early embryo exhibit homologous chromosome pairing (*57-59, 74*). Pairing increases over developmental time, reaching a maximum level of 90-95% by the third instar larval stage (*57-59, 74*). In addition to its variability across developmental stages, pairing differs widely between cell types. Two copies of a transgene containing the Mcp element pair in >90% of larval eye disc nuclei but only 20-60% of larval wing disc and brain cells, and Mcp transgene pairing is not observed in polytene nuclei or early spermatocytes (*71*). Similarly, transgenes encompassing entire TAD buttons drive pairing with their endogenous loci in the larval eye disc but not in the larval antennal disc (*72*). Intriguingly, pairing between endogenous homologous loci occurs at similar rates in the larval eye and antennal disc (*72*). These studies support a model in which certain tissues,

especially the eye disc, are more amenable to pairing than others. Intact homologous chromosomes can likely drive pairing in both strong and weak pairing cell types because they have a large number of buttons interspersed along their entire lengths that work together to bring the chromosomes into proximity (**Fig. 4A-B**). Transgene studies provide a sensitized system for examining variation in pairing levels between cell types: transgenes only encompass a single button, and therefore may only drive pairing efficiently in strong pairing cell types (**Fig. 4A-B**).

A number of factors may contribute to differences in pairing levels between cell types. It is possible that mitotic state plays a role in pairing efficiency; pairing occurs at a much higher level in post-mitotic larval photoreceptors than in the actively dividing antennal disc and embryo (*57-59, 72, 74*). Additionally, differences in TAD formation between cell types may contribute to cell-type-specific pairing; TAD boundaries or entire TADs might be more strongly defined in tissues where pairing occurs more efficiently. Differences in *trans* factor expression between cell types may also contribute to cell-type-specific pairing. Comparison of RNA-seq data between cell types such as the eye and antennal disc could provide candidate *trans* factors that contribute to the stronger pairing observed in the eye disc. The advent of single-cell HiC and single-cell RNA-seq offer the opportunity for a more precise comparison of the cell-to-cell differences that contribute to variation in pairing levels.

1.5: Homologous chromosome interactions facilitate interchromosomal gene regulation

In addition to physical colocalization, interactions between chromosomes frequently involve regulatory crosstalk between different alleles of the same gene. Often, chromosome pairing is a prerequisite for this interchromosomal gene regulation. Interactions between different alleles of the same gene can be divided into two main categories, heritable and non-heritable, as discussed below.

Paramutation involves heritable interactions between chromosomes

One major category of interchromosomal gene regulation involves the phenomenon of paramutation, in which interactions between different alleles of the same gene lead to heritable changes in gene expression. Paramutation was first described in maize by Alexander Brink, and has since been identified in a number of other systems, including mice, worms, and potentially humans (*75*)(reviewed in (*76*)). The phenomenon is defined by three main features: 1) an allele's new expression state is transmitted to future generations even though the allele that originally transmitted the change in expression state is not passed on; 2) the inherited altered allele is subsequently able to transmit expression changes to homologous sequences; and 3) there are no sequence changes in the altered allele, indicating that the change in expression is regulated epigenetically (*77*). The prevailing model in the field suggests that paramutation is mediated by small RNAs, which are transcribed by one allele to direct the

epigenetic modification of another allele, typically through DNA methylation (*76*). This DNA methylation is then passed on to the next generation and is to new alleles through the same mechanism (*76*).

A well-studied example of paramutation occurs in maize at the *b1* locus, which is involved in the production of purple anthocyanin pigments. Through paramutation, the B' allele of the *b1* locus, which is weakly transcribed, converts the B-I allele, which is more strongly transcribed, to a B' state, leading to a lighter pigmentation (**Fig. 5A**)(*78*). The newly converted allele can transmit the B' state to subsequent generations (*78*). *b1* expression is controlled by a long-distance enhancer consisting of seven tandem repeats of an ~850 bp sequence (*79-81*). As in many cases of maize paramutation, transcription of siRNAs from tandem repeats is required for the B' allele to silence the B-1 allele (*79, 82, 83*).

Paramutation has also been described in mammals. In mice, animals that are heterozygous for the Kit^{tm1Alf} allele of the *Kit* gene have white tails and feet and display reduced *Kit* mRNA levels (*84*). When $Kit^{tm1Alf/+}$ mice are crossed with $Kit^{+/+}$ mice, many of their genetically wild type progeny continue to display the Kit^{tm1Alf} phenotype and can transmit the phenotype to subsequent generations (**Fig. 5B**) (*84*). Injection of wild type one-cell embryos with RNAs isolated from $Kit^{tm1Alf/+}$ sperm and brain heritably decreases *Kit* expression, indicating that *Kit* paramutation involves siRNA transmission in the germline (*84*).

Large-cohort diabetes studies have raised the possibility that paramutation also occurs in humans. Class I alleles of the *insulin* (*INS*) gene have 25-63 5' tandem repeats and cause a predisposition to type I diabetes, while Class III *INS*

alleles, which have 140-200 5' tandem repeats, do not cause a diabetes predisposition (*85*). When a father is heterozygous for the Class I *INS-814* allele and a Class III allele, offspring who inherit the *INS-814* allele do not have an increased risk for diabetes (**Fig. 5C**)(*85*). This work suggests that paramutation between Class III and Class I alleles alters the heritable traits of Class I alleles. This paramutation may be mediated by tandem repeats, as is observed in maize (*76, 85*). Thus, paramutation is a conserved mechanism of interchromosomal gene regulation.

Non-heritable interactions between chromosomes through trans-activation and trans-repression

Another category of interchromosomal gene regulation involves non-heritable interactions mediated by the phenomenon of transvection. Via transvection, DNA elements on one copy of a gene act between chromosomes on another copy of a gene to increase (trans-activation) or decrease (trans-repression) gene expression (**Fig. 6B**)(*55*). Unlike paramutation, transvection-related changes in gene expression are not passed on to subsequent generations; once two interacting alleles are segregated away from each other, their expression levels return to their original states. Additionally, rather than an siRNA-based mechanism, transvection typically involves physical interactions between alleles on different chromosomes, mediated by homologous pairing (**Fig. 6A-B**)(*55*).

Transvection has been most thoroughly studied in flies (see below), but examples of tran-activation and trans-repression have also been described in other organisms. In the fungus *Neurospora crassa*, trans-activation controls the expression of the gene *Asm-1*$^{+}$, which is involved in female structure formation and ascospore maturation (*86*). Fungi that are hemizygous for *Asm-1*$^{+}$ have no gene function, and inserting an additional copy of the gene at an ectopic location does not rescue gene expression (*86*). However, when *Asm-1*$^{+}$ hemizygotes have two additional copies of *Asm-1*$^{+}$ at the same ectopic location (one on each copy of the chromosome), *Asm-1*$^{+}$ expression is rescued, suggesting that two copies of *Asm-1*$^{+}$ must be in the same genomic location and homologously pair to activate *Asm-1*$^{+}$ expression (*86*). Trans-activation has also been observed in human lymphoma and myeloma cell lines (*87*). In these cell lines, translocations of the IgH locus cause the IgH enhancer to incorrectly interact with the *CCND1* gene, leading to pairing of *CCND1* and changes in its DNA methylation state (*87*).

In *Antirrhinum majus* (snapdragon), trans-repression occurs between different alleles of the *nivea* gene. The mutant allele *niv*-525, which contains a duplicated promoter region, acts between chromosomes on the wild type *nivea* locus to decrease expression four- to five-fold (*88, 89*). This repression is likely pairing-dependent, and may be due to the extra copy of the promoter present on the *niv*-525 locus titrating away transcription factors from the wild type allele (*89*). Together, these studies indicate that pairing-dependent communication between alleles occurs across organisms to regulate gene expression.

Transvection in Drosophila: a paradigm for interchromosomal gene regulation

Transvection was first discovered by Ed Lewis in *Drosophila* (*55*), and has since been described for a large number of fly genes (*90*). Studies of transvection in fruit flies have provided an in-depth understanding of the DNA elements involved in this phenomenon. These studies have provided the following general "rules" for transvection, which apply to the majority, though not all, of transvecting genes.

Rule 1: Transvection involves two mutant copies of a gene: In the majority of cases, neither copy of a transvecting gene can produce functional protein on its own (**Fig. 4B**)(*55, 90*). DNA elements on one mutant allele act between chromosomes on the other mutant allele to restore normal gene expression (**Fig. 4B**).

Rule 2: Transvection is more efficient without a promoter in *cis*: Transvection often involves the enhancer of one mutant allele activating the promtoer of the other mutant allele (*90*)(**Fig. 6B**). In many cases, enhancer action between chromosomes is more efficient in the absence of a promoter in *cis (90)*. This improved enhancer efficiency has been observed at the *Abd-b, yellow, apterous,* and *Ubx* loci (*91-95*).

Rule 3: Chromosome pairing is required for transvection: In nearly all described cases, chromosome rearrangements that disrupt homologous chromosome

pairing also ablate transvection (*55, 90*). Therefore, intact homologous chromosomes that are properly aligned along their entire lengths are typically necessary for transvection. One likely model to explain this requirement is that homologous pairing brings two alleles of a gene close enough together within the nucleus that their regulatory elements can act between chromosomes (**Fig. 6A-B**).

The transvection "rule-breakers"

Certain cases of transvection do not follow the rules described above. These "rule-breakers" provide important insight into the mechanisms controlling nuclear organization, homologous chromosome pairing, and interchromosomal gene regulation.

Unlike the large majority of transvection-competent genes, transvection at the *Abd-b* and *ss* loci is not disrupted by chromosome rearrangements (*59, 72, 91, 96, 97*). Interestingly, both loci overlap with button TADs (*72*). This evidence suggests a general model in which buttons hold homologous chromosomes together at specific sites to facilitate transvection (**Fig. 6C**). In most cases, chromosome rearrangements would disrupt transvection by breaking buttons in half or relocating buttons far away from transvecting loci, preventing those loci from pairing and interacting (**Fig. 6D-E**). Because *Abd-b* and *ss* encompass buttons, only rearrangements with breakpoints within the genes themselves would be predicted to disrupt their pairing and transvection (**Fig. 6E**). In the case of *ss*, two additional buttons lie directly downstream of the locus and can mediate

pairing and transvection even when *ss* itself is split between two chromosomes (**Fig. 6E**)(*72, 98*). Therefore, the presence of buttons maintains nuclear organization and keeps gene expression robust in the presence of drastic chromosomal aberrations.

Studies of transvection at the *ss* locus also suggest an important feature of the relationship between homologous pairing and transvection. While copies of *ss* must be in close physical proximity to perform transvection, pairing alone is not sufficient for interchromosomal gene regulation (*72, 98*). An intact insulator element is required on both *ss* copies for trans-activation and trans-repression between alleles (*72, 98*). This work suggests that pairing is necessary but not sufficient for transvection.

The *ss* locus breaks the typical rules of transvection in additional ways. Unlike the enhancers of many other transvection-competent loci, the *ss* enhancers cannot perform transvection in the absence of a promoter in *cis (98)*. These data suggest that there may be two classes of transvection-competent genes: one in which transvection is **more** efficient in the absence of a promoter in *cis*, and one in which transvection is **less** efficient in the absence of a promoter in *cis*.

Furthermore, transvection at *ss* occurs between wild-type alleles. The *ss* gene is typically expressed in 67% of R7 photoreceptors in the fly retina to drive expression of UV-detecting photopigments (*99*). However, *ss* expression frequency varies widely across wild-derived *Drosophila* alleles (*100, 101*). When an allele with an expression frequency above 67% is crossed with an allele with

an expression frequency below 67%, the resulting progeny display an intermediate expression frequency, in between that of either parent (*98*). Thus, communication between *ss* copies to control proper photoreceptor patterning provides one of the first examples of transvection in a wild type context (*98*).

Trans-activation versus trans-repression

Transvection can be divided into two phenomena: trans-activation, in which alleles interact to increase gene expression, and trans-repression, in which alleles interact to decrease gene expression. Trans-activation typically involves an enhancer acting between chromosomes on a promoter, as described above (*90-95*). Insulator elements have also been implicated in trans-activation. At the *ss* locus, an insulator on both copies of the gene is required to activate gene expression between chromosomes (*72*). Similarly, the Homie and Nhomie insulators of the gene *eve* allow trans-activation of Homie- and Nhomie-containing transgenes by the *eve* enhancer (*102, 103*), and an insulator is required for trans-activation between transgenes containing the *snail* enhancer and the *even skipped* promoter (*104*).

Trans-repression has also been described for a number of loci, and includes the phenomena of *brown* trans-inactivation, the *zeste-white* interaction, and pairing-sensitive silencing. *Brown* (*bw*) trans-inactivation involves the *brown-dominant* (bw^D) allele, in which a chromosome rearrangement brings centromeric heterochromatin into close proximity with the *bw* locus, leading to position-effect variegation and silencing of *bw* (**Fig. 7A**)(*105*). When crossed with a wild-type *bw*

allele, bw^D asserts a dominant effect, leading to heterochromatinization and silencing of the wild-type allele in *trans* (**Fig. 7A**)(*105*). Certain chromosome rearrangements disrupt this process, suggesting that *brown* trans-inactivation is pairing-dependent (*105*).

In the *zeste-white* interaction, the neomorphic z^1 allele of the gene *zeste* (*z*) represses expression of the *white* (*w*) locus (*106*). *w* repression by z^1 only occurs in flies in which two copies of *w* are present at the same genomic location, indicating that the *zeste-white* interaction is pairing-dependent (**Fig. 7B**)(*106*). The z^1 mutant protein self-associates to form aggregates with a higher DNA affinity, and z^1 overexpression can repress *w* that is present in a single copy, suggesting that pairing of *w* alleles is required to concentrate a high enough level of z^1 protein in the vicinity to overcome a minimum threshold for repression (*107*).

Pairing-sensitive silencing (PSS) (reviewed in (*108*)) involves interactions between Polycomb response elements (PREs), which bind to the repressive Polycomb Group Complex. PSS is typically identified by its effects on a *white* (*w*) reporter transgene. If a sequence of interest does not contain a PRE, then a transgene containing the sequence and the *w* reporter will cause darker eye pigmentation in flies homozygous for the transgene than in flies heterozygous for the transgene due to dosage of the *w* gene (**Fig. 7C**). If a sequence of interest does contain a PRE, then flies homozygous for the transgene will have lighter eye pigmentation than flies heterozygous for the transgene, as the two copies of the PRE will interact with each other between chromosomes to repress *w* expression (**Fig. 7C**). PSS has been described for many PREs across the

genome, including elements in the *engrailed*, *escargot*, *polyhomeotic*, *even skipped, sex combs reduced, proboscipedia*, *Abd-a,* and *Abd-b* loci (*67, 109-115*).

ss is one of the only a few for which both trans-activation and trans-repression phenomena have been described. Two PREs within an upstream silencer can act between chromosomes to repress other copies of *ss*, even in the presence of chromosome rearrangements with breakpoints that lie within the *ss* locus (*98*). Like *ss* trans-activation, *ss* trans-repression requires an intact insulator on each allele (*98*).

In summary, transvection can be separated into activating and repressing functions. While these functions are well-understood in *Drosophila*, further investigation is required to determine whether similar DNA elements are involved in cases of interchromosomal gene regulation in other organisms.

1.6: Conclusions

The study of homologous chromosome pairing and transvection provides insight into the general mechanisms that drive nuclear organization across organisms. Major outstanding questions in the field involve the biological purpose of TADs and the structures driving interchromosomal interactions. Examination of the features that bring homologous chromosomes together suggests a novel role for TADs in driving interactions between chromosomes, in addition to their hypothesized function in isolating transcriptional units. In flies, these interactions between TADs allow homologous sequences to interact and cross-regulate gene

expression. In mammals, TAD interactions may account for the interchromosomal contacts observed in processes such as imprinting and X-inactivation. It is also possible that mammalian TADs have been adapted to interact with heterologous sequences, potentially assisting in the formation of segregated chromosome territories. Thus, TADs may serve a structural role in organizing chromatin in addition to their role in controlling gene expression.

The study of transvection provides insight into the biological function of interchromosomal contacts. In *Drosophila*, these contacts maintain gene expression in the presence of mutations, often through enhancer action between chromosomes. Instances of enhancer action between chromosomes also exist in mammals, such as in the mouse olfactory system, where multiple enhancers on different chromosomes act in tandem to regulate a single olfactory receptor gene (*33*). Examples from mammals suggest that interchromosomal contacts have evolved from more simple instances of a single enhancer acting in *trans* on a specific gene to complex cases in which multiple enhancers act together between chromosomes to regulate of multiple genes simultaneously. The ability of enhancers to act between chromosomes suggests the intriguing hypothesis that spatially separated chromosome territories evolved to prevent improper interchromosomal action of DNA regulatory elements.

Even as we begin to elucidate the mechanisms controlling pairing and transvection, many questions remain. The biological role of homologous chromosome pairing remains unclear. We propose that button-driven pairing allows the genome to remain robust to chromosome rearrangements by

“reconstituting” genes that are split by rearrangement breakpoints (**Fig. 6C-D**)(*98*). Additionally, by facilitating transvection, pairing allows the rescue of gene expression from mutant alleles. However, if pairing serves such a beneficial biological role, it is difficult to explain why this phenomenon does not occur more broadly across organisms. It has previously been proposed that chromosome pairing represents a balancing act between pairing and anti-pairing factors, with the scales tipped towards pairing in *Drosophila* and towards anti-pairing in mammals (*41*). Improper homologous pairing in mammals is associated with deleterious effects including loss of heterozygosity, imprinting defects, and the development of cancer (*22, 23, 26, 41, 52*). Therefore, one hypothesis in the field is that mammals favor anti-pairing to avoid the disadvantages of homologous chromosome interactions, while *Drosophila* take advantage of the positive consequences of pairing and have evolved mechanisms to mitigate its negative effects (*41*).

Another outstanding question involves the role of transvection in the wild. While transvection between mutant alleles has been described at many *Drosophila* loci, very few studies have focused on the role of transvection between wild-type alleles. In many cases, enhancers act more efficiently between chromosomes in the absence of a promoter in *cis*, suggesting that some instances of transvection may only occur in mutant backgrounds. However, the recent discovery that the enhancer of *ss* only performs transvection in the presence of a promoter in *cis* (*98*) indicates that certain transvection-competent genes can perform transvection without DNA regulatory element deletions. The

action of enhancers, silencers, and other regulatory elements between wild-type loci may allow the fine-tuning of expression levels between wild-type alleles, preventing excessive or insufficient levels of transcription.

How widespread is transvection? The phenomenon has been described for approximately 15-20 *Drosophila* genes (*55, 56, 70, 91, 93, 96, 105, 106, 116-139*), but transgene studies suggest that the ability to act between chromosomes may be a general feature of *Drosophila* enhancers (*140, 141*). Certain genes may have a greater capacity for transvection due to their proximity to buttons; genes located at a significant distance from buttons may have weaker pairing, preventing communication between chromosomes. Outside of flies, transvection has only been detected at a handful of loci. The application of new technologies such as haplotype-resolved HiC (*62, 63*) to mammalian systems will facilitate the identification of new candidate loci whose close colocalization might allow transvection.

Our understanding of homologous chromosome pairing and nuclear architecture has grown enormously in recent years with the assistance of assays including next-generation sequencing, HiC, and high-throughput FISH technologies. Investigating the links between chromosome organization, gene expression, and disease should ultimately provide a comprehensive map of the localization of each gene in the genome over developmental time, along with functional information about how each gene's localization affects its expression. An understanding of the detailed organization of the genome across development will provide a means of disease diagnosis through detection of

changes in chromosome architecture, with the ultimate goal of manipulating nuclear organization and therefore gene expression to treat disease phenotypes.

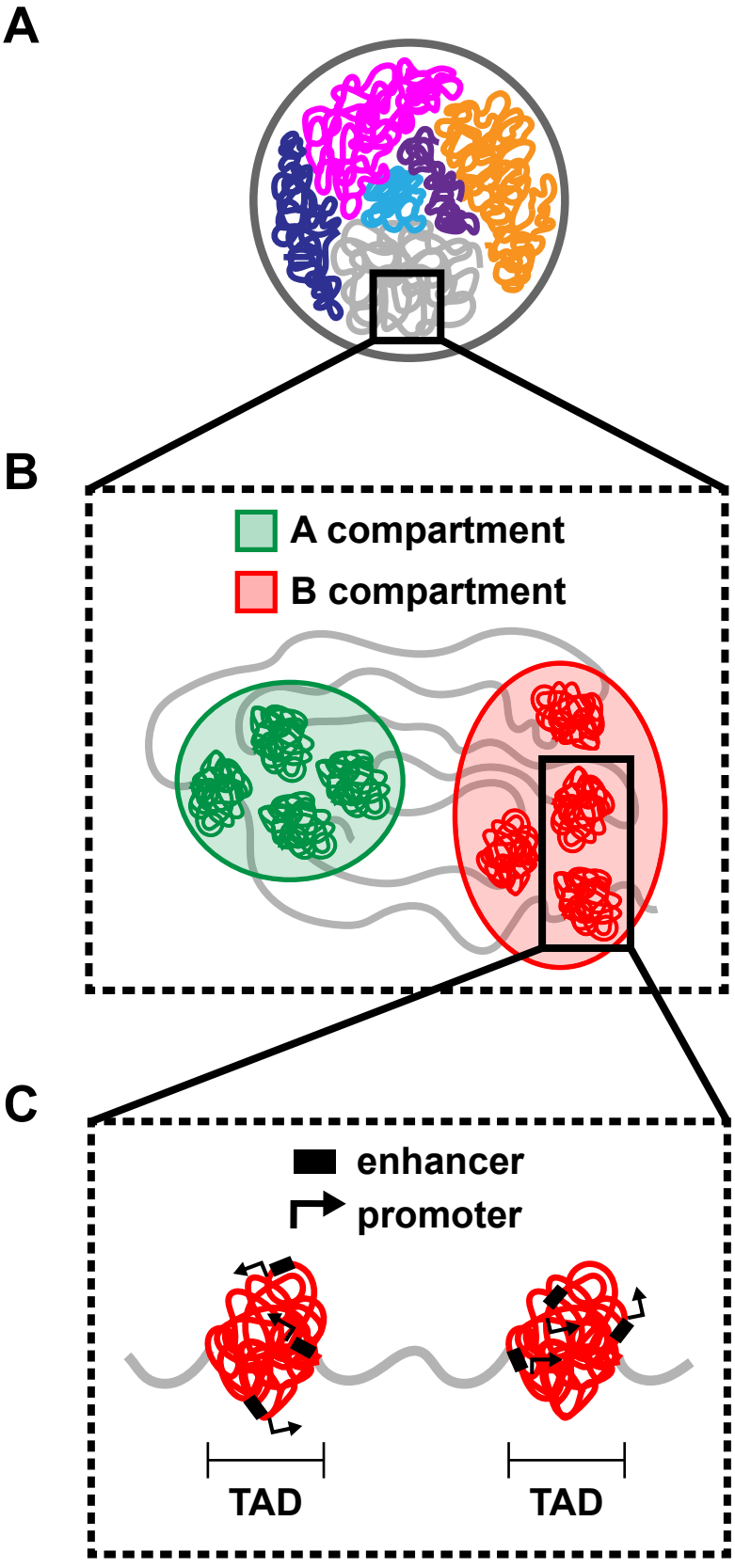


Figure 1

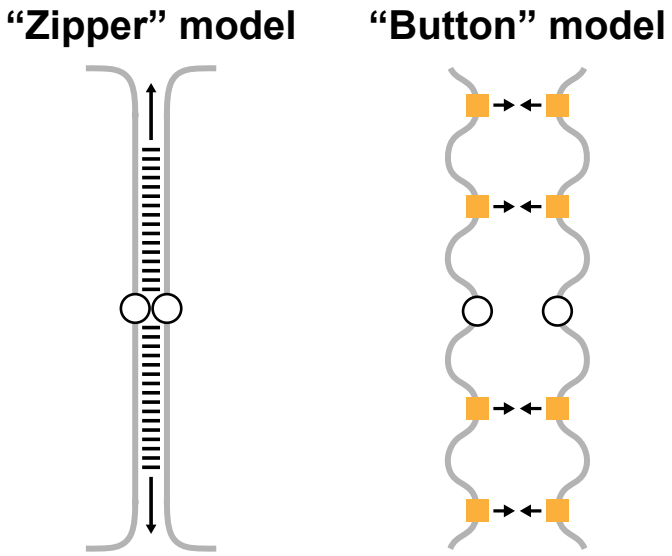
Figure 1: Chromosome territories, compartments, and TADs organize the nucleus.

A. Chromosomes segregate into individual chromosome territories. Larger chromosomes and gene-poor chromosomes tend to localize near the nuclear periphery, while smaller chromosomes and gene-rich chromosomes tend to localize to the nuclear interior.

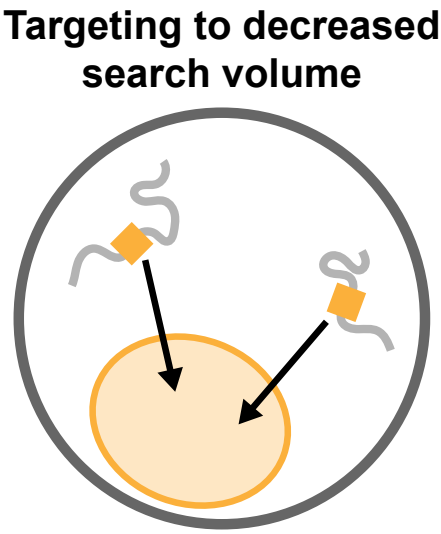
B. Regions along chromosomes interact to form compartments, which are segregated by gene expression state. The A compartment corresponds to active genes, and the B compartment corresponds to inactive genes.

C. Chromosomes are further organized into topologically associated domains (TADs), regions of close self-association that are hypothesized to isolate genes from activation by incorrect DNA regulatory elements.

A


“Zipper” model
“Button” model

B


Targeting to decreased search volume

C

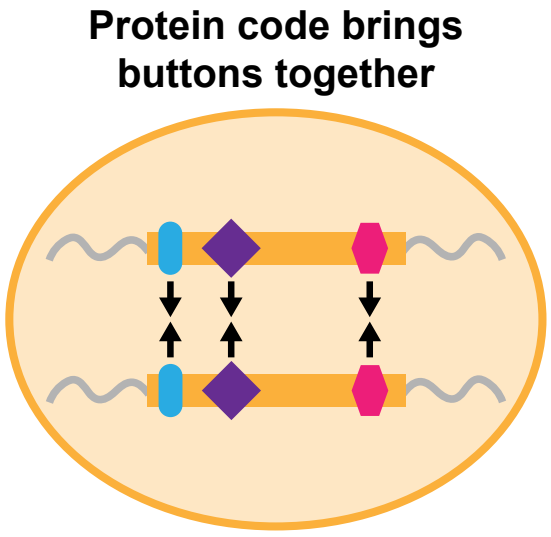

Protein code brings buttons together

Figure 2

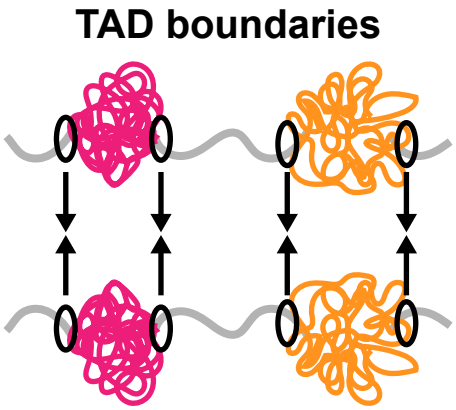
Figure 2: Buttons drive homologous chromosome pairing.

A. Two main models have been proposed to describe the initiation of homologous chromosome pairing. The zipper model proposes that chromosomes pair evenly along their entire lengths, with synapsis beginning at the centromere and proceeding to the telomeres. The more recent button model, which is now favored by the field, hypothesizes that button loci interspersed across chromosome arms drive pairing at a higher affinity than their surrounding regions. Orange boxes: buttons.

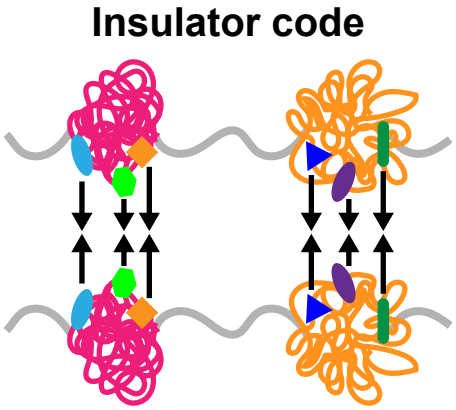
B. To facilitate pairing, each allele of a gene may be targeted to the same nuclear subcompartment based on transcriptional status, providing a button with a decreased search volume in which to find its homologue. Orange boxes: buttons, orange oval: nuclear subcompartment.

C. After targeting to a nuclear subcompartment, the binding of unique proteins in a specific order and at specific intervals may provide a button with a unique "code," allowing it to precisely pair with its homologous partner. Orange rectangle: button, orange oval: nuclear subcompartment, blue ovals, purple diamonds, and pink hexagons: unique protein code.

A



B



C

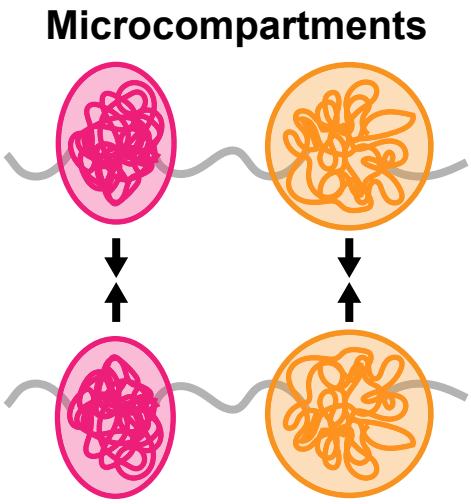


Figure 3

Figure 3: Models for TAD-driven pairing.

A. The boundaries of TADs drive pairing, after which TAD interiors associate with each other. Black ovals indicate the locations of TAD boundaries.

B. The binding of unique combinations of insulator proteins at a precise spacing within each TAD region provides an “insulator code” that allows a TAD to pair with its homologue. Blue ovals and triangles, green hexagons and rounded rectangles, purple ovals, and orange diamonds: insulators.

C. The unique chromosome conformation of each TAD allows it to create a microcompartment within the nucleus and associate with its unique homologue to initiate pairing.

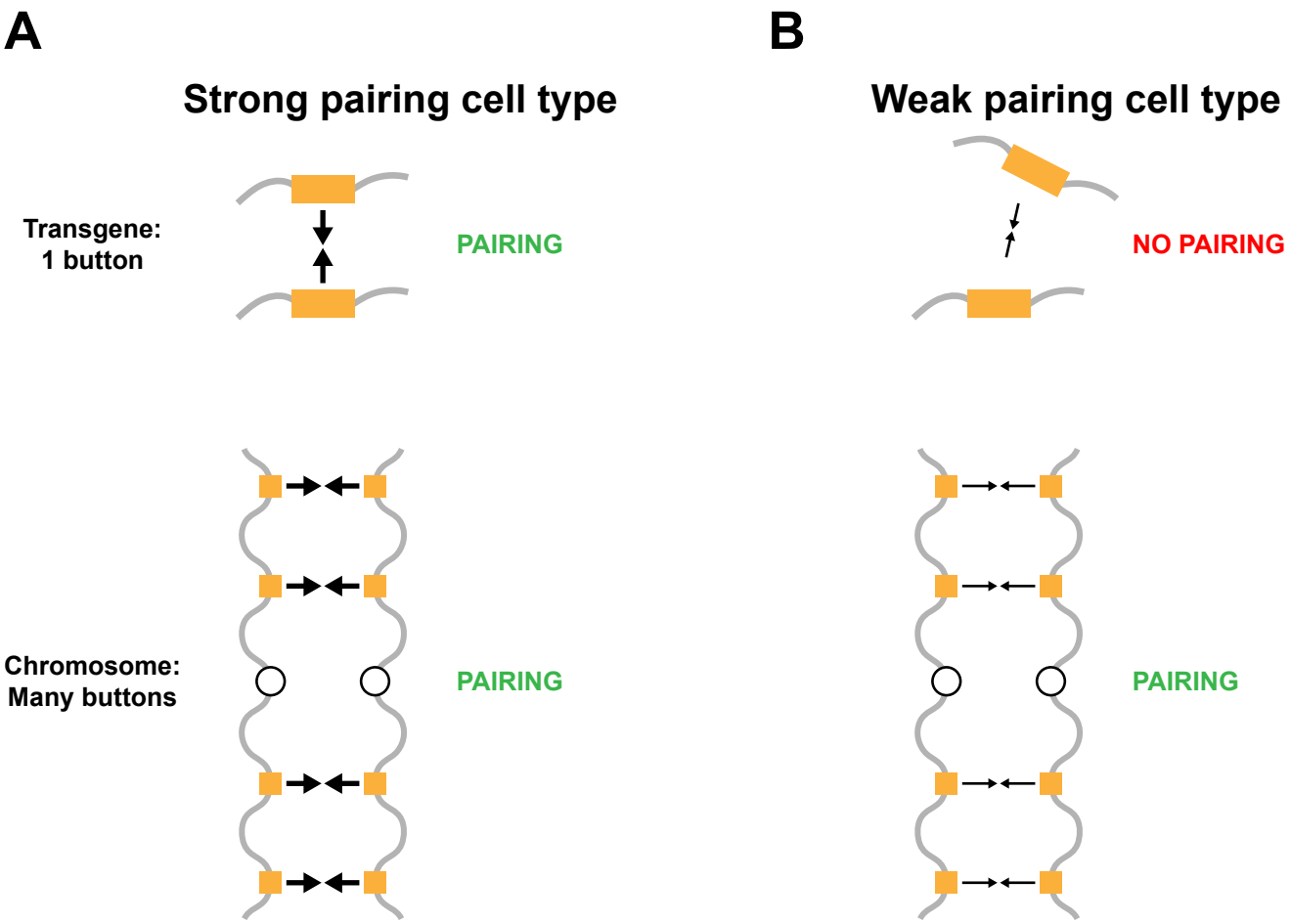

A
Strong pairing cell type
Transgene:
1 button
PAIRING
Chromosome:
Many buttons
PAIRING
B
Weak pairing cell type
NO PAIRING
PAIRING

Figure 4

Figure 4: Transgenes provide a sensitized system to compare pairing between cell types.

A. In a strong pairing cell type, which provides an environment conducive to pairing, both transgenes, which encompass a single button, and whole chromosomes, which encompass many buttons, are able to pair. Orange boxes: buttons, large black arrows: strong pairing.

B. In a weak pairing cell type, which does not provide an environment conducive to pairing, transgenes that encompass a single button do not have the pairing strength required to find their homologous sequence. The strength of a large number of buttons combined over the length of a whole chromosome allows intact chromosomes to successfully pair in this cell type. Orange boxes: buttons, small black arrows: weak pairing.

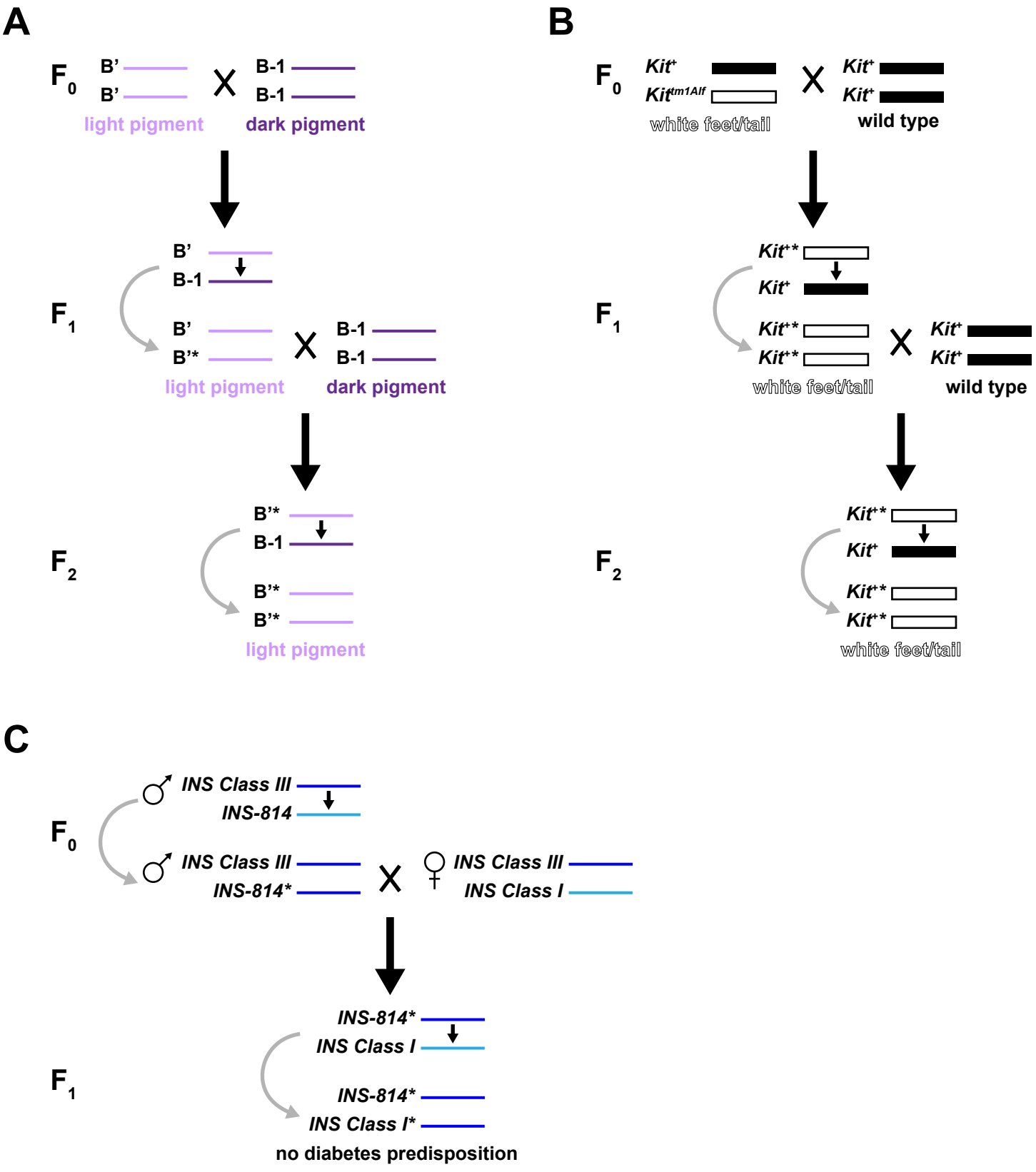

A
F0
B'
B'
B-1
B-1
light pigment
dark pigment
F1
B'
B-1
B'
B'*
B-1
B-1
light pigment
dark pigment
F2
B'*
B-1
B'*
B'*
light pigment
B
F0
Kit+
Kittm1Alf
Kit+
Kit+
white feet/tail
wild type
F1
Kit+*
Kit+
Kit+*
Kit+*
Kit+
Kit+
white feet/tail
wild type
F2
Kit+*
Kit+
Kit+*
Kit+*
white feet/tail
C
F0
INS Class III
INS-814
INS Class III
INS-814*
INS Class III
INS Class I
F1
INS-814*
INS Class I
INS-814*
INS Class I*
no diabetes predisposition

Figure 5

Figure 5: Paramutation occurs in maize, mice, and humans.

A. Paramutation at the *b1* locus in maize. * indicates paramutated allele.

B. Paramutation at the *Kit* locus in mice. * indicates paramutated allele.

C. Paramutation at the *insulin* locus in humans. * indicates paramutated allele.

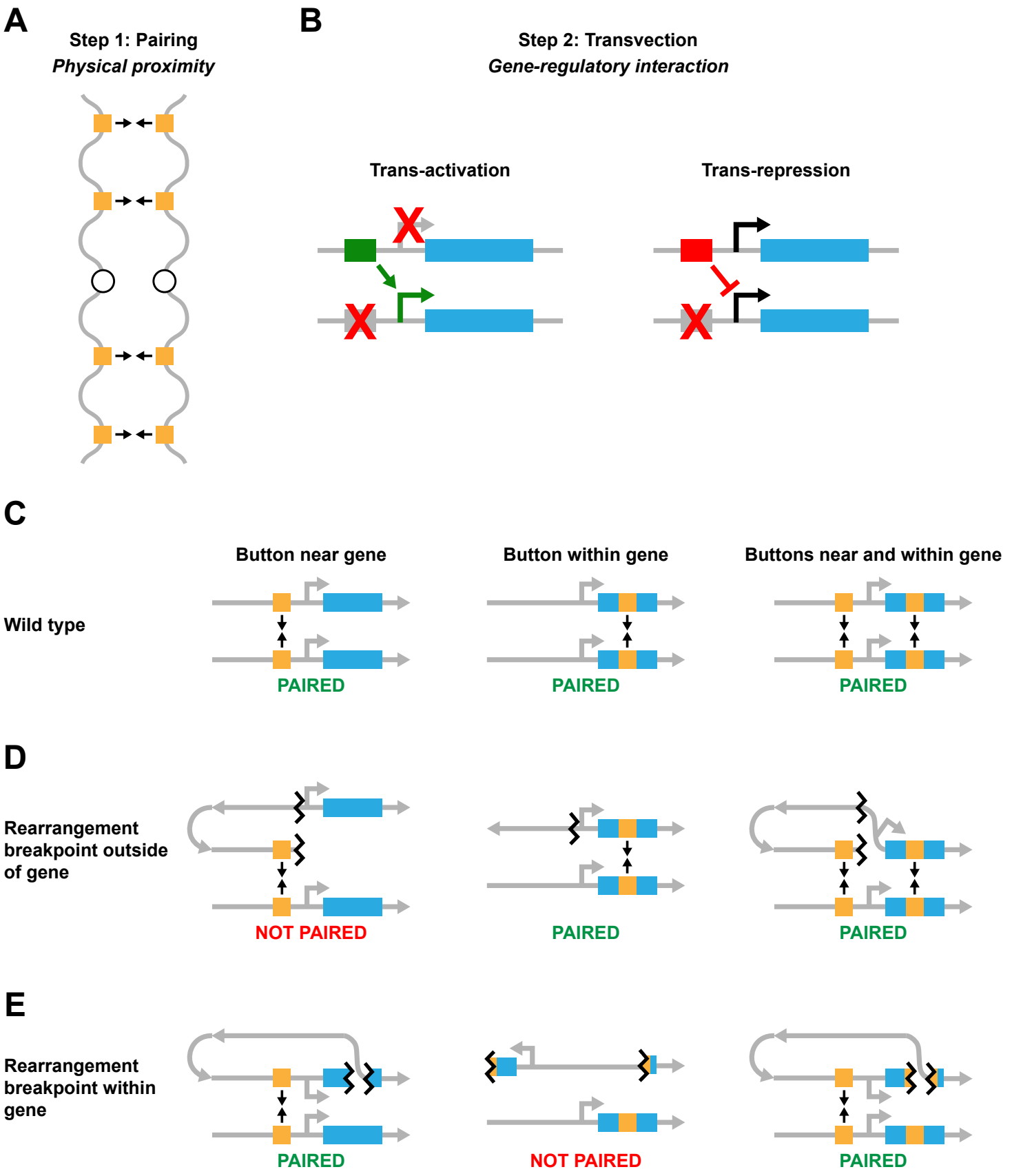

A
Step 1: Pairing
Physical proximity
B
Step 2: Transvection
Gene-regulatory interaction
Trans-activation
Trans-repression
C
Button near gene
Button within gene
Buttons near and within gene
Wild type
PAIRED
PAIRED
PAIRED
D
Rearrangement breakpoint outside of gene
NOT PAIRED
PAIRED
PAIRED
E
Rearrangement breakpoint within gene
PAIRED
NOT PAIRED
PAIRED

Figure 6

Figure 6: Buttons maintain transvection and pairing in the presence of chromosome rearrangements.

A. Pairing brings different alleles of the same gene into close physical proximity, allowing transvection to occur.

B. In the gene-regulatory process of transvection, DNA regulatory elements act between chromosomes to increase (trans-activation) or decrease (trans-repression) gene expression between chromosomes. Under trans-activation, green rectangle: active enhancer, green bent arrow: activated promoter, blue rectangles: protein coding region, gray rectangle with red “X”: mutated enhancer, gray bent arrow with red “X”: mutated promoter. Green arrow between active enhancer and promoter indicates that the enhancer is acting between chromosomes to rescue gene expression. Under trans-repression, red box: functional silencer, black bent arrows: promoter, blue rectangles: protein coding region, gray rectangle with red “X”: mutated silencer. Red flat arrow between functional silencer and promoter indicates that the silencer is acting between chromosomes to repress gene expression.

C-E: Models indicating how buttons in different locations (near a gene, within a gene, or both near and within a gene) maintain pairing in the presence of chromosome rearrangements. Orange boxes: buttons, gray arrows: promoters, blue rectangles: protein coding regions. Black arrows indicate locations where buttons are driving pairing.

C. Potential button pairing states in a wild-type background. Pairing will occur for all three button locations.

D. Potential button pairing states when there is a rearrangement breakpoint outside of a gene. When there is a button near a gene, the button is rearranged to a location far away from the gene and no longer drives pairing between the gene and its homologue. When there is a button within the gene, the rearrangement has no effect on button location and pairing still occurs. When there are buttons near and within a gene, the button near the gene is rearranged to a location far away from the gene and no longer drives pairing between the gene and its homologue, but the button within the gene, whose location is not affected by the rearrangement, ensures that pairing between the gene and its homologue is maintained. Black zig-zag lines: rearrangement breakpoints.
E. Potential button pairing states when there is a rearrangement breakpoint within a gene. When there is a button near a gene, it brings the rearranged portion of the gene back into proximity with the wild-type copy of the gene, maintaining pairing. When there is a button within a gene, the button is split in half by the rearrangement and cannot drive pairing. When there are buttons near and within a gene, the button within the gene is split by the rearrangement and cannot pair, but the button near the gene brings the rearranged portion of the gene back into proximity with the wild-type copy of the gene, maintaining pairing. Black zig-zag lines: rearrangement breakpoints.

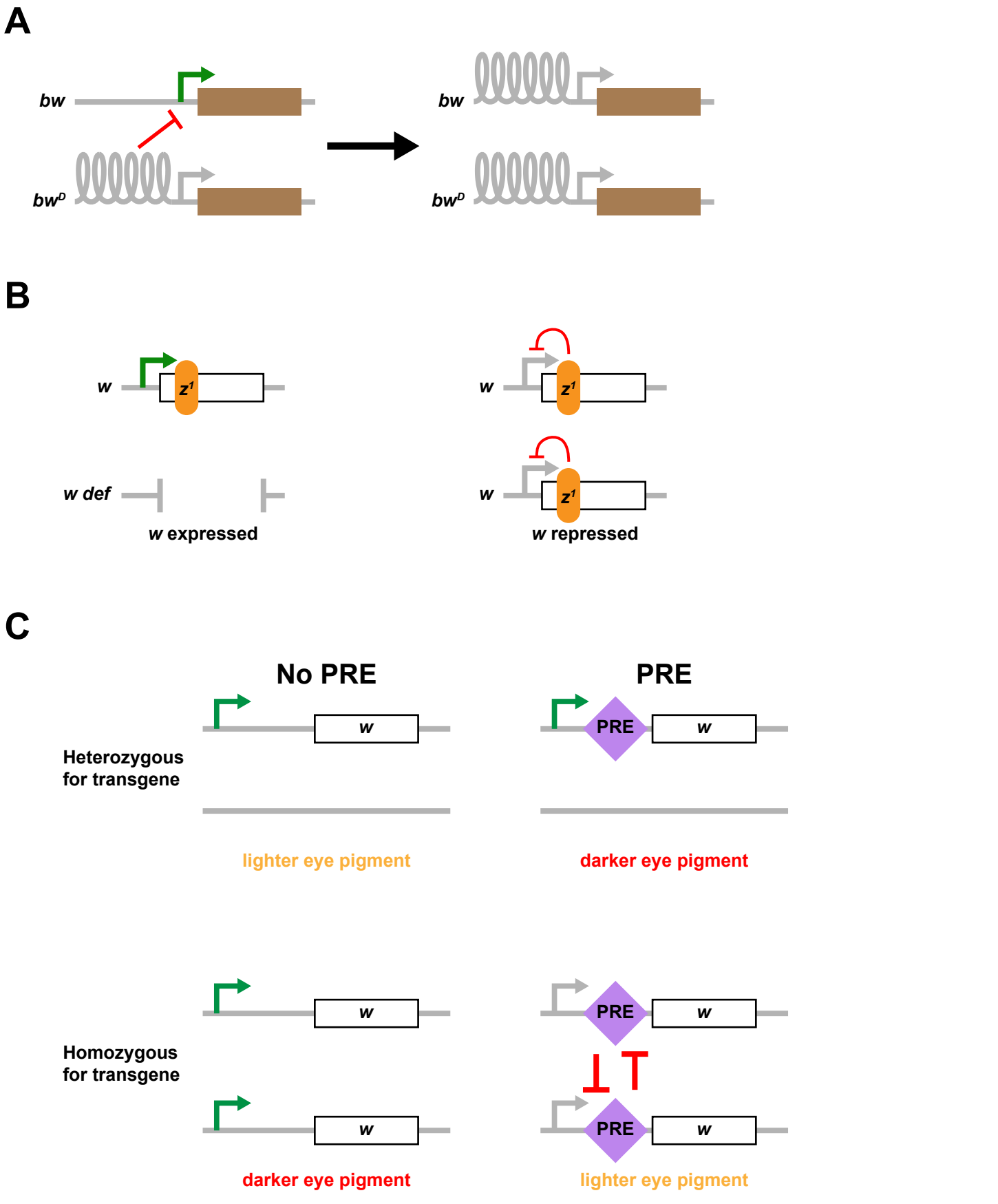

A
bw
bw^D
bw
bw^D
B
w
z1
w def
w expressed
w
z1
w
z1
w repressed
C
No PRE
PRE
Heterozygous for transgene
w
PRE
w
lighter eye pigment
darker eye pigment
Homozygous for transgene
w
w
PRE
w
PRE
w
darker eye pigment
lighter eye pigment

Figure 7

Figure 7: Examples of trans-repression in *Drosophila*.

A. Brown-dominant trans-inactivation. Green arrow: active promoter, gray arrows: inactive promoter, brown rectangles: protein-coding region, loops: repressive heterochromatin. Red flattened arrow indicates that bw^D is acting between chromosomes on wild type *bw* to repress gene expression.

B. The zeste-white interaction. Green arrow: active promoter, gray arrows: inactive promoter, white rectangle: protein coding region, orange oval: neomorphic *zeste* protein. Red flattened arrows indicate that z^1 protein represses gene expression when *w* is homozygous.

C. Pairing-sensitive silencing. Green arrows: active promoter, gray arrows: inactive promoter, white rectangle: *w* protein coding region, purple diamond: PRE. Red flat arrows indicate that PREs are acting between transgenes to repress *w* expression.

Chapter 2: Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates

CellPress

Review

Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates

Kayla Viets,[1,2] Kiara C. Eldred,[1,2] and Robert J. Johnston Jr[1,*]

Across the animal kingdom, visual systems have evolved to be uniquely suited to the environments and behavioral patterns of different species. Visual acuity and color perception depend on the distribution of photoreceptor (PR) subtypes within the retina. Retinal mosaics can be organized into three broad categories: stochastic/regionalized, regionalized, and ordered. We describe here the retinal mosaics of flies, zebrafish, chickens, mice, and humans, and the gene regulatory networks controlling proper PR specification in each. By drawing parallels in eye development between these divergent species, we identify a set of conserved organizing principles and transcriptional networks that govern PR subtype differentiation.

Trends

Retinal mosaics can be organized into three broad categories: stochastic/regionalized, regionalized, and ordered.

Networks of transcription factors with sequence-level and/or functional homologies are utilized to determine PR fates.

Between diverse species, there are amazing similarities in regulatory processes, including waves of differentiation, gradients of signaling molecules, and DNA element looping.

Retinas Are Patterned in Stochastic/Regionalized, Regionalized, and Ordered Mosaics

Evolution has produced highly tuned opsin proteins that enable organisms to detect wavelengths of light specific to their environments. For instance, humans can differentiate colors most precisely in the yellow to red range of the color spectrum, which corresponds to the colors of ripening fruit [1–3], while flies are sensitive to polarized light, which assists in navigation during flight [4–6]. In this review we describe the patterns of PR mosaics and the gene regulatory networks that lead to diverse PR subtype fates across several commonly studied organisms: fruit flies, zebrafish, chickens, mice, and humans. The retinal mosaics of these organisms can be grouped into three classes: stochastic/regionalized, regionalized, and ordered. These species share numerous similarities in retinal development, revealing surprising conservation in the gene regulatory mechanisms and developmental patterns that form diverse visual systems.

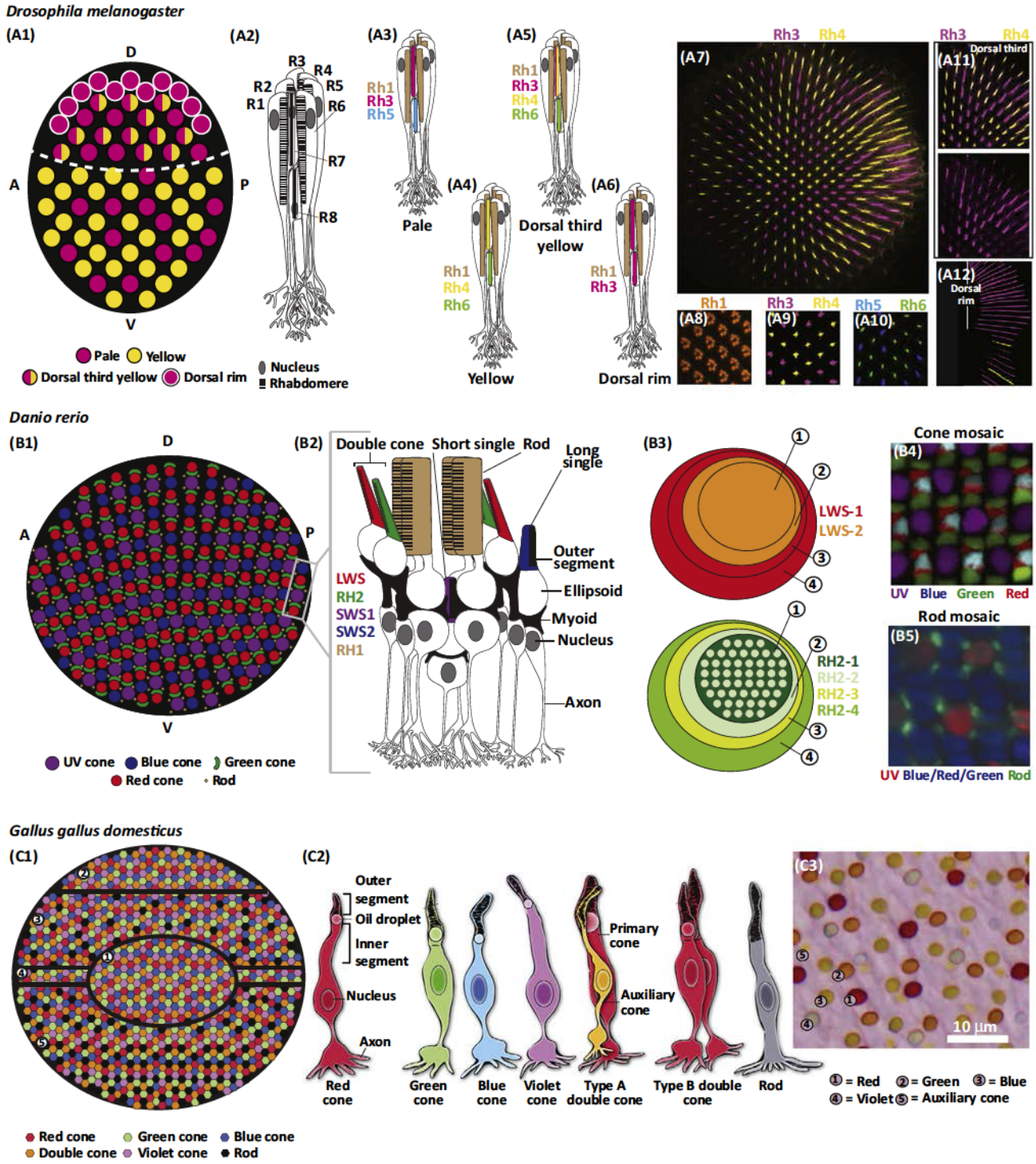
The Stochastic/Regionalized Mosaic of the *Drosophila melanogaster* Retina

The *D. melanogaster* (fruit fly) retina is composed of approximately 800 ommatidia (i.e., unit eyes) each of which contains eight PRs, R1–R8 (Figure 1A2). These PRs can be divided into two groups: the outer PRs, R1–R6, and the inner PRs, R7 and R8. The outer PRs encircle the inner PRs, and the R7 is located above the R8 relative to the apical surface of the retina (Figure 1A2) [7]. A rhabdomere, a series of thousands of microvilli containing a high concentration of photopigment, extends the full length of each PR cell body (Figure 1A2) [8,9].

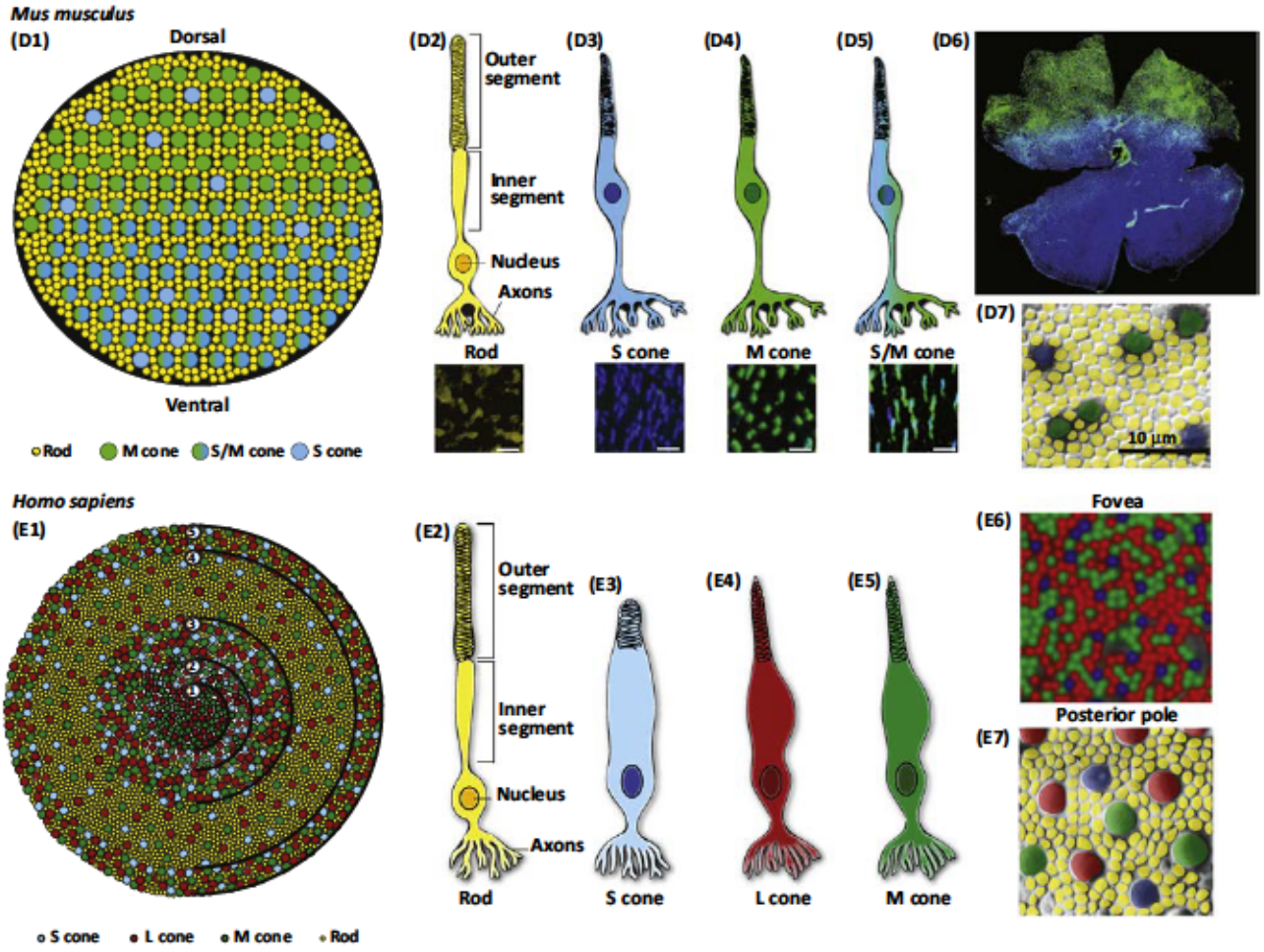
All outer PRs express the motion-detecting photopigment Rhodopsin 1 (Rh1) (Figures 1A3–1A6, 1A8) [10,11]. Expression of different Rhodopsins in the inner PRs defines four subtypes of ommatidia: pale (Figure 1A3), yellow (Figure 1A4), dorsal third yellow (Figure 1A5), and dorsal rim (Figure 1A6) [4,12–20]. In pale ommatidia, pR7s express UV-detecting Rhodopsin 3 (Rh3) and pR8s express blue-detecting Rhodopsin 5 (Rh5) (Figures 1A3, 1A9, 1A10) [12,13]. In yellow ommatidia, yR7s express UV-detecting Rhodopsin 4 (Rh4) and yR8s express green-detecting

[1]Department of Biology, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218-2685, USA
[2]These authors contributed equally to this work.

*Correspondence: robertjohnston@jhu.edu (R.J. Johnston Jr).



(See figure legend on the bottom of the next page.)



Trends in Genetics

Figure 1. Retinas Are Patterned in Stochastic/Regionalized, Regionalized, and Ordered Mosaics. (A1) Schematic of the *Drosophila melanogaster* (fruit fly) PR mosaic (not to scale). Above white dotted line: dorsal third. (A2) Schematic of a *Drosophila* ommatidium. (A3) Schematic of a pale ommatidium. (A4) Schematic of a yellow ommatidium. (A5) Schematic of a dorsal third yellow ommatidium. (A6) Schematic of a dorsal rim ommatidium. (A7) Whole-mount immunostain of a *Drosophila* retina showing the stochastic distribution of ommatidial subtypes. (A8) Immunostain showing Rhodopsin 1 (Rh1) expression in the outer PRs. (A9) Immunostain showing the stochastic patterning of Rh3 and Rh4 in a section of the *Drosophila* retina. (A10) Immunostain showing the stochastic patterning of Rh5 and Rh6 in a section of the *Drosophila* retina. (A11) Immunostain of the dorsal third of the *Drosophila* retina, showing coexpression of Rh3 and Rh4 in dorsal third yR7s. (A12) Immunostain of the dorsal rim of the *Drosophila* retina, showing expression of Rh3 in R7s and R8s. (B1) Schematic of the *Danio rerio* (zebrafish) PR mosaic (not to scale). (B2) Schematic side-view of a single unit of the zebrafish retinal pattern. (B3) Schematic showing the overlapping, regionalized expression patterns of zebrafish LWS and RH2 opsin subtypes (not to scale). 1, inner central/dorsal area; 2, outer central/dorsal area; 3, inner periphery/ventral area; 4, outer periphery/ventral area. (B4) Immunostain of a section of the zebrafish cone mosaic; reproduced, with permission, from [254]. (B5) Immunostain of a section of the zebrafish rod mosaic; reprinted, with permission, from [37] (C1) Schematic of the *Gallus gallus domesticus* (chicken) PR mosaic (not to scale). 1, area centralis; 2, dorsal rod free zone; 3, dorsal rod zone; 4, central meridian; 5, ventral rod rich zone. (C2) The chicken has five different types of cone cells: red, green, blue, violet, and double cones. Type A double cones contain an auxiliary cone lacking an oil droplet. Type B double cones both have oil droplets. Images adapted from [46,255]. (C3) Light microscope image of oil droplets in the chicken retina. Adapted, with permission, from [43]. (D1) Schematic of the *Mus musculus* (mouse) PR mosaic (not to scale). D2–D5, Labeled depiction and immunostaining of mouse PRs. Rods shown in yellow (D2), S-cones in blue (D3), M-cones in green (D4), and S/M-cones in blue/green (D5). (D6) Immunostain of a whole-mount mouse retina. Blue, S-opsin; green, M-opsin. (D7) Pseudocolored differential interference contrast (DIC) section of whole-mount mouse retina, showing cone and rod distribution. Rods are shown in yellow. Blue and green are arbitrarily chosen to represent S- and M-cones, respectively, but each cell could express S-opsin only, M-opsin only, or both S- and M-opsins. Adapted, with permission, from [58]. (E1) Schematic of the *Homo sapiens* (human) PR mosaic (not to scale): 1, foveola; 2, fovea; 3, macula; 4, posterior pole; 5, peripheral rim. (E2–E5) Labeled depiction of human PRs. Rods (E2), S-cone (E3), L-cone (E4), M-cone (E5). (E6) Pseudocolored adaptive optics image of the human fovea. Blue, S-cones; green, M-cones; red, L-cones. Adapted, with permission, from [77]. (E7) Pseudocolored image of human cones in the posterior pole. Blue, S-cones; yellow, rods. Red and green are arbitrarily chosen to represent L- and M-cones, respectively, but each cell could be either red or green. Adapted, with permission, from [67]. Note: in *Danio rerio* and *Mus musculus* the optic disc is located temporal to the central retina, whereas in *Gallus gallus domesticus* and *Homo sapiens* retinas it is located temporal to the foveal center. This area is devoid of photoreceptors and is not represented in the included mosaics. Abbreviations: A, anterior, D, dorsal; P, posterior; V, ventral.

Rhodopsin 6 (Rh6) [12,13] (Figures 1A4, 1A9, 1A10). PRs in the ventral two-thirds of the retina are arranged in a stochastic mosaic: pale and yellow ommatidia in this region are randomly patterned in a ratio of 35:65 [12] (Figures 1A1, 1A7). Specialized ommatidial subtypes occur in the dorsal region of the retina. In the dorsal third of the retina, Rh3 is coexpressed with Rh4 in stochastically distributed yR7s [4] (Figures 1A1, 1A5, 1A11). Dorsal rim ommatidia are found only at the extreme dorsal edge of the retina and express Rh3 in both R7s and R8s (Figures 1A1, 1A6, 1A12) [20].

An Ordered Array of Cones and Rods in the Retina of *Danio rerio*

As in flies, *D. rerio* (zebrafish) PRs contain a ciliated region with a high concentration of photopigment (Figure 1B2) [21,22]. In zebrafish, this region is known as the outer segment and is located at the apical end of the PR (Figure 1B2). Outer segments connect to the ellipsoid, which refracts light onto the outer segment (Figure 1B2) [21,23–26]. The ellipsoid is joined to the myoid region, which contracts to extend and retract PRs in response to changes in light (Figure 1B2) [27–29]. Below the myoid lies the cell soma which contains the nucleus (Figure 1B2) [21,23].

Zebrafish retinas contain four PR classes: rods, which express motion-detecting Rhodopsin (RH1); short single cones, which express UV opsin (Short-Wavelength Sensitive 1; SWS1); long single cones, which express blue opsin (Short-Wavelength Sensitive 2; SWS2); and double cone pairs, in which one cone expresses red opsin (Long-Wavelength Sensitive; LWS) and the other cone expresses green opsin (Rhodopsin-like; RH2) (Figure 1B2) [30–33]. Zebrafish PRs are arranged in a repetitive pattern throughout the retina [21]. Rows of double cones alternate with rows of interdigitated UV and blue cones (Figures 1B1, 1B4). Within double cone rows, each red–green pair is turned 180° with respect to the previous double cone (Figures 1B1, 1B4). Each row of double cones is shifted one half cycle with respect to the previous row such that each UV cone is flanked by two green cones and each blue cone is flanked by two red cones (Figures 1B1, 1B4). Rods are interspersed evenly between the rows of cones, forming a square pattern around UV cones (Figures 1B1, 1B5) [22,33–37].

Within this highly ordered mosaic, regionalized expression of two subtypes of LWS (LWS-1 and LWS-2) and four subtypes of RH2 (RH2-1, RH2-2, RH2-3, and RH2-4) in double cones defines distinct areas of the zebrafish retina. In the inner central/dorsal area, double cones expressing LWS-2 and RH2-1 are interspersed with double cones expressing LWS-2 and RH2-2 (Figure 1B3). The outer central/dorsal area surrounds the inner central/dorsal area, and all double cones in this region express LWS-2 and RH2-2 (Figure 1B3). The next ring of expression, the inner periphery/ventral area, contains double cones expressing LWS-1 and RH2-3 (Figure 1B3). Finally, double cones in the outer periphery/ventral area express LWS-1 and RH2-4 (Figure 1B3) [38,39].

Overlapping Regular Spacing of PR Subtypes Forms a Semi-Random Mosaic in the *Gallus gallus domesticus* Retina

Similarly to zebrafish, the *G. g. domesticus* (chicken) retina contains Rh1-expressing rods, specialized for night vision, and multiple single and double cone types. The four single cone types in the chicken retina are sensitive to red, green, blue, and violet wavelengths of light (expressing LWS, Rh2, SWS2, and SWS1 opsins respectively) (Figure 1C2) [40,41]. These cone types have been identified chiefly by differently colored oil droplets located between the inner and outer segments, which may act as a filter for specific wavelengths of light, as well as focusing photons onto the outer segment [42–45] (Figure 1C2). Two morphologically different sets of double cones in chickens are sensitive to long wavelengths of light [46]. In the more common double cone pair, both cones have an oil droplet [46,47] (Figure 1C2, Type B). In the other pair, only the larger (primary) cone contains an oil droplet (Figure 1C2, Type A) [44,46].

These double cones may be specialized for motion detection rather than for color vision because they appear to contain the same photopigments and synapse onto one another [3,48,49].

Double cones cover about 40% of the chicken retina, with the majority positioned ventrally [43]. Green and red single cones each comprise about 20% of total cone cells. Blue and violet cones make up the remaining 12% and 8%, respectively, and are more abundant dorsally [43]. Each cone in the chicken retina is positioned at a regular distance from other cones of the same subtype (e.g., each red cone is at a specific distance from its neighboring red cone cell) [43]. However, the relative positions of different cone cell subtypes (e.g., red vs green) are not regular. Thus, the final retinal pattern in chickens is semi-random (Figures 1C1, 1C3, 4D) rather than the perfectly ordered pattern seen in zebrafish (Figure 1B1) [43].

Chickens and other birds have an afoveate structure, meaning the most central part of the retina is densely packed with cones and lacks rods [40,50] (area centralis, Figure 1C1). Further from the foveal center, cone packing becomes less dense [43,45,51,52]. In addition to the area centralis, rod numbers are reduced in a lateral stripe through the center of the retina [40] and in the dorsal retina (central meridian and dorsal rod free zone, Figure 1C1). The rod population has a pattern distinct from cones, forming a ventral to dorsal gradient [40] (dorsal rod zone and ventral rod-rich zone, Figure 1C1).

Regionalized Patterning of Cones in the Retina of *Mus musculus*

The *M. musculus* (mouse) retina has fewer PR types than zebrafish and chickens, containing motion-detecting rods that express rhodopsin and three subtypes of color-detecting cones that express S-opsin (UV-detecting), M-opsin (green-detecting), or both S- and M-opsins (Figures 1D2–1D5). These PRs are patterned in a regionalized mosaic, with cones being arranged in opposing dorsal to ventral gradients (Figures 1D1, 1D6). M-opsin is expressed most highly in the dorsal third, and S-opsin is expressed in the ventral two-thirds [53,54] (Figures 1D1, 1D6). In the region in which these opposing gradients meet, single cone cells have varying levels of M- and S-opsin coexpression [54,55] (Figures 1D1, 1D6). A subset of S-opsin-expressing cones appear to be stochastically arranged throughout the retina [56] (Figure 1D1). These cones may be part of a primordial S-cone color system that synapses onto a dedicated population of bipolar cells [57]. Rods are evenly interspersed throughout the retina and vastly outnumber cones, making up about 97% of the PR population [58] (Figures 1D1, 1D7).

Stochastic/Regionalized Patterning of Cones and Rods in the *Homo sapiens* Retina

The human retinal mosaic contains four types of PRs: rods for night vision, and blue (S-opsin), red (L-opsin), and green (M-opsin) cones for color and daytime vision [59–63] (Figures 1E2–1E5). Human retinal patterning is mostly random, with a few areas of organization. Similarly to chickens, the central area of the human retina is densely packed with cones [64] (Figure 1E1). This area can be divided into three regions: the foveola, the fovea, and the macula (Figure 1E1). The foveola contains only L- and M-opsin-expressing cones arranged in a stochastic pattern [65,66] (Figure 1E1). S-cones become integrated into the mosaic outside the foveola within the fovea and the macula (Figures 1E1, 1E6) [67]. It is unclear whether the S-cone mosaic is also random [67] or if it is distributed in a lattice pattern separate from the L/M-cone cell pattern [67–69]. Cones in the foveola and fovea are smaller than those found in the macula and in the posterior pole [70] (Figure 1E1). Rods are integrated into the mosaic starting in the macula region [67] (Figure 1E1). The posterior pole of the retina is rod-dominated, with a random pattern of L-, M-, and S-cones scattered throughout [71] (Figures 1E1, 1E7). One other densely packed cone region exists along the peripheral rim of the retina [72] (Figure 1E1).

L- and M-cones are very similar, and until recently it was almost impossible to distinguish between the two [65,66,71,73–77]. It is widely believed that the only difference between L- and M-cones is the type of opsin expressed. However, evidence from monkeys suggests that the two populations have different numbers of synapses between the cone and the midget bipolar cell [78]. S-cones are easily distinguished by their short, stubby outer segments, while L/M-cones produce long, skinny outer segments [67,68,79,80]. S- and L/M-cones also have distinct patterns of connectivity with other retinal cell types [68].

The Unique Retinal Patterning of Different Organisms Has Evolved To Suit their Environments and Behaviors

Evolution has optimized stochastic/regionalized, regionalized, and ordered retinal patterns to fit the needs of diverse organisms. For example, regionalization of specialized ommatidia within an overall stochastic mosaic provides the fly with the optimal light-detecting abilities to respond to its environment. The dorsal rim ommatidia detect polarized light to allow proper navigation during flight, while the coexpression of Rh3 and Rh4 in dorsal third yR7s may assist in detecting the location of the sun [4–6]. The evolutionary advantage of a stochastic rather than patterned distribution of PRs remains unclear. Random placement of yellow and pale ommatidia that results in similar 65:35 ratios throughout the eye may be the simplest evolutionary mechanism to ensure that all regions of the retina detect multiple wavelengths of light with the same efficiency.

The ordered distribution of zebrafish cones is uniquely suited to its aquatic environment, preventing under- or over-sampling of specific light wavelengths in different areas of the retina [37]. The ability to detect such a broad spectrum of wavelengths may allow the zebrafish to see efficiently when light conditions vary as a result of water turbidity, seasonal changes, or fluctuations in water microorganism and mineral content [23].

The semi-random mosaic of the chicken retina is tuned to perceive many wavelengths of light with high visual acuity. The chicken's cone-rich retina and densely packed area centralis, which also has a greater ganglion cell density [81,82], likely provides high-acuity color vision in daylight to allow identification of prey and predators. Different bird species display different ratios of cone subtypes. For example, sea birds generally have fewer long-wavelength opsin cones compared to blue and green, possibly because long wavelengths are filtered out by water [44,83]. This implies that genetic mechanisms governing cone subtype specification are highly tunable to the environmental niche that an avian species inhabits.

Because the regionalized mouse retina contains two color-detecting opsins that are mostly separated into the dorsal third and ventral retina, mouse vision is believed to be largely monochromatic. Ventral expression of S-opsin and dorsal expression of M-opsin allows the mouse to maximize sampling of UV (sky) and terrestrial light sources with the most appropriate PRs. In the center of the retina, where S- and M-opsin expression converge, differing levels of opsin coexpression between neighboring cells may give the mouse dichromatic vision [84,85].

The random distribution of the three human cone types allows efficient spectral sampling of the visual field and maximizes contrast sensitivity [1–3]. The dense packing of cones in the fovea provides maximal visual acuity in daylight, and the rod-dominated retina outside of the macula allows efficient night vision.

The Gene Regulatory Networks Controlling PR Specification Share Functional and Sequence-Level Homologs

The gene regulatory networks controlling PR specification are extremely complex and in many cases are still being elucidated. We provide here simplified networks to highlight the proteins that play conserved roles in PR fate at either the functional or sequence level, focusing mainly on flies,

zebrafish, and mice, whose gene regulatory networks are better characterized than those of chickens and humans. PR differentiation occurs in four basic decision steps (Figure 2A): (Step 1) PR versus non-PR fate, (Step 2) rod versus cone fate, (Step 3) cone subtype, and (Step 4) opsin subtype.

Step 1: PR Versus Non-PR Fate Choice

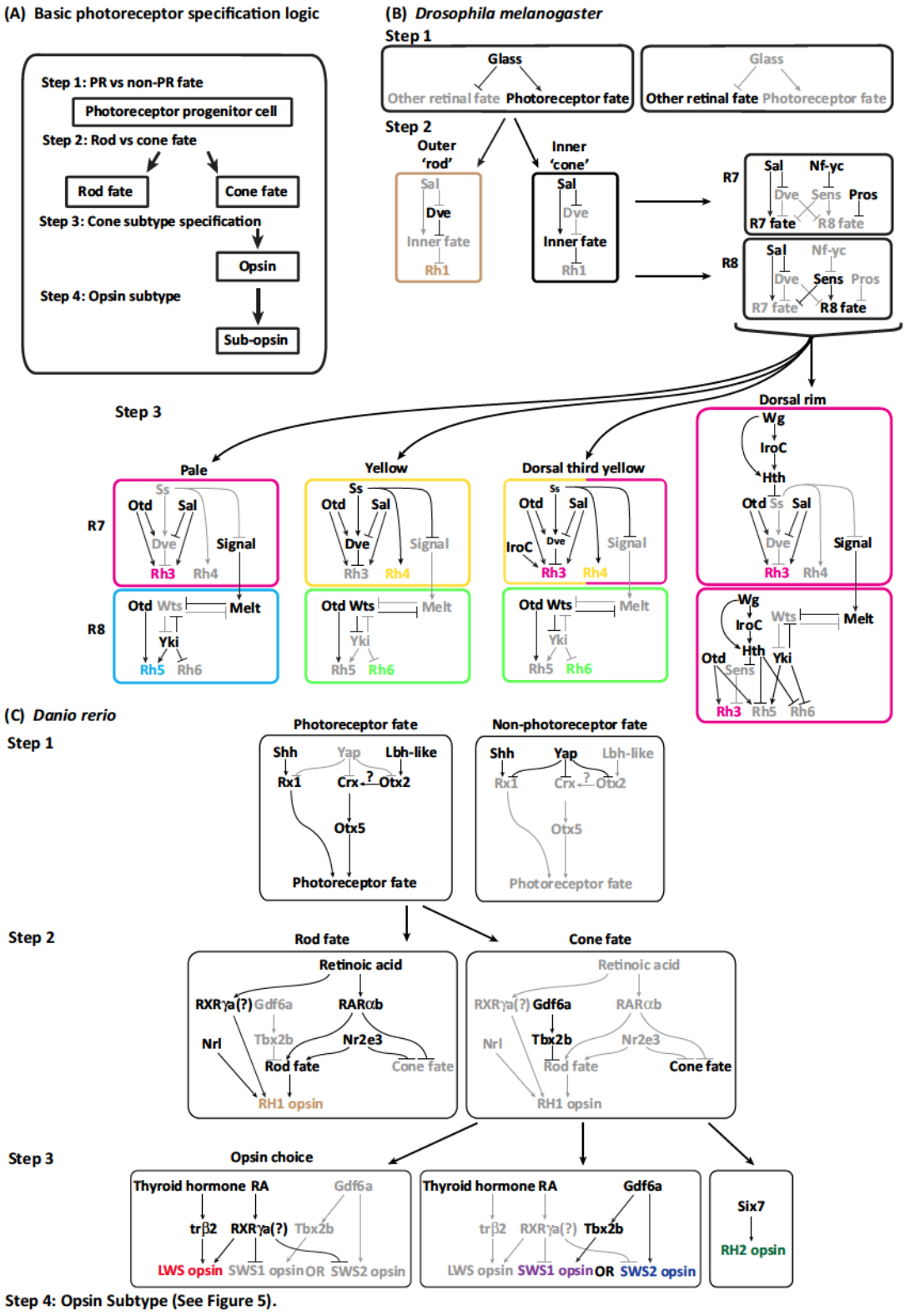
Step 1 of PR specification involves the expression of factors that distinguish differentiating PRs from other cell fates. In flies, the zinc-finger transcription factor Glass plays this role [86] (Figure 2B, Step 1). Vertebrate PR differentiation involves a core set of conserved transcription factors, including Cone-rod homeobox (Crx), the Orthodenticle homeobox proteins (Otx2 and Otx5), and the Retinal homeobox proteins (Rx1, RaxL, Rax) [87–118] (Figures 2C–E, and Figure S1 in the supplemental information online, Step 1). Species-specific inputs have emerged to regulate these conserved factors. In zebrafish, the Hippo pathway transcriptional activator Yes-associated protein (Yap) represses these core transcription factors (Figure 2C, Step 1) while, in mice, the Notch-1 transmembrane receptor plays this role (Figure 2D, Step 1) [119–121]. The core PR factors are also activated by species-specific inputs: in zebrafish, the signaling molecule Sonic hedgehog (Shh) and the transcription cofactor Lbh-like activate Rx1 and Otx2, respectively (Figure 2C, Step 1) [87,122]. The network topology between these conserved factors varies between organisms; in mice, Rax activates Otx2 [92] (Figure 2D, Step 1), while in zebrafish no link between Rx1 and Otx2 has been established (Figure 2C, Step 1) [87,89–92,100]. In both mice and zebrafish, Otx2 likely activates Crx (Figure 2C,D, Step 1) [87,101]. Other regulators complement these core factors: for example, in zebrafish Crx activates the species-specific Otx homolog Otx5 to drive PR fate (Figure 2C, Step 1) [97–99,119].

Step 2: Rod Versus Cone Fate Choice

In Step 2, PR precursors select either rod or cone fate. In *Drosophila*, outer PRs (rods) are specified by the presence of the homeodomain protein Defective Proventriculus (Dve), which represses the expression of color-detecting Rhodopsins (Figure 2B, Step 2) [123]. In zebrafish, mice, and humans, the bZIP transcription factor Neural retina leucine zipper protein (Nrl) and the orphan nuclear receptor Nuclear Receptor Subfamily 2 Group E Member 3 (Nr2e3) play important roles in rod fate (Figure 2C–E, Step 2) [88,98,124–142]. Nrl activates Nr2e3 in mice and may play a similar role in humans and zebrafish (Figure 2C–E, Step 2) [143,144]. In zebrafish and possibly chickens, retinoic acid (RA) signaling is also involved in rod development (Figures 2C, S1, Step 2); in zebrafish, RA signals through the RARαb receptor and possibly the RXRγa receptor to specify rods (Figure 2C, Step 2) [145,146]. In addition, the growth factor glial cell line-derived neurotrophic factor (GDNF) is expressed specifically in rods in both chickens and mice, and may also play a role in zebrafish (Figures 2D, S1, Step 2) [147–151]. Two non-conserved factors, the SUMO-E3 ligase/transcription factor Pias3 and the orphan nuclear receptor RORβ, are also involved in rod fate in mice (Figure 2D, Step 2) [132,152–154]. Recent evolutionary studies suggest that mammalian S-cone and rod PRs may have similar lineages, and may temporally switch from S-cone precursors to rods [155].

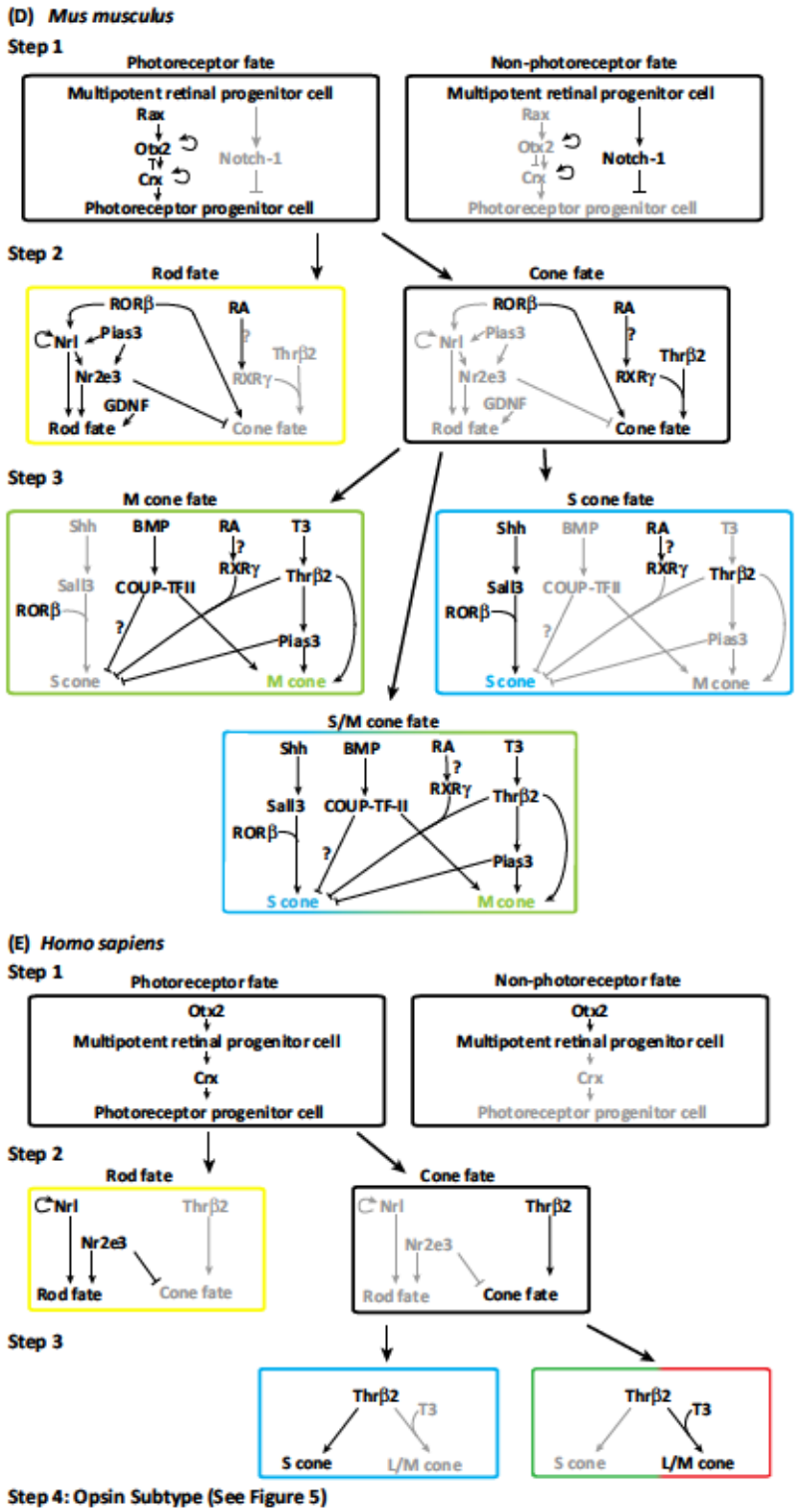
In *Drosophila*, the zinc-finger transcription factor Spalt (Sal) drives inner PR ('cone') fate by repressing Dve (Figure 2B, Step 2) [123]. In an additional step, not conserved in higher organisms, inner PR 'cones' differentiate further into two types: R7s, specified by the homeodomain transcription factor Prospero (Pros) and the transcription factor subunit Nf-yc, and R8s, specified by the zinc-finger transcription factor Senseless (Sens) (Figure 2B, Step 2) [156–158].

In zebrafish, the BMP family ligand Gdf6a induces the transcription factor Tbx2b to repress rod fate and allow cone development (Figure 2C, Step 2) [159–161]. Tbx2b does not appear to play a conserved role in cone specification; it is involved in dorsal–ventral retinal development in chickens, mice, and humans, but its expression is not restricted to cones [162,163]. In chickens



Step 4: Opsin Subtype (See Figure 5).

(See figure legend on the bottom of the next page.)



Trends in Genetics

Figure 2. The Gene Regulatory Networks Controlling PR Specification. All gene regulatory networks have been simplified to emphasize PR factors that are conserved between species. Arrows within gene networks solely represent our current understanding of network relationships and do not imply genetic mechanisms such as direct or indirect transcriptional regulation. (A) The basic steps of PR differentiation, which are largely conserved between organisms. (B) *Drosophila melanogaster*. (C) *Danio rerio*. (D) *Mus musculus*. (E) *Homo sapiens*.

and mice, RA signaling through the RXRγ receptor may be important for cone fate (Figures 2D, S1, Step 2) [164–167]. In addition, the Thrβ2 receptor plays a role in cone specification in chickens, mice, and humans (Figures 2D, 2E, S1, Step 2) [163,168–179].

Step 3: Cone Subtype Choice

In Step 3, cone precursors are specified into subtypes, marked by the expression of specific color-detecting opsins. Interestingly, several of the proteins required for cone subtype selection in flies are conserved in vertebrate PRs, although they have been adapted to play different roles. Selection between yellow and pale ommatidial subtypes in *Drosophila* is based on the stochastic expression of the PAS-bHLH transcription factor Spineless (Ss) in 65% of R7s [180]. In yR7s, Ss activates the expression of Rh4 and Dve, which represses Rh3 (Figure 2B, Step 3) [123,180]. In pR7s lacking Ss, Rh4 and Dve are not expressed, leading to activation of Rh3 by Sal and Orthodenticle (Otd), a homolog of vertebrate Crx, Otx2, and Otx5 (Figure 2B, Step 3) [180,181]. Intriguingly, the mammalian homolog of Sal, Sall3, has been conserved at both the sequence and functional levels; it also activates opsins in mice (Figure 2D, Step 3) [182].

In yR7s, Ss represses an unknown signal to R8s (Figure 2B, Step 3). In the absence of this signal, the Warts (Wts) serine/threonine kinase is activated, causing repression of the transcriptional coactivator Yorkie (Yki), a homolog of zebrafish Yap, in yR8s (Figure 2B, Step 3). Repression of Yki induces the activation of Rh6 and loss of Rh5 (Figure 2B, Step 3) [13,14,180,183,184]. In pR7s, the unknown signal activates the PH domain-containing protein Melted (Melt), which represses Wts to allow Yki activation and Rh5 expression in pR8s (Figure 2B, Step 3) [183,184]. In addition, Otd acts permissively in pR8s to activate Rh5 (Figure 2B, Step 3) [123].

In dorsal third yR7s, reduced Ss and Dve levels, combined with activation by the Iroquois complex of transcription factors (IroC), induces coexpression of Rh3 with Rh4 (Figure 2B, Step 3) [4,123,185]. In the dorsal rim, high local concentrations of the diffusible morphogen Wingless (Wg) act with IroC to drive the expression of the homeodomain transcription factor Homothorax (Hth) in R7s and R8s (Figure 2B, Step 3, and Figure 3A) [19,20]. Hth represses Ss in R7s, and Rh5, Rh6, and Sens in R8s, causing Rh3 expression in R7s and R8s (Figure 2B, Step 3) [19,20,186].

In zebrafish, mice, and humans, T3 thyroid hormone signals through the trβ2/Thrβ2 receptor to drive expression of specific opsins. In zebrafish, T3 activates LWS opsin, in mice, it activates M-opsin and represses S-opsin, and in humans, it may select L/M-opsins over S-opsin (Figure 2C–E, Step 3) [163,168–179,187]. RA signaling through the RXRγa/RXRγ receptor also controls opsin expression in vertebrates; in zebrafish, RA signaling activates LWS opsin and represses SWS1 and SWS2 opsins, while in mice it may repress S-opsin (Figure 2C,D, Step 3) [166,188,189].

Because T3 and RA are also involved in earlier steps of PR specification, additional factors likely work with them to specify cone subtypes. In mice, Pias3, BMP, and COUP-TFII work in addition to T3 to activate M-opsin (Figure 2D, Step 3) [152,170,190]. BMP and COUP-TFII may also assist RA and T3 in repressing mouse S-opsin (Figure 2D, Step 3) [190].

Additional factors have been implicated in vertebrate opsin expression, although it is currently unclear if they are conserved between species. In zebrafish, Gdf6a drives SWS2 expression and works in combination with Tbx2b to activate SWS1 (Figure 2C, Step 3) [159–161]. In addition, the fish-specific transcription factor Sine oculis homeobox homolog 7 (Six7) drives activation of RH2 (Figure 2C, Step 3) [191]. In mice, Shh signaling may activate Sall3, which acts with RORβ to activate S-opsin (Figure 2D, Step 3) [182,192,193].

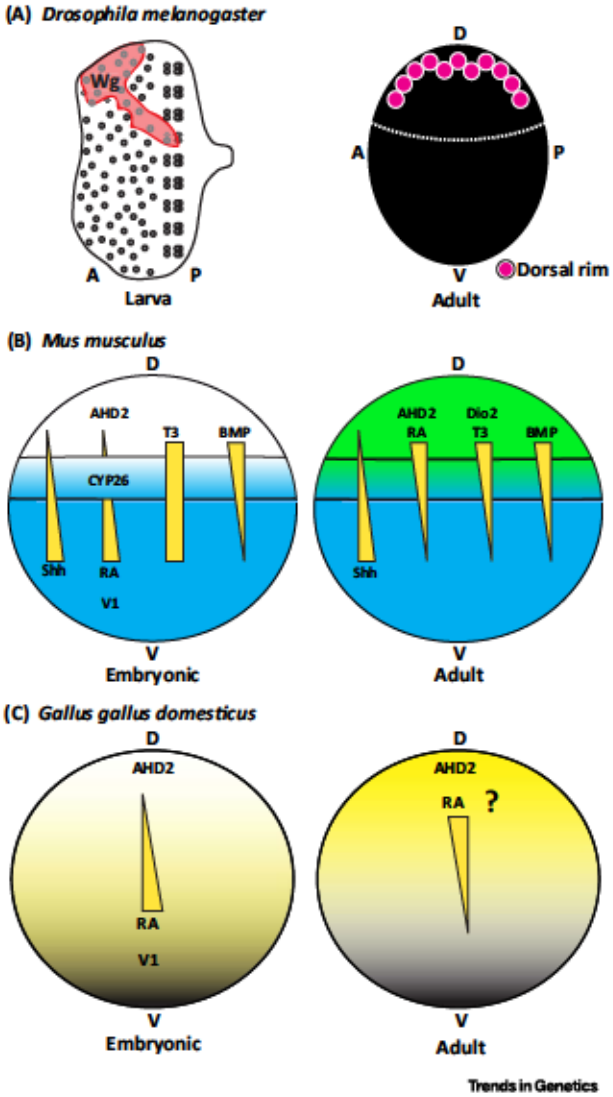


Figure 3. Gradients of Signaling Molecules Determine Regionalized Retinal Development. (A) In *Drosophila*, the diffusible morphogen Wg is expressed in a dorsal patch of the larval eye-antennal disc, beginning the signaling cascade leading to expression of Rh3 in the dorsal rim in the adult (Figure 2B). (B) Gradients of signaling molecules in the mouse retina leading to M (green) and S (blue) opsin expression. Sonic Hedgehog (Shh) is expressed in a ventral to dorsal gradient in both the embryo and the adult. Retinoic acid (RA) is expressed in a ventral to dorsal gradient at embryonic stages, and is produced by the enzymes V1 (ventral, high enzymatic activity) and AHD2 (dorsal, low enzymatic activity). CYP26 degrades RA in a strip through the middle of the retina. In the adult neither V1 nor CYP26 are expressed, and RA is therefore present in a dorsal to ventral gradient. Thyroid hormone (T3) is present throughout the embryonic retina. In the adult, T3 is present in a dorsal to ventral gradient, presumably governed by the presence of the T3 synthesizing enzyme Dio2. BMP is present in a dorsal to ventral gradient in both the embryonic and adult mouse retina. (C) In the chicken, RA is expressed in a ventral to dorsal gradient at embryonic stages and is produced by V1 and AHD2, as in mice. This mirrors the ventral to dorsal gradient of rods (black) within the chick retina. In the adult, V1 is not expressed, and RA is therefore present in a dorsal to ventral gradient. Abbreviations: A, anterior, D, dorsal; P, posterior; V, ventral.

Step 4: Opsin Subtype Choice

In zebrafish and humans, a final choice further differentiates cone subtypes based on opsin subtype expression (Figure 2C,E, Step 4). Zebrafish red- and green-detecting cones select between multiple LWS and RH2 opsin subtypes, respectively (Figure 5B) [194–196]. RA, potentially acting through RXRga, directs the expression of LWS-1 over LWS-2 (Figure 5B) [189]. Human L/M cones select between the closely related L- and M-opsins (Figure 5C) [197]. In

Table 1. PR Proteins with Functional, But Not Sequence-Level, Homology

Function	Fly	Zebrafish	Mouse
PR fate	Glass	Lbh-like	N/A
Rod fate	Dve	N/A	Pias3, RORβ
Cone fate	Pros, Nf-yc, Sens	Tbx2b, Gdf6a	N/A
Opsin choice	Ss, Dve, IroC, Wg, Hth	Gdf6a, Tbx2b, Six7	COUP-TFII, Pias3, BMP, RORβ

Table 2. PR Proteins with Sequence-Level, But Not Functional, Homology

Protein	Fly	Zebrafish	Chicken	Mouse	Human
Otd/Otx2/Otx5	Opsin choice	PR fate	PR fate	PR fate	Retinal cell fate
Yki/Yap	Opsin choice	PR fate repression	N/A	N/A	N/A
Shh	N/A	PR fate	N/A	Retinal regionalization	N/A

Table 3. PR Proteins with Functional and Sequence-Level Homology

Protein	Fly	Zebrafish	Chicken	Mouse	Human
Crx	N/A	PR fate	PR fate	PR fate	PR fate
Otx2/Otx5	N/A	PR fate	PR fate	PR fate	N/A
Rx1/RaxL/Rax	N/A	PR fate	PR fate	PR fate	N/A
Nrl	N/A	Rod fate	N/A	Rod fate	Rod fate
Nr2e3	N/A	Rod fate	N/A	Rod fate	Rod fate
RA	N/A	Rod fate, opsin choice, opsin subtype choice	Rod fate, cone fate	Cone fate, opsin choice	N/A
GDNF	N/A	Rod fate (?)	Rod fate	Rod fate	N/A
Sal/Sall3	Cone fate, opsin choice	N/A	N/A	Opsin choice	N/A
trβ2/Thrβ2	N/A	Opsin choice	Cone fate	Cone fate, opsin choice	Cone fate, opsin choice

both zebrafish and humans, locus control regions (LCRs) have evolved to regulate opsin subtype choice at the *cis* level (see below; Figure 5) [194–196].

Functional and Sequence-Level Homology

The proteins controlling PR specification can be divided into three main categories based on their functional and/or sequence-level homology. The first category involves factors that serve similar developmental roles but share no sequence homology (Table 1). A second category includes factors that are conserved on the sequence level but perform unique roles in different organisms (Table 2). The third category contains factors with functional and sequence-level homology (Table 3). In some cases, factors in this category may drive further, species-specific processes in addition to their conserved role.

Gradients of Signaling Molecules Determine Regionalized Retinal Development

In addition to conserved gene regulatory networks, diverse organisms share a common mechanism for delineating retinal regions, involving gradients of signaling molecules. There are two models for how such gradients are established. The first, more traditional model

suggests that gradients arise from diffusion of signaling molecules from a specific source. This occurs in *Drosophila*, where Wg is secreted from a stripe called the dorsal margin to create a dorsal-to-ventral gradient in the larval eye-antennal disc that specifies the location of dorsal rim ommatidia in adults (Figure 3A) [19,198].

An alternative 'gradient-free' model proposes that enzymes that produce or degrade signaling molecules are expressed in a regionalized pattern, and thereby regulate local levels of small molecules to create a gradient throughout the tissue [199]. This gradient-free mechanism may establish ventral to dorsal gradients of RA involved in retinal development and patterning in zebrafish, chickens, and mice [40,145,146,170,188,200–202]. In the developing chicken and mouse retina, the dorsally expressed aldehyde dehydrogenase AHD2 produces low RA, and the ventrally expressed aldehyde dehydrogenase V1 produces high RA, creating regional 'gradients' of RA in the retina (Figure 3B,C) [200,201,203–205]. In the chick, the ventral-to-dorsal gradient of RA mirrors the rod gradient, suggesting that regionalized RA processing enzymes drive gradients of PR subtypes (Figure 3C) [200,206,207]. In the mouse, an additional enzyme, the oxidase CYP26, causes RA degradation and a potential breakdown in the gradient in the central retina (Figure 3B) [200,208]. Together, these conserved patterns delineate different retinal regions during development.

Interestingly, in the adult retinas of both chick and mouse, ventral V1 dehydrogenase expression is lost, leaving dorsal AHD2 as the only RA synthesizing enzyme and causing a reversal of the gradient to higher RA levels in the dorsal retina (Figure 3B,C) [200,201]. In mice, this reversal may promote dorsal S-opsin repression after postnatal day 8 by activation of RXRγ (Figure 3B) [166,200,209]. In the chicken, it is unclear how this reversal affects PR fate specification (Figure 3C).

Deiodinases play a similar role in thyroid hormone gradient formation. They are expressed in regionalized areas and/or at different timepoints in the chick [169], mouse [210,211], and zebrafish retinas [212–214]. In mice, Deiodinase 2 (Dio2), which converts thyroid hormone from the inactive T4 to the active T3 form, is expressed at higher levels in the dorsal retina [210,215] and likely establishes a T3 gradient [216] (Figure 3B). High dorsal T3 signaling promotes expression of M-opsin and repression of S-opsin [152,166,170,171,217] (Figure 3B). Low T3 signaling in the ventral retina allows expression of S-opsin (Figure 3B) [170].

Dorsal–ventral BMP gradients and ventral–dorsal Shh gradients in the mouse retina activate M- and S-opsin, respectively, but the sources of these gradients remain unclear (Figure 3B) [182,190,192,218].

Retinal Development Proceeds Through Waves of Differentiation

Despite significant differences in morphology, regionalization, and sensitivity between organisms, retinal development in many species involves waves of differentiation. Within the developing fly eye-antennal disc, a wave of differentiation known as the morphogenetic furrow moves from the posterior to the anterior of the retina (Figure 4A), driven partially by the signaling molecule Hedgehog (Hh) and the bHLH transcription factor Atonal (Ato) [7,219–223]. Undifferentiated PR precursors lie anterior to the furrow whereas, posterior to the furrow, PRs differentiate in a specific order (Figure 4A) [7,219]. The R8 PR serves as a 'founder' cell, recruiting undifferentiated PR precursors and driving their stepwise differentiation into a complete ommatidium via multiple signaling pathways (well-reviewed in [9,224–230]). The initial differentiation of R2, R5, R3, and R4 is followed by the second mitotic wave (Figure 4A), after which R1, R6, and R7 sequentially differentiate [7,219,226].

Despite their separation by over 800 million years of evolution, zebrafish retinal differentiation shares much in common with the processes observed in flies. As in flies, a wave of neural

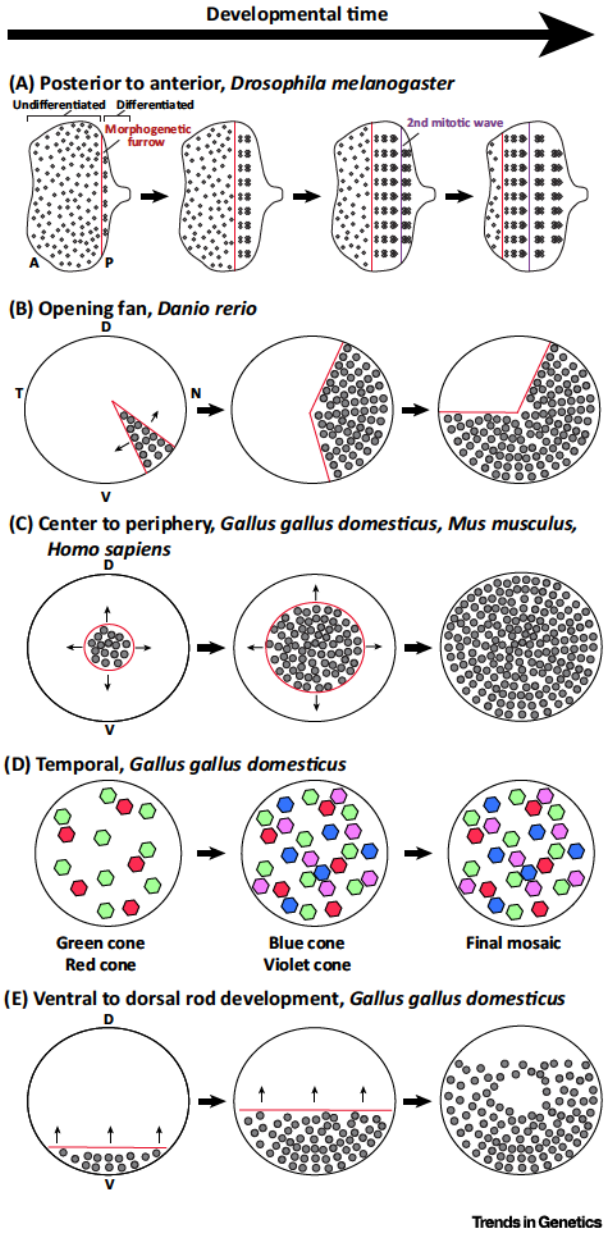


Figure 4. Retinal Development Proceeds Through Waves of Differentiation. (A) In *Drosophila*, waves of differentiation and mitosis move from posterior to anterior. (B) In zebrafish, differentiation proceeds from ventral–nasal to dorsal–temporal in a wave resembling an opening fan. (C) In chickens, mice, and humans, differentiation begins in the center of the retina and expands towards the periphery. (D) Chicken retinal development also involves a temporal wave of cone maturation. Green and red cones are the earliest to mature, followed by blue and violet cones. (E) A ventral-to-dorsal wave of differentiation patterns rods in the chicken retina in a density gradient, excluding the area centralis. Abbreviations; A, anterior, D, dorsal; N, nasal; P, posterior; T, temporal; V, ventral.

differentiation driven in part by Hedgehog signaling and *ath5*, a zebrafish homolog of Ato, spreads across the developing retina (Figure 4B) [231,232]. In zebrafish, PRs differentiate from an initial patch [233–235]. Cones spread from this patch in a wave resembling an opening fan, with differentiation sweeping from ventral–nasal to dorsal–temporal [21,233,234] (Figure 4B). A mitotic wave follows the initial fan gradient to complete cone differentiation [233,235]. Early-differentiating red cones may act similarly to R8 PRs in *Drosophila*, functioning as 'founders' to recruit

undifferentiated cone precursors and drive their differentiation [21]. While rods are also found initially in the ventral patch, they differentiate separately from cones. Clusters of rod precursors scattered throughout the retina undergo multiple rounds of mitosis before differentiating into rods and migrating to their final positions around UV cones [37,233,234,236].

The chicken retina is similar to that of zebrafish in that differentiation begins at a central patch, the area centralis [40,46,169]. Sequential waves of transcription factor expression emanate from the center to the periphery to drive cell differentiation and retinal patterning (Figure 4C). First, a wave of cone precursor transcription factors is expressed, including Thrβ2 and Otx2 [169]. Individual cone subtypes then express opsins in temporal waves. Green and red opsins are expressed first, followed by blue and violet [40] (Figure 4D). An additional wave of differentiation sweeps linearly across the retina from the ventral to dorsal region to pattern rods (Figure 4E) [40].

Although mice do not have a fovea, retinal differentiation follows the same central-to-peripheral pattern that is seen in chickens (Figure 4C) [237,238]. Generation of different retinal cell types is coincident with temporal waves, and this phenomenon has been used to identify important factors in retinal generation [239].

The developmental pattern of the human and other primate retinas closely resembles the chick and mouse retina, with differentiation following sequential waves emanating from the optic disc, near the fovea, outward [240–242] (Figure 4C). S-cones are seen first in the foveal area, followed by L/M-cones. Rods are seen later outside the fovea [242,243]. The fetal fovea is not packed as tightly as the adult fovea [64], suggesting that differentiated cones migrate toward the central fovea later in development to create a densely packed array [69,244,245].

Looping of DNA Elements Regulates Cone Subtypes

Beyond retina-wide signaling gradients and waves of differentiation, conserved mechanisms control retinal development at an individual PR level. Looping of regulatory DNA elements plays a crucial role in opsin choice across organisms. In *Drosophila*, DNA looping may regulate the stochastic expression of *spineless* (*ss*), the key determinant of R7 ('cone') subtype fate. The *ss* locus contains an enhancer, which activates *ss* in 100% of R7s, and two silencers, which randomly repress *ss* in 35% of R7s (Figure 5A) [246]. Because the two silencers are located at a significant distance from the *ss* promoter, it is likely that they regulate *ss* through a looping-based mechanism. An enticing hypothesis is that the enhancer and silencers compete for looping to the *ss* promoter, resulting in activation or repression of *ss* and regulation of downstream Rhodopsins (Figure 5A).

In a striking example of convergent evolution between zebrafish and humans, DNA elements known as locus control regions (LCRs) likely regulate opsin expression through looping-based mechanisms. In both cases, ancestral enhancers that regulated the expression of a single opsin gene were adapted in response to an opsin gene duplication [62,194–196,247–250].

Zebrafish opsin genes are regulated by two LCRs, one that selects between LWS subtypes and one that selects between RH2 subtypes (Figure 5B) [194–196]. LCR-mediated regulation of opsin subtypes is controlled in a temporal progression [39,194,195]. RH2-1, RH2-2, and LWS-2 are expressed earliest and are present in the central and dorsal regions of the zebrafish retina, which develop first (Figures 1B3, 5B) [39]. RH2-3, RH2-4, and LWS-1 are expressed later and thus localize to the later-developing retinal periphery (Figures 1B3, 5B) [39].

Human opsin genes are regulated by one LCR that selects between L- and M-opsin expression (Figure 5C) [251]. It is hypothesized that the LCR loops randomly to the promoter of either the L- or

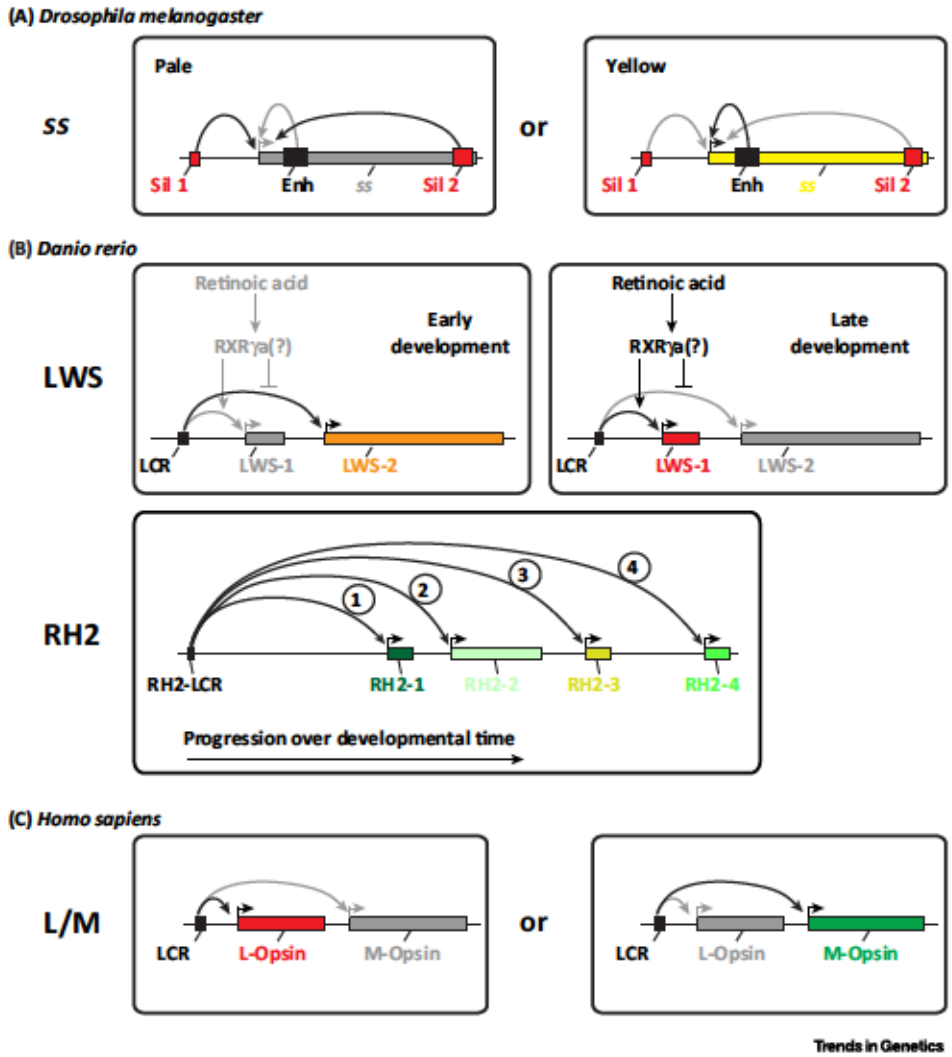


Figure 5. Looping of DNA Elements Regulates Cone Subtypes. (A) In *Drosophila*, looping of regulatory elements may cause activation or repression of *ss*, the key determinant of R7 subtype fate. (B) RA signaling and LCR looping select between opsin subtypes in zebrafish. Numbers in the RH2 box indicate the temporal order of RH2 subtype expression. (C) LCR looping selects between L- and M-opsin for expression in human L/M-cones. Abbreviations: Enh, enhancer; LCR, locus control region; LWS, Long-Wavelength Sensitive; RA, retinoic acid; RH1, Rhodopsin; RH2, Rhodopsin-like; Sil 1, silencer 1; Sil 2, silencer 2; *ss*, *Drosophila spineless*; SWS1/2, Short-Wavelength Sensitive 1/2.

M-opsin gene to drive opsin expression [63,252,253]. Alternatively, the human LCR might activate opsins in a temporal progression, after which L- and M-opsin-expressing cones might migrate to their final, random positions in the human retina [244,245].

The zebrafish and human LCRs are all about 0.5 kb in size, perhaps reflecting a common sequence length that is required for robust activation of opsin expression [194,195,251,252]. Despite their common sizes, the RH2, LWS, and human LCRs have little sequence similarity other than shared binding sites for the transcription factor Crx [194,195].

Concluding Remarks

Many questions about the gene regulatory and evolutionary mechanisms governing retinal development remain unanswered (see Outstanding Questions). Further study of PR

Outstanding Questions

What are the functional roles of stochastic/regionalized, regionalized, and ordered retinal mosaics?

What are the missing regulatory nodes controlling PR fate?

What are the mechanisms controlling stochastic cell fate specification?

What signaling or migration mechanisms determine the ordered pattern of PRs in the fish eye?

How are multiple gradients integrated to dictate regionalized PR patterns?

How do functionally homologous gene networks evolve to control similar PR specification processes?

development and maintenance will provide insight into the evolutionary advantages of different retinal mosaics and uncover additional conserved and species-specific gene regulatory networks required for retinal patterning. A deeper understanding of these mechanisms may ultimately lead to new treatments for many developmental disorders of the visual system and the development of effective PR regenerative therapies.

Acknowledgments

We would like to thank Seth Blackshaw and members of the laboratory of R.J.J. for helpful comments and suggestions on the manuscript. K.C.E. is a Howard Hughes Medical Institute (HHMI) Gilliam fellow. This work was supported by a Pew Scholar Award from Pew Charitable Trusts (00027373 to R.J.J.), a Basil O'Connor Scholar Award from the March of Dimes Foundation (5-FY15-21 to R.J.J.), and a grant from the National Institutes of Health (R01EY025598 to R.J.J.).

Supplemental Information

Supplemental information associated with this article can be found online at doi:10.1016/j.tig.2016.07.004.

References

1. Julius, D. and Nathans, K. (2012) Signaling by sensory receptors. *Cold Spring Harb. Perspect. Biol.* 4, a005991
2. Osorio, D. and Vorobyev, M. (2008) A review of the evolution of animal colour vision and visual communication signals. *Vision Res.* 48, 2042–2051
3. Osorio, D. and Vorobyev, M. (2005) Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. *Proc. Biol. Sci.* 272, 1745–1752
4. Mazzoni, E.O. *et al.* (2008) Iroquois complex genes induce co-expression of rhodopsins in *Drosophila*. *PLoS Biol.* 6, e97
5. Wernet, M.F. *et al.* (2012) Genetic dissection reveals two separate retinal substrates for polarization vision in *Drosophila*. *Curr. Biol.* 22, 12–20
6. Weir, P.T. and Dickinson, M.H. (2012) Flying *Drosophila* orient to sky polarization. *Curr. Biol.* 22, 21–27
7. Wolff, T. and Ready, D.F. (1991) The beginning of pattern formation in the *Drosophila* compound eye: the morphogenetic furrow and the second mitotic wave. *Development* 113, 841–850
8. Wolff, T. and Ready, D.F. (1993) Pattern formation in the *Drosophila* retina. In *The Development of Drosophila melanogaster (Vol. 2)* (Bate, M. and Martinez Arias, A., eds), pp. 1277–1325, Cold Spring Harbor Laboratory Press
9. Pichaud, F. (2014) Transcriptional regulation of tissue organization and cell morphogenesis: the fly retina as a case study. *Dev. Biol.* 385, 168–178
10. O'Tousa, J.E. *et al.* (1985) The *Drosophila* ninaE gene encodes an opsin. *Cell* 40, 839–850
11. Zuker, C.S. *et al.* (1985) Isolation and structure of a rhodopsin gene from D. *melanogaster*. *Cell* 40, 851–858
12. Bell, M.L. *et al.* (2007) Two types of *Drosophila* R7 photoreceptor cells are arranged randomly: a model for stochastic cell-fate determination. *J. Comp. Neurol.* 502, 75–85
13. Chou, W.H. *et al.* (1996) Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells. *Neuron* 17, 1101–1115
14. Chou, W.H. *et al.* (1999) Patterning of the R7 and R8 photoreceptor cells of *Drosophila*: evidence for induced and default cell-fate specification. *Development* 126, 607–616
15. Montell, C. *et al.* (1987) A second opsin gene expressed in the ultraviolet-sensitive R7 photoreceptor cells of *Drosophila* melanogaster. *J. Neurosci.* 7, 1558–1566
16. Papatsenko, D. *et al.* (1997) A new rhodopsin in R8 photoreceptors of *Drosophila*: evidence for coordinate expression with Rh3 in R7 cells. *Development* 124, 1665–1673
17. Zuker, C.S. *et al.* (1987) A rhodopsin gene expressed in photoreceptor cell R7 of the *Drosophila* eye: homologies with other signal-transducing molecules. *J. Neurosci.* 7, 1550–1557
18. Fortini, M.E. and Rubin, G.M. (1990) Analysis of cis-acting requirements of the Rh3 and Rh4 genes reveals a bipartite organization to rhodopsin promoters in *Drosophila* melanogaster. *Genes Dev.* 4, 444–463
19. Tomlinson, A. (2003) Patterning the peripheral retina of the fly: decoding a gradient. *Dev. Cell* 5, 799–809
20. Wernet, M.F. *et al.* (2003) Homothorax switches function of *Drosophila* photoreceptors from color to polarized light sensors. *Cell* 115, 267–279
21. Raymond, P.A. and Barthel, L.K. (2004) A moving wave patterns the cone photoreceptor mosaic array in the zebrafish retina. *Int. J. Dev. Biol.* 48, 935–945
22. Allison, W.T. *et al.* (2010) Ontogeny of cone photoreceptor mosaics in zebrafish. *J. Comp. Neurol.* 518, 4182–4195
23. Wheeler, T.G. (1982) Color vision and retinal chromatic information processing in teleost: a review. *Brain Res.* 257, 177–235
24. MacNichol, E.F., Jr *et al.* (1978) Ellipsosomes: organelles containing a cytochrome-like pigment in the retinal cones of certain fishes. *Science* 200, 549–552
25. Hoang, Q.V. *et al.* (2002) Photoreceptor inner segments in monkey and human retina: mitochondrial density, optics, and regional variation. *Visual Neurosci.* 19, 395–407
26. Tarboush, R. *et al.* (2014) Variability in mitochondria of zebrafish photoreceptor ellipsoids. *Visual Neurosci.* 31, 11–23
27. Troutt, L.L. and Burnside, B. (1988) Microtubule polarity and distribution in teleost photoreceptors. *J. Neurosci.* 8, 2371–2380
28. Nagle, B. and Burnside, B. (1983) Retinomotor movements of photoreceptors and retinal pigment epithelium: mechanisms and regulation. In *Progress in Retinal Research (Vol. 2)* (Osborne, N. *et al.*, eds), pp. 67–109
29. Ali, M.A. (1975) Retinomotor responses. In *Vision in Fishes, New Approaches in Research* (Ali, M.A., ed.), pp. 313–355, Plenum
30. Nawrocki, L. *et al.* (1985) Larval and adult visual pigments of the zebrafish, *Brachydanio rerio*. *Vision Res.* 25, 1569–1576
31. Raymond, P.A. *et al.* (1993) Expression of rod and cone visual pigments in goldfish and zebrafish: a rhodopsin-like gene is expressed in cones. *Neuron* 10, 1161–1174
32. Vihtelic, T.S. *et al.* (1999) Cloning and characterization of six zebrafish photoreceptor opsin cDNAs and immunolocalization of their corresponding proteins. *Visual Neurosci.* 16, 571–585
33. Branchek, T. and Bremiller, R. (1984) The development of photoreceptors in the zebrafish, *Brachydanio rerio*. I. Structure. *J. Comp. Neurol.* 224, 107–115
34. Engstrom, K. (1963) Cone types and cone arrangements in teleost retinae. *Acta Zool.* 44, 179–243
35. Larison, K.D. and Bremiller, R. (1990) Early onset of phenotype and cell patterning in the embryonic zebrafish retina. *Development* 109, 567–576
36. Raymond, P.A. (1995) . In *Neurobiology and clinical Aspects of the Outer Retina* (Djamgoz, M.B.A. *et al.*, eds), pp. 1–23, Chapman & Hall
37. Fadool, J.M. (2003) Development of a rod photoreceptor mosaic revealed in transgenic zebrafish. *Dev. Biol.* 258, 277–290

38. Chinen, A. *et al.* (2003) Gene duplication and spectral diversification of cone visual pigments of zebrafish. *Genetics* 163, 663–675
39. Takechi, M. and Kawamura, S. (2005) Temporal and spatial changes in the expression pattern of multiple red and green subtype opsin genes during zebrafish development. *J. Exp. Biol.* 208, 1337–1345
40. Bruhn, S.L. and Cepko, C.L. (1996) Development of the pattern of photoreceptors in the chick retina. *J. Neurosci.* 16, 1430–1439
41. Govardovskii, V.I. and Zueva, L.V. (1977) Visual pigments of chicken and pigeon. *Vision Res.* 17, 537–543
42. Meyer, D.B. and May, H.C., Jr (1973) The topographical distribution of rods and cones in the adult chicken retina. *Exp. Eye Res.* 17, 347–355
43. Kram, Y.A. *et al.* (2010) Avian cone photoreceptors tile the retina as five independent, self-organizing mosaics. *PLoS ONE* 5, e8992
44. Hart, N.S. (2001) Variations in cone photoreceptor abundance and the visual ecology of birds. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* 187, 685–697
45. Wilby, D. *et al.* (2015) Optics of cone photoreceptors in the chicken (*Gallus gallus domesticus*). *J. R. Soc. Interface* 12, 20150591
46. Wai, M.S. *et al.* (2006) Morphogenesis of the different types of photoreceptors of the chicken (*Gallus domesticus*) retina and the effect of amblyopia in neonatal chicken. *Microsc. Res. Tech.* 69, 99–107
47. Araki, M. *et al.* (1990) Localization of iodopsin in the chick retina during in vivo and in vitro cone differentiation. *Invest. Ophthalmol. Vis. Sci.* 31, 1466–1473
48. Campenhausen, M.V. and Kirschfeld, K. (1998) Spectral sensitivity of the accessory optic system of the pigeon. *J. Comp. Physiol. A* 183, 1–6
49. Campenhausen, M. and Kirschfeld, K. (1998) Spectral sensitivity of the accessory optic system of the pigeon. *J. Comp. Physiol. A* 183, 1–6
50. Morris, V.B. (1982) An afoveate area centralis in the chick retina. *J. Comp. Neurol.* 210, 198–203
51. Bueno, J.M. *et al.* (2011) Analysis of the chicken retina with an adaptive optics multiphoton microscope. *Biomed. Opt. Express* 2, 1637–1648
52. Headington, K. *et al.* (2011) Single cell imaging of the chick retina with adaptive optics. *Curr. Eye Res.* 36, 947–957
53. Szel, A. *et al.* (1996) Distribution of cone photoreceptors in the mammalian retina. *Microsc. Res. Tech.* 35, 445–462
54. Lukats, A. *et al.* (2005) Photopigment coexpression in mammals: comparative and developmental aspects. *Histol. Histopathol.* 20, 551–574
55. Rohlich, P. *et al.* (1994) Two different visual pigments in one retinal cone cell. *Neuron* 13, 1159–1166
56. Applebury, M.L. *et al.* (2000) The murine cone photoreceptor: a single cone type expresses both S and M opsins with retinal spatial patterning. *Neuron* 27, 513–523
57. Haverkamp, S. *et al.* (2005) The primordial, blue-cone color system of the mouse retina. *J. Neurosci.* 25, 5438–5445
58. Jeon, C.J. *et al.* (1998) The major cell populations of the mouse retina. *J. Neurosci.* 18, 8936–8946
59. Deb, S.S. (2005) The molecular basis of variation in human color vision. *Clin. Genet.* 67, 369–377
60. Hunt, D.M. (2001) Seeing rainbows. *Biologist* 48, 67–71
61. Kainz, P.M. *et al.* (1998) Recent evolution of uniform trichromacy in a New World monkey. *Vision Res.* 38, 3315–3320
62. Nathans, J. (1999) The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments. *Neuron* 24, 299–312
63. Nathans, J. *et al.* (1986) Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232, 193–202
64. Hendrickson, A. (1992) A morphological comparison of foveal development in man and monkey. *Eye* 6, 136–144
65. Roorda, A. *et al.* (2001) Packing arrangement of the three cone classes in primate retina. *Vision Res.* 41, 1291–1306
66. Roorda, A. and Williams, D.R. (1999) The arrangement of the three cone classes in the living human eye. *Nature* 397, 520–522
67. Curcio, C.A. *et al.* (1991) Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J. Comp. Neurol.* 312, 610–624
68. Ahnelt, P.K. (1998) The photoreceptor mosaic. *Eye* 12, 531–540
69. Cornish, E.E. *et al.* (2004) Distribution of short-wavelength-sensitive cones in human fetal and postnatal retina: early development of spatial order and density profiles. *Vision Res.* 44, 2019–2026
70. Curcio, C.A. *et al.* (1990) Human photoreceptor topography. *J. Comp. Neurol.* 292, 497–523
71. Hofer, H. *et al.* (2005) Organization of the human trichromatic cone mosaic. *J. Neurosci.* 25, 9669–9679
72. Williams, R.W. (1991) The human retina has a cone-enriched rim. *Visual Neurosci.* 6, 403–406
73. Gowdy, P.D. and Cicerone, C.M. (1998) The spatial arrangement of the L and M cones in the central fovea of the living human eye. *Vision Res.* 38, 2575–2589
74. Otake, S. *et al.* (2000) The spatial arrangement of L and M cones in the peripheral human retina. *Vision Res.* 40, 677–693
75. Rossi, E.A. *et al.* (2011) Imaging retinal mosaics in the living eye. *Eye* 25, 301–308
76. Li, K.Y. and Roorda, A. (2007) Automated identification of cone photoreceptors in adaptive optics retinal images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 24, 1358–1363
77. Williams, D.R. (2011) Imaging single cells in the living retina. *Vision Res.* 51, 1379–1396
78. Calkins, D.J. *et al.* (1994) M and L cones in macaque fovea connect to midget ganglion cells by different numbers of excitatory synapses. *Nature* 371, 70–72
79. Mustafi, D. *et al.* (2009) Structure of cone photoreceptors. *Prog. Retin. Eye Res.* 28, 289–302
80. Ahnelt, P.K. *et al.* (1987) Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina. *J. Comp. Neurol.* 255, 18–34
81. Ehrlich, D. (1981) Regional specialization of the chick retina as revealed by the size and density of neurons in the ganglion cell layer. *J. Comp. Neurol.* 195, 643–657
82. Straznicky, C. and Chehade, M. (1987) The formation of the area centralis of the retinal ganglion cell layer in the chick. *Development* 100, 411–420
83. Lythgoe, J. (1979) *The Ecology of Vision,* Claredon Press
84. Chang, L. *et al.* (2013) Chromatic coding from cone-type unselective circuits in the mouse retina. *Neuron* 77, 559–571
85. Baden, T. *et al.* (2013) A tale of two retinal domains: near-optimal sampling of achromatic contrasts in natural scenes through asymmetric photoreceptor distribution. *Neuron* 80, 1206–1217
86. Moses, K. *et al.* (1989) The glass gene encodes a zinc-finger protein required by *Drosophila* photoreceptor cells. *Nature* 340, 531–536
87. Li, W.H. *et al.* (2015) Zebrafish Lbh-like is required for Otx2-mediated photoreceptor differentiation. *Int. J. Biol. Sci.* 11, 688–700
88. Nelson, S.M. *et al.* (2008) The developmental sequence of gene expression within the rod photoreceptor lineage in embryonic zebrafish. *Dev. Dynam.* 237, 2903–2917
89. Nelson, S.M. *et al.* (2009) Retinal homeobox 1 is required for retinal neurogenesis and photoreceptor differentiation in embryonic zebrafish. *Dev. Biol.* 328, 24–39
90. Furukawa, T. *et al.* (1997) rax, a novel paired-type homeobox gene, shows expression in the anterior neural fold and developing retina. *Proc. Natl. Acad. Sci. U.S.A.* 94, 3088–3093
91. Furukawa, T. *et al.* (2000) rax, Hes1, and notch1 promote the formation of Muller glia by postnatal retinal progenitor cells. *Neuron* 26, 383–394
92. Muranishi, Y. *et al.* (2011) An essential role for RAX homeoprotein and NOTCH–HES signaling in Otx2 expression in embryonic

retinal photoreceptor cell fate determination. *J. Neurosci.* 31, 16792–16807

93. Nishida, A. *et al.* (2003) Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. *Nature Neurosci.* 6, 1255–1263
94. Omori, Y. *et al.* (2011) Analysis of transcriptional regulatory pathways of photoreceptor genes by expression profiling of the Otx2-deficient retina. *PLoS ONE* 6, e19685
95. Furukawa, T. *et al.* (1997) Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell* 91, 531–541
96. Furukawa, T. *et al.* (1999) Retinopathy and attenuated circadian entrainment in Crx-deficient mice. *Nat. Genet.* 23, 466–470
97. Shen, Y.C. and Raymond, P.A. (2004) Zebrafish cone-rod (crx) homeobox gene promotes retinogenesis. *Dev. Biol.* 269, 237–251
98. Liu, Y. *et al.* (2001) Isolation and characterization of a zebrafish homologue of the cone rod homeobox gene. *Invest. Ophthalmol. Vis. Sci.* 42, 481–487
99. Gamse, J.T. *et al.* (2002) Otx5 regulates genes that show circadian expression in the zebrafish pineal complex. *Nat. Genet.* 30, 117–121
100. Chuang, J.C. *et al.* (1999) Expression of three Rx homeobox genes in embryonic and adult zebrafish. *Mech. Dev.* 84, 195–198
101. Hennig, A.K. *et al.* (2008) Regulation of photoreceptor gene expression by Crx-associated transcription factor network. *Brain Res.* 1192, 114–133
102. Slavotinek, A.M. *et al.* (2015) Exome sequencing in 32 patients with anophthalmia/microphthalmia and developmental eye defects. *Clin. Genet.* 88, 468–473
103. Takagi, M. *et al.* (2015) Heterozygous defects in PAX6 gene and congenital hypopituitarism. *Eur. J. Endocrinol.* 172, 37–45
104. Vincent, A. *et al.* (2014) OTX2 mutations cause autosomal dominant pattern dystrophy of the retinal pigment epithelium. *J. Med. Genet.* 51, 797–805
105. Sohocki, M.M. *et al.* (1998) A range of clinical phenotypes associated with mutations in CRX, a photoreceptor transcription-factor gene. *Am. J. Hum. Genet.* 63, 1307–1315
106. Akagi, T. *et al.* (2005) Iris-derived cells from adult rodents and primates adopt photoreceptor-specific phenotypes. *Invest. Ophthalmol. Vis. Sci.* 46, 3411–3419
107. Chen, C.M. and Cepko, C.L. (2002) The chicken RaxL gene plays a role in the initiation of photoreceptor differentiation. *Development* 129, 5363–5375
108. Chen, S. *et al.* (1997) Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. *Neuron* 19, 1017–1030
109. Ochi, H. *et al.* (2004) Temporal expression of L-Maf and RaxL in developing chicken retina are arranged into mosaic pattern. *Gene Expr. Patterns* 4, 489–494
110. Emerson, M.M. and Cepko, C.L. (2011) Identification of a retina-specific Otx2 enhancer element active in immature developing photoreceptors. *Dev. Biol.* 360, 241–255
111. Emerson, M.M. *et al.* (2013) Otx2 and Onecut1 promote the fates of cone photoreceptors and horizontal cells and repress rod photoreceptors. *Dev. Cell* 26, 59–72
112. Bovolenta, P. *et al.* (1997) Implication of OTX2 in pigment epithelium determination and neural retina differentiation. *J. Neurosci.* 17, 4243–4252
113. Freund, C.L. *et al.* (1997) Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* 91, 543–553
114. Ragge, N.K. *et al.* (2005) Heterozygous mutations of OTX2 cause severe ocular malformations. *Am. J. Hum. Genet.* 76, 1008–1022
115. Jacobson, S.G. *et al.* (1998) Retinal degenerations with truncation mutations in the cone-rod homeobox (CRX) gene. *Invest. Ophthalmol. Vis. Sci.* 39, 2417–2426
116. Swain, P.K. *et al.* (1997) Mutations in the cone-rod homeobox gene are associated with the cone-rod dystrophy photoreceptor degeneration. *Neuron* 19, 1329–1336
117. Rivolta, C. *et al.* (2001) Dominant Leber congenital amaurosis, cone-rod degeneration, and retinitis pigmentosa caused by mutant versions of the transcription factor CRX. *Human Mutat.* 18, 488–498
118. Wang, S. *et al.* (2014) A gene regulatory network controls the binary fate decision of rod and bipolar cells in the vertebrate retina. *Dev. Cell* 30, 513–527
119. Asaoka, Y. *et al.* (2014) The Hippo pathway controls a switch between retinal progenitor cell proliferation and photoreceptor cell differentiation in zebrafish. *PLoS ONE* 9, e97365
120. Yaron, O. *et al.* (2006) Notch1 functions to suppress cone-photoreceptor fate specification in the developing mouse retina. *Development* 133, 1367–1378
121. Jadhav, A.P. *et al.* (2006) Notch 1 inhibits photoreceptor production in the developing mammalian retina. *Development* 133, 913–923
122. Stenkamp, D.L. *et al.* (2002) Embryonic retinal gene expression in sonic-you mutant zebrafish. *Dev. Dynam.* 225, 344–350
123. Johnston, R.J., Jr *et al.* (2011) Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the *Drosophila* eye. *Cell* 145, 956–968
124. Chen, J. *et al.* (2005) The rod photoreceptor-specific nuclear receptor Nr2e3 represses transcription of multiple cone-specific genes. *J. Neurosci.* 25, 118–129
125. Kitambi, S.S. and Hauptmann, G. (2007) The zebrafish orphan nuclear receptor genes nr2e1 and nr2e3 are expressed in developing eye and forebrain. *Gene Expr. Patterns* 7, 521–528
126. Bessant, D.A. *et al.* (1999) A mutation in NRL is associated with autosomal dominant retinitis pigmentosa. *Nat. Genet.* 21, 355–356
127. DeAngelis, M.M. *et al.* (2002) Novel mutations in the NRL gene and associated clinical findings in patients with dominant retinitis pigmentosa. *Arch Ophthalmol.* 120, 369–375
128. Haider, N.B. *et al.* (2006) The transcription factor Nr2e3 functions in retinal progenitors to suppress cone cell generation. *Visual Neurosci.* 23, 917–929
129. Jacobson, S.G. *et al.* (2004) Nuclear receptor NR2E3 gene mutations distort human retinal laminar architecture and cause an unusual degeneration. *Hum. Mol. Genet.* 13, 1893–1902
130. Mears, A.J. *et al.* (2001) Nrl is required for rod photoreceptor development. *Nat. Genet.* 29, 447–452
131. Mitton, K.P. *et al.* (2000) The leucine zipper of NRL interacts with the CRX homeodomain. A possible mechanism of transcriptional synergy in rhodopsin regulation. *J. Biol. Chem.* 275, 29794–29799
132. Montana, C.L. *et al.* (2011) Transcriptional regulation of neural retina leucine zipper (Nrl), a photoreceptor cell fate determinant. *J. Biol. Chem.* 286, 36921–36931
133. Nishiguchi, K.M. *et al.* (2004) Recessive NRL mutations in patients with clumped pigmentary retinal degeneration and relative preservation of blue cone function. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17819–17824
134. Oh, E.C. *et al.* (2007) Transformation of cone precursors to functional rod photoreceptors by bZIP transcription factor NRL. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1679–1684
135. Rehemtulla, A. *et al.* (1996) The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 93, 191–195
136. Roger, J.E. *et al.* (2010) Sumoylation of bZIP transcription factor NRL modulates target gene expression during photoreceptor differentiation. *J. Biol. Chem.* 285, 25637–25644
137. Swain, P.K. *et al.* (2001) Multiple phosphorylated isoforms of NRL are expressed in rod photoreceptors. *J. Biol. Chem.* 276, 36824–36830
138. Wright, A.F. *et al.* (2004) Mutation analysis of NR2E3 and NRL genes in Enhanced S Cone Syndrome. *Human Mutat.* 24, 439
139. Yoshida, S. *et al.* (2004) Expression profiling of the developing and mature $Nrl^{-/-}$ mouse retina: identification of retinal disease candidates and transcriptional regulatory targets of Nrl. *Hum. Mol. Genet.* 13, 1487–1503
140. Haider, N.B. *et al.* (2000) Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. *Nat. Genet.* 24, 127–131

141. Haider, N.B. *et al.* (2001) Excess cone cell proliferation due to lack of a functional NR2E3 causes retinal dysplasia and degeneration in rd7/rd7 mice. *Hum. Mol. Genet.* 10, 1619–1626
142. Milam, A.H. *et al.* (2002) The nuclear receptor NR2E3 plays a role in human retinal photoreceptor differentiation and degeneration. *Proc. Natl. Acad. Sci. U.S.A.* 99, 473–478
143. Oh, E.C. *et al.* (2008) Rod differentiation factor NRL activates the expression of nuclear receptor NR2E3 to suppress the development of cone photoreceptors. *Brain Res.* 1236, 16–29
144. Hao, H. *et al.* (2012) Transcriptional regulation of rod photoreceptor homeostasis revealed by in vivo NRL targetome analysis. *PLoS Genet.* 8, e1002649
145. Stevens, C.B. *et al.* (2011) Plasticity of photoreceptor-generating retinal progenitors revealed by prolonged retinoic acid exposure. *BMC Dev. Biol.* 11, 51
146. Hyatt, G.A. *et al.* (1996) Retinoic acid alters photoreceptor development in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13298–13303
147. Lucini, C. *et al.* (2007) Glial cell line-derived neurotrophic factor expression in the retina of adult zebrafish (*Danio rerio*). *Neurosci. Lett.* 429, 156–160
148. Rothermel, A. and Layer, P.G. (2003) GDNF regulates chicken rod photoreceptor development and survival in reaggregated histotypic retinal spheres. *Invest. Ophthalmol. Vis. Sci.* 44, 2221–2228
149. Ogilvie, J.M. *et al.* (2000) Growth factors in combination, but not individually, rescue rd mouse photoreceptors in organ culture. *Exp. Neurol.* 161, 676–685
150. Frasson, M. *et al.* (1999) Glial cell line-derived neurotrophic factor induces histologic and functional protection of rod photoreceptors in the rd/rd mouse. *Invest. Ophthalmol. Vis. Sci.* 40, 2724–2734
151. Volpert, K.N. *et al.* (2007) GDNF stimulates rod photoreceptors and dopaminergic amacrine cells in chicken retinal reaggregates. *Invest. Ophthalmol. Vis. Sci.* 48, 5306–5314
152. Onishi, A. *et al.* (2010) Pias3-dependent SUMOylation controls mammalian cone photoreceptor differentiation. *Nature Neurosci.* 13, 1059–1065
153. Onishi, A. *et al.* (2009) Pias3-dependent SUMOylation directs rod photoreceptor development. *Neuron* 61, 234–246
154. Kautzmann, M.A. *et al.* (2011) Combinatorial regulation of photoreceptor differentiation factor, neural retina leucine zipper gene NRL, revealed by in vivo promoter analysis. *J. Biol. Chem.* 286, 28247–28255
155. Kim, J.W. *et al.* (2016) Recruitment of rod photoreceptors from short-wavelength-sensitive cones during the evolution of nocturnal vision in mammals. *Dev. Cell* 37, 520–532
156. Cook, T. *et al.* (2003) Distinction between color photoreceptor cell fates is controlled by Prospero in *Drosophila*. *Dev. Cell* 4, 853–864
157. Morey, M. *et al.* (2008) Coordinate control of synaptic-layer specificity and rhodopsins in photoreceptor neurons. *Nature* 456, 795–799
158. Xie, B. *et al.* (2007) Senseless functions as a molecular switch for color photoreceptor differentiation in *Drosophila*. *Development* 134, 4243–4253
159. Alvarez-Delfin, K. *et al.* (2009) Tbx2b is required for ultraviolet photoreceptor cell specification during zebrafish retinal development. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2023–2028
160. Raymond, P.A. *et al.* (2014) Patterning the cone mosaic array in zebrafish retina requires specification of ultraviolet-sensitive cones. *PLoS ONE* 9, e85325
161. Duval, M.G. *et al.* (2014) gdf6a is required for cone photoreceptor subtype differentiation and for the actions of tbx2b in determining rod versus cone photoreceptor fate. *PLoS ONE* 9, e92991
162. Sowden, J.C. *et al.* (2001) Expression of *Drosophila* omb-related T-box genes in the developing human and mouse neural retina. *Invest. Ophthalmol. Vis. Sci.* 42, 3095–3102
163. Gibson-Brown, J.J. *et al.* (1998) Expression of T-box genes Tbx2–Tbx5 during chick organogenesis. *Mech. Dev.* 74, 165–169
164. Kelley, M.W. *et al.* (1995) Ligands of steroid/thyroid receptors induce cone photoreceptors in vertebrate retina. *Development* 121, 3777–3785
165. Hoover, F. *et al.* (1998) Retinoid X receptor gamma gene transcripts are expressed by a subset of early generated retinal cells and eventually restricted to photoreceptors. *J. Comp. Neurol.* 391, 204–213
166. Roberts, M.R. *et al.* (2005) Retinoid X receptor (gamma) is necessary to establish the S-opsin gradient in cone photoreceptors of the developing mouse retina. *Invest. Ophthalmol. Vis. Sci.* 46, 2897–2904
167. Mori, M. *et al.* (2001) Systematic immunolocalization of retinoid receptors in developing and adult mouse eyes. *Invest. Ophthalmol. Vis. Sci.* 42, 1312–1318
168. Sjoberg, M. *et al.* (1992) Thyroid hormone receptors in chick retinal development: differential expression of mRNAs for alpha and N-terminal variant beta receptors. *Development* 114, 39–47
169. Trimarchi, J.M. *et al.* (2008) Thyroid hormone components are expressed in three sequential waves during development of the chick retina. *BMC Dev. Biol.* 8, 101
170. Roberts, M.R. *et al.* (2006) Making the gradient: thyroid hormone regulates cone opsin expression in the developing mouse retina. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6218–6223
171. Shibusawa, N. *et al.* (2003) Thyroid hormone action in the absence of thyroid hormone receptor DNA-binding in vivo. *J. Clin. Invest.* 112, 588–597
172. Applebury, M.L. *et al.* (2007) Transient expression of thyroid hormone nuclear receptor TRbeta2 sets S opsin patterning during cone photoreceptor genesis. *Dev. Dynam.* 236, 1203–1212
173. Cakir, M. *et al.* (2015) The effect of hypothyroidism on color contrast sensitivity: a prospective study. *Eur. Thyroid J.* 4, 43–47
174. Liu, Y. *et al.* (2007) Identification of novel retinal target genes of thyroid hormone in the human WERI cells by expression microarray analysis. *Vision Res.* 47, 2314–2326
175. Weiss, A.H. *et al.* (2012) Reduced L- and M- and increased S-cone functions in an infant with thyroid hormone resistance due to mutations in the THRbeta2 gene. *Ophthalmic Genet.* 33, 187–195
176. Zhou, S. *et al.* (2015) Differentiation of human embryonic stem cells into cone photoreceptors through simultaneous inhibition of BMP, TGFbeta and Wnt signaling. *Development* 142, 3294–3306
177. Yanagi, Y. *et al.* (2002) Distinct functions of photoreceptor cell-specific nuclear receptor, thyroid hormone receptor beta2 and CRX in one photoreceptor development. *Invest. Ophthalmol. Vis. Sci.* 43, 3489–3494
178. Ng, L. *et al.* (2001) A thyroid hormone receptor that is required for the development of green cone photoreceptors. *Nat. Genet.* 27, 94–98
179. Ng, L. *et al.* (2011) Two transcription factors can direct three photoreceptor outcomes from rod precursor cells in mouse retinal development. *J. Neurosci.* 31, 11118–11125
180. Wernet, M.F. *et al.* (2006) Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* 440, 174–180
181. Tahayato, A. *et al.* (2003) Otd/Crx, a dual regulator for the specification of ommatidia subtypes in the *Drosophila* retina. *Dev. Cell* 5, 391–402
182. de Melo, J. *et al.* (2011) The Spalt family transcription factor Sall3 regulates the development of cone photoreceptors and retinal horizontal interneurons. *Development* 138, 2325–2336
183. Mikeladze-Dvali, T. *et al.* (2005) The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in *Drosophila* R8 photoreceptors. *Cell* 122, 775–787
184. Jukam, D. and Desplan, C. (2011) Binary regulation of Hippo pathway by Merlin/NF2, Kibra, Lgl, and Melted specifies and maintains postmitotic neuronal fate. *Dev. Cell* 21, 874–887
185. Thanawala, S.U. *et al.* (2013) Regional modulation of a stochastically expressed factor determines photoreceptor subtypes in the *Drosophila* retina. *Dev. Cell* 25, 93–105
186. Johnston, R.J., Jr (2013) Lessons about terminal differentiation from the specification of color-detecting photoreceptors in the Drosophila retina. *Ann. N.Y. Acad. Sci.* 1293, 33–44
187. Suzuki, S.C. *et al.* (2013) Cone photoreceptor types in zebrafish are generated by symmetric terminal divisions of dedicated precursors. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15109–15114

188. Prabhudesai, S.N. *et al.* (2005) Targeted effects of retinoic acid signaling upon photoreceptor development in zebrafish. *Dev. Biol.* 287, 157–167

189. Mitchell, D.M. *et al.* (2015) Retinoic acid signaling regulates differential expression of the tandemly-duplicated long wavelength-sensitive cone opsin genes in zebrafish. *PLoS Genet.* 11, e1005483

190. Satoh, S. *et al.* (2009) The spatial patterning of mouse cone opsin expression is regulated by bone morphogenetic protein signaling through downstream effector COUP-TF nuclear receptors. *J. Neurosci.* 29, 12401–12411

191. Ogawa, Y. *et al.* (2015) Homeobox transcription factor Six7 governs expression of green opsin genes in zebrafish. *Proc. Biol. Sci.* 282, 20150659

192. Kawakami, Y. *et al.* (2009) Sall genes regulate region-specific morphogenesis in the mouse limb by modulating Hox activities. *Development* 136, 585–594

193. Srinivas, M. *et al.* (2006) Activation of the blue opsin gene in cone photoreceptor development by retinoid-related orphan receptor beta. *Mol. Endocrinol.* 20, 1728–1741

194. Tsujimura, T. *et al.* (2007) Identification of a locus control region for quadruplicated green-sensitive opsin genes in zebrafish. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12813–12818

195. Tsujimura, T. *et al.* (2010) A single enhancer regulating the differential expression of duplicated red-sensitive opsin genes in zebrafish. *PLoS Genet.* 6, e1001245

196. Tsujimura, T. *et al.* (2015) Spatially differentiated expression of quadruplicated green-sensitive RH2 opsin genes in zebrafish is determined by proximal regulatory regions and gene order to the locus control region. *BMC Genet.* 16, 130

197. Wang, Y. *et al.* (1992) A locus control region adjacent to the human red and green visual pigment genes. *Neuron* 9, 429–440

198. Legent, K. and Treisman, J.E. (2008) Wingless signaling in *Drosophila* eye development. *Methods Mol. Biol.* 469, 141–161

199. Hernandez, R.E. *et al.* (2007) Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development* 134, 177–187

200. Mey, J. *et al.* (1997) Retinoic acid synthesis in the developing chick retina. *J. Neurosci.* 17, 7441–7449

201. McCaffrery, P. *et al.* (1993) Changing patterns of the retinoic acid system in the developing retina. *Dev. Biol.* 158, 390–399

202. Perz-Edwards, A. *et al.* (2001) Retinoic acid-mediated gene expression in transgenic reporter zebrafish. *Dev. Biol.* 229, 89–101

203. McCaffery, P. *et al.* (1999) Dorsal and ventral retinal territories defined by retinoic acid synthesis, break-down and nuclear receptor expression. *Mech. Dev.* 82, 119–130

204. McCaffery, P. *et al.* (1992) Asymmetrical retinoic acid synthesis in the dorsoventral axis of the retina. *Development* 115, 371–382

205. McCaffery, P. *et al.* (1999) Dorsal and ventral rentinoic territories defined by retinoic acid synthesis, break-down and nuclear receptor expression. *Mech. Dev.* 85, 203–214

206. Nicotra, C.M. *et al.* (1994) Retinoid dynamics in chicken eye during pre- and postnatal development. *Mol. Cell. Biochem.* 132, 45–55

207. Stenkamp, D.L. *et al.* (1993) Retinoid effects in purified cultures of chick embryo retina neurons and photoreceptors. *Invest. Ophthalmol. Vis. Sci.* 34, 2425–2436

208. Sakai, Y. *et al.* (2004) CYP26A1 and CYP26C1 cooperate in degrading retinoic acid within the equatorial retina during later eye development. *Dev. Biol.* 276, 143–157

209. Niederreither, K. *et al.* (1997) Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. *Mech. Dev.* 62, 67–78

210. Corbo, J.C. *et al.* (2007) A typology of photoreceptor gene expression patterns in the mouse. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12069–12074

211. Ng, L. *et al.* (2010) Type 3 deiodinase, a thyroid-hormone-inactivating enzyme, controls survival and maturation of cone photoreceptors. *J. Neurosci.* 30, 3347–3357

212. Bagci, E. *et al.* (2015) Deiodinase knockdown during early zebrafish development affects growth, development, energy metabolism, motility and phototransduction. *PLoS ONE* 10, e0123285

213. Thisse, C. *et al.* (2003) Spatial and temporal expression patterns of selenoprotein genes during embryogenesis in zebrafish. *Gene Expr. Patterns* 3, 525–532

214. Guo, C. *et al.* (2014) Intrinsic expression of a multiexon type 3 deiodinase gene controls zebrafish embryo size. *Endocrinology* 155, 4069–4080

215. Bedolla, D.E. and Torre, V. (2011) A component of retinal light adaptation mediated by the thyroid hormone cascade. *PLoS ONE* 6, e26334

216. Dentice, M. *et al.* (2013) The deiodinases and the control of intracellular thyroid hormone signaling during cellular differentiation. *Biochim. Biophys. Acta* 1830, 3937–3945

217. Glaschke, A. *et al.* (2011) Thyroid hormone controls cone opsin expression in the retina of adult rodents. *J. Neurosci.* 31, 4844–4851

218. Inoue, M. *et al.* (2010) COUP-TFI and -TFII nuclear receptors are expressed in amacrine cells and play roles in regulating the differentiation of retinal progenitor cells. *Exp. Eye Res.* 90, 49–56

219. Ready, D.F. *et al.* (1976) Development of the *Drosophila* retina, a neurocrystalline lattice. *Dev. Biol.* 53, 217–240

220. Ma, C. *et al.* (1993) The segment polarity gene hedgehog is required for progression of the morphogenetic furrow in the developing *Drosophila* eye. *Cell* 75, 927–938

221. Heberlein, U. *et al.* (1993) The TGF beta homolog dpp and the segment polarity gene hedgehog are required for propagation of a morphogenetic wave in the *Drosophila* retina. *Cell* 75, 913–926

222. Jarman, A.P. *et al.* (1994) Atonal is the proneural gene for *Drosophila* photoreceptors. *Nature* 369, 398–400

223. Jarman, A.P. *et al.* (1995) Role of the proneural gene, atonal, in formation of *Drosophila* chordotonal organs and photoreceptors. *Development* 121, 2019–2030

224. Treisman, J.E. (2013) Retinal differentiation in *Drosophila*. *Wiley Interdiscip. Rev. Dev. Biol.* 2, 545–557

225. Quan, X.J. *et al.* (2012) Transcriptional control of cell fate specification: lessons from the fly retina. *Curr. Top. Dev. Biol.* 98, 259–276

226. Tomlinson, A. and Ready, D.F. (1987) Neuronal differentiation in *Drosophila* ommatidium. *Dev. Biol.* 120, 366–376

227. Baonza, A. *et al.* (2001) A primary role for the epidermal growth factor receptor in ommatidial spacing in the *Drosophila* eye. *Curr. Biol.* 11, 396–404

228. Freeman, M. (1994) The spitz gene is required for photoreceptor determination in the *Drosophila* eye where it interacts with the EGF receptor. *Mech. Dev.* 48, 25–33

229. Freeman, M. (1996) Reiterative use of the EGF receptor triggers differentiation of all cell types in the *Drosophila* eye. *Cell* 87, 651–660

230. Tio, M. *et al.* (1994) Spitz, a *Drosophila* homolog of transforming growth factor-alpha, is required in the founding photoreceptor cells of the compound eye facets. *Mech. Dev.* 48, 13–23

231. Masai, I. *et al.* (2000) Midline signals regulate retinal neurogenesis in zebrafish. *Neuron* 27, 251–263

232. Neumann, C.J. and Nuesslein-Volhard, C. (2000) Patterning of the zebrafish retina by a wave of sonic hedgehog activity. *Science* 289, 2137–2139

233. Raymond, P.A. *et al.* (1995) Developmental patterning of rod and cone photoreceptors in embryonic zebrafish. *J. Comp. Neurol.* 395, 537–550

234. Schmitt, E.A. and Dowling, J.E. (1996) Comparison of topographical patterns of ganglion and photoreceptor cell differentiation in the retina of the zebrafish. *Danio rerio*. *J. Comp. Neurol.* 371, 222–234

235. Hu, M. and Easter, S.S. (1999) Retinal neurogenesis: the formation of the initial central patch of postmitotic cells. *Dev. Biol.* 207, 309–321

236. Johns, P.R. and Fernald, R.D. (1981) Genesis of rods in teleost fish retina. *Nature* 293, 141–142

237. Carter-Dawson, L.D. and LaVail, M.M. (1979) Rods and cones in the mouse retina. II. Autoradiographic analysis of cell generation using tritiated thymidine. *J. Comp. Neurol.* 188, 263–272

238. Young, R.W. (1985) Cell differentiation in the retina of the mouse. *Anat. Rec.* 212, 199–205

239. Dinet, V. *et al.* (2011) APP involvement in retinogenesis of mice. *Acta Neuropathol.* 121, 351–363

240. La Vail, M.M. *et al.* (1991) Cytogenesis in the monkey retina. *J. Comp. Neurol.* 309, 86–114

241. Cornish, E.E. *et al.* (2005) Gradients of cone differentiation and FGF expression during development of the foveal depression in macaque retina. *Visual Neurosci.* 22, 447–459

242. Hendrickson, A. *et al.* (2008) Rod photoreceptor differentiation in fetal and infant human retina. *Exp. Eye Res.* 87, 415–426

243. Xiao, M. and Hendrickson, A. (2000) Spatial and temporal expression of short, long/medium, or both opsins in human fetal cones. *J. Comp. Neurol.* 425, 545–559

244. Diaz-Araya, C. and Provis, J.M. (1992) Evidence of photoreceptor migration during early foveal development: a quantitative analysis of human fetal retinae. *Visual Neurosci.* 8, 505–514

245. Packer, O. *et al.* (1990) Development redistribution of photoreceptors across the *Macaca nemestrina* (pigtail macaque) retina. *J. Comp. Neurol.* 298, 472–493

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

247. Trezise, A.E. and Collin, S.P. (2005) Opsins: evolution in waiting. *Curr. Biol.* 15, R794–R796

248. Bowmaker, J.K. (2008) Evolution of vertebrate visual pigments. *Vision Res.* 48, 2022–2041

249. Hofmann, C.M. and Carleton, K.L. (2009) Gene duplication and differential gene expression play an important role in the diversification of visual pigments in fish. *Integr. Comp. Biol.* 49, 630–643

250. Brainard, D.H. *et al.* (2000) Functional consequences of the relative numbers of L and M cones. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 17, 607–614

251. Nathans, J. *et al.* (1989) Molecular genetics of human blue cone monochromacy. *Science* 245, 831–838

252. Smallwood, P.M. *et al.* (2002) Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1008–1011

253. Peng, G.H. and Chen, S. (2011) Active opsin loci adopt intrachromosomal loops that depend on the photoreceptor transcription factor network. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17821–17826

254. Hoon, M. *et al.* (2014) Functional architecture of the retina: development and disease. *Prog. Retin. Eye Res.* 42, 44–84

255. Santiago Ramon, C. (2000) *Texture of the Nervous System of Man and the Vertebrates (Vol. II and III),* Springer-Verlag/Wein

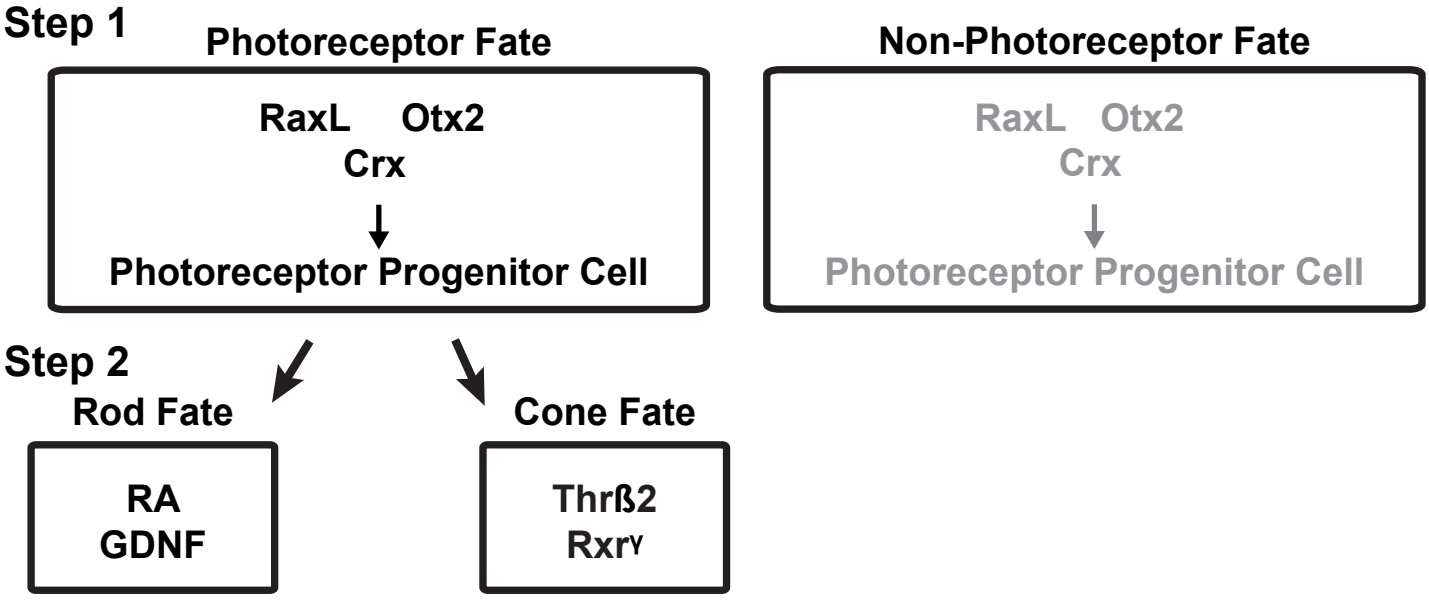
2.9: **Author Contributions**

The author list for this paper was as follows: Kayla Viets, Kiara C. Eldred, and Robert J. Johnston Jr. I was a co-first author with Kiara C. Eldred for this review. We were jointly responsible for writing the text of the article. I designed the *Drosophila melanogaster* and *Danio rerio* panels for Figures 1-5, and K.E. designed the *Gallus gallus domesticus, Mus musculus,* and *Homo sapiens* panels. I was the main author responsible for editing of the text. Robert J. Johnston was corresponding author and was responsible for generating ideas and editing the text.

Supplemental Figure 1

A

Gallus gallus domesticus



Supplemental Figure 1: The gene-regulatory network controlling PR specification in *Gallus gallus domesticus*. The gene-regulatory network has been simplified to emphasize PR factors that are conserved between species. Arrows within gene networks solely represent our current understanding of network relationships and do not imply genetic mechanisms such as direct or indirect transcriptional regulation.

Chapter 3: TADs pair homologous chromosomes to promote interchromosomal gene regulation

3.1: Abstract

Homologous chromosomes colocalize to regulate gene expression in processes including genomic imprinting and X-inactivation, but the mechanisms driving these interactions are poorly understood. In *Drosophila*, homologous chromosomes pair throughout development, promoting an interchromosomal gene regulatory mechanism called transvection. Despite over a century of study, the molecular features that facilitate chromosome-wide pairing are unknown. The "button" model of pairing proposes that specific regions along chromosomes pair with a higher affinity than their surrounding regions, but only a handful of DNA elements that drive homologous pairing between chromosomes have been described. Here, we identify button loci interspersed across the fly genome that have the ability to pair with their homologous sequences. Buttons are characterized by topologically associated domains (TADs), which drive pairing with their endogenous loci from multiple locations in the genome. Fragments of TADs do not pair, suggesting a model in which combinations of elements interspersed along the full length of a TAD are required for pairing. Though DNA-binding insulator proteins are not associated with pairing, buttons are enriched for insulator cofactors, suggesting that these proteins may mediate higher order interactions between homologous TADs. Using a TAD spanning the *spineless* gene as a paradigm, we find that pairing is necessary but not sufficient for transvection. *spineless* pairing and transvection are cell-type-specific, suggesting

that local buttoning and unbuttoning regulates transvection efficiency between cell types. Together, our data support a model in which specialized TADs button homologous chromosomes together to facilitate cell-type-specific interchromosomal gene regulation.

3.2: Introduction

Chromosomes are organized in a complex manner in the nucleus. In higher organisms, they localize to distinct territories (*142*). Regions of chromosomes interact to form compartments, which are segregated based on gene expression states (*143*). Chromosomes are further organized into TADs, regions of self-association that are hypothesized to isolate genes into regulatory domains and ensure their activation by the correct *cis*-regulatory elements (*143*). TADs vary in size from ~100 kb in *Drosophila melanogaster* to ~1 Mb in mammals *(11, 143)*. Disruptions of nuclear organization, such as alteration of TAD structure and localization of genes to incorrect nuclear compartments, have major effects on gene expression (*13, 19, 20, 34*). However, it is unclear how elements within the genome interact between chromosomes to organize chromatin and regulate gene expression.

One key aspect of nuclear architecture involves the tight colocalization, or "pairing," of homologous chromosomes to facilitate regulatory interactions between different alleles of the same gene (*41*). In *Drosophila melanogaster*, homologous chromosomes are paired throughout interphase in somatic cells

(*54*). This stable pairing provides an excellent paradigm to study the mechanisms driving interactions between chromosomes.

Despite over a century of study, it is poorly understood how homologous chromosomes come into close physical proximity. Classical studies of pairing proposed a “zipper” model, in which all regions of the genome have an equal ability to pair based on sequence homology, and chromosome pairing begins at the centromere and proceeds distally to the telomeres (*55, 56*). Studies of chromosome pairing initiation during development led to a shift in thinking towards the “button” model, which proposes that regions of high pairing affinity are interspersed along chromosome arms and come together through a random walk to initiate pairing (**Fig. 1A**) (*57-59*).

The nature of the high-affinity “buttons” that bring homologous chromosomes together is still unclear. Many elements, including insulators, Polycomb Response Elements (PREs), and heterochromatic repeats, drive looping interactions in *cis* along the same chromosome arm or clustering between non-homologous sequences on different chromosomes (*65, 69, 74, 102, 103, 144*). However, only a handful of small DNA elements, including the gypsy retrotransposon and the Fab-7, Mcp, and TMR regions of the *Abd-B* locus, are known to drive pairing between identical homologous sequences on different chromosomes (*65, 67-71*). As three of these elements are within the same locus, the sequence and structural features that contribute to genome-wide homologous chromosome pairing are unknown. The scarcity of small DNA elements that are known to drive interchromosomal homologous pairing suggests that

combinations of elements and/or higher order chromatin structures are required to button homologous chromosomes together.

Pairing of homologous chromosomes facilitates a gene-regulatory mechanism known as transvection, in which two different mutant alleles interact between chromosomes to rescue gene expression (**Fig. 1B**) (*55*). Transvection has been described for a number of *Drosophila* genes (*90*). With the exceptions of *Abd-B* and certain transgenes containing the Homie, gypsy, Mcp, TMR, and Fab7 sequences, transvection requires homologous chromosome pairing and is disrupted by chromosome rearrangements (*55, 59, 67, 69, 71, 90, 91, 96, 102, 127, 145-147*). DNA elements such as insulators and PREs contribute to transvection and similar phenomena at many loci across the genome (*67, 69, 71, 102, 103, 109-115, 127, 145-147*), but it is unclear if the same DNA elements are always involved in both homologous chromosome pairing and transvection or if pairing and transvection are mechanistically separable.

Homologous chromosome pairing occurs more strongly in some cell types than in others. Pairing occurs in 15-30% of nuclei in the early embryo, gradually increases throughout embryonic development, and reaches a peak of 90-95% by the third instar larval stage (*57-59, 74*). Similarly, transvection efficiency varies widely between cell types (*109, 140, 141, 148*). However, a direct link between the level of pairing in a given cell type and the strength of transvection in that cell type has not been established.

Here, we develop a method to screen for DNA elements that pair and identify multiple button sites interspersed across the *Drosophila* genome,

allowing us to examine the features that determine button activity. We find that a subset of TADs are associated with buttons and can drive pairing from different positions in the genome. By testing mutant alleles and transgenes of the *spineless* gene, we find that pairing and transvection are mechanistically separable and cell-type-specific. Together, our data suggest that TADs are one key feature of buttons that drive homologous chromosome pairing to promote cell-type-specific interchromosomal gene regulation.

3.3: Results

Identification of new button loci interspersed across the fly genome

Only a few elements are known to drive homologous pairing, and a majority of these elements are located within the *Abd-B* locus (*65, 67-71*), limiting the identification of general features that drive pairing throughout the genome. To look more broadly for elements that bring homologous chromosomes together, we selected transgenes from multiple locations in the genome, inserted them on heterologous chromosomes, and tested if they drove pairing with their endogenous loci. This assay provided a sensitized system for identifying regions of the genome that have an especially high affinity for homologous sequences, since it detected DNA sequences that pair outside of the context of chromosome-wide pairing.

We first screened a set of ~80-110 kb transgenes tiling a ~1 Mb region on chromosome 3R (**Fig. 2E**). We inserted individual transgenes into a site on chromosome 2L (site 1; **Fig. 2A**) and visualized their nuclear position using

Oligopaints DNA FISH (*149*). As the endogenous and transgenic sequences were identical, we distinguished between them by labeling the sequence neighboring the endogenous locus with red probes and the sequence neighboring the transgene insertion site with green probes ("2 color strategy")(**Fig. 2A**). We examined pairing in post-mitotic larval photoreceptors to avoid disruptions caused by cell division.

To determine whether a transgene drove pairing, we compared the 3-D distance between the transgene and its endogenous site to the 3-D distance of a negative control. If the distance between the insertion site and the endogenous site was significantly lower with a transgene present than in the negative control, then the transgene drove pairing (**Fig. 2A**). The negative control measured the distance between the transgene insertion site on chromosome 2L and a site on chromosome 3R in a wild type background. Since different loci along a chromosome arm might vary in their distance from other chromosomes, we examined additional negative controls in multiple cell types and observed no statistical differences (**Fig. S1A**). Thus, for all transgene pairing experiments described in this paper, we used single negative controls for each chromosome arm tested.

To determine the statistical significance of transgene pairing relative to a control, we first tested each transgene dataset for a Gaussian distribution. In instances where transgene distance distributions were non-Gaussian, we tested for statistical significance by comparing the median of each transgene distribution to the negative control (**Fig. 2F**). Gaussian transgene distance distributions were

tested for significance using a t-test to compare means. In all cases where more than one transgene was compared to a control, we corrected for multiple comparisons.

Only a subset (5/17) of transgenes ("pairers") in our initial screen drove pairing between chromosomes 2L and 3R in our initial screen, bringing the distances between the red and green FISH signals significantly closer than in the negative control (**Fig. 2B-C, F; Fig. S2A**). The red and green signals did not completely overlap, likely because they did not directly label the paired sites (**Fig. S3A-C**). For the remaining 12/17 transgenes ("non-pairers"), the distances between the red and green signals were not significantly different from the negative control, indicating that they did not drive pairing (**Fig. 2B, D, F; Fig. S2B**).

To confirm our assignment of pairers vs. non-pairers, we used maximum likelihood estimation to fit our transgene data to single or double Gaussian distributions. 4/5 pairers fit a double Gaussian distribution, indicating a split between paired and unpaired populations of nuclei. The remaining pairer, *Transgene E*, fit a single Gaussian distribution, but the mean of the *Transgene E* distribution was significantly lower than the mean of the negative control distribution, indicating a high degree of pairing (**Fig. S1B**). Conversely, only 2/12 non-pairers (*Transgenes L* and *N*) fit a double Gaussian distribution. Since the median distances for *Transgenes L* and *N* were not significantly different from the negative control (**Fig. 2F**), these transgenes may drive "weak" pairing at a level significantly lower than the other pairers we identified. The remaining 10/12 non-

pairers fit a single Gaussian distribution, and their means did not significantly differ from the negative control (**Fig. S1B**). Thus, our screen identified multiple new button elements.

The pairing observed between transgenes on chromosome 2L and endogenous sequences on chromosome 3R could be affected by the transgene insertion site. To test the position independence of button pairing, we inserted *Transgene E* onto chromosome 3L (site 3; **Fig. S4A**) and found that it paired with its homologous endogenous locus on chromosome 3R (**Fig. S4B-D**), showing that buttons can drive pairing from different sites in the genome.

Thus, we identified multiple button loci along chromosome 3R that overcome endogenous nuclear architecture to drive pairing between non-homologous chromosomes.

TADs are features of buttons

We next sought to determine common features between the new buttons we identified, focusing first on chromatin structure. We examined 14 publicly available Hi-C datasets to determine the relationship between buttons and topologically associated domains (TADs), genomic regions of self-association. We defined TADs using directionality indices, which measure the bias of a genomic region towards upstream or downstream interactions along the chromosome (*10*). TADs on a directionality index are read from the beginning of a positive peak, which indicates downstream interactions, to the end of a

negative peak, which indicates upstream interactions (**Fig. S5A; Fig. S6A-E; Fig. S7A-E**).

We found that 60% of pairers encompassed a complete TAD, including both TAD boundaries, compared to only 8% of non-pairers (**Fig. 2F; Fig. S5A; Fig. S8A-B**), suggesting that specific TADs contribute to button function. To test the hypothesis that TADs are a feature of buttons, we selected six transgenes encompassing entire TADs on chromosomes X, 2L, 2R, and 3R (**Fig. 3E; Fig. S6A-E; Fig. S7C; Fig. S8A**) and compared them to four transgenes that did not encompass entire TADs, taken from chromosomes X, 2L, and 3R (**Fig. 3E; Fig. S2E; Fig. S7A-B, D-E; Fig. S8B**). Based on the availability of Oligopaints probes, we used an alternative FISH strategy for a subset of these transgenes, in which the identical transgene and endogenous sequences were labeled with the same red fluorescent probes (**Fig. 3A**). With this 1-color strategy, FISH punctae $\leq$0.4 μm apart could not be distinguished as separate and were assigned a distance of 0.4 μm apart (see Materials and Methods).

5/6 transgenes that spanned entire TADs drove pairing, while 0/4 transgenes that did not span a TAD drove pairing (**Fig. 3B-E; Fig. S2C-E**), further supporting the importance of TADs for button activity. In total for all transgenes tested in **Fig. 2F**, **Fig. 3E**, and **Fig. S2E**, 80% of pairers spanned a TAD (8/10) while only 12% of non-pairers spanned a TAD (2/17) (**Fig 3F**), indicating that specific TADs contribute to button activity and drive pairing. Transgenes of near identical lengths had different pairing abilities, further

suggesting that the content of a transgene (i.e. TADs) determines pairing (**Fig. 3G**).

The ~80-110 kb size limitation of publicly available transgenes prevented testing larger TADs for pairing with our transgene assay. Transgenes that covered only parts of a large TAD on chromosome 3R did not drive pairing (**Fig. S9A**). To test this large TAD for pairing, we utilized a 460-kb duplication of chromosome 3R onto chromosome 2R (**Fig. S9B**), which encompassed the entire TAD (**Fig. S5A; Fig. S9A**). We found that the duplication drove pairing with its homologous endogenous site (**Fig. S9C-E**), further supporting a role for TADs in pairing.

Entire TADs could drive pairing, or smaller elements contained within TADs could bring homologous regions together. We therefore examined the effects of "splitting" a TAD, focusing on the TAD spanned by *Transgene E*. *Transgene D,* which covers the 5' end of *Transgene E*, did not drive pairing (**Fig. 3H-I; Fig. S2B**). *Transgenes G* and *S*, which cover the 3' end of *Transgene E*, also did not drive pairing (**Fig. 3H-I; Fig. S2B**). Because there is a ~19 kb gap between *Transgenes D* and *G*, we hypothesized that this 19 kb region might contain an element required for pairing. However, *Transgenes F* and *R*, which both contained this 19 kb region, did not drive pairing. Thus, on their own, the 5', middle, or 3' regions of the *Transgene E* TAD are not sufficient to drive pairing. Together, these observations suggest that a combination of elements interspersed across the complete *Transgene E* TAD drive pairing.

Examining the relationship between pairing, chromatin factors, and gene expression state

We next examined other features that could contribute to homologous chromosome pairing. As insulators have been linked to long-distance chromosome interactions(*65, 69, 74, 102, 103, 144, 150*), we tested whether the number of binding sites for individual *Drosophila* insulator proteins was higher in pairers than in non-pairers. Using previously published ChIP-chip and ChIP-seq data(*151-155*), we found no association between pairing and any of the DNA-binding insulator proteins (BEAF, Su(Hw), CTCF, and GAF)(**Fig. 4A-D**), suggesting that complex combinations of bound insulators, rather than individual proteins, might contribute to homologous pairing or stabilize interactions between already paired TADs(*104*). Intriguingly, both Cp190 and Mod(mdg4), insulator co-factors that do not directly bind to DNA, were associated with pairing (**Fig. 4E-F**). These proteins may mediate interactions between complex clusters of DNA-binding insulator proteins or play additional roles independent of insulator function to bring homologous chromosomes together.

We next investigated the relationship between pairing and insulator clustering, defining a cluster as any region bound by ≥4 unique insulator proteins, and observed no association with pairing (**Fig. 4G-H**). As *Drosophila* TAD boundaries are enriched for CTCF, BEAF, and Cp190 (*11*), we examined the association between pairers and clusters of these insulators. Again, we observed no relationship with pairing (**Fig. 4I**), suggesting that a complete TAD, rather than the insulators clustered at its boundaries, is needed for pairing. Thus, we do not

find a strong relationship between insulators and pairing. However, it is possible that even larger clusters of insulators, in combination with other architectural proteins, mediate pairing between TADs, or that each button has a unique "code" of insulators bound across the entire TAD region that assist in driving pairing.

Because the small DNA elements Mcp, Fab-7, and TMR drive pairing between homologous chromosomes (*65, 67, 69-71*), we next examined whether these elements drove pairing in our assay. Unexpectedly, *Transgene M*, which contained Mcp, and *Transgene N*, which contained Fab-7 and TMR, were both non-pairers (**Fig. 2E-F; Fig. S2B**), suggesting that pairing driven by these elements may be context-specific, or that these elements drive pairing at a low level that cannot be distinguished by this assay.

To investigate additional elements that contribute to pairing, we examined modENCODE ChIP data and found no association between pairing and Polycomb Group (PcG) binding sites, repressing epigenetic marks, or non-coding RNAs (ncRNAs)(**Fig. S10A-F**).

Gene activity plays a critical role in nuclear architecture: individual TADs interact with each other in compartments, which are partitioned by expression state(*143*). The pairing we observe between transgenes and their endogenous loci might simply be a result of segregation of the genome into active (A) and repressed (B) compartments. We hypothesized that if compartmentalization alone drives pairing, then the distance between any two active regions or any two repressed regions should be less than the distance between an active region and a repressed region. To test this hypothesis, we performed RNA-seq on larval eye

discs, the tissue we used in our pairing experiments, to identify active or repressed regions. We selected two loci on different chromosomes that were highly expressed (A1 and A2; **Fig. S11A, C**), and two loci that were expressed at low levels (B1 and B2; **Fig. S11B-C**). The level of A1-A2 and B1-B2 interaction did not differ from a negative control or from the level of A1-B2 interactions (**Fig. S11D-H**). Moreover, we found no association between pairing and active transcription (**Fig. S11I-J**). Our data suggest that pairing is driven by specific interactions between homologous TADs, rather than general interactions based on expression state alone.

Pairing and transvection occur despite chromosomal rearrangements

We next interrogated the relationship between pairing and the gene regulatory process of transvection. Chromosomal rearrangements have been shown to disrupt pairing of genes located near rearrangement breakpoints (*55, 90*). However, we observed pairing of ~100 kb transgenes with their endogenous loci, suggesting that intact homologous chromosomes are not required for pairing and that pairing driven by TADs tolerates nearby breakpoints. Supporting our observations, the *Abd-B* locus pairs in the presence of chromosomal rearrangements (*59, 91, 96*). We therefore reexamined how rearrangements affect pairing, focusing on a button defined by a TAD spanning the *spineless (ss)* locus ("*ss* button"; **Fig. 5A**).

To assess the effects of local rearrangements on *ss* button pairing, we examined a naturally occurring chromosomal inversion with a breakpoint

immediately upstream of *ss* (*ss*^{inversion}) and a duplication with a breakpoint immediately downstream of *ss* (**Fig. 5E**). Both *ss*^{inversion} and the duplication paired with endogenous *ss* (**Fig. S9C-E; Fig. S12A-B**), showing that *ss* button pairing occurs despite chromosomal rearrangements. Consistent with these findings, pairing also occurred at the *ss* locus in flies with balancer chromosomes containing numerous large inversions and rearrangements (**Fig. S12F-J**). Thus, similar to *Abd-B*, pairing of *ss* occurs despite chromosomal rearrangements, consistent with a model in which homologous TADs find each other in the nucleus independent of chromosome-wide homology.

Pairing is required for the genetic phenomenon of transvection, in which DNA elements on a mutant allele of a gene act between chromosomes to rescue expression of a different mutant allele (**Fig. 1B**). In cases where chromosomal rearrangements perturb pairing, transvection is also disrupted (*55, 90*). Since chromosomal rearrangements did not ablate pairing at the *ss* button, we hypothesized that transvection would occur at the *ss* locus in these genetic conditions.

In the fly eye, Ss is normally expressed in ~70% of R7 photoreceptors to activate expression of Rhodopsin 4 (Rh4) and repress Rhodopsin 3 (Rh3; **Fig. 5B-D**). Ss is absent in the remaining 30% of R7s, allowing Rh3 expression (**Fig. 5B-D**) (*99*). Regulatory mutations in the *ss* gene cause decreases or increases in the ratio of Ss^{ON}: Ss^{OFF} cells. When two *ss* alleles with different ratios are heterozygous, transvection between chromosomes (also known as Interchromosomal Communication) determines the final ratio of Ss^{ON}: Ss^{OFF} R7s

(*97*). Thus, the Ss^{ON}: Ss^{OFF} ratio is a phenotype that allows for quantitative assessment of transvection. Throughout our *ss* transvection experiments, we evaluated Rh3 and Rh4 expression, as they faithfully report Ss expression in R7s (i.e. Ss^{ON} = Rh4; Ss^{OFF} = Rh3). We previously observed transvection at the *ss* locus for the duplication and balancer chromosome alleles (*97*). We similarly observed transvection at the *ss* locus for the $ss^{inversion}$ allele (**Fig. S12C-E**). Together, these data suggested that buttons can drive pairing and transvection despite chromosomal rearrangements.

Pairing is necessary but not sufficient for transvection

As chromosomal rearrangements did not impair *ss* pairing or transvection, we further investigated the relationship between pairing and transvection using *ss* transgenes. Both *Transgene S* and *Transgene T* are expressed in 100% of R7s because they lack a silencer DNA element, but do not produce functional Ss protein because they lack critical coding exons (**Fig. 5E; Fig. S13A-J**)(*97*). *Transgene T* differs from *Transgene S* in that it lacks 6 kb at its 5' end (**Fig. 5E**). We predicted that if *Transgenes S* and *T* performed transvection, they would upregulate expression of endogenous *ss*.

When inserted onto chromosomes 2L or 3L (sites 1 and 3; **Fig. 5F; Fig. S14A**), *Transgenes S* and *T* did not drive pairing with the endogenous *ss* locus on chromosome 3R (**Fig. 5G-I; Fig. S14B-C, E, Q**). At these sites, *Transgenes S* and *T* did not upregulate *ss* expression, indicating that they could not perform transvection when unpaired (**Fig. 5J-L; Fig. S14D, F**).

We next wondered whether *Transgenes S* and *T* could perform transvection if we mimicked pairing by forcing them into close physical proximity with endogenous *ss*. We performed a FISH screen to identify genomic sites that naturally loop to endogenous *ss* (**Fig. 5M**) and identified three such sites, located 4.8 Mb upstream of *ss*, 0.4 Mb upstream of *ss*, and 4.6 Mb downstream of *ss* (sites 2, 4, and 5; **Fig. 5M-O; Fig. S14G-H, M-N, Q**).

When we inserted *Transgene S* at these sites, it was forced into close proximity with endogenous *ss* (**Fig. 5P; Fig. S14I, O, Q**) and upregulated Ss (Rh4) into nearly 100% of R7s (**Fig. 5R-S; Fig. S14J, P**)(*97*). Thus, natural chromosome looping can force loci into proximity and, like pairing, facilitate transvection. In contrast, when we forced *Transgene T* into close proximity with endogenous *ss* (**Fig. 5Q; Fig. S14K, Q**), it did not upregulate Ss (Rh4) expression (**Fig. 5T; Fig. S14L**), indicating that it could not perform transvection even when paired. Thus, pairing is necessary but not sufficient for transvection.

We compared the DNA sequence of *Transgene T*, which does not perform transvection, to *Transgene S,* the $ss^{inversion}$, and the duplication, which perform transvection. An upstream region of ~1.6 kb is present in *Transgene S,* the $ss^{inversion}$, and the duplication, but missing from *Transgene T*, suggesting that this region contains a critical element for transvection (**Fig. 5E**). ModENCODE ChIP data showed that this region was bound by the *Drosophila* insulator proteins CTCF, BEAF, Mod(Mdg4), and Cp190. Additionally, this DNA sequence performed P-element homing (*97*), an indicator of insulator activity. Together,

these data suggested that the DNA element required for transvection is an insulator.

To further test whether this insulator was required for transvection, we examined *Transgene E*, which drove pairing and contained the complete *ss* locus, except for the insulator element (**Fig. 2F; Fig. 5E; Fig. S2A; Fig. S4B-D**). We utilized genetic backgrounds in which *Transgene E* was the only source of Ss protein, so that any changes in Ss (Rh4) expression would indicate transvection effects on *Transgene E*. As a control, we examined *Transgene E* expression when the endogenous *ss* locus was hemizygous for a protein null allele ($ss^{protein\ null}$) that did not perform transvection (**Fig. S15A-B**). In this background, *Transgene E* expressed Ss in 52% of R7s (**Fig. S15A-B**). We next tested *Transgene E* for transvection with a high-frequency protein null allele ($ss^{high\ freq\ null}$), which can perform transvection to increase *ss* expression (*97*). When the endogenous *ss* locus was hemizygous for the $ss^{high\ freq\ null}$, we observed no increase in *Transgene E* expression, indicating that it did not perform transvection (51% Ss (Rh4); **Fig. S15A, C**). Moreover, *Transgene E* did not perform transvection in other genetic conditions (**Fig. S15D-E**). Thus, *Transgene E* paired with the endogenous *ss* locus but failed to perform transvection. These data show that an insulator is required for transvection but not for pairing, indicating that *ss* transvection and pairing are mechanistically separable.

ss pairing and transvection are cell-type-specific

It is poorly understood how pairing impacts transvection in a cell-type-specific manner. We propose two models: constitutive and cell-type-specific buttoning. In the constitutive model, all buttons drive pairing in all cell types, and differences in transvection would occur due to variation in transcription factor binding or chromatin state between cell types (**Fig. 6A**). In the cell-type-specific model, different buttons drive pairing in each cell type, bringing different regions into physical proximity to control transvection efficiency (**Fig. 6A**).

We tested these models by investigating pairing and transvection of *ss* in two different tissues. In addition to its role in R7 photoreceptors, *ss* is required for the development of the arista, a structure on the antenna (**Fig. 6D-E**) (*99, 156*). *Transgene E,* which contains the *ss* button, drove pairing in the eye but not the antenna from two different insertion sites (sites 1 and 3; **Fig. 6B-C; Fig. S4A-D; Fig. S16A-F**), suggesting that button pairing is cell-type-specific.

As pairing is required for transvection and the *ss* button pairs in a cell-type-specific manner, we hypothesized that transvection at the *ss* locus is cell-type-specific. To test this hypothesis, we examined an allele of *ss* that specifically affects arista development ($ss^{arista\ 1}$) (**Fig. 6H; Fig. S17A-F**). In flies transheterozygous for $ss^{arista\ 1}$ and a *ss* deficiency (ss^{def}), aristae were transformed into legs (i.e. aristapedia) (**Fig. 6H-I; Fig. S17A, C**). Aristapedia was also observed for $ss^{protein\ null}$ / ss^{def} flies (**Fig. 6F-G**). In the eye, $ss^{protein\ null}$ performed transvection to rescue *ss* expression (**Fig. S18A-D**). However, the aristapedia mutant phenotype persisted in $ss^{arista\ 1}$ / $ss^{protein\ null}$ flies (**Fig. 6J-K;**

Fig. S17D, F), suggesting that, unlike in the eye, transvection does not rescue *ss* expression in the arista. Cell-type-specific transvection of the *ss* gene in the eye but not the arista was also observed in other genetic conditions (**Fig. S17G-L; S19A-L**).

As *ss* button pairing and transvection are cell-type-specific and pairing is required for transvection, our data support the cell-type-specific model, in which local buttoning and unbuttoning occur in a cell-type-specific manner to determine transvection efficiency (**Fig. 6A**).

3.4: Discussion

Despite the discovery of homologous chromosome pairing in flies over 100 years ago (*54*), the mechanisms that facilitate pairing have remained unclear. We identified multiple button loci interspersed across the genome that drive pairing with their homologous sequences. Specific TADs are responsible for button activity and can pair from multiple locations in the genome. Consistent with our findings, homologous TADs have been observed to interact between paired chromosomes using super resolution microscopy (*73*). Additionally, homologous chromosome pairing initiates at the same embryonic stage as TAD formation (*57-59, 74, 157*), consistent with a model in which TADs drive homologous chromosomes together.

Our data suggest that complete TADs are sufficient to drive pairing, and that in general, smaller DNA elements such as single insulators do not have strong pairing activity. We propose a model in which complex networks of

insulator elements work within the context of a TAD, allowing the domain to take on unique chromatin conformations or bind specific combinations of insulator proteins to create nuclear microcompartments that enable homologous TAD association and pairing.

While TADs are strongly associated with pairing (**Fig. 3F**), a small subset of pairers do not span a TAD (*Transgenes A* and *O*; **Fig. 2F; Fig. S2A; Fig. S5A**), raising the possibility that additional mechanisms work in tandem with TADs to drive homologous chromosomes together(*62, 63*). Additionally, certain transgenes that span TADs do not pair (*Transgenes P* and *BB*; **Fig. 2F; Fig. 3E; Fig. S2B-C; Fig. S5A; Fig. S7C**). These transgenes might span TADs that drive cell-type-specific pairing in non-retinal cell types, similar to *Transgene E*, which drives pairing in photoreceptors but not the antennal disc (**Fig. 6B-C; Fig. S4A-D; Fig. S16A-F**). Alternatively, a subset of TADs may not drive pairing in any cell type.

Our data indicate that pairing and transvection are mechanistically separable: TADs facilitate pairing, while an insulator element facilitates transvection to the endogenous *spineless* locus. Consistent with our findings using endogenous alleles, an insulator is required for transvection but not pairing between transgenes containing the *snail* enhancer and the *eve* promoter (*104*).

We find that the *ss* locus drives pairing and performs transvection in the eye but not in the antenna. Our results support a model in which different buttons drive pairing in different cell types. In this model, local buttoning or unbuttoning at a specific gene determines its transvection efficiency in a given cell type.

Variation in levels of pairing or transvection across cell types has been observed for a number of loci (*58, 59, 141*), suggesting that differences in pairing between cell types may be a general mechanism regulating gene expression.

The mechanisms driving chromosome pairing and transvection have remained a mystery of fly genetics since their initial discoveries by Nettie Stevens and Ed Lewis (*54, 55*). Our results provide strong support for the button model of pairing initiation and offer the first evidence of a general feature, specialized TADs, that drives homologous chromosomes together. Furthermore, we find that pairing is necessary but not sufficient for transvection and that distinct elements are required for these processes. Both pairing and transvection are cell-type-specific, suggesting that tighter pairing in a given cell type enables more efficient transvection in that cell type. Our findings suggest a general mechanism in which TADs drive homologous chromosome pairing and interchromosomal gene regulation across organisms to facilitate processes including X-inactivation and imprinting.

3.5: Materials and Methods

Drosophila lines

Flies were raised on standard cornmeal-molasses-agar medium and grown at 25° C.

Table 1: Genotypes of *Drosophila* lines

Fly line	Full genotype	Source	Figures
wild type control	*yw; +; +* or *yw; pm181>Gal4, UAS>mcd8GFP/CyO; +* or *yw; sp/CyO; +*	*(158)*	2B, F; 3B, E, I; 5G, J, O, R; 6C-E; S1A-B; S2C-E; S4B-D; S9C-E; S11D-H; S12B, F-G, J; S14B, H, N, Q; S16A-B, D, F
Transgene B site 1	*yw; pBac{CH321-38G20}VK00037; +*	*(159)**	2C, E-F; 3F, G; 4A-I; S5A; S8A; S10A-F; S11I-J
Transgene D site 1	*yw; pBac{CH321-25M02}VK00037; +*	*(159)**	2D-F; 3F-I; 4A-I; S1B; S5A; S8B; S10A-F; S11I-J
Transgene A site 1	*yw; pBac{CH321-94A21}VK00037; +*	*(159)**	2E-F, 3F-G; 4A-I; S2A; S5A; S8A; S10A-F; S11I-J
Transgene C site 1	*yw; pBac{CH321-86F17}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S2A; S5A; S8A; S10A-F; S11I-J
Transgene E site 1	*yw; pBac{CH321-28L15}VK00037; +*	*(97, 159)*	2E-F; 3F-I; 4A-I; 5E; 6B-C; S1B; S2A; S5A; S8A; S10A-F; S11I-J
Transgene F site 1	*yw; pBac{CH321-23C04}VK00037; +*	*(159)**	2E-F; 3F-I; 4A-I; S1B; S2B; S5A; S8B; S10A-F; S11I-J
Transgene G site 1	*yw; pBAC{CH321-02A24}VK00037; +*	*(159)**	2E-F; 3F-I; 4A-I; S1B; S2B; S5A; S8B; S10A-F; S11I-J
Transgene H site 1	*yw; pBAC{CH321-92J22}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S1B; S2B; S5A; S8B; S9A; S10A-F; S11I-J
Transgene I site 1	*yw; pBAC{CH321-95F12}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S1B; S2B; S5A; S8B; S9A; S10A-F; S11I-J
Transgene J site 1	*yw; pBAC{CH321-71G17}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S1B; S2B; S5A; S8B; S9A; S10A-F; S11I-J
Transgene K site 1	*yw; pBAC{CH321-50E16}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S1B; S2B; S5A; S8B; S9A; S10A-F; S11I-J
Transgene L site 1	*yw; pBAC{CH321-60D22}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S2B; S5A; S8B; S10A-F; S11I-J
Transgene M site 1	*yw; pBAC{CH321-58G7}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S1B; S2B; S5A; S8B; S10A-F; S11I-J
Transgene N site 1	*yw; pBAC{CH321-96A10}VK00037/+; +*	*(159)**	2E-F; 3F-G; 4A-I; S2B; S5A; S8B; S10A-F; S11I-J
Transgene O site 1	*yw; pBAC{CH321-45F07}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S2A; S5A; S8A; S10A-F; S11I-J
Transgene P site 1	*yw; pBAC{CH321-58J11}VK00037; +*	*(159)**	2E-F; 3F-G; 4A-I; S1B; S2B; S5A; S8B; S10A-F; S11I-J
Transgene Q site 1	*yw; pBAC{CH321-52D18}VK00037/+; +*	*(159)**	2E-F; 3F-G; 4A-I; S1B; S2B; S5A; S8B; S10A-F; S11I-J
Transgene Y site 1	*yw; pBAC{CH321-47D18}VK00037; +*	*(159)**	3C, E-G; 4A-I; S6E; S8A; S10A-F; S11I-J

Transgene Z site 1	*w^{1118}; PBac{y[+mDint2]w[+mc]=pros-GFP.FPTB}VK00037; +*	Bloomington, *(159)*	3D-G; 4A-I; S3B; S7A; S8B; S10A-F; S11I-J
Transgene U site 3	*w^{1118}; +; Dp(1;3)DC212, PBac{y[+mDint2]w[+mC]=DC212} VK00033*	Bloomington, *(160)*	3E-G; 4A-I; S2C; S6A; S8A; S10A-F; S11I-J
Transgene V site 3	*w^{1118}; +; Dp(1;3)DC550, PBac{y[+mDint2]w[+mC]=DC550} VK00033*	Bloomington, *(160)*	3E-G; 4A-I; S2C; S6B; S8A; S10A-F; S11I-J
Transgene W site 3	*w^{1118}; +; Dp(1;3)DC305, PBac{y[+mDint2]w[+mC]=DC305} VK00033*	Bloomington, *(160)*	3E-G; 4A-I; S2C; S6C; S8A; S10A-F; S11I-J
Transgene X site 3	*yw; +; pBAC{CH321-43H12}VK00033*	*(159)**	3E-G; 4A-I; S2C; S6D; S8A; S10A-F; S11I-J
Transgene AA site 3	*yw; +; pBAC{CH321-16A21}VK00033*	*(159)**	3E-G; 4A-I; S2C; S7B; S8B; S10A-F; S11I-J
Transgene BB site 3	*yw; +; pBAC{CH321-68L02}VK00033*	*(159)**	3E-G; 4A-I; S2C; S7C; S8B; S10A-F; S11I-J
Transgene CC site 3	*w^{1118}; +; Dp(1;3)DC129, PBac{y[+mDint2]w[+mC]=DC129} VK00033*	Bloomington, *(160)*	3F-G; 4A-I; S2D-E; S7D; S8B; S10A-F; S11I-J
Transgene DD site 3	*w^{1118}; +; Dp(1;3)DC372, PBac{y[+mDint2]w[+mC]=DC372} VK00033*	Bloomington, *(160)*	3F-G; 4A-I; S2D-E; S7E; S8B; S10A-F; S11I-J
Transgene R site 1	*yw; pBAC{CH321-48C04}VK00037; +*	*(159)**	3H-I; S2B
Transgene S site 1	*yw; pBAC{pBJ250}VK00037; +*	*$(97)^{+}$*	3H-I; 5E, H, K; S13B; S14Q
Transgene T site 1	*yw; pBAC{pBJ205}VK00037; +*	*$(97)^{+}$*	5E, I, L; S13G; S14Q
duplication	*yw; Dp(3;2)P10/CyO; +*	Bloomington, *(161)*	5E; S5A; S9A-B, D-E
$ss^{inversion}$/ +	*yw; +/CyO; In(3R)P/+*	Bloomington, *(162)*	5E; S12A-C
$ss^{high\ freq\ null}$/ $ss^{inversion}$	*yw; +; ss^{52}/In(3R)P*	Bloomington, *(97, 162)*	5E; S12D
Transgene S-ss^{def}/ $ss^{inversion}$	*yw; +; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/In(3R)P*	Bloomington, *(97, 162, 163)*	5E; S12E
Transgene S site 2	*yw; +; pBAC{pBJ250}VK00027*	*$(97)^{+}$*	5E, P-S; S13C; S14Q
Transgene T site 2 for third instar larvae	*yw; pm181>Gal4, UAS>mcd8GFP/CyO; pBAC{pBJ205}VK00027/TM6B*	*$(97, 158)^{+}$*	5E, Q; S14Q
Transgene T site 2 for pupae and adults	*yw; +; pBAC{pBJ205}VK00027*	*$(97)^{+}$*	5E, T; S13H
Transgene E site 3	*yw; +; pBac{CH321-28L15}VK00033*	*(159)**	5E; S4A, C-D; S16C, E-G
Transgene S site 3	*yw; +; pBac{pBJ250}VK00033*	*$(97)^{+}$*	5E; S13D; S14C-D, Q
Transgene S site 4 for pupae and adults	*yw; +; pBAC{pBJ250}ZH-86Fb*	*$(97)^{+}$*	5E; S13E; S14J

Transgene S site 5	*yw; +; pBAC{pBJ250}VK00028*	*(97)*$^{+}$	5E; S13F; S14O-Q
Transgene T site 3	*yw; +; pBAC{pBJ205}VK00033*	*(97)*$^{+}$	5E; S13I; S14E-F, Q
Transgene T site 4	*yw; +; pBAC{pBJ205}ZH-86Fb*	*(97)*	5E; S13J; S14K-L, Q
Transgene S site 4 for third instar larvae	*yw; pm181>Gal4, UAS>mcd8GFP/CyO; pBAC{pBJ250}ZH-86Fb/TM6B*	*(97, 158)*	5E; S14I, Q
Transgene E+ ss$^{protein\ null}$	*yw; pBAC{CH321-28L15}VK0037/+; ss$^{d115.7}$/Df(3R)Exel7330*	Bloomington, *(159, 163-165)*	5E; S15A-B
Transgene E+ ss$^{high\ freq\ null}$	*yw; pBAC{CH321-28L15}VK0037/+; ss^{52}/Df(3R)Exel7330*	Bloomington, *(97, 159, 163, 164)*	5E; S15A, C
Transgene E+ Transgene S-ss^{def}/ ss$^{protein\ null}$	*yw; pBAC{CH321-28L15}VK00037/CyO; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/ ss$^{d115.7}$*	Bloomington, *(97, 159, 163, 164)*	5E; S15D-E
Transgene S-ss^{def}/ss$^{upstream\ del}$	*yw; +; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/ss$^{upstream\ deletion}$*	Bloomington, *(97, 163)*	5E; S19D-F
Transgene S-ss^{def}/ss$^{arista\ 1}$	*yw; +; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/ss^{a}*	Bloomington, *(97, 156, 163, 166)*	5E; S19G-I
Transgene S-ss^{def}/ss$^{arista\ 2}$	*yw; +; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/ss^{a40a}*	Bloomington, *(97, 163, 166)*	5E; S19J-L
ss$^{protein\ null}$/ss^{def}	*yw; +; ss$^{d115.7}$/Df(3R)Exel6269*	Bloomington, *(163, 164)*	6F-G
ss$^{arista\ 1}$/ss^{def}	*yw; +; ss^{a}/Df(3R)Exel6269*	Bloomington, *(156, 163, 166)*	6H-I; S17A-C
ss$^{arista\ 1}$/ss$^{protein\ null}$	*yw; +; ss^{a}/ss$^{d115.7}$*	Bloomington, *(156, 164, 166)*	6J-K; S17D-F
rearrangements	*yw; +; TM2/TM6B*	N/A	S12H-J
ss$^{arista\ 2}$/ss^{def}	*yw; +; ss^{a40a}/Df(3R)Exel6269*	Bloomington, *(163, 166)*	S17G-I
ss$^{arista\ 2}$/ss$^{protein\ null}$	*yw; +; ss^{a40a}/ss$^{d115.7}$*	Bloomington, *(164, 166)*	S17J-L
ss$^{enh\ del}$/ss^{def}	*yw; +; ss$^{enhancer\ deletion}$/Df(3R)Exel6269*	Bloomington, *(163)*	S18A-B
ss$^{enh\ del}$ /ss$^{protein\ null}$	*yw; +; ss$^{enhancer\ deletion}$/ss$^{d115.7}$*	*(164)*	S18C-D
ss$^{upstream\ del}$ / ss^{def}	*yw; +; ss$^{upstream\ deletion}$/ Df(3R)Exel6269*	Bloomington, *(163)*	S19A-C

*Constructs were purchased from the CHORI *Drosophila melanogaster* BAC library collection (*159*) and sent to BestGene Inc. (Chino Hills, CA) or Rainbow Transgenic Flies, Inc. (Camarillo, CA) for injection.

[+]Constructs were generated in (*97*) and sent to BestGene Inc. (Chino Hills, CA) or Rainbow Transgenic Flies, Inc. (Camarillo, CA) for injection.

Constructs were inserted via PhiC31 integration at the following landing sites:

Table 2: Transgene landing site coordinates

Landing site	Cytological coordinates	Genome coordinates
site 1 (VK00037)	22A3	2L: 1,582,820
site 2 (VK00027)	89E11	3R: 17,052,863
site 3 (VK00033)	65B2	3L: 6,442,676
site 4 (ZH-86Fb)	86F8	3R: 11,808,607
site 5 (VK00028)	92F1	3R: 20,549,650

Oligopaints probe libraries

Table 3: Genome coordinates targeted by Oligopaints probe libraries

Probe set	Oligopaints library name	Genome coordinates targeted	Conjugated fluorophore	Figures
site 1 neighboring sequence	right of 2L>22A3 transgene insertion site	2L: 1,582,821-1,642,821	Cy5	2B-D; 3B; 5G-I; S2A-B; S3B; S16A-B
ss	old *ss* 90K library	3R: 16,374,660-16,430,430	Cy3	2B; 3B; 5G; 6B; S12A, G, I; S16A-B
Transgene B neighboring endogenous sequence	downstream of 38G20	3R: 16,263,284-16,313,284	Cy3	2C
Transgene D neighboring endogenous sequence	downstream of 25M02	3R: 16,381,436-16,431,436	Cy3	2D
Transgene Y	*bicoid* 25-kb left extension+*bicoid* DNA+*bicoid* 25-kb right extension	3R: 6,729,194-6,787,593	Cy3	3C
Transgene Z	*prospero* DNA	3R: 11,246,862-11,407,918	Cy3	3D; S3B
upstream of *Transgenes S* and *T*	upstream of pBJ250 and pBJ205	3R: 16,340,760-16,390,760	Cy3	5H-I, O-P
site 2 neighboring sequence	pBJ250>3R(89E11) insertion site	3R: 16,992,863-17,052,863	Cy5	5O, P, Q

upstream of *ss*	*spineless* 50-kb extension (left)	3R: 16,320,533-16,370,533	Cy3	5Q; S9C-D; S14H-I, K, N-O
downstream of *ss*	*spineless* 50-kb extension (right)	3R: 16,435,681-16,485,681	Cy3	S2A; S4B-C; S11D; S14B-C, E; S16D-E
Transgene A neighboring endogenous sequence	downstream of 94A21	3R: 16,240,324-16,290,324	Cy3	S2A
Transgene C neighboring endogenous sequence	downstream of 86F17	3R: 16,324,960-16,374,960	Cy3	S2A
Transgene O neighboring endogenous sequence	downstream of 45F07	3R: 17,026,709-17,076,709	Cy3	S2A
Transgene F neighboring endogenous sequence	downstream of 23C04	3R: 16,455,152-16,505,152	Cy3	S2B
Transgene G neighboring endogenous sequence	upstream of 02A24	3R: 16,350,218-16,400,218	Cy3	S2B
Transgene H neighboring endogenous sequence	upstream of 92J22	3R: 16,390,309-16,440,309	Cy3	S2B
Transgene I neighboring endogenous sequence	upstream of 95F12	3R: 16,459,720-16,509,720	Cy3	S2B
Transgene J neighboring endogenous sequence	upstream of 71G17	3R: 16,511,320-16,561,320	Cy3	S2B
Transgene K neighboring endogenous sequence	downstream of 50E16	3R:16,691,689-16,741,689	Cy3	S2B
Transgene L neighboring endogenous sequence	downstream of 60D22	3R: 16,844,756-16,894,756	Cy3	S2B
Transgene M neighboring endogenous sequence	upstream of 58G07	3R: 16,739,235-16,789,235	Cy3	S2B; S11F-G
Transgene N neighboring endogenous sequence	upstream of 96A10	3R: 16,844,621-16,894,621	Cy3	S2B

Transgene P neighboring endogenous sequence	upstream of 58J11	3R: 16,967,427-17,017,427	Cy3	S2B
Transgene Q neighboring endogenous sequence	upstream of 52D18	3R: 17,043,366-17,093,366	Cy3	S2B
Transgene R neighboring endogenous sequence	downstream of 48C04	3R: 16,424,440-16,474,440	Cy3	S2B
site 3 neighboring sequence	pBJ250>3L(65B2) insertion site	3L: 6,442,676-6,502,676	Cy5	S2C-D; S4B-C; S11D; S14B-C, E; S16D-E
3L-2R control probe	*egfr* DNA	2R: 21,520,393-21,560,246	Cy3	S2C
Transgene U	*sp1* DNA	X: 9,697,559-9,778,741	Cy3	S2C-D
Transgene V	*merlin* 25-kb left extension+*merlin* DNA+*merlin* 25-kb right extension	X: 19,663,948-19,718,977	Cy3	S2C
Transgene W	*scalloped* 25-kb left extension+*scalloped* DNA	X: 15,778,880-15,827,986	Cy3	S2C; S11E, G
Transgene X	*yki* 25-kb left extension+*yki* DNA+*yki* 25-kb right extension	2R: 24,040,405-24,093,757	Cy3	S2C; S11E
Transgene AA	upstream of *clamp* DNA	2L: 22,115,720-22,165,720	Cy3	S2C
Transgene BB	downstream of *smo* DNA	2L: 282,167-332,167	Cy3	S2C
Transgene CC	*CG15930* 25-kb left extension+*CG15930* DNA+*CG15930* 25-kb right extension	X: 5,288,125-5,342,409	Cy3	S2D; S11F
Transgene DD	*phf7* 25-kb left extension+*phf7* DNA+*phf7* 25-kb right extension	X: 20,134,872-20,191,696	Cy3	S2D
neighboring duplication breakpoint	*spineless* duplication onto chromosome 2	2R: 14,522,912-14,582,912	Cy5	S9C-D
site 4 neighboring sequence	pBJ250>J36 insertion site	3R: 11,748,607-11,808,607	Cy5	S14H-I, K
site 5 neighboring sequence	downstream of 92F1 insertion site	3R: 20,549,650-20,599,650	Cy5	S14N-O
secondary sequence 1	sec 1	N/A	Cy3	Targets all Cy3-conjugated probes

secondary sequence 2	sec 2	N/A	Cy5	Targets all Cy5-conjugated probes

Pairing controls

Table 4: Controls used for FISH experiments. PRs: photoreceptors.

Control	Transgene insertion site chromosome	Endogenous site chromosome	Cell type	FISH assay	Probe sets used	Figures
2L-3R (2-color)	2L	3R	PRs	2-color	site 1 neighboring sequence and *ss*	2B, F; 3I; 5G; S1A-B; S14Q
3L-X (1-color)	3L	X	PRs	1-color	site 3 neighboring sequence and *Transgene U*	3E; S1A; S2C
3L-2R	3L	2R	PRs	1-color	and site 3 neighboring sequence and 3L-2R control probe	3E; S1A; S2C
2L-3R (1-color)	2L	3R	PRs	1-color	site 1 neighboring sequence and *ss*	3B, E; 6C; S1A; S12J; S16A
3L-2L	3L	2L	PRs	1-color	site 3 neighboring sequence and *Transgene AA*	3E; S1A; S2C
site 2	3R	3R	PRs	2-color	site 2 neighboring sequence and upstream of *Transgenes S* and *T*	5O; S14Q
3L-3R	3L	3R	PRs	2-color	site 3 neighboring sequence and downstream of *ss*	S1A; S4B-D; S11D, H
site 4	3R	3R	PRs	2-color	site 4 neighboring sequence and upstream of *ss*	S14H, Q
site 5	3R	3R	PRs	2-color	site 5 neighboring sequence and upstream of *ss*	S14N, Q
2L-3R (antenna)	2L	3R	antenna	1-color	site 1 neighboring sequence and *ss*	6C; S1A; S16B
3L-X (2-color)	3L	X	PRs	2-color	site 3 neighboring sequence and *Transgene U*	S1A; S2D-E
duplication	2R	3R	PRs	2-color	neighboring duplication breakpoint and upstream of *ss*	S9C, E

3L-3R (antenna)	3L	3R	antenna	2-color	site 3 neighboring sequence and downstream of *ss*	S1A; S16D, F

Compartments

Table 5: Loci examined in compartmentalization experiments

Compartment	Chromosome	Genes in locus	Probe set used
A1	X	*CG8191, CG12379, MagR, Arp6, CG11679, CG8206, CCT6, CG8239, Graf, CG8260, CG17209, CG8952, CG33172, CG43673, Efhc1.1, Paf-AHalpha, mRpS30, Rhp, CR44383, CG32581, CG15602, PGRP-LE, CR44384, CG8944, mRpL3, CG43672, CG8974, sd, CG8509*	*Transgene W*
A2	2R	*IntS1, tsr, gammaSnap1, ppk29, CG13563, eEF5, CR44814, RpL12, RpL39, Rap2l, CG3209, Mlp60A, CG10339, CG13564, snama, gek, CG4049, CG3253, CG43775, CG43776, CG43777, tamo, Zfrp8, CG4065, CR44826, sei, yki, CG16786, enok, Dat, CG3257*	*Transgene X*
B1	X	*CG42749, CG3323, CG15465, rg*	*Transgene CC*
B2	3R	*CG31275, CR44945, Glut3, CR45750, lncRNA:TS16, bxd, abd-A, iab-8*	*Transgene M* neighboring endogenous sequence

Antibodies

Antibodies and dilutions were as follows: mouse anti-Lamin B (DSHB ADL67.10 and ADL84.12), 1:100; rabbit anti-GFP (Invitrogen), 1:500; rabbit anti-Rh4 (gift from C. Zuker, Columbia University), 1:50; mouse anti-Rh3 (gift from S. Britt, University of Texas at Austin), 1:50; mouse anti-Prospero (DSHB MR1A), 1:10; rat anti-Elav (DSHB 7E8A10), 1:50; guinea pig anti-Ss (gift from Y.N. Jan, University of California, San Francisco), 1:500. All secondary antibodies (Molecular Probes) were Alexa Fluor-conjugated and used at a dilution of 1:400.

Antibody staining (pupal and adult eyes)

Dissections were performed as described in references (*165, 167-169*). Eyes were dissected and fixed at room temperature for 15 minutes in 4% formaldehyde diluted in 1X PBX (PBS+0.3% Triton-X), then washed three times in 1X PBX. Eyes were incubated overnight at room temperature in primary antibody diluted in 1X PBX, then washed three times in 1X PBX and incubated in PBX at room temperature for ≥3 hours. Secondary antibody diluted in 1X PBX was added and incubated overnight at room temperature. Eyes were then washed three times in 1X PBX and incubated in PBX at room temperature for ≥2 hours. Adult eyes were mounted in SlowFade Gold (Invitrogen), and pupal eyes were mounted in Vectashield (Vector Laboratories, Inc.). Images were acquired on a Zeiss LSM700 confocal microscope.

The adult eye dissection protocol was used for **Fig. 5D, J-L, R-T; Fig. S12C-E; Fig. S14D, F, J, L, P; Fig. S15B-C, E; Fig. S17B, E, H, K; Fig. S18B, D;** and **Fig. S19B, E, H, K**. The pupal dissection protocol was used for **Fig. 5C** and **Fig. S13B-J**.

Oligopaints probe design

Probes for DNA FISH were designed using the Oligopaints technique (*149, 170*). Target sequences were run through the bioinformatics pipeline available at http://genetics.med.harvard.edu/oligopaints/ to identify sets of 42-bp (for old ss 90K probes) or 50-bp (for all other probes) optimized probe sequences (i.e. “libraries”) tiled across the DNA sequence of interest. Five 19-bp barcoding

primers, gene F and R; universal (univ) F and R, and either sublibrary (sub) F or random (rando) R, were appended to the 5' and 3' ends of each probe sequence (**Fig. S20A-B**). To ensure that all probes were the same length, an additional 8-bp random sequence was added to the 3' end of the old ss 90K probes. The gene F and R primers allowed PCR amplification of a probe library of interest out of the total oligo pool, and the univ F and R primers allowed conjugation of fluorophores, generation of single-stranded DNA probes, and PCR addition of secondary sequences to amplify probe signal. The *ss* 50-kb left and right extension libraries had a sub F primer between the gene and universal forward primers to allow PCR amplification of probes targeting a specific sub-region of the locus of interest (**Fig. S20A**). All other probe libraries had a rando R primer appended at the 3' end to maintain a constant sequence length between all probes (**Fig. S20B**).

Barcoding primer sequences were taken from a set of 240,000 randomly generated, orthogonal 25-bp sequences (*171*) and run through a custom script to select 19-bp sequences with ≤15-bp homology to the *Drosophila* genome. Primers were appended to probe sequences using the orderFile.py script available at http://genetics.med.harvard.edu/oligopaints/. Completed probe libraries were synthesized as custom oligo pools by Custom Array, Inc. (Bothell, WA), and fluorescent FISH probes were generated as described in references (*149, 170*).

DNA FISH

DNA FISH was performed using modified versions of the protocols described in references (*149, 170*). 20-50 eye-antennal discs attached to mouth hooks from third instar larvae were collected on ice and fixed in 129 μL ultrapure water, 20 μL 10X PBS, 1 μL Tergitol NP-40, 600 μL heptane, and 50 μL fresh 16% formaldehyde. Tubes containing the fixative and eye discs were shaken vigorously by hand, then fixed for 10 minutes at room temperature with nutation. Eye discs were then given three quick washes in 1X PBX, followed by three five-minute washes in PBX at room temperature with nutation. Eye discs were then removed from the mouth hooks and blocked for 1 hour in 1X PBX+1% BSA at room temperature with nutation. They were then incubated in primary antibody diluted in 1X PBX overnight at 4°C with nutation. Next, eye discs were washed three times in 1X PBX for 20 minutes and incubated in secondary antibody diluted in 1X PBX for two hours at room temperature with nutation. Eye discs were then washed two times for 20 minutes in 1X PBX, followed by a 20-minute wash in 1X PBS. Next, discs were given one 10-minute wash in 20% formamide+2X SSCT (2X SSC+.001% Tween-20), one 10-minute wash in 40% formamide+2X SSCT, and two 10-minute washes in 50% formamide+2X SSCT. Discs were then predenatured by incubating for four hours at 37°C, three minutes at 92°C, and 20 minutes at 60°C. Primary probes were added in 45 μL hybridization buffer consisting of 50% formamide+2X SSCT+2% dextran sulfate (w/v), + 1 μL RNAse A. All probes were added at a concentration of ≥5 pmol fluorophore/μL. For FISH experiments in which a single probe was used, 4 μL of

probe was added. For FISH experiments in which two probes were used, 2 μL of each probe was added. After addition of probes, eye discs were incubated at 91°C for three minutes and at 37°C for 16-20 hours with shaking. Eye discs were then washed for 1 hour at 37°C with shaking in 50% formamide+2X SSCT. 1 μL of each secondary probe was added at a concentration of 100 pmol/μL in 50 μL of 50% formamide+2X SSCT. Secondary probes were hybridized for 1 hour at 37°C with shaking. Eye discs were then washed twice for 30 minutes in 50% formamide+2X SSCT at 37°C with shaking, followed by three 10-minute washes at room temperature in 20% formamide+2X SSCT, 2X SSCT, and 2X SSC with nutation. Discs were mounted in SlowFade Gold immediately after the final 2X SSC wash, and imaged using a Zeiss LSM700 confocal microscope.

Generation of CRISPR lines

CRISPR lines were generated as described in references (*101, 172-174*). For both $ss^{enh\ del}$ and $ss^{upstream\ del}$, sense and antisense DNA oligos for the forward and reverse strands of four gRNAs were designed to generate BbsI restriction site overhangs. The oligos were annealed and cloned into the pCFD3 cloning vector (Addgene, Cambridge, MA). A single-stranded DNA homology bridge was generated with 60-bp homologous regions flanking each side of the predicted cleavage site and an EcoRI (for $ss^{enh\ del}$) or NaeI (for $ss^{upstream\ del}$) restriction site to aid in genotyping. The gRNA constructs (125 ng/μl) and homologous bridge oligo (100 ng/μl) were injected into *Drosophila* embryos (BestGene, Inc., Chino Hills, CA). Single males were crossed with a balancer

stock (*yw; +; TM2/TM6B*), and F1 female progeny were screened for the insertion via PCR, restriction digest, and sequencing. Single F1 males whose siblings were positive for the deletion were crossed to the balancer stock (*yw; +; TM2/TM6B*), and the F2 progeny were screened for the deletion via PCR, restriction digest, and sequencing. Deletion-positive flies from multiple founders were used to establish independent stable stocks.

The following oligos were used for the *ss*$^{enh\ del}$ CRISPR:

Table 6: Oligos used for *ss*$^{enh\ del}$ CRISPR

Oligo name	Sequence
Homologous bridge	CAATTTAATTGAGCTCCCAAGTGCTGGGAAGCAGCTGCCCTTTGAATTGGGCTTCTCACCGAATTC TGGCCTGGCTTTGGAGCTCCTTTTGGTGAGAGACCAAAAGAGATTCCGCTGCGCGAATCG
gRNA 1F	GTCGTAATATTCGCTAGGACCTA
gRNA 1R	AAACTAGGTCCTAGCGAATATTAC
gRNA 2F	GTCGAATTGGGCTTCTCACCCCT
gRNA 2R	AAACAGGGGTGAGAAGCCCAATTC
gRNA 3F	GTCGCCAGGCCATGTGGGCATTT
gRNA 3R	AAACAAATGCCCACATGGCCTGGC
gRNA 4F	GTCGCTCCAAAGCCAGGCCATGT
gRNA 4R	AAACACATGGCCTGGCTTTGGAGC
genotype F	CTTAGCTTCAAGCGGCTCCG
genotype R	GAATAACGTCAACTGTGCCA

The following oligos were used for the *ss*$^{upstream\ del}$ CRISPR:

Table 7: Oligos used for *ss*$^{upstream\ del}$ CRISPR

Oligo name	Sequence
Homologous bridge	TGAGTTGATTGAAGGCTGTAAGAGCAGATTACAGTGGGGCGGAGGCCCAAGTCTGGATCT GCCGGCCTCTGGGTATTCATTTTTTTCGACTTGGCAATTGCAAATGCAAAACCATTTCATTTGCCG
gRNA 1F	GTCGTCGTCTAGCCTAGAAGCGTT
gRNA 1R	AAACAACGCTTCTAGGCTAGACGA
gRNA 2F	GTCGGGCCCAAGTCTGGATCTCCC

gRNA 2R	AAACGGGAGATCCAGACTTGGGCC
gRNA 3F	GTCGCAAAACAATATGAGGTCTAA
gRNA 3R	AAACTTAGACCTCATATTGTTTTGC
gRNA 4F	GTCGAAGTGGCCTGGGCTTATCTC
gRNA 4R	AAACGAGATAAGCCCAGGCCACTT
genotype F	GACCATTTAAGCGGCTACAAA
genotype R	GGTGGTCAGTCGGCAAATGAA

Scanning electron microscopy

Adult *Drosophila* heads were removed and immediately mounted on a pin stub without fixation or sputtering. Heads were imaged at high vacuum at a voltage of 1.5 kV. All SEM was performed on a FEI Quanta ESEM 200 scanning electron microscope. SEM was used for **Fig. 6E, G, I, K; Fig. S17C, F, I, L;** and **Fig. S19C, F, I, L.**

Pairing quantifications

All quantifications were performed in 3D on z-stacks with a slice thickness of 0.2 μm. Quantifications were performed manually using Fiji (*175, 176*). To chart the z position of each FISH dot, a line was drawn through the dot and the Plot Profile tool was used to assess the stack in which the dot was brightest. To determine the x-y distance between the two FISH dots, a line was drawn from the center of one dot to the center of the other dot and the length of the line was measured with the Plot Profile tool. The distance between the FISH dots was then calculated in 3D. A total of 50 nuclei from three eye discs were quantified for each genotype (i.e. N=3, n=50).

For experiments in which the transgene and endogenous site were both labeled with red fluorescent probes, FISH punctae ≤0.4 μm apart could not be

distinguished as separate and were assigned a distance of 0.4 μm apart. For all controls in **Fig. 3E**, **6C**, and **S12J**, green probes labeling the transgene insertion site were pseudocolored red and data were quantified in the same way as experiments in which the transgene and endogenous site were both labeled with red probes. 3L-X control data in **Fig. 3E** are taken from the same experiment as in **Fig. S2E**, but the data were re-quantified with the green probes pseudocolored red. Similarly, 2L-3R eye control data in **Fig. 3E, 6C**, and **S12J** are taken from the same experiment as in **Fig. 2F, 3I, S1A-B,** and **S14Q**, but the data were re-quantified with the green probes pseudocolored red.

Adult eye quantifications

The frequencies of Rh4- and Rh3-expressing R7s were scored manually for at least eight eyes per genotype. R7s co-expressing Rh3 and Rh4 were scored as Rh4-positive. 100 or more R7s were scored for each eye. For **Fig. S19E, H,** and **K**, only the ventral half of each eye was scored.

Hi-C mapping and TAD calling

Directionality index scores were calculated across 15-kb windows, stepping every 5 kb, by finding the log2 transform of the difference in the ratios of downstream versus upstream summed observed over expected interactions ranging from 15 kb to 100 kb in size. The expected value of a bin was defined as the sum of the product of fragment corrections for each valid fragment pair with both interaction fragments falling within the bin.

Directionality indices were generated using 14 published Hi-C datasets (*177-180*):

Table 8: NCBI accession numbers for analyzed HiC datasets

Dataset	NCBI Accession Number
1	GSE38468
2	GSE38468
3	GSE61471
4	GSE61471
5	GSE61471
6	GSE63515
7	GSE63515
8	GSM2679637
9	GSM2679640
10	GSM2679641
11	GSM2679642
12	GSM2679643
13	GSM2679644
14	GSM2679645

TADs were read from the beginning of a positive directionality index peak to the end of a negative directionality index peak. Parameters for calling a TAD were as follows: **1)** The positive peak must have a signal of $\geq$0.8; **2)** The negative peak must have a signal of $\leq$-0.8; and **3)** The TAD must be present in at least two datasets. Any transgene covering $\geq$95% of a TAD was considered to span a TAD.

mRNA sequencing and analysis

RNA-seq was performed on three biological replicates, each consisting of 30 third instar larval eye discs. Eye discs were dissected in 1X PBS, separated from the mouth hooks and antennal discs, and placed directly into 300 μL of Trizol. RNA was purified using a Zymo Direct-zol RNA MicroPrep kit (catalog

number R2062). mRNA libraries were prepared using an Illumina TruSeq Stranded mRNA LT Sample Prep Kit (catalog number RS-122-2101). Sequencing was performed using an Illumina NextSeq 500 (75 bp, paired end). Sequencing returned an average of 23,048,349 reads per replicate.

The following pipeline was used for mRNA-sequencing analysis: 1) FASTQ sequencing datasets were assessed for quality using FastQC; 2) Pseudoalignment with the *Drosophila* dm6 transcriptome and read quantifications were performed using kallisto (*181*); 3) Transcript abundance files generated by kallisto were joined to a file containing the genomic coordinates of all *Drosophila* mRNA transcripts (dmel-all-r6.20.gtf, available from Flybase); 4) The joined transcript coordinate file was compared to a file containing the coordinates of all tested transgenes using the bedtools intersect tool (http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html)(*182*). The output file contained a list of all TPMs for each gene contained in each transgene.

Assessment of chromatin marks and ncRNA, Polycomb Group Complex, and insulator density

ncRNA content of transgenes was assessed manually using the GBrowse tool on FlyBase. tRNAs, miRNAs, snoRNAs, and lncRNAs were included in the analysis of ncRNA content.

Transgenes were evaluated for insulator binding sites, Polycomb Group Complex binding sites, and the presence of chromatin marks using publicly

available ChIP-chip and ChIP-seq datasets(*151-155*). The following datasets were used for this analysis:

Table 9: modENCODE and NCBI accession number information for analyzed ChIP data

Protein/chromatin mark	modENCODE dataset ID(s)	NCBI Accession Number(s)
BEAF-32	21	GSM762845
Su(Hw)	27, 901, 4104, 4105	GSM762839, GSM1015406
CTCF	769, 770, 908, 2638, 2639	GSM762842, GSM1015410
Cp190	22	GSM762836, GSM1015404, GSM1261702
Mod(Mdg4)	24, 4094	GSM892322, GSM1015408
GAF	23, 2568, 3238, 3245, 3397, 3814, 3830, 5028	N/A
Pcl	3237, 3813, 3816, 3960	N/A
Pc	325, 326, 816, 948, 3791, 3957, 5064	N/A
dRING	927, 928, 3750, 5071, 5255	N/A
Pho	3894	N/A
H3K27me3	346, 767, 869, 919	N/A

To ensure a higher likelihood of selecting true ChIP peaks rather than false positives, only those insulators present in data from multiple cell types were considered when assessing the number of insulator sites per transgene. .bed files containing the genomic coordinates of all ChIP peaks in each dataset were downloaded and classified by cell type. All datasets from the same cell type were merged into one file using the bedtools merge tool (http://bedtools.readthedocs.io/en/latest/content/tools/merge.html)(*182*). Insulator ChIP peaks present in more than one cell type were identified using the bedtools multiIntersectBed tool(*182*). The coordinates of insulator ChIP peaks present in

multiple cell types were compared to a .bed file containing the genomic coordinates of all transgenes using the bedtools intersect tool (http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html)(*182*). This pipeline output the number of insulator ChIP peaks contained in each transgene.

To identify clusters of insulators, files containing the ChIP peak coordinates for each insulator were compared using the the bedtools multiIntersectBed tool(*182*), which output a .bed file listing all of the locations of overlap between insulator binding sites. Insulators were considered to cluster if their binding site coordinates overlapped or were directly adjacent to each other. The coordinates of clusters containing specific numbers or combinations of insulators were selected from the intersected .bed file and compared to a .bed file containing the genomic coordinates of all transgenes using the bedtools intersect tool (http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html)(*182*). This pipeline output the number of insulator clusters contained in each transgene.

For Polycomb Group Complex proteins and chromatin marks, .bed files containing the genomic coordinates of all ChIP peaks in each dataset were downloaded and merged into one file using the bedtools merge tool (http://bedtools.readthedocs.io/en/latest/content/tools/merge.html)(*182*). The merged file was compared to a .bed file containing the genomic coordinates of all transgenes using the bedtools intersect tool (http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html)(*182*). This

pipeline output the number of protein or chromatin mark ChIP peaks contained in each transgene.

Statistical analysis

All datasets were tested for a Gaussian distribution using a D'Agostino and Pearson omnibus normality test and a Shapiro-Wilk normality test. If either test indicated a non-Gaussian distribution for any of the datasets in an experiment, datasets were tested for statistical significance using a Wilcoxon rank-sum test (for single comparisons) or a one-way ANOVA on ranks with Dunn's multiple comparisons test (for multiple comparisons). If both the D'Agostino and Pearson and the Shapiro-Wilk tests indicated a Gaussian distribution for all datasets in an experiment, datasets were tested for statistical significance using an unpaired t-test with Welch's correction (for single comparisons) or an ordinary one-way ANOVA with Dunnett's multiple comparisons test (for multiple comparisons).

Maximum likelihood calculations: Parameters for either a single or double Gaussian distribution were estimated using maximum likelihood, and model selection was subsequently performed using the Bayesian Information Criterion (BIC). For the double Gaussian, parameters were estimated using a nonlinear recursion (Levenberg-Marquardt) algorithm to maximize the log likelihood of the distribution.

Maximum likelihood calculations were performed for all transgenes tested in **Fig. 2F.** Maximum likelihood estimation was not possible for transgenes in **Fig.**

3E, as the 0.4 μm distance cutoff for 1-color FISH was too high to allow separation of paired and unpaired distributions into two Gaussian distributions.

3.6: Author Contributions

The author list for this paper was as follows: Kayla Viets, Michael Sauria, Chaim Chernoff, Caitlin Anderson, Sang Tran, Abigail Dove, Raghav Goyal, Lukas Voortman, Andrew Gordus, James Taylor, and Robert J. Johnston Jr. I was the sole first author on this publication. I was responsible for writing and editing the text, conceiving and performing experiments, and analyzing data. Michael Sauria performed HiC data analysis. Chaim Chernoff performed adult retina dissections for Supplemental Figure 12E and Supplemental Figure 15E. Caitlin Anderson and Sang Tran generated CRISPR lines. Abigail Dove wrote a custom script used for Oligopaints probe generation. Raghav Goyal performed preliminary experiments examining *Transgene S* transvection. Lukas Voortman assisted in preparing RNA-seq samples. James Taylor assisted with HiC and RNA-seq data analysis. As corresponding author, Robert J. Johnston Jr. conceived experiments and edited the manuscript.

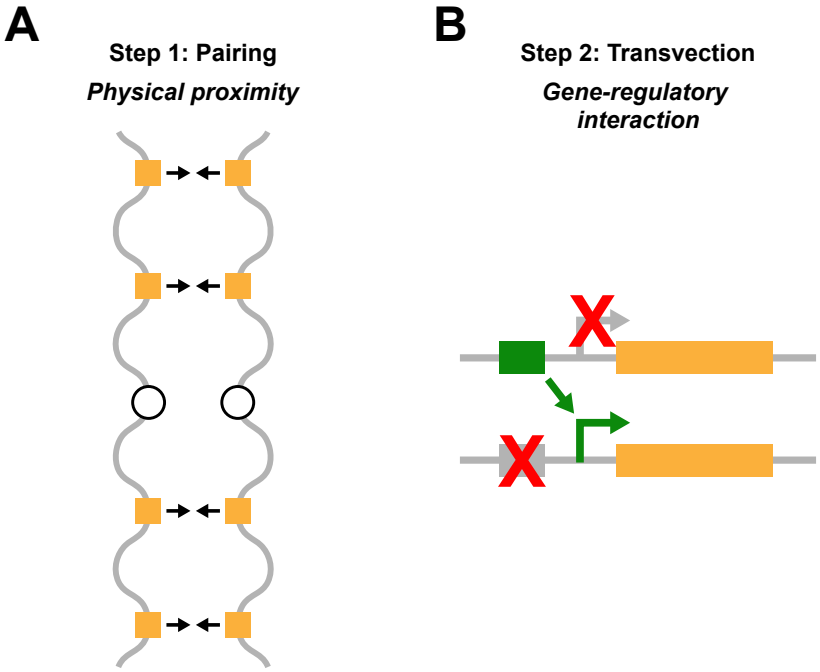

A
Step 1: Pairing
Physical proximity
B
Step 2: Transvection
Gene-regulatory interaction

Figure 1

Figure 1: Homologous chromosomes “button” together to facilitate transvection.

A. Button model of chromosome pairing. Yellow squares: button loci (high pairing affinity).

B. During transvection, two different mutant alleles interact between chromosomes to rescue gene expression. Green box: functional enhancer. Gray box with red X: mutated enhancer. Green arrow: functional promoter. Gray arrow with red X: mutated promoter.

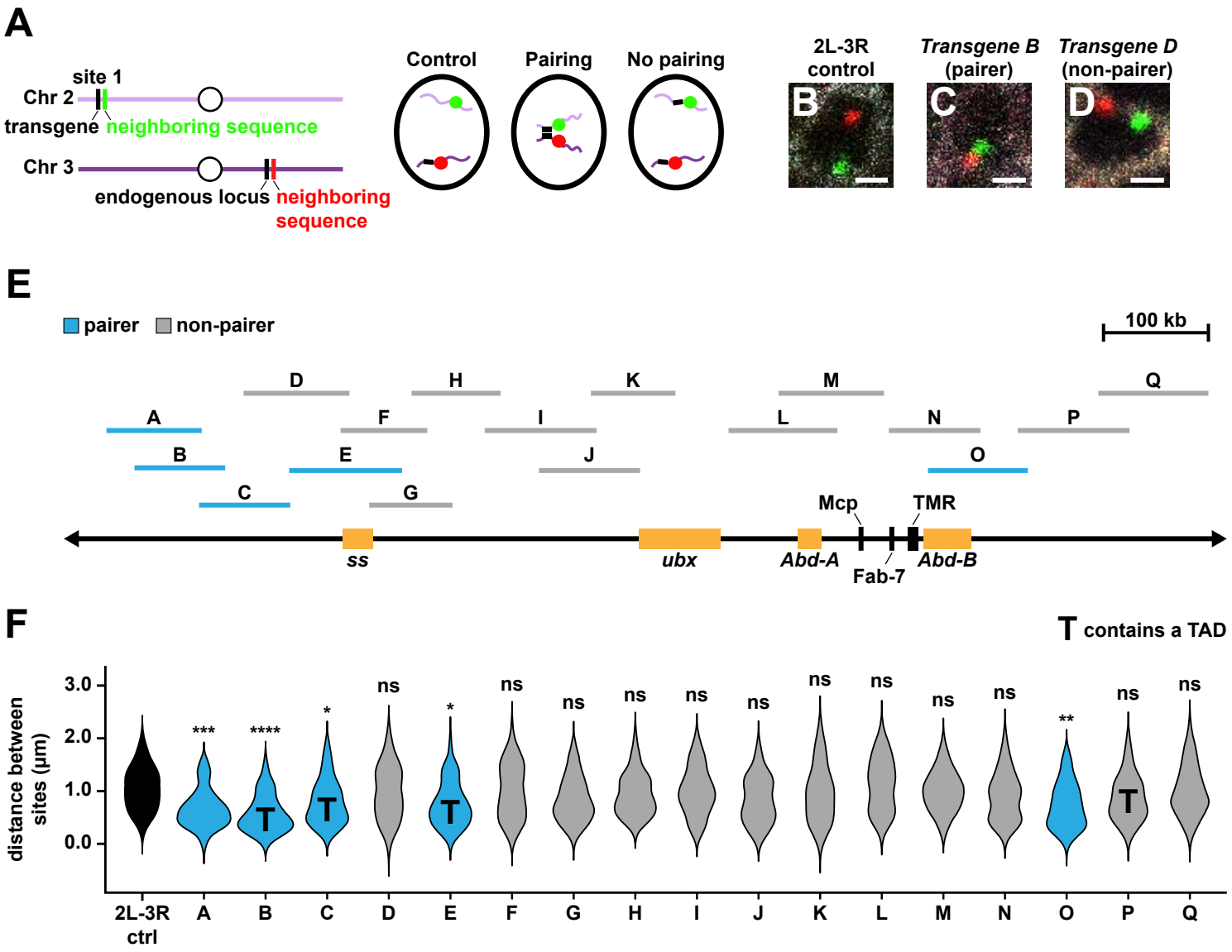

A
site 1
Chr 2
transgene
neighboring sequence
Chr 3
endogenous locus
neighboring sequence
Control
Pairing
No pairing
2L-3R control
Transgene B (pairer)
Transgene D (non-pairer)
B
C
D
E
pairer
non-pairer
100 kb
D
H
K
M
Q
A
F
I
L
N
P
B
E
J
O
C
G
Mcp
TMR
ss
ubx
Abd-A
Fab-7
Abd-B
F
T contains a TAD
distance between sites (µm)
3.0
2.0
1.0
0.0
***
****
*
ns
*
ns
ns
ns
ns
ns
ns
ns
ns
ns
**
ns
ns
T
T
T
T
2L-3R ctrl
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q

Figure 2

Figure 2: A screen for pairing elements identifies buttons interspersed along chromosome 3R.

A. Two-color DNA FISH strategy. In controls, red and green FISH punctae on heterologous chromosomes are far apart in the nucleus. If a transgene drives pairing, red and green FISH punctae are close together in the nucleus. If a transgene does not drive pairing, red and green FISH punctae are far apart in the nucleus, similar to a control.

B-D. Control, pairer, and non-pairer examples. Image for 2L-3R control is from the same experiment as in **Fig. 3B, 5G,** and **S16A**. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.

E. ~1 Mb region of chromosome 3R used for pairing screen. Orange boxes indicate locations of the major developmental genes *spineless* (*ss*), *ultrabithorax* (*ubx*), *Abdominal-A* (*Abd-A*), and *Abdominal-B* (*Abd-B*). Black lines indicate the locations of the pairing elements Mcp, Fab-7, and TMR.

F. Quantifications for all transgenes from the initial screen. Black: control, blue: pairers, gray: non-pairers. **T:** contains a TAD. Control data are the same as in **Fig. 3E** (2L-3R control), **3I**, **6C** (2L-3R eye control), **S1A-B, S12J,** and **S14Q** (2L-3R control). *Transgene D-G* data are the same as in **Fig. 3I** and **S1B**. *Transgene H-K, M, P, Q* data are the same as in **Fig. S1B**. ****=p<0.0001, ***=p<0.001, **=p<0.005, *=p<0.05, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test.

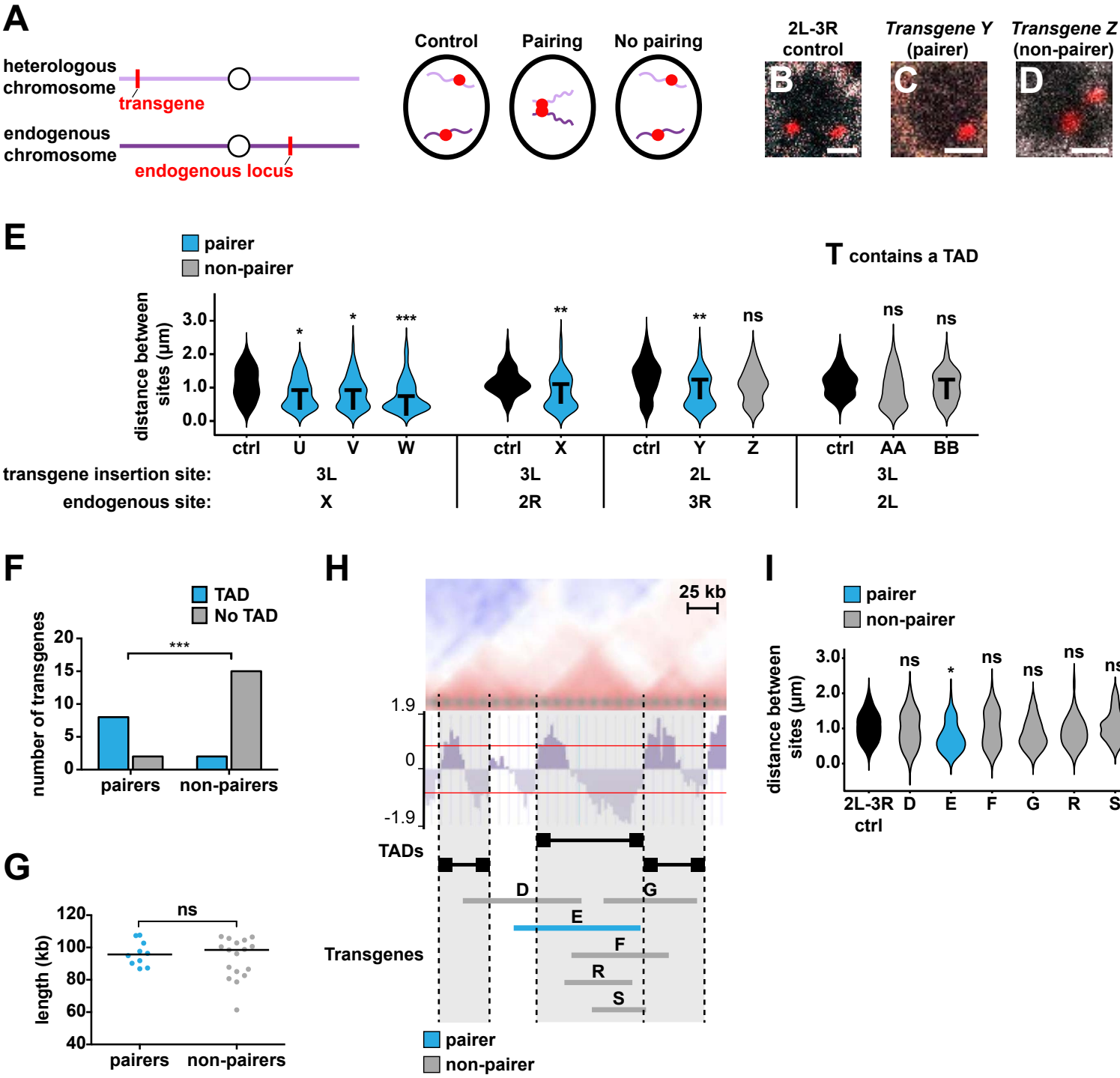

A
heterologous chromosome
transgene
endogenous chromosome
endogenous locus
Control
Pairing
No pairing
2L-3R control
Transgene Y (pairer)
Transgene Z (non-pairer)
B
C
D
E
pairer
non-pairer
T contains a TAD
distance between sites (µm)
3.0
2.0
1.0
0.0
*
*
***
**
**
ns
ns
ns
T
ctrl
U
V
W
ctrl
X
ctrl
Y
Z
ctrl
AA
BB
transgene insertion site:
3L
3L
2L
3L
endogenous site:
X
2R
3R
2L
F
TAD
No TAD
number of transgenes
20
15
10
5
0
***
pairers
non-pairers
G
length (kb)
120
100
80
60
40
ns
pairers
non-pairers
H
25 kb
1.9
0
-1.9
TADs
D
G
E
F
R
S
Transgenes
pairer
non-pairer
I
pairer
non-pairer
distance between sites (µm)
3.0
2.0
1.0
0.0
ns
*
ns
ns
ns
ns
2L-3R ctrl
D
E
F
G
R
S

Figure 3

Figure 3: Specialized TADs contribute to button activity and drive pairing.

A. One-color DNA FISH strategy: In controls, two red FISH punctae on heterologous chromosomes are far apart in the nucleus. If a transgene drives pairing, the two red FISH punctae are close together in the nucleus and indistinguishable as separate dots. If a transgene does not drive pairing, the two red FISH punctae are far apart in the nucleus, similar to a control.

B-D. Control, pairer, and non-pairer examples. Scale bars=1 μm. White: Lamin B, red: probes against endogenous sequence and transgene. Image for 2L-3R control is from the same experiment as **Fig. 2B**, **5G**, and **S16A**. Image for *Transgene Z* is from the same experiment as **Fig. S3B**.

E. Quantifications for additional transgenes. **T:** contains a TAD. Black: controls, blue: pairers, gray: non-pairers. ***=p<0.001, **=p<0.005, *=p<0.05, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test (for *Transgenes U-W*, *Y-Z*, *AA-BB*) or Wilcoxon rank-sum test (for *Transgene X*). 3L-X control data are the same as in **Fig. S1A-B** and **S2E**. 3L-2R control data and 3L-2L control data are the same as in **Fig. S1A.** 2L-3R control data are the same as in **Fig. 2F, 3I, 6C** (2L-3R eye control), **S1A-B, S12J,** and **S14Q** (2L-3R control). Controls were imaged in two colors, then pseudocolored red and re-scored in one color.

F. Number of pairers vs. non-pairers spanning TADs. Blue: spans a TAD. Gray: does not span a TAD. ***=p<0.001, Fisher's exact test.

G. Comparison of length for all pairers vs. non-pairers tested in **Fig. 2F, 3E,** and **S2E**. Blue: pairers, gray: non-pairers. ns=p>0.05, Wilcoxon rank-sum test. Black lines indicate medians.

H. Representative Hi-C heat map and directionality index (NCBI GSE38468) showing TADs in the region covered by *Transgenes D-G*. Dotted lines: TAD boundaries. Black bars: TADs. Red lines indicate a directionality index signal of 0.8 or -0.8, the cutoff for a TAD. See **Fig. S5A** for TAD assessment.

I. Quantifications for transgenes that "split" the TAD covered by *Transgene E* in **Fig. 3H**. Black: control, blue: pairers, gray: non-pairers. *=p<0.05, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test. Control data are the same as in **Fig. 2F, 3E** (2L-3R control), **6C** (2L-3R eye control), **S1A-B, S12J,** and **S14Q** (2L-3R control). *Transgene D, F,* and *G* data are the same as in **Fig. 2F** and **S1B**. *Transgene E* data are the same as in **Fig. 2F** and **S1B**. *Transgene S* data are the same as in **Fig. S14Q** (*Transgene S* site 1).

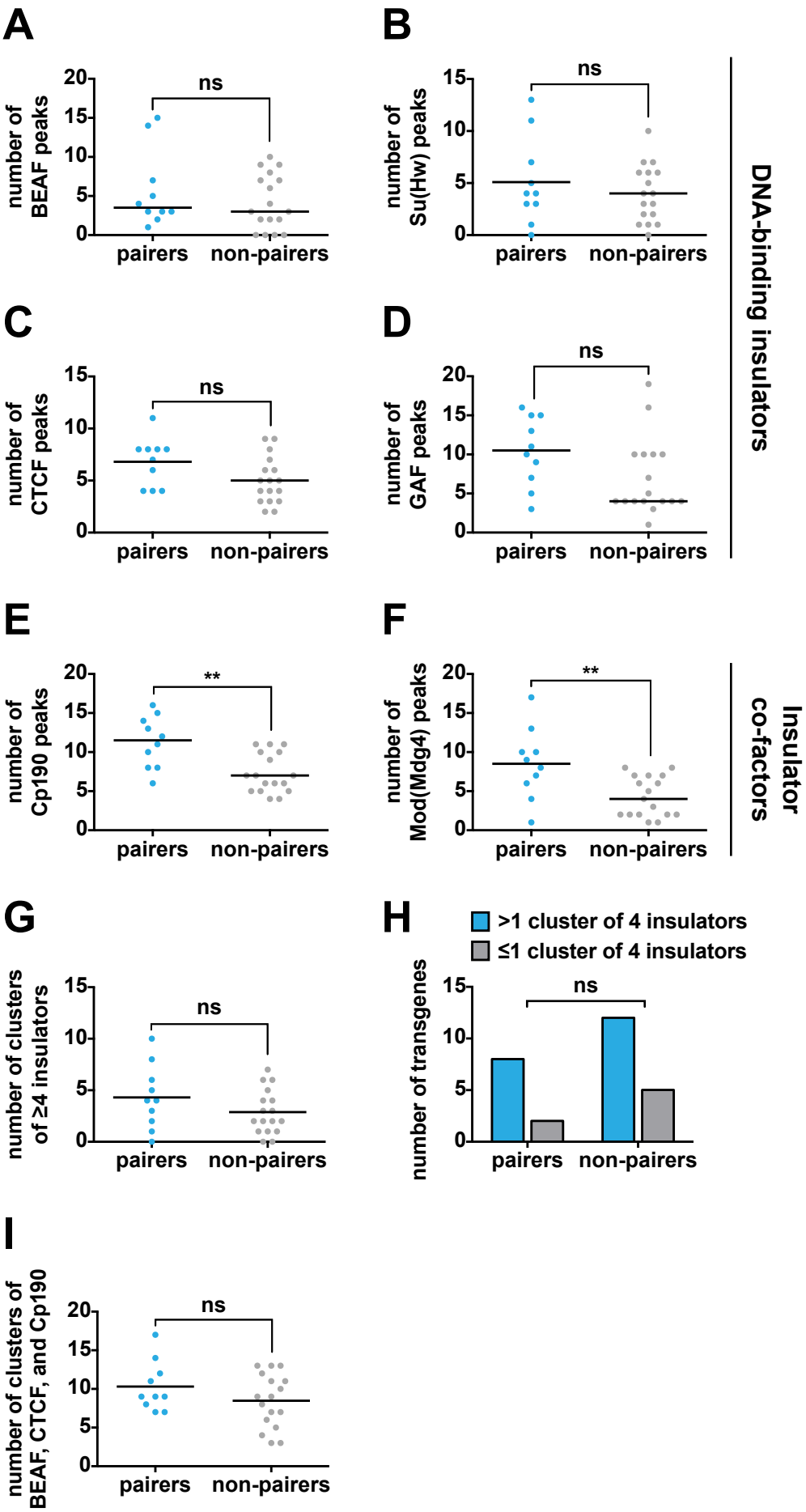

A
number of BEAF peaks
ns
pairers
non-pairers
B
number of Su(Hw) peaks
ns
pairers
non-pairers
DNA-binding insulators
C
number of CTCF peaks
ns
pairers
non-pairers
D
number of GAF peaks
ns
pairers
non-pairers
E
number of Cp190 peaks
**
pairers
non-pairers
F
number of Mod(Mdg4) peaks
**
pairers
non-pairers
Insulator co-factors
G
number of clusters of ≥4 insulators
ns
pairers
non-pairers
H
>1 cluster of 4 insulators
≤1 cluster of 4 insulators
number of transgenes
ns
pairers
non-pairers
I
number of clusters of BEAF, CTCF, and Cp190
ns
pairers
non-pairers

Figure 4

Figure 4: Examining the relationship between insulator binding sites and pairing.

A-I. Quantifications for pairers and non-pairers tested in **Fig. 2F, 3E,** and **S2E**. **=p<0.01, ns=p>0.05, Wilcoxon rank-sum test (**A, D-F**), unpaired t-test with Welch's correction (**B-C, G, I**), or Fisher's exact test (**H**). The Wilcoxon rank-sum test compares medians, while the unpaired t-test with Welch's correction compares means. Therefore, the black lines in **A, D-F** indicate medians, and the black lines in **B-C, G, I** indicate means.

A-G, I: Blue: pairers, gray: non-pairers.

A. BEAF ChIP peaks.

B. Su(Hw) ChIP peaks.

C. CTCF ChIP peaks.

D. GAF ChIP peaks.

E. Cp190 ChIP peaks.

F. Mod(mdg4) ChIP peaks.

G. Number of clusters of ≥4 insulators.

H. Number of pairers vs. non-pairers with >1 or ≤1 cluster of 4 insulators. Blue: >1 cluster, gray: ≤1 cluster.

I. Number of clusters containing any combination of BEAF, CTCF, and Cp190.

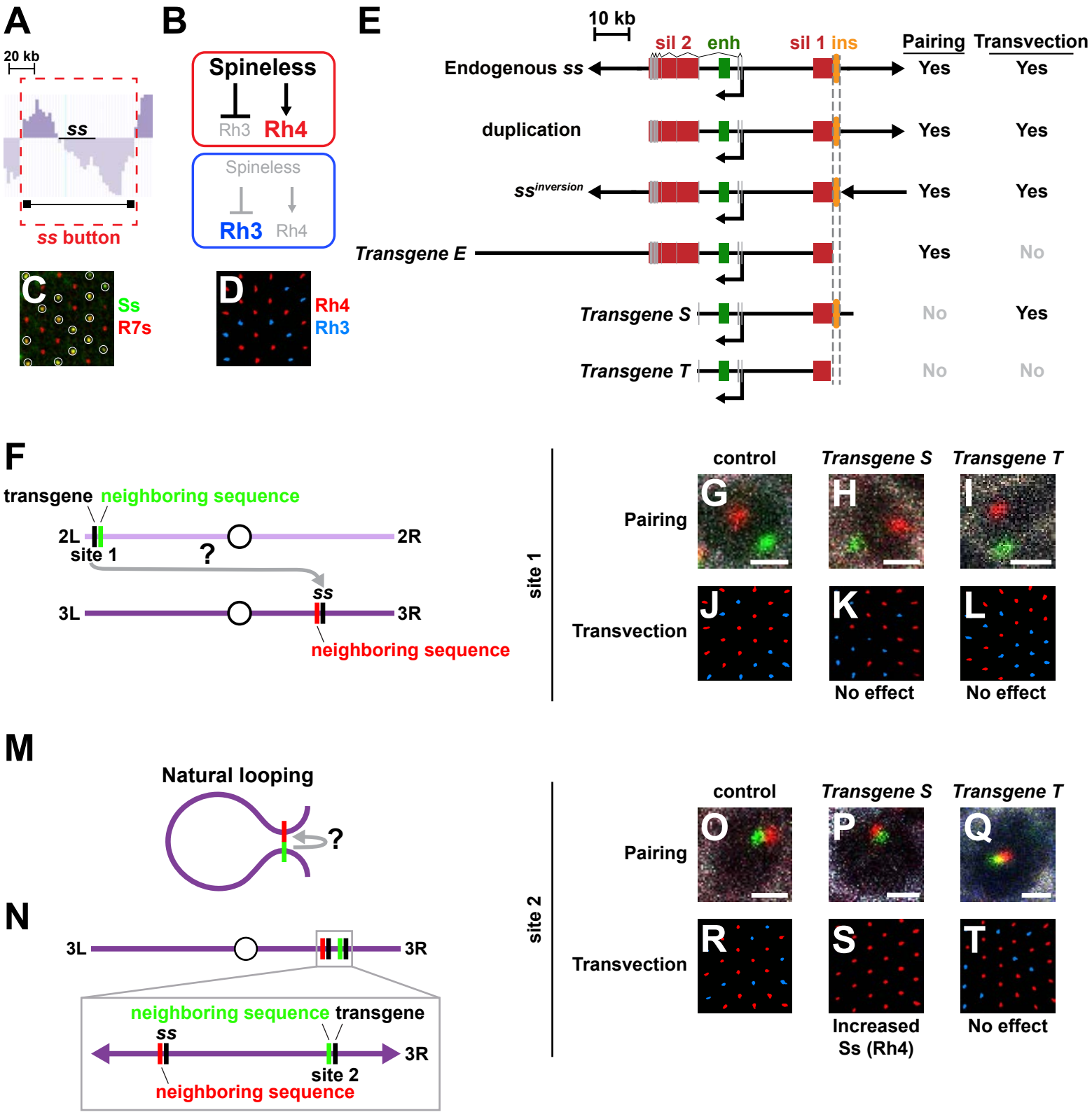

A
20 kb
ss
ss button
B
Spineless
Rh3
Rh4
Spineless
Rh3
Rh4
C
Ss
R7s
D
Rh4
Rh3
E
10 kb
sil 2
enh
sil 1 ins
Pairing
Transvection
Endogenous ss
Yes
Yes
duplication
Yes
Yes
ss inversion
Yes
Yes
Transgene E
Yes
No
Transgene S
No
Yes
Transgene T
No
No
F
transgene
neighboring sequence
2L
site 1
?
2R
3L
ss
3R
neighboring sequence
site 1
control
Transgene S
Transgene T
Pairing
G
H
I
Transvection
J
K
L
No effect
No effect
M
Natural looping
?
N
3L
3R
neighboring sequence
transgene
ss
site 2
3R
neighboring sequence
site 2
control
Transgene S
Transgene T
Pairing
O
P
Q
Transvection
R
S
T
Increased
Ss (Rh4)
No effect

Figure 5

Figure 5: Pairing is necessary but not sufficient for transvection.

A. Representative directionality index (NCBI GSE38468) showing the TAD that defines the *ss* button. Black bar: TAD. See **Fig. S5A** for TAD assessment.

B. Spineless (Ss) activates Rh4 and represses Rh3.

C. Ss is expressed in ~70% of R7s. Green: Ss, red: Prospero (R7 marker), white circles: Ss-expressing R7s.

D. Rh3 (blue) and Rh4 (red) expression in wild type R7s.

E. *ss* alleles and transgenes. ins: insulator, sil 1: silencer 1, enh: enhancer, sil 2: silencer 2. Smaller black arrows: transcription start sites. Gray rectangles: exons. Dotted gray lines: region required for transvection.

F. Strategy used to assess pairing and transvection from site 1 in **Fig. 5G-L**. Gray arrow with "?" indicates that *Transgenes S* and *T* were tested for transvection.

G-I, O-Q. Scale bars= 1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.

J-L, R-T. Red: Rh4, blue: Rh3.

G-I. Pairing assay images of *2L-3R control*, *Transgene S site 1,* and *Transgene T site 1*. See **Fig. S14Q** for quantifications. Image for 2L-3R control is from the same experiment as **Fig. 2B, 3B,** and **S16A**.

J-L. Rh3 and Rh4 expression in *wild type control* (Ss(Rh4)=70%), *Transgene S site 1* (Ss(Rh4)=57%), and *Transgene T site 1* (Ss(Rh4)=55%). The slight decrease in Rh4 frequency for *Transgene S site 1* and *Transgene T site 1* is likely due to background genetic effects.

M. Natural chromosome looping forces transgenes into close proximity with endogenous *ss*, mimicking pairing and facilitating transvection. Gray arrow with "?" indicates that *Transgenes S* and *T* were tested for transvection.

N. Strategy used to assess pairing and transvection from site 2 in **Fig. 5O-T**.

O-Q. Pairing assay images of *site 2 control*, *Transgene S site 2*, and *Transgene T site 2*. See **Fig. S14Q** for quantifications.

R-T. Rh3 and Rh4 expression in *wild type control* (Ss(Rh4)=70%), *Transgene S site 2* (Ss(Rh4)=98%) and *Transgene T site 2* (Ss(Rh4)=78%).

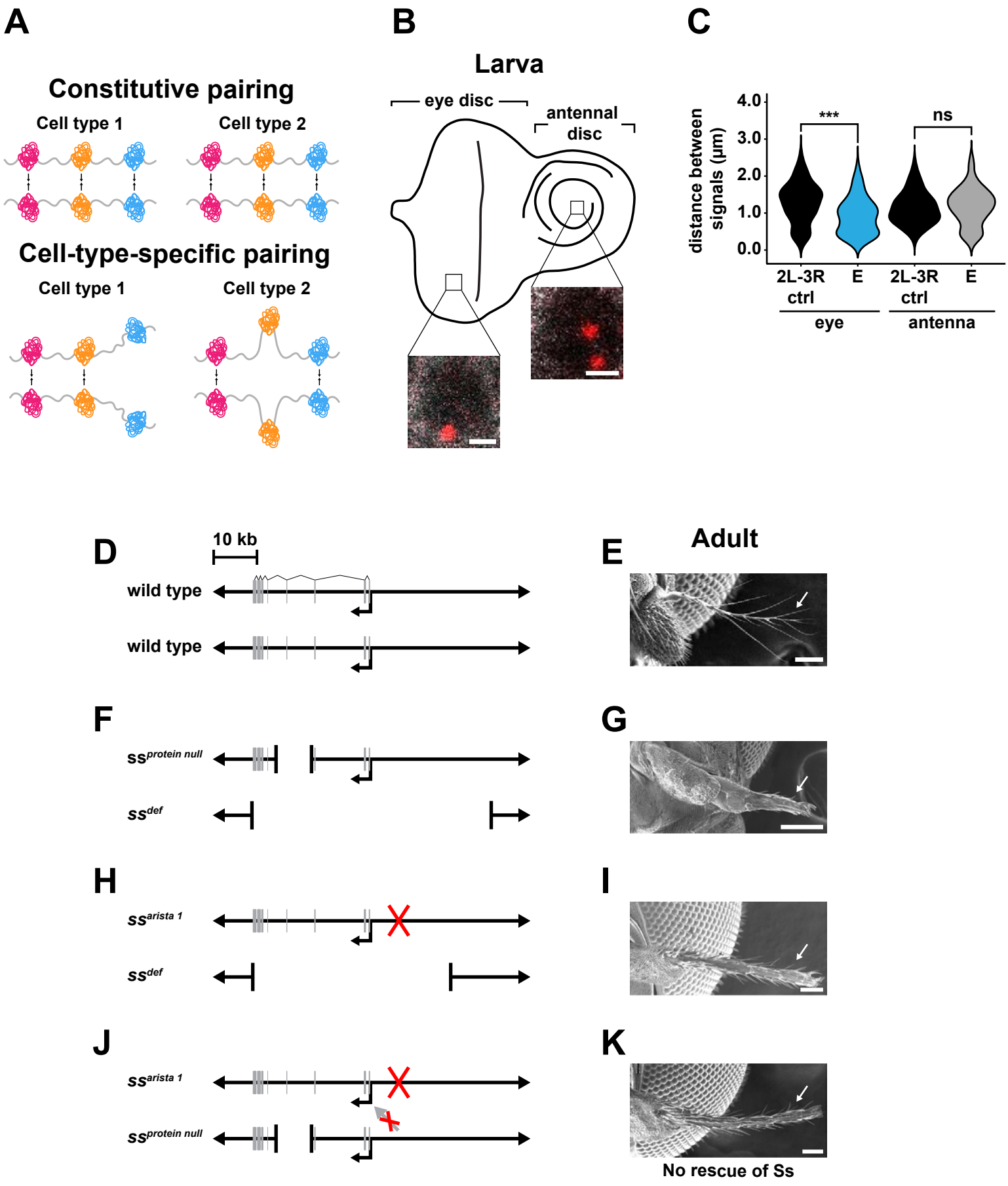

A
Constitutive pairing
Cell type 1
Cell type 2
Cell-type-specific pairing
Cell type 1
Cell type 2
B
Larva
eye disc
antennal disc
C
distance between signals (µm)
4.0
3.0
2.0
1.0
0.0
***
ns
2L-3R
E
2L-3R
E
ctrl
ctrl
eye
antenna
D
10 kb
wild type
wild type
E
Adult
F
ss protein null
ss def
G
H
ss arista 1
ss def
I
J
ss arista 1
ss protein null
K
No rescue of Ss

Figure 6

Figure 6: *ss* pairing and transvection are cell-type-specific.

A. Constitutive vs. cell-type-specific pairing models.

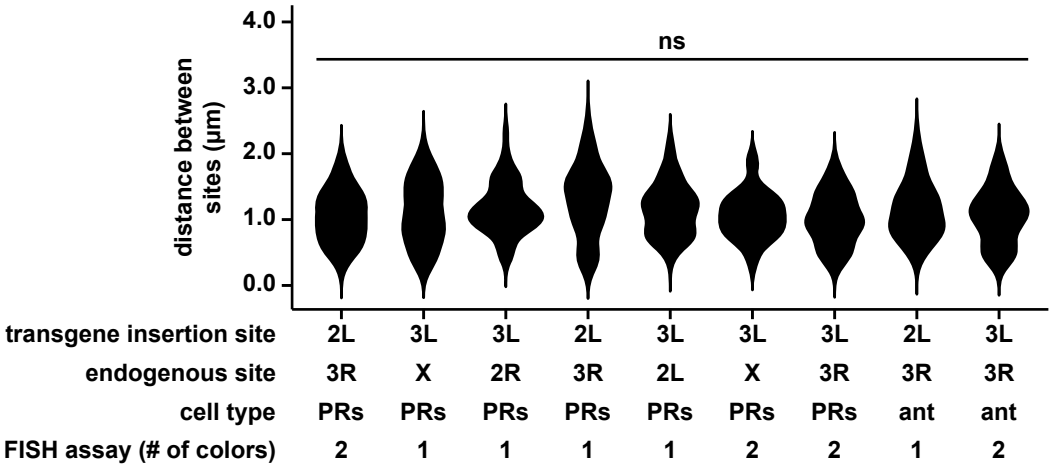
B. Third instar larval eye-antennal disc. The *ss* button drove pairing in the larval eye but not the larval antenna. Scale bars=1 μm. White: Lamin B, red: probes against endogenous *ss* and *Transgene E*.

C. Quantifications for **Fig. 6B**. Black: control, blue: pairer, gray: non-pairer. ***=p<0.05, Wilcoxon rank-sum test; ns=p>0.05, unpaired t-test with Welch's correction. Data for 2L-3R eye control are the same as in **Fig. 2F, 3E** (2L-3R control), **3I, S1A-B, S12J,** and **S14Q**. Data for 2L-3R antenna control are the same as in **Fig. S1A** (2L-3R antenna control). Controls were imaged in two colors, then pseudocolored red and scored in one color.

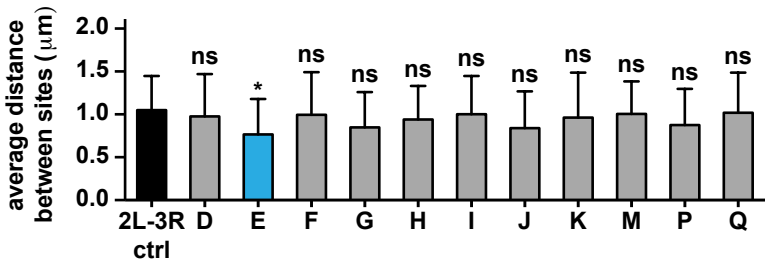
D, F, H, J. Genotypes tested for transvection. Gray rectangles: exons. Smaller black arrows: transcription start sites. Red X indicates an uncharacterized mutation in the $ss^{arista\ 1}$ sequence. Red X over gray arrow indicates an absence of transvection between alleles in the arista.

E, G, I, K. Arista phenotype. Scale bars=50 μm. White arrows indicate arista. Image for **Fig. 6I** is from the same experiment as **Fig. S17C**.

A



B

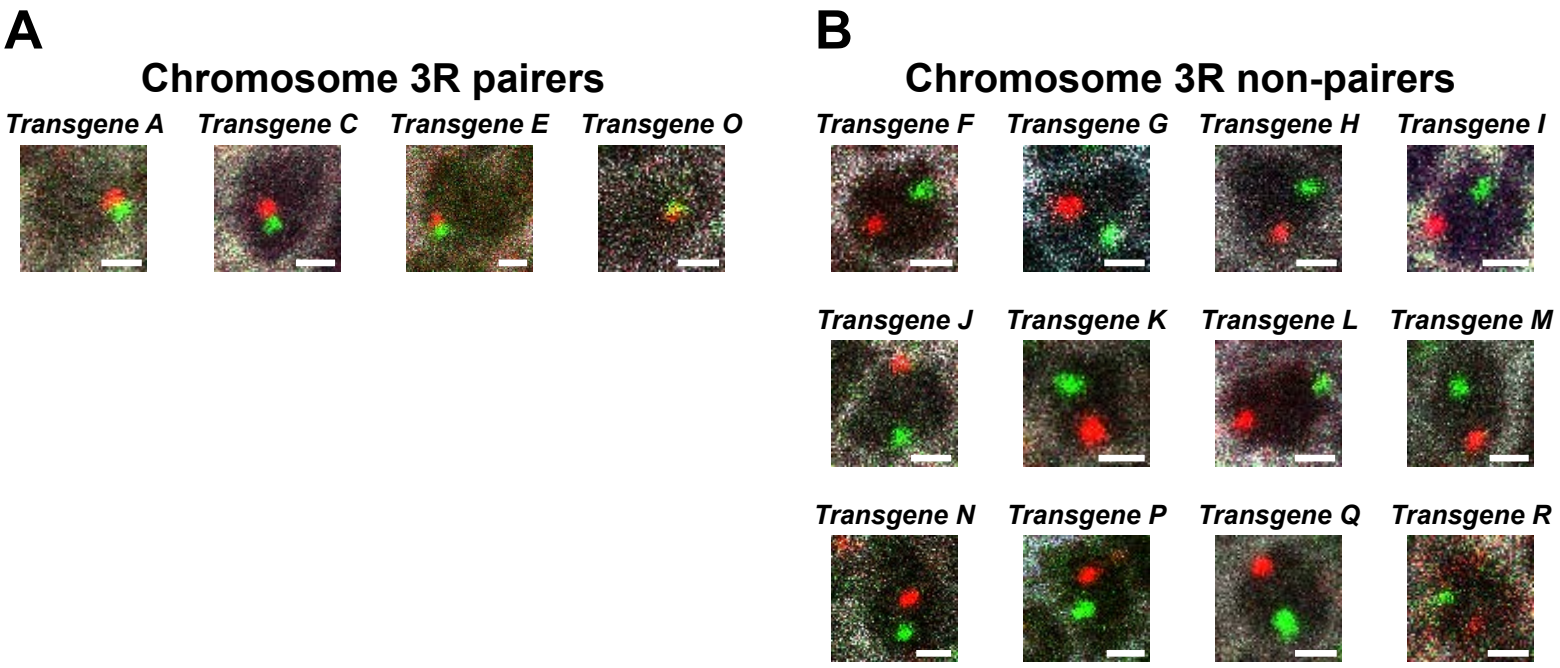


Supplemental Fig. 1

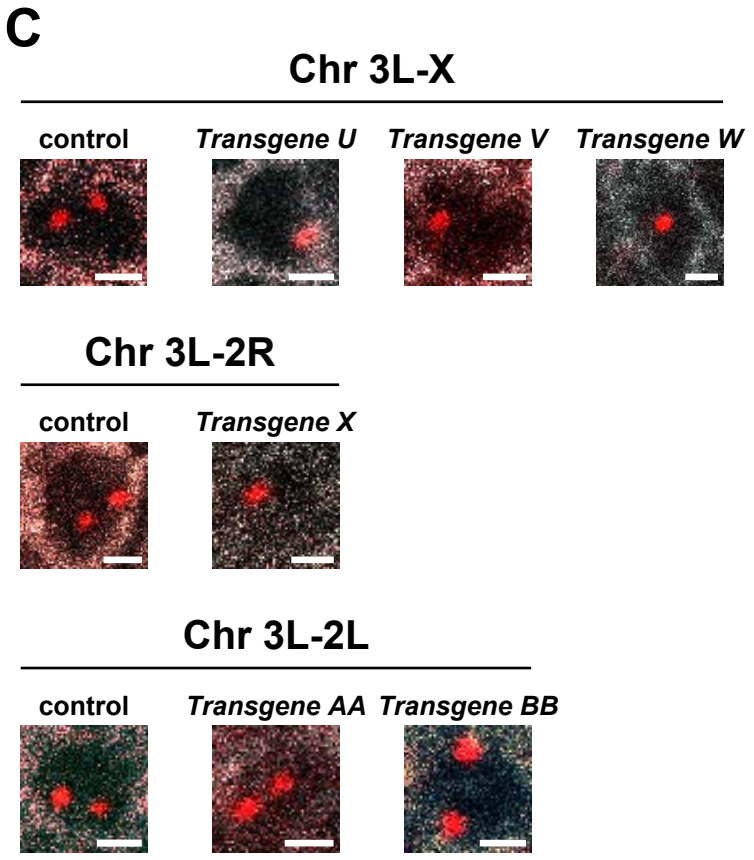
Supplemental Figure 1: Comparisons between pairing controls and additional statistical tests confirm identification of pairers.

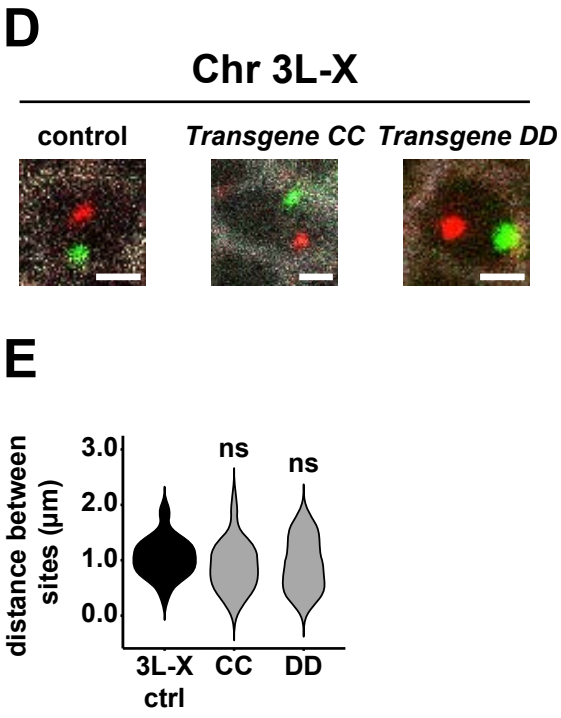
A. Quantifications for all negative controls used to assess pairing between a transgene and its endogenous site in this study. PRs: photoreceptors, ant: antenna. ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test. Data for 2L-3R PRs 2-color control are the same as in **Fig. 2F, 3I, S1B,** and **S14Q** (2L-3R control). Data for 3L-X 1-color control are the same as in **Fig. 3E** (3L-X control). Data for 3L-2R control are the same as in **Fig. 3E** (3L-2R control). Data for 2L-3R PRs 1-color control are the same as in **Fig. 3E** (2L-3R control), **6C** (2L-3R eye control), and **S12J.** Data for 3L-2L control are the same as in **Fig. 3E** (3L-2L control). Data for 3L-X 2-color control are the same as in **Fig. S2E**. Data for 3L-3R control are the same as in **Fig. S4D, S11H, S14Q** (3L-3R control). Data for 2L-3R antenna control are the same as in **Fig. 6C** (2L-3R antenna control). Data for 3L-3R antenna are the same as in **S16F**. All 1-color controls were imaged in two colors, then pseudocolored red and scored in one color.

B. Comparison of the means for all datasets in **Fig. 2F** fit to a single Gaussian distribution by maximum likelihood estimation. Black: control, blue: pairers, gray: non-pairers. *=p<0.05, ns=p>0.05, ordinary one-way ANOVA with Dunnett's multiple comparisons test. Data for 2L-3R control are the same as in **Fig. 2F, 3E** (2L-3R control), **6C** (2L-3R eye control), **S1A, S12J,** and **S14Q** (2L-3R control). Data for *Transgene D-G* are the same as in **Fig. 2A** and **3I**. Data for *Transgene H-K, M, P,* and *Q* are the same as in **Fig. 2A**.



Additional transgenes





Supplemental Fig. 2

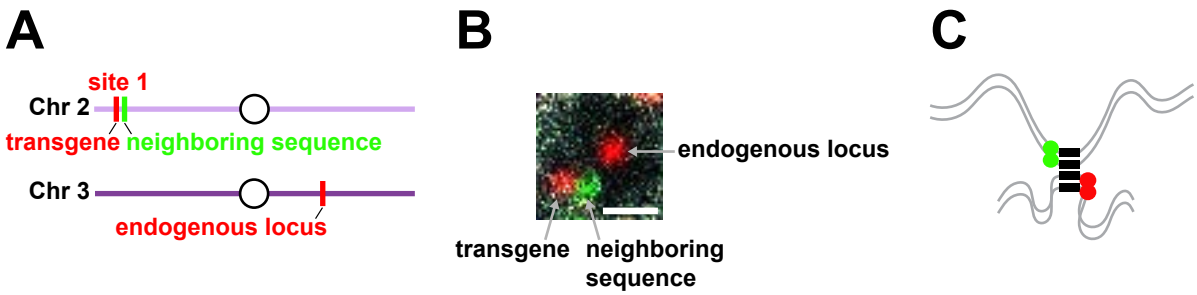
Supplemental Figure 2: A subset of transgenes interspersed across the genome drive pairing.

A-B. Pairer and non-pairer transgenes from the initial screen of a ~1 Mb region of chromosome 3R. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.

C. Additional transgenes taken from chromosomes X (*Transgenes U-W*), 2R (*Transgene X*), and 2L (*Transgenes AA and BB*) and inserted at site 3 on Chr 3L. 3L-X control image is from the same experiment as **Fig. S2D** control.

D. Additional transgenes taken from chromosome X and inserted at site 3. Pairing was assessed with a two-color FISH strategy. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site. Control image is from the same experiment as **Fig. S2C** (3L-X control).

E. Quantifications for transgenes in **Fig. S2D**. Neither transgene contained a TAD. Black: control, gray: non-pairers. ns=$p>0.05$, one way ANOVA on ranks with Dunn's multiple comparisons test. Control data are the same as in **Fig. 3E** (3L-X control) and **S1A**.



Supplemental Fig. 3

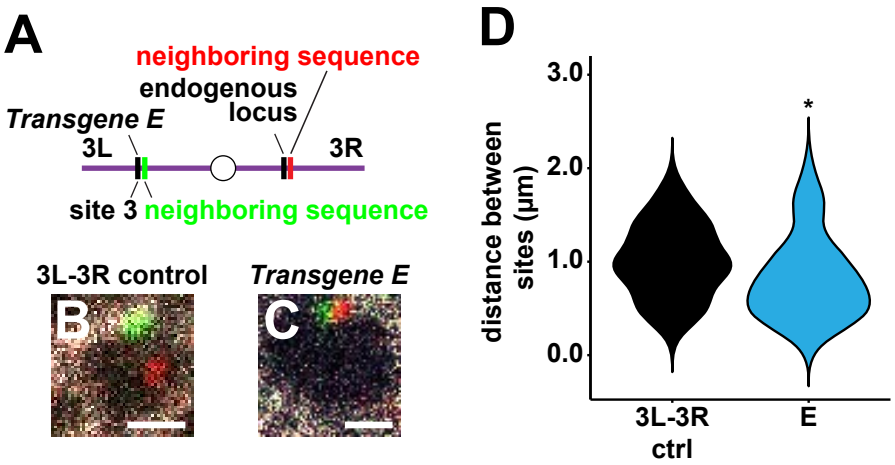
Supplemental Figure 3: Probes neighboring paired sequences give offset probe signals.

A. Schematic of FISH strategy used to label transgene and endogenous sequences and the region directly neighboring the transgene insertion site.

B. Probes directly neighboring the transgene insertion site could be distinguished from probes labeling the transgene sequence itself, despite being immediately downstream on the DNA. Image is from the same experiment as **Fig. 3D**.

C. When a transgene drives pairing with its endogenous site, the two copies of the transgene are paired with each other and the two copies of the endogenous site are paired with each other due to homologous chromosome pairing. Therefore, one green FISH puncta (neighboring the transgene insertion site) and one red FISH puncta (neighboring the endogenous site) are observed.

The experiment in **Fig. S3B** showed that two sets of probes targeting neighboring regions on the DNA could be distinguished from each other. As the red and green probes in **Fig. S3C** are neighboring the paired sites, not directly targeting the paired sites, it is therefore feasible that their signals do not completely overlap, despite being close in 3D space.



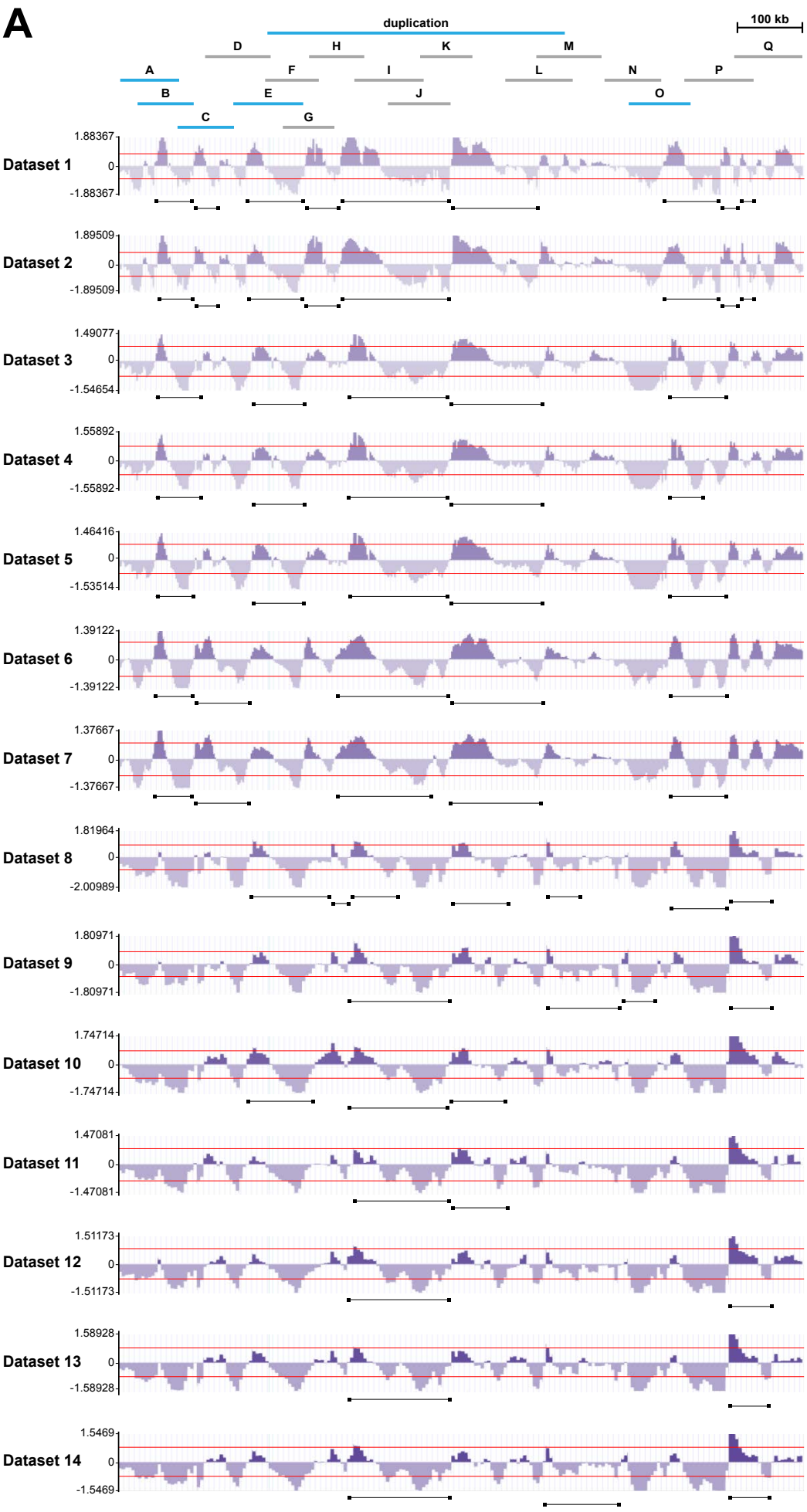
Supplemental Fig. 4

Supplemental Figure 4: Buttons drive pairing in a position-independent manner.

A. DNA FISH strategy used to assess pairing of *Transgene E* inserted at site 3 with its endogenous locus.

B-C. Control and *Transgene E* at site 3. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site. Image for **S4B** is from the same experiment as in **S11D** and **S14B**.

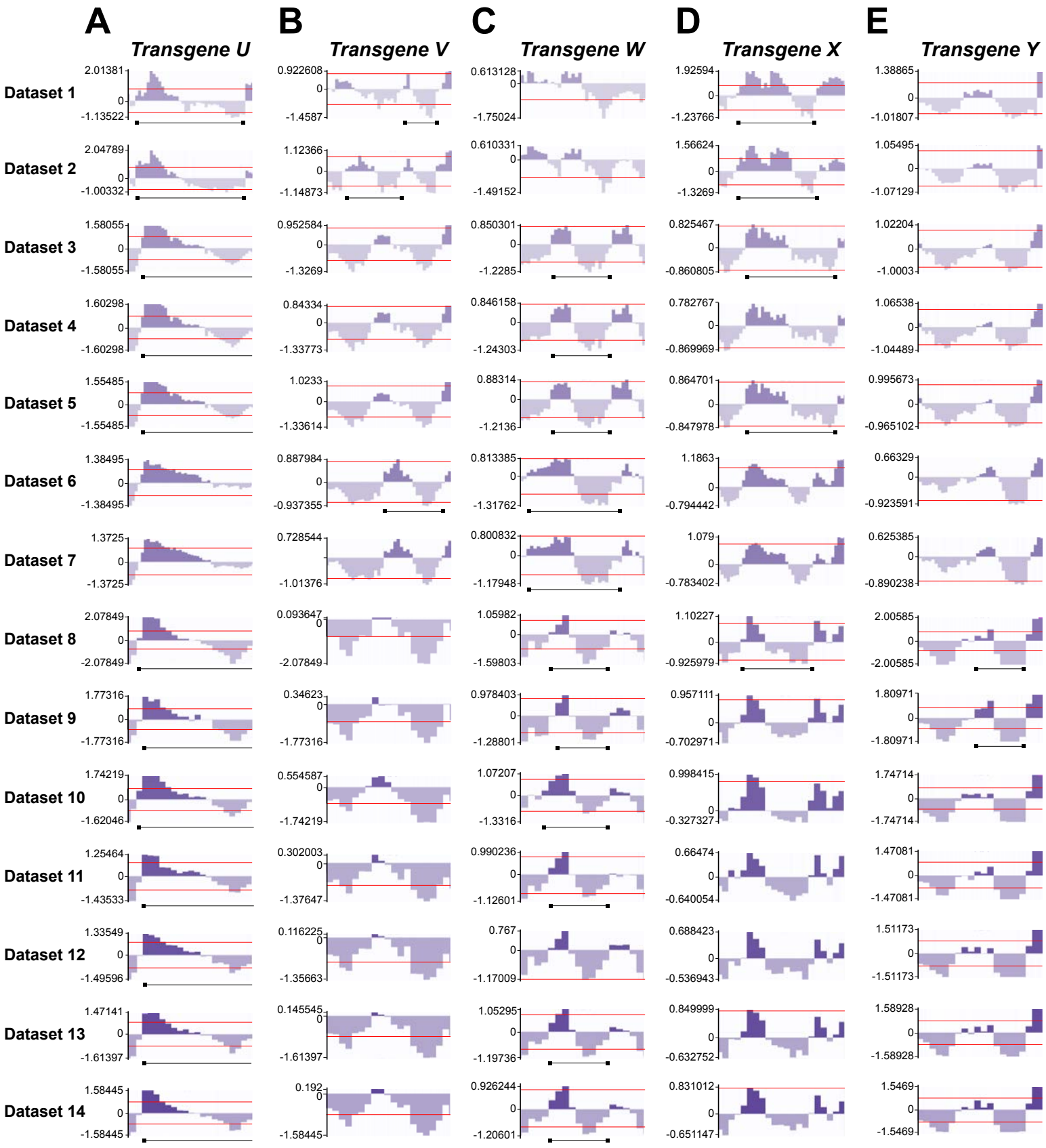
D. Quantifications for **Fig. S4B-C**. Black: control, blue: pairer. *=$p<0.05$, Wilcoxon rank-sum test. Data for control are the same as in **Fig. S1A** (3L-3R control), **S11H,** and **S14Q** (3L-3R control).



Supplemental Fig. 5

Supplemental Figure 5: TAD calls across 14 Hi-C datasets for the region on chromosome 3R used for the initial pairing screen.

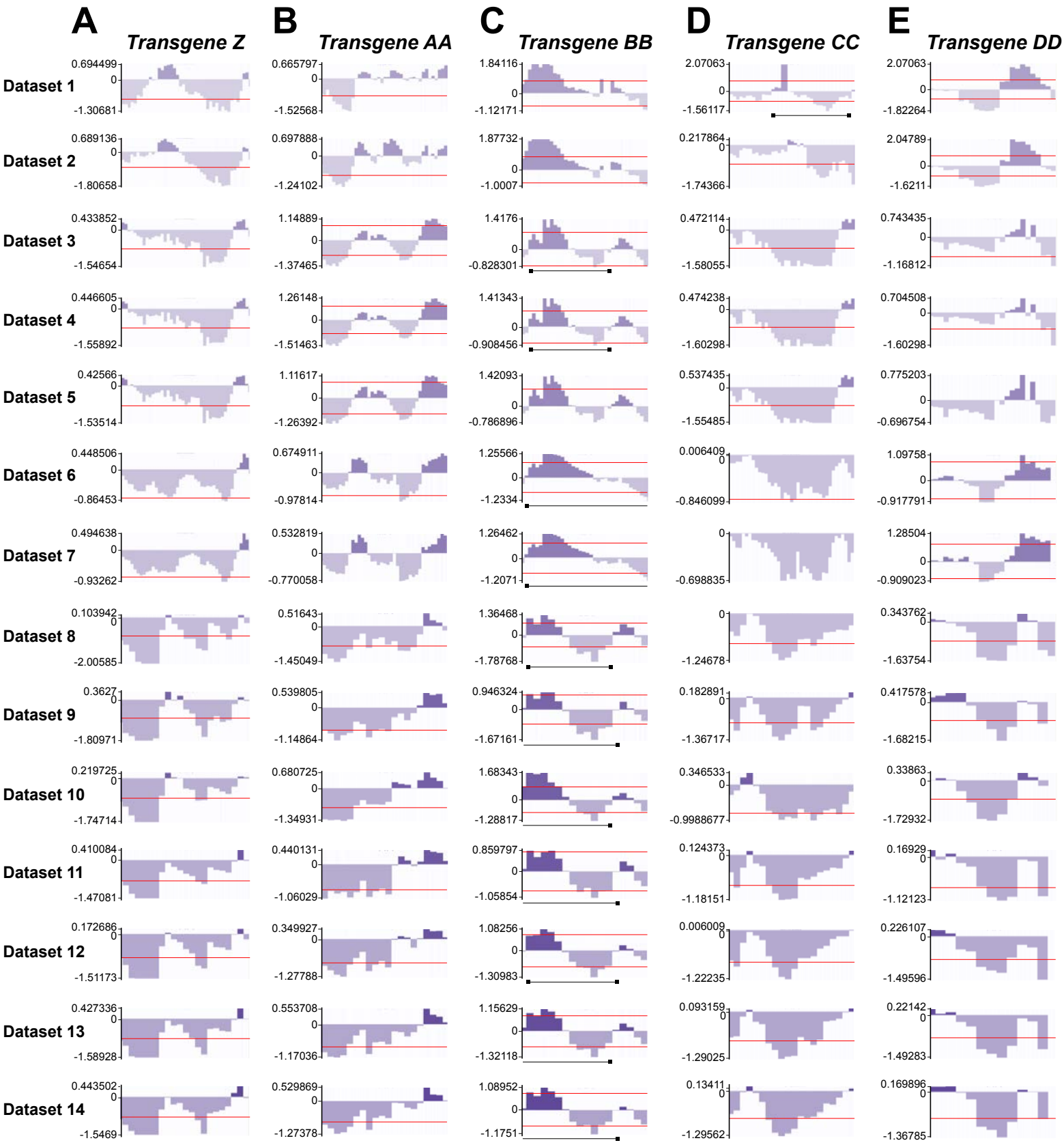
A. Red lines indicate a directionality index signal of 0.8 or -0.8, the cutoff for a TAD. Black bars indicate TAD calls.



Supplemental Fig. 6

Supplemental Figure 6: TAD calls across 14 Hi-C datasets for Transgenes U-Y.

A-E. Red lines indicate a directionality index signal of 0.8 or -0.8, the cutoff for a TAD. Black bars indicate TAD calls. Directionality indices are shown for the entire region spanned by each transgene.



Supplemental Fig. 7

Supplemental Figure 7: TAD calls across 14 Hi-C datasets for Transgenes Z-DD.

A-E. Red lines indicate a directionality index signal of 0.8 or -0.8, the cutoff for a TAD. Black bars indicate TAD calls. Directionality indices are shown for the entire region spanned by each transgene.

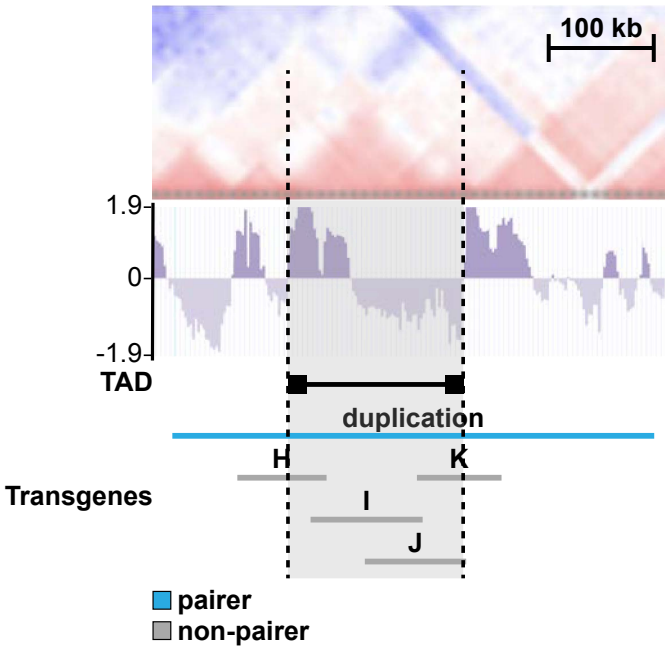
Supplemental Fig. 8

Supplemental Figure 8: A higher percentage of pairers encompass entire TADs than non-pairers.

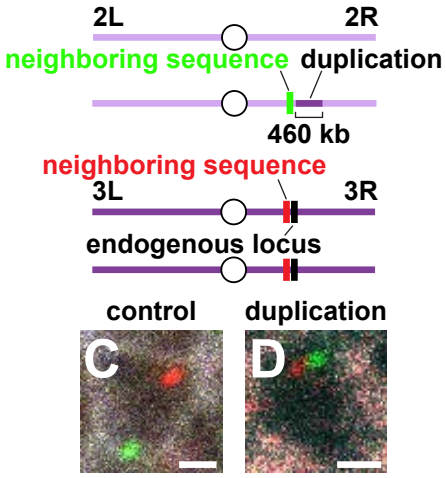
A. Representative directionality indices showing the percentage of a TAD covered by each pairing transgene. Black bars indicate consensus TAD calls generated from analysis of 14 Hi-C datasets (*177-180*). Representative directionality indices are from NCBI accession numbers GSE38468 (*Transgenes C, U)*, GSE61471 (*Transgenes A, B, E, O, W, X*), GSM2679637 (*Transgene Y*), and GSE63515 (*Transgene V*). Directionality indices are shown for the entire region spanned by each transgene.

B. Representative directionality indices showing the percentage of a TAD covered by each non-pairing transgene. Black bars indicate consensus TAD calls generated from analysis of 14 Hi-C datasets(*177-180*). Representative directionality indices are from NCBI accession numbers GSE38468 (*Transgenes G, H, N, P*), GSE61471 (*Transgenes D, F, I, J, K, L, M, Z, BB, CC*), GSE63515 (*Transgene DD*), GSM2679644 (*Transgene Q*), and GSM2679637 (*Transgene AA*). Directionality indices are shown for the entire region spanned by each transgene.

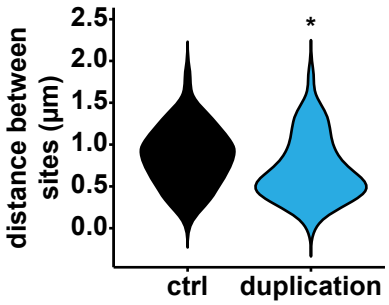
A



B



E



Supplemental Fig. 9

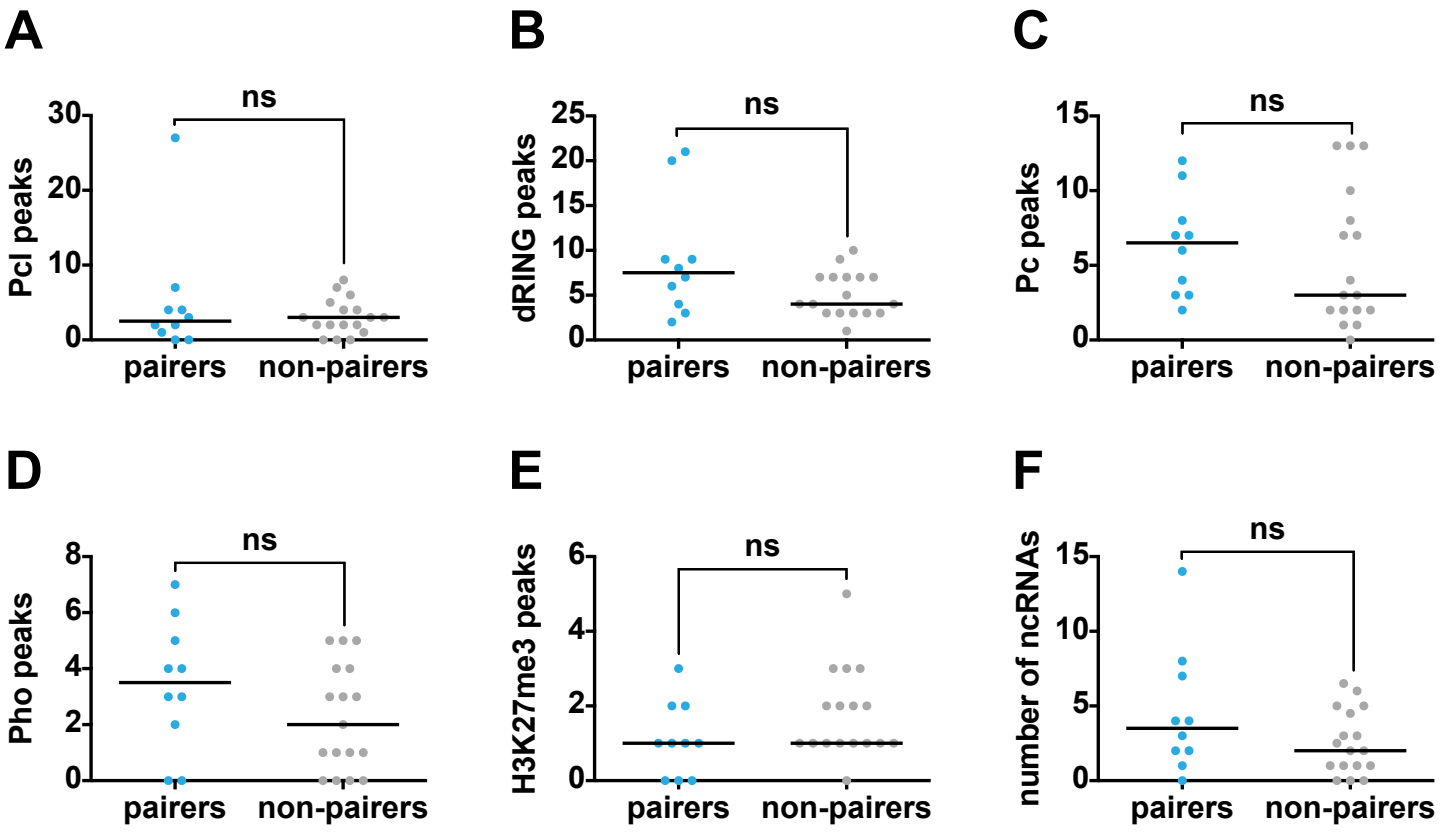
Supplemental Figure 9: A 460-kb duplication encompassing a TAD drives pairing.

A. Representative HiC heat map and directionality index (NCBI GSE38468) showing large TAD covered by the duplication. Dotted lines: TAD boundaries. Black bar: TAD. See **Fig. S5A** for TAD assessment.

B. DNA FISH strategy used to assess pairing of the duplication with its endogenous locus.

C-D. Control and duplication. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous locus, green: probes neighboring duplication breakpoint.

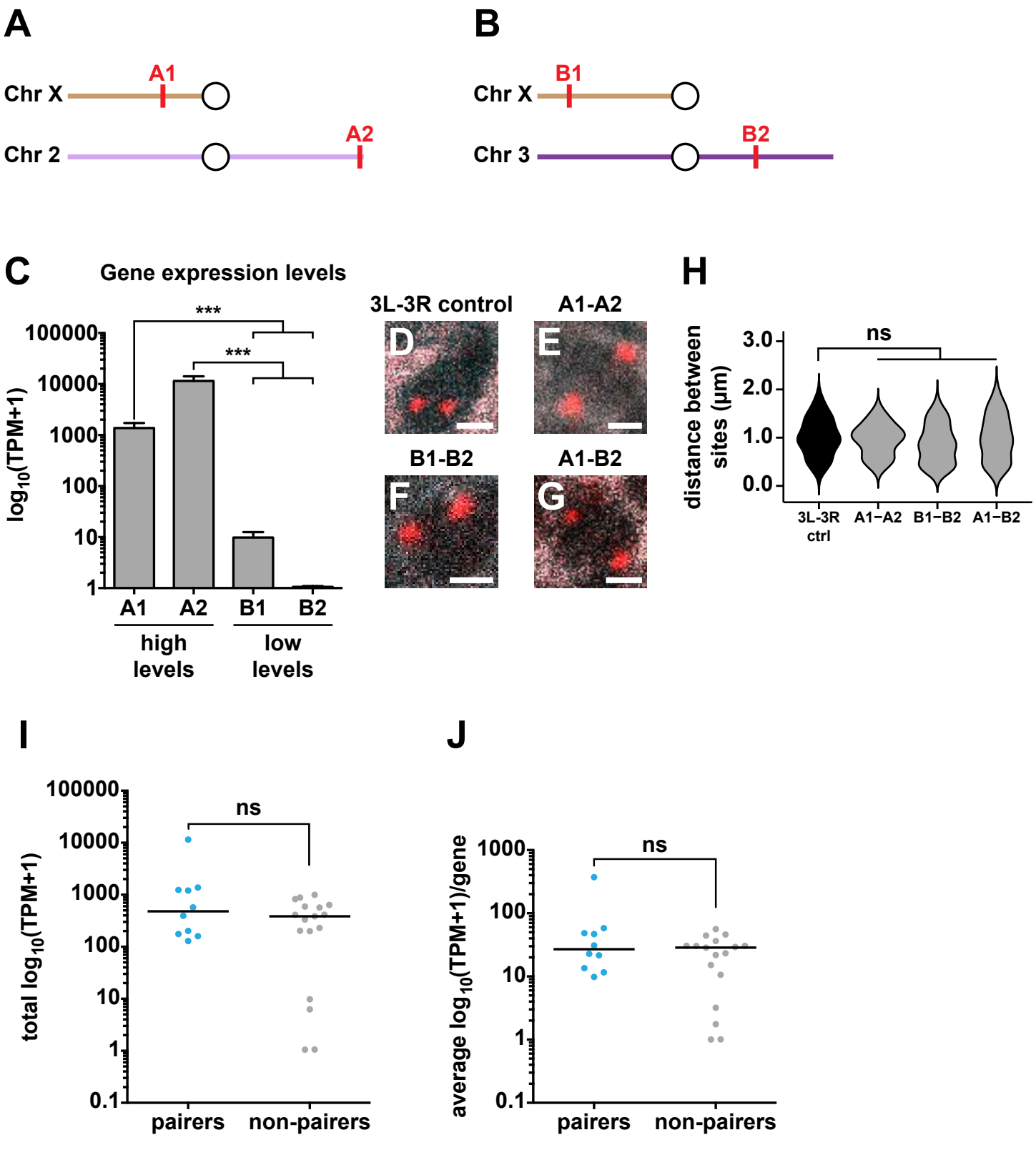
E. Quantifications for **Fig. S9C-D**. Black: control, blue: pairer. *=p<0.05, Wilcoxon rank-sum test.



Supplemental Fig. 10

Supplemental Figure 10: Polycomb Group Complex binding sites, repressive chromatin marks, and ncRNAs do not account for pairing.

A-F. Graphs showing the number of Polycomb Group Complex or H3K27me3 ChIP peaks or the number of ncRNAs per transgene for all pairers and non-pairers tested in **Fig. 2F**, **3E**, and **S2E**. Blue: pairers, gray: non-pairers. ns=p>0.05, Wilcoxon rank-sum test. Black lines indicate medians.



Supplemental Fig. 11

Supplemental Figure 11: Compartmentalization and gene expression state do not account for pairing.

A-B. Schematics indicating genomic locations of high-expressing loci A1 and A2 (**A**) and low-expressing loci B1 and B2 (**B**).

C. TPM+1 values for A and B compartment regions tested for interactions.

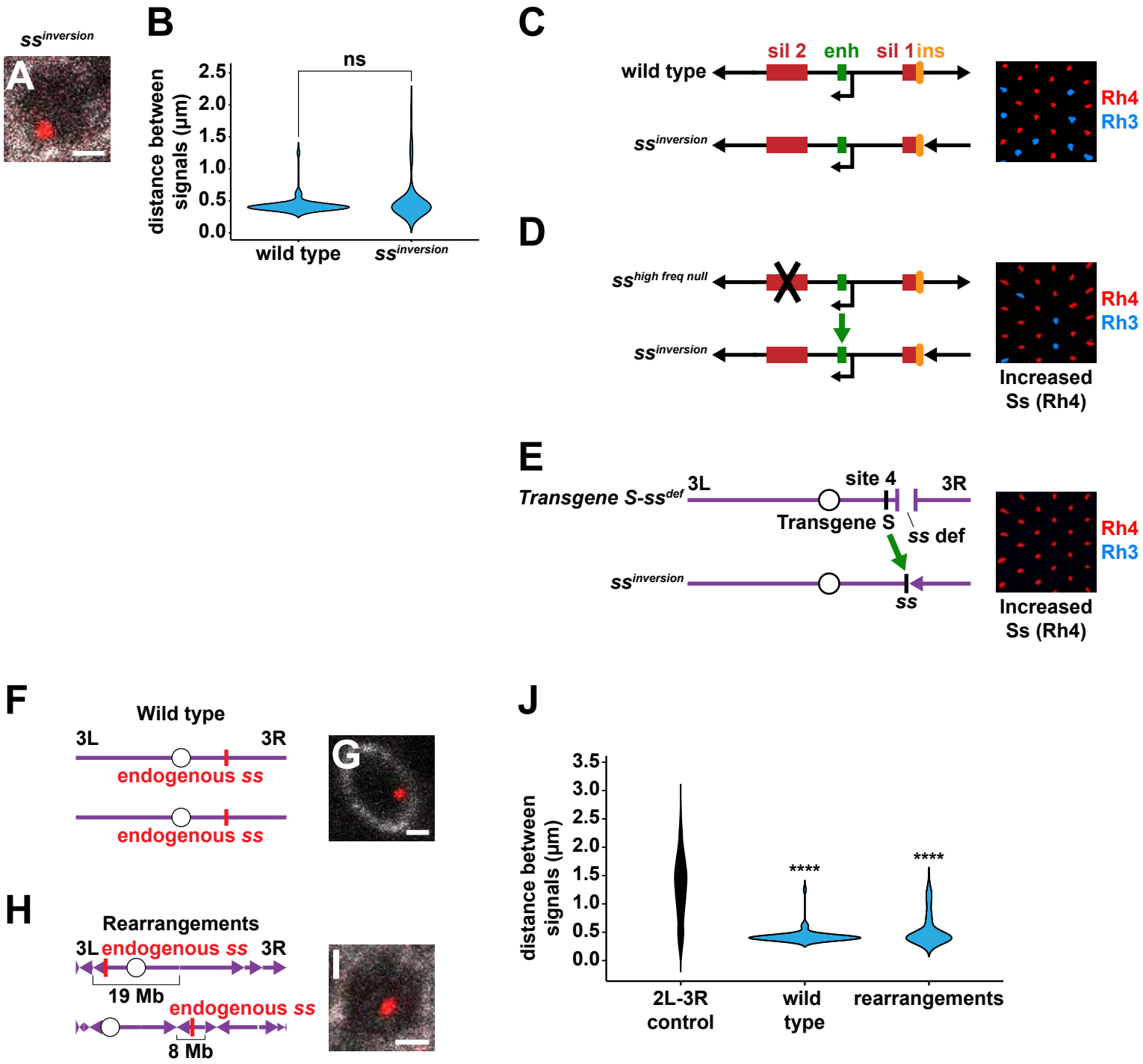
D-G. Example images for 3L-3R control, A1-A2, B1-B2, and A1-B2 compartment experiments. Scale bars=1 μm. White: Lamin B, red: probes against transgene insertion site and endogenous site (**D**) or two compartment sites (**E-G**). 3L-3R control image is from the same experiment as in **Fig. S4B** and **S14D.**

H. Quantifications for **S11D-G**. Black: control, gray: non-pairers. ns=p>0.05, one way ANOVA on ranks with Dunn's multiple comparisons test. 3L-3R control data are the same as in **Fig. S1A** (3L-3R control), **S4D,** and **S14Q** (3L-3R control).

I-J. ns=p>0.05, Wilcoxon rank-sum test. Black lines indicate medians.

I. Total TPM+1 per transgene.

J. Average TPM+1/gene for each transgene.



Supplemental Fig. 12

Supplemental Figure 12: The *ss* button drives pairing and transvection despite chromosome rearrangements.

A. $ss^{inversion}$ / + example. Scale bar=1 μm. White: Lamin B, red: probes against endogenous *ss*.

B. Quantification of **Fig. S12A**. ns=p>0.05, Wilcoxon rank-sum test. Data for wild type are the same as in **Fig. S12J**. Blue: pairers.

C. Schematic and representative image of $ss^{inversion}$ / + adult R7s. Ss(Rh4)=63%. $ss^{inversion}$ / + had no effect on the normal Rh3:Rh4 ratio. ins: insulator, sil 1: silencer 1, enh: enhancer, sil 2: silencer 2. Smaller black arrows: transcription start sites. Red: Rh4, blue: Rh3.

D. Schematic and representative image of $ss^{inversion}$ / $ss^{high\ freq\ null}$ adult R7s. Ss(Rh4)=80%. $ss^{high\ freq\ null}$ produces no functional Ss protein, but it performs transvection to increase the expression frequency of *ss* on other chromosomes (*97*). $ss^{high\ freq\ null}$ upregulated expression frequency from $ss^{inversion}$, indicating that $ss^{inversion}$ performed transvection. Black X indicates that there is a mutation in the second silencer of *ss* that disrupts the protein coding sequence of $ss^{high\ freq\ null}$. Smaller black arrows: transcription start sites. Green arrow indicates upregulation of *ss* between chromosomes. Red: Rh4, blue: Rh3.

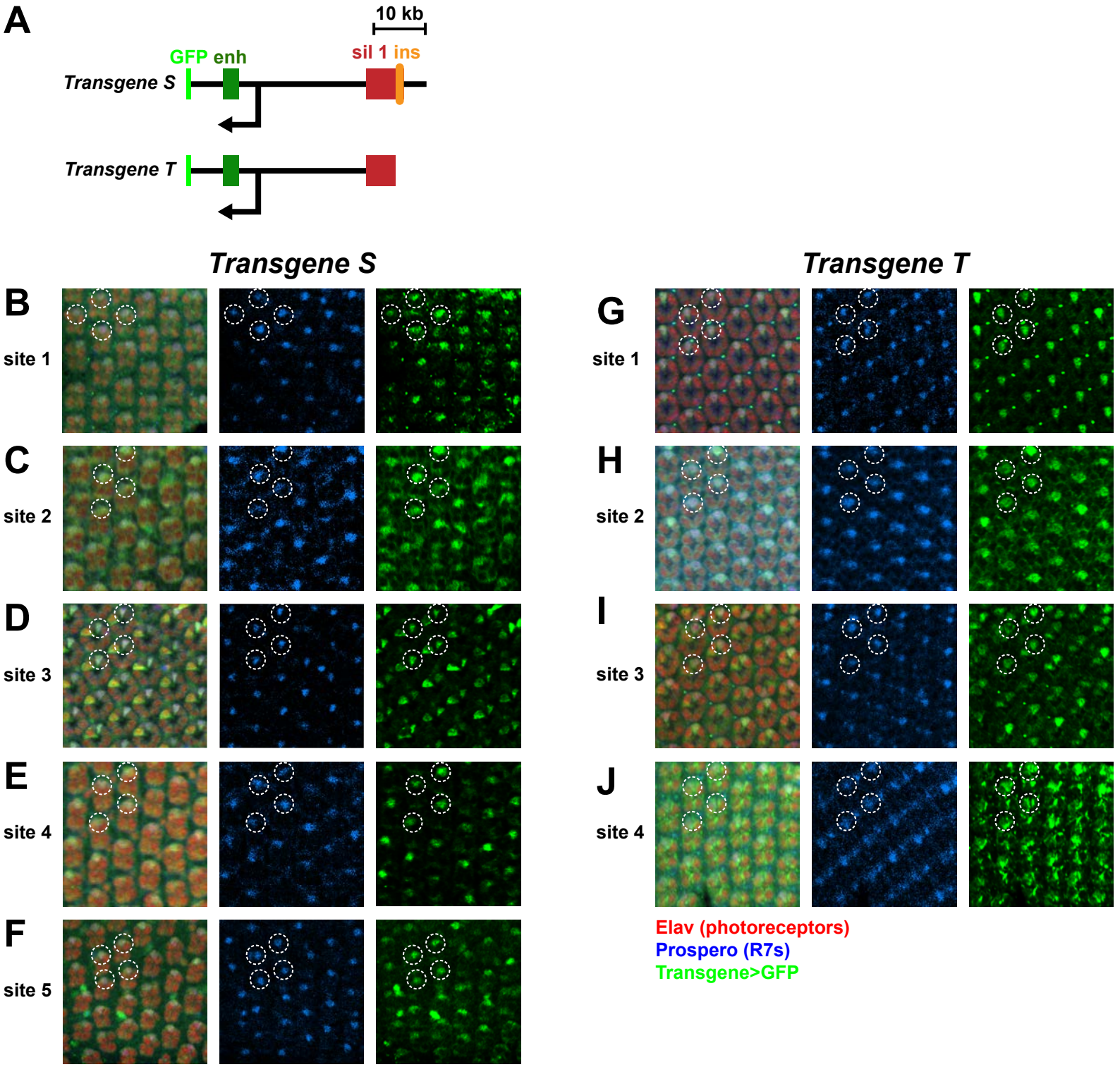
E. Schematic and representative image of *Transgene S*-ss^{def} / $ss^{inversion}$ adult R7s. Ss(Rh4)=99%. *Transgene S* was recombined onto a chromosome with a *ss* deficiency to examine *Transgene S* transvection with mutant *ss* alleles. *Transgene S* performs transvection to upregulate expression of wild type, endogenous *ss* (*97*). *Transgene S* upregulated expression of *ss* on the $ss^{inversion}$

allele, indicating that *ss*^{inversion} performed transvection. Red: Rh4, blue: Rh3. Green arrow indicates upregulation of *ss* between chromosomes.

F&H. DNA FISH strategies used to assess endogenous *ss* pairing in wild type and chromosome rearrangement backgrounds.

G&I. Wild type and chromosome rearrangement examples. Scale bars=1 μm. White: pm181 (R7 marker)>GFP in **G**, Lamin B in **I**. Red: probes against endogenous *ss.*

J. Quantifications for **Fig. S12G, I**. ****=p<0.0001, one-way ANOVA on ranks with Dunn's multiple comparisons test. 2L-3R control data are the same as in **Fig. 2F, 3E** (2L-3R control), **3I, 6C** (2L-3R eye control), **S1A-B, S14Q** (2L-3R control). Data for wild type are the same as in **Fig. S11B**. Black: control, blue: pairers.



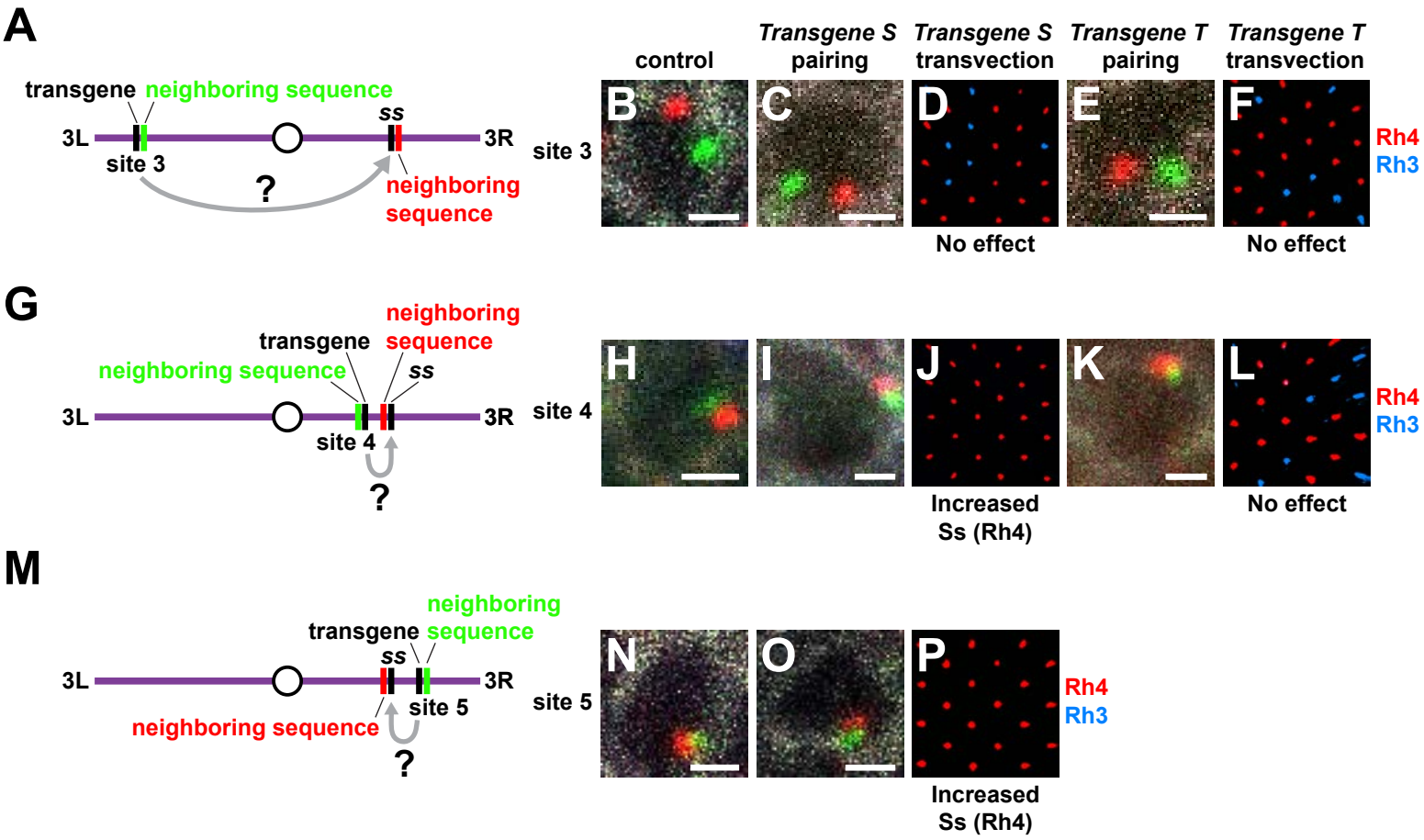
Supplemental Fig. 13

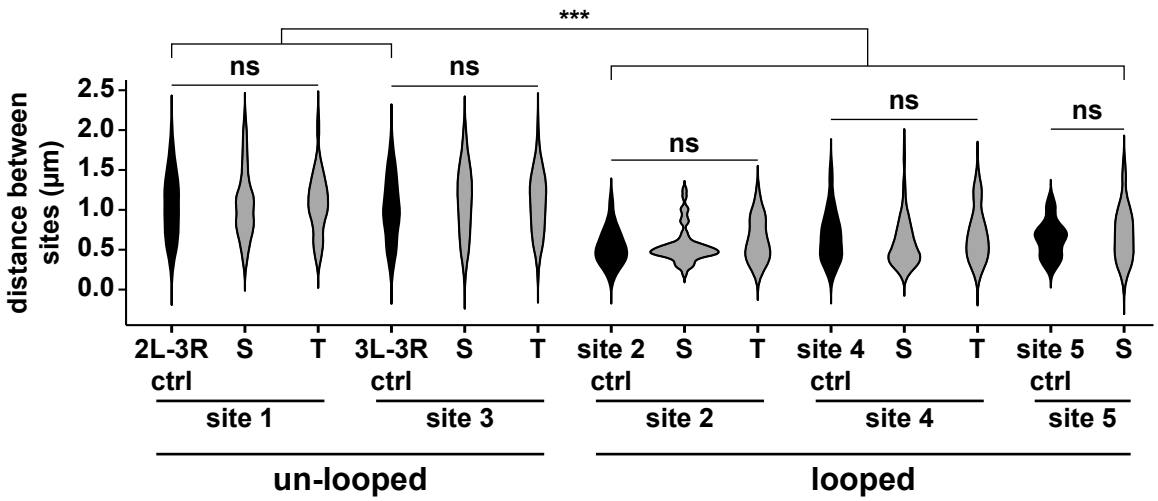
Supplemental Figure 13: *Transgenes S* and *T* are expressed in 100% of R7 photoreceptors at all insertion sites.

A. Schematic of *Transgenes S* and *T* with GFP tags. ins: insulator, sil1: silencer, enh: enhancer. Black arrows: transcription start sites.

B-F. Representative images of *Transgene S*>GFP expression in mid-pupal R7 photoreceptors for all insertion sites. Red: Elav (photoreceptors), blue: Prospero (R7s), green: *Transgene S*>GFP. White circles indicate representative R7s.

G-J. Representative images of *Transgene T*>GFP expression in mid-pupal R7 photoreceptors for all insertion sites. Red: Elav (photoreceptors), blue: Prospero (R7s), green: *Transgene T*>GFP. White circles indicate representative R7s.





Supplemental Fig. 14

Supplemental Figure 14: Pairing is necessary but not sufficient for *ss* transvection.

A, G, M. DNA FISH strategies used to test pairing and transvection of *Transgenes S* and *T* at each insertion site. Gray arrow with "?" indicates that *Transgenes S* and *T* were tested for transvection.

B-C, E, H-I, K, N-O. DNA FISH examples for control, *Transgene S*, and *Transgene T* at each insertion site. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site. Image for site 3 control is from the same experiment as in **S4B** and **S11D**.

D, F, J, L, P. Representative images of adult eyes for *Transgene S* and *Transgene T* at each insertion site. Red: Rh4, blue: Rh3.

D. Ss(Rh4)=71%

F. Ss(Rh4)=76%

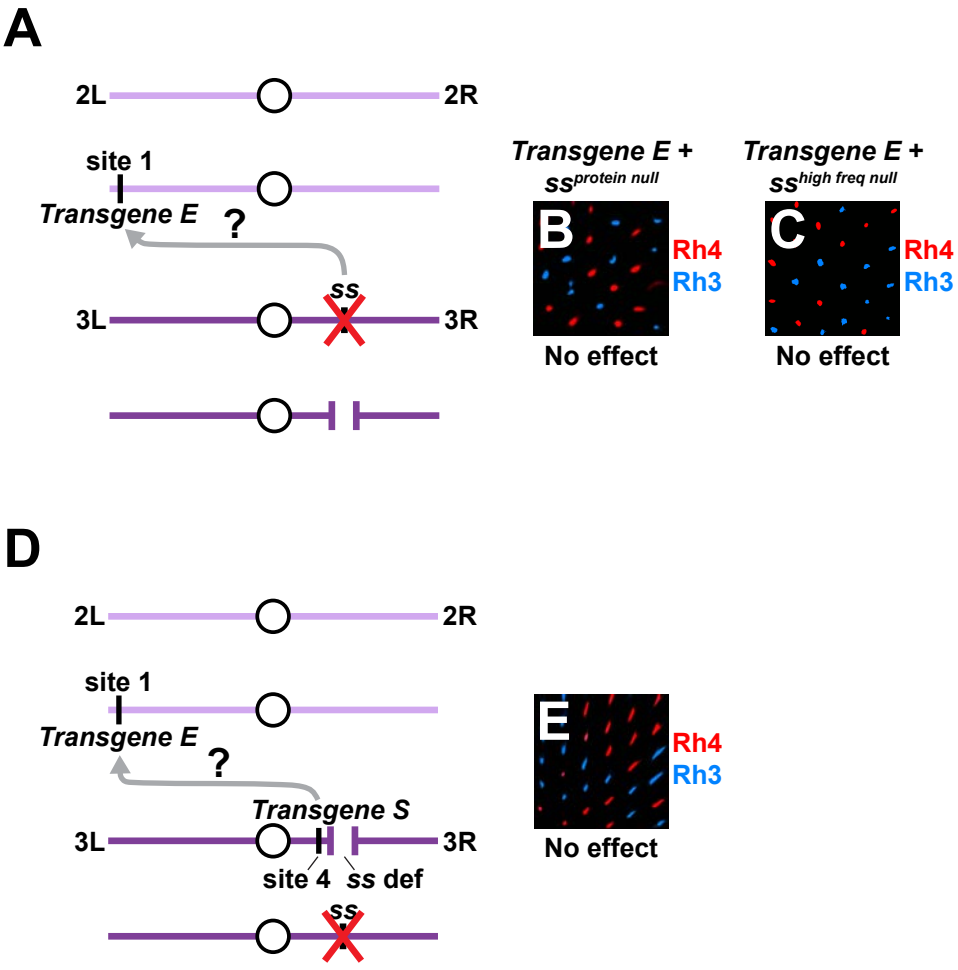
J. Ss(Rh4)=98%

L. Ss(Rh4)=74%

P. Ss(Rh4)=98%

Q. Quantifications for controls, *Transgene S*, and *Transgene T* at each insertion site. Black: control, gray: non-pairer. ***=p<0.0005, ns=p>0.05, ordinary one-way ANOVA with Dunnett's multiple comparisons test (*Transgenes S* and *T site 1* vs. *2L-3R ctrl*, *Transgenes S* and *T site 3* vs. *3L-3R ctrl*), one-way ANOVA on ranks with Dunn's multiple comparisons test (*Transgenes S* and *T site 2* vs. *site 2 control*; *Transgenes S* and *T site 4* vs. *site 4 control*; *site 2 control, Transgenes S*

and *T site 2* vs. *2L-3R control* and vs. *3L-3R control*; *site 4 control, Transgene S* and *T site 4* vs. *2L-3R control* and *3L-3R control*; *site 5 control* and *Transgene S site 5* vs. *2L-3R control* and *3L-3R control*), or unpaired t-test with Welch's correction (*Transgene S site 5*). Data for 2L-3R control are the same as in **Fig. 2F, 3E** (2L-3R control), **3I, 6C** (2L-3R eye control), **S1A-B,** and **S12J**. Data for site 3 control are the same as in **Fig. S4D** and **S11H**. Data for *Transgene S site 1* are the same as in **Fig. 3I.**



Supplemental Fig. 15

Supplemental Figure 15: *Transgene E* does not perform transvection.

A. Transvection assay to test whether mutant *ss* alleles alter expression from *Transgene E.*

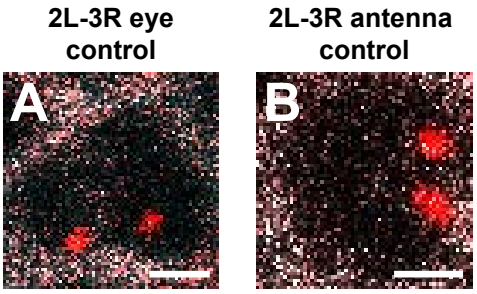
Red X indicates *ss* mutant allele, which is $ss^{protein\ null}$ from **Fig. S15B** or $ss^{high\ freq\ null}$ from **Fig. S15C**. Gray arrow with "?" indicates that *Transgene E* was tested for transvection.

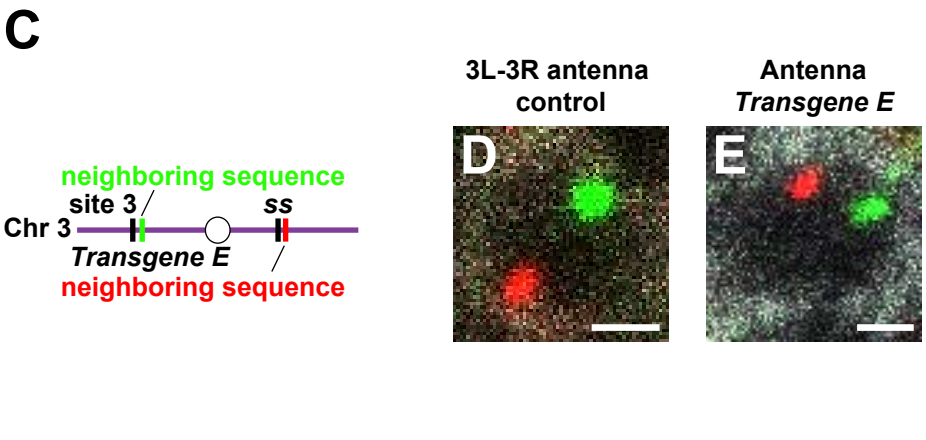
B-C. Rh3 and Rh4 expression in *Transgene E* + $ss^{protein\ null}$ / ss^{def} (Ss(Rh4)=52%) and *Transgene E* + $ss^{high\ freq\ null}$ / ss^{def} (Ss(Rh4)=51%).

D. Schematic of *Transgene E* + *Transgene S*-ss^{def} / $ss^{protein\ null}$ genotype. *Transgene S* was recombined onto a chromosome with a *ss* deficiency to examine *Transgene S* transvection with mutant *ss* alleles. *Transgene S* performed transvection to upregulate expression of wild type, endogenous *ss* (*97*). In the *Transgene E* + *Transgene S*-ss^{def} / $ss^{protein\ null}$ genotype, the endogenous *ss* locus was hemizygous for a protein coding null allele of *ss*, which produced no functional Ss protein. Therefore, any functional Ss protein in this genotype was produced by *Transgene E*, and an increase in Ss (Rh4) expression frequency indicated that *Transgene S* was performing transvection to upregulate Ss expression from *Transgene E*. Red X over the *ss* locus indicates that the *ss* allele is a protein coding null. Gray arrow with "?" indicates that *Transgene E* was tested for transvection.

E. Adult eye for the *Transgene E* + *Transgene S*-ss^{def} / $ss^{protein\ null}$ genotype. *ss* expression frequency was not upregulated (Ss(Rh4)=53%), indicating that

Transgene S did not perform transvection with *Transgene E*. Red: Rh4, blue: Rh3.





Supplemental Fig. 16

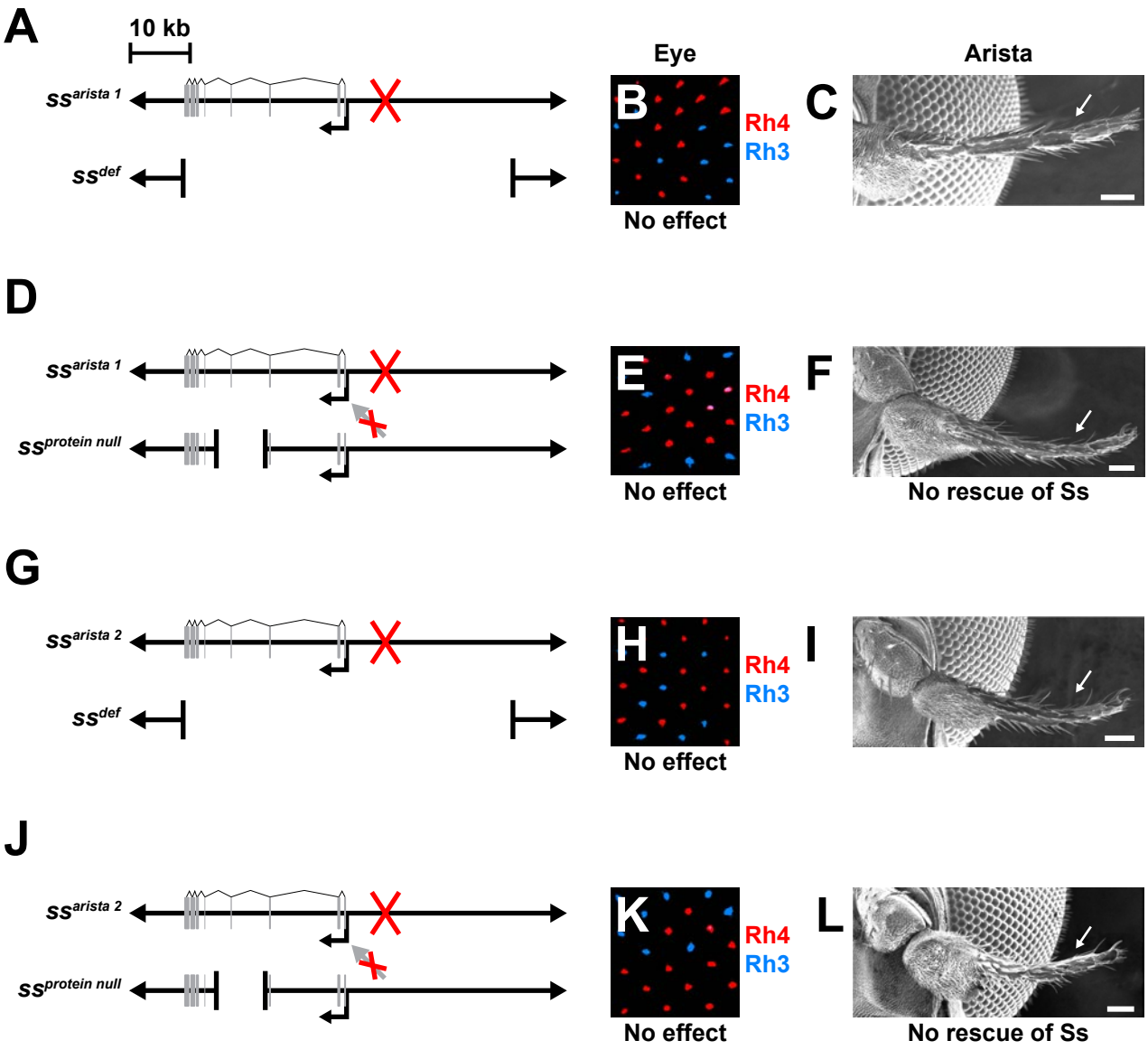
Supplemental Figure 16: Pairing driven by the *ss* button is cell-type specific.

A-B. *Site 1 control* images in the eye and antenna. Scale bars=1 μm. White: Lamin B, red: probes against endogenous *ss* and site 1 insertion site. Site 1 probes were pseudocolored red to allow direct comparison with single-color FISH experiments.

C. DNA FISH strategy used to assess pairing of *Transgene E* at site 3 with its endogenous locus in the larval antenna.

D-E. *Site 3 control* and *Transgene E site 3* examples in the larval antenna. Scale bar=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.

F. Quantifications for **Fig. S16B-C**. ns=p>0.05, unpaired t-test with Welch's correction.



Supplemental Fig. 17

Supplemental Figure 17: *ss* mutant alleles with arista-specific phenotypes do not perform transvection in the arista.

A, D, G, J. Genotypes tested for transvection. Gray rectangles: exons. Smaller black arrows: transcription start sites. Red X indicates an uncharacterized mutation in the $ss^{arista\ 1}$ or $ss^{arista\ 2}$ sequence. Red X over gray arrow indicates an absence of transvection between alleles in the arista.

B. $ss^{arista\ 1}$ / ss^{def} adult eye. $ss^{arista\ 1}$ / ss^{def} had no effect on eye development. Red: Rh4, blue: Rh3. Ss(Rh4)=53%.

C. $ss^{arista\ 1}$ / ss^{def} arista. $ss^{arista\ 1}$ / ss^{def} caused aristapedia. Scale bar= 50 μm. White arrow indicates arista. Image is from the same experiment as in **Fig. 6I**.

E. $ss^{arista\ 1}$ / $ss^{protein\ null}$ adult eye. $ss^{arista\ 1}$ / $ss^{protein\ null}$ had no effect on eye development. Red: Rh4, blue: Rh3. Ss(Rh4)=53%.

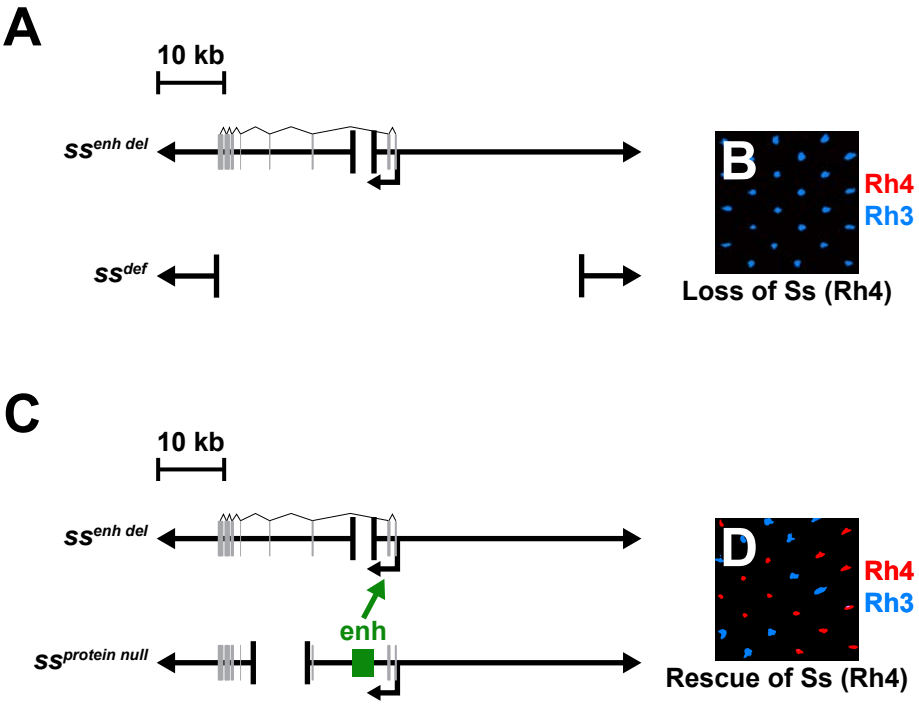
F. $ss^{arista\ 1}$ / $ss^{protein\ null}$ arista. $ss^{arista\ 1}$ / $ss^{protein\ null}$ had aristapedia, indicating that transvection did not occur to rescue the mutant *ss* phenotype. Scale bar= 50 μm. White arrow indicates arista.

H. $ss^{arista\ 2}$ / ss^{def} adult eye. $ss^{arista\ 2}$ / ss^{def} had no effect on eye development. Red: Rh4, blue: Rh3. Ss(Rh4)=60%.

I. $ss^{arista\ 2}$ / ss^{def} arista. $ss^{arista\ 2}$ / ss^{def} caused aristapedia. Scale bar= 50 μm. White arrow indicates arista.

K. $ss^{arista\ 2}$ / $ss^{protein\ null}$ adult eye. $ss^{arista\ 2}$ / $ss^{protein\ null}$ had no effect on eye development. Red: Rh4, blue: Rh3. Ss(Rh4)=62%.

L. *ss*$^{arista\ 2}$ / *ss*$^{protein\ null}$ arista. *ss*$^{arista\ 2}$ / *ss*$^{protein\ null}$ had aristapedia, indicating that transvection did not occur to rescue the mutant *ss* phenotype. Scale bar= 50 μm. White arrow indicates arista.



Supplemental Fig. 18

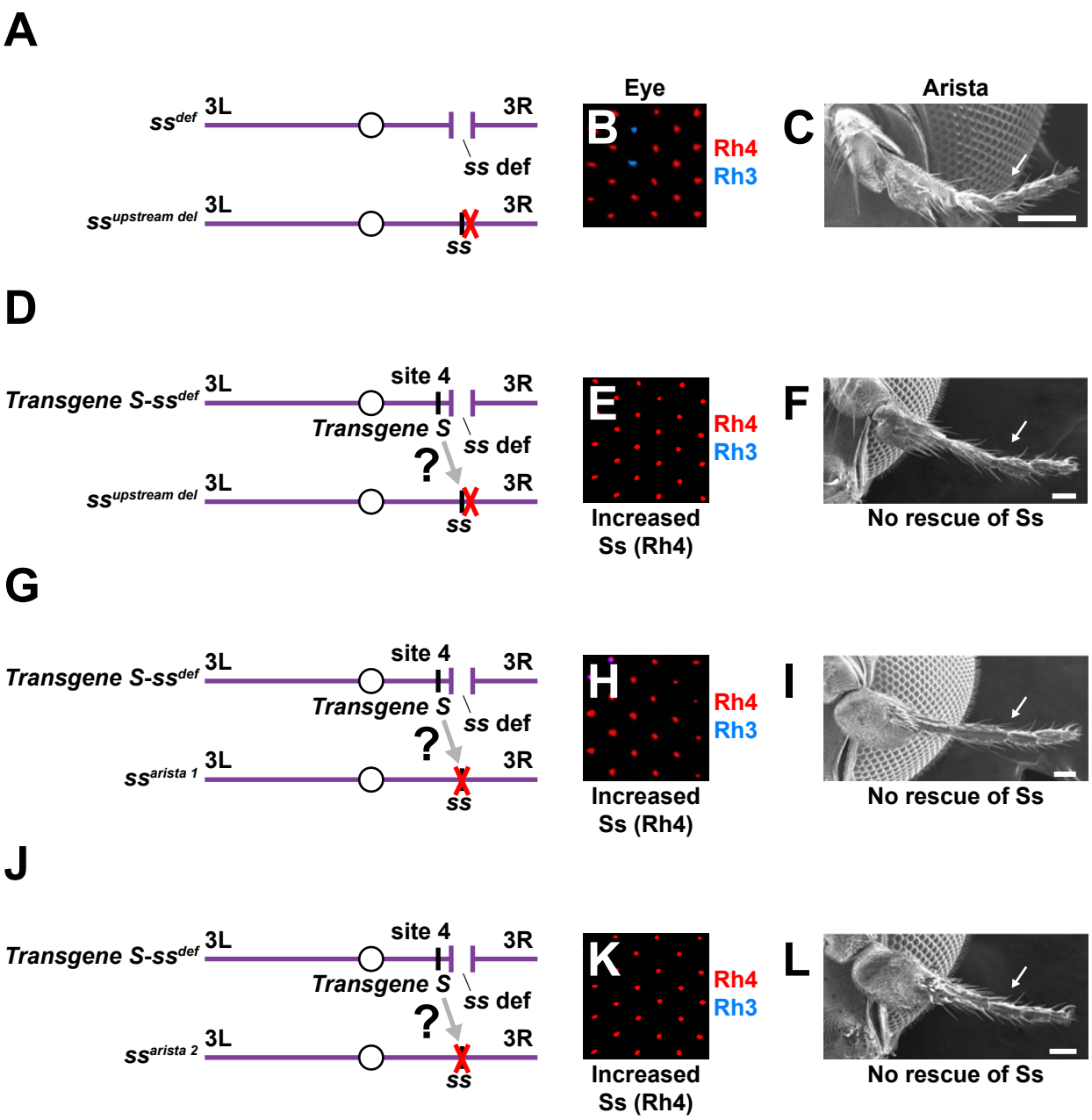
Supplemental Figure 18: *ss*$^{protein\ null}$ performs transvection in the eye.

A. Schematic of *ss* enhancer deletion (*ss*$^{enh\ del}$) allele over a *ss* deficiency (*ss*^{def}). Gray rectangles: exons. Smaller black arrow: transcription start site.

B. *ss*$^{enh\ del}$ / *ss*^{def} adult eye. *ss*$^{enh\ del}$ / *ss*^{def} caused a near complete loss of Ss/Rh4 expression. Red: Rh4, blue: Rh3. Ss(Rh4)=0.1%.

C. Schematic of *ss*$^{enh\ del}$ over *ss*$^{protein\ null}$. Through transvection, the functional enhancer of *ss*$^{protein\ null}$ acted on the functional protein coding region of *ss*$^{enh\ del}$ to rescue Ss expression (green arrow). Enh: enhancer, gray rectangles: exons. Smaller black arrows: transcription start sites.

D. *ss*$^{enh\ del}$ / *ss*$^{protein\ null}$ adult eye. *ss*$^{protein\ null}$ rescued Ss expression from the *ss*$^{enh\ del}$ allele. Red: Rh4, blue: Rh3. Ss(Rh4)=59%.



Supplemental Fig. 19

Supplemental Figure 19: Transgene S performs transvection in the eye but not in the arista.

A. Schematic of the *ss*$^{upstream\ del}$ allele over the *ss*^{def} allele. *ss*$^{upstream\ del}$ is a CRISPR allele in which 12.7 kb of the upstream regulatory regions of *ss* are deleted. Red X indicates the deletion of regulatory regions directly upstream of the *ss* locus.

B. *ss*$^{upstream\ del}$ / *ss*^{def} adult eye. *ss*$^{upstream\ del}$ displayed Ss(Rh4) expression in 85% of R7s. Red: Rh4, blue: Rh3.

C. *ss*$^{upstream\ del}$ / *ss*^{def} arista. *ss*$^{upstream\ del}$ caused aristapedia. Scale bar= 50 μm. White arrow indicates arista.

D. Schematic of the *Transgene S-ss*^{def} allele over the *ss*$^{upstream\ del}$ allele. *Transgene S* was recombined onto a chromosome with a *ss* deficiency to examine *Transgene S* transvection with mutant *ss* alleles. Red X indicates the deletion of regulatory regions directly upstream of the *ss* locus. Gray arrow with a "?" indicates that *Transgene S* was tested for transvection with the *ss*$^{upstream\ del}$ allele.

E. *Transgene S-ss*^{def} / *ss*$^{upstream\ del}$ adult eye. *Transgene S-ss*^{def} upregulated Ss(Rh4) expression from the *ss*$^{upstream\ del}$ allele into 100% of R7s, indicating that *Transgene S* performed transvection with *ss* $^{upstream\ del}$ in the eye. Red: Rh4, blue: Rh3.

F. *Transgene S-ss*^{def} / *ss*$^{upstream\ del}$ arista. *Transgene S-ss*^{def} / *ss*$^{upstream\ del}$ had aristapedia, indicating that *Transgene S* did not perform transvection to rescue

ss$^{upstream\ del}$ expression in the arista. Scale bar= 50 μm. White arrow indicates arista.

G. Schematic of the *Transgene S-ss*^{def} allele over the *ss*$^{arista\ 1}$ allele. Red X indicates an uncharacterized mutation in the *ss*$^{arista\ 1}$ allele. Gray arrow with "?" indicates that *Transgene S* was tested for transvection with the *ss*$^{arista\ 1}$ allele.

H. *Transgene S-ss*^{def} / *ss*$^{arista\ 1}$ adult eye. *Transgene S-ss*^{def} upregulated Ss(Rh4) expression from the *ss*$^{arista\ 1}$ allele into 99% of R7s, indicating that *Transgene S* performed transvection with *ss*$^{arista\ 1}$ in the eye. Red: Rh4, blue: Rh3.

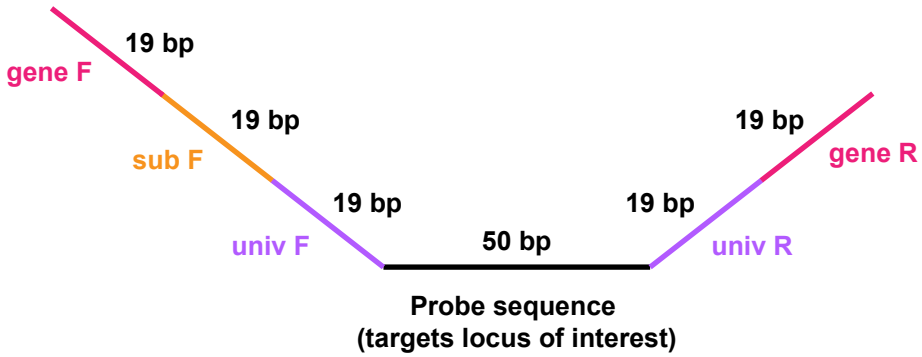
I. *Transgene S-ss*^{def} / *ss*$^{arista\ 1}$ arista. *Transgene S-ss*^{def} / *ss*$^{arista\ 1}$ had aristapedia, indicating that *Transgene S* did not perform transvection to rescue *ss*$^{arista\ 1}$ expression in the arista. Scale bar= 50 μm. White arrow indicates arista.

J. Schematic of the *Transgene S-ss*^{def} allele over the *ss*$^{arista\ 2}$ allele. Red X indicates an uncharacterized mutation in the *ss*$^{arista\ 2}$ allele. Gray arrow with "?" indicates that *Transgene S* was tested for transvection with the *ss*$^{arista\ 2}$ allele.

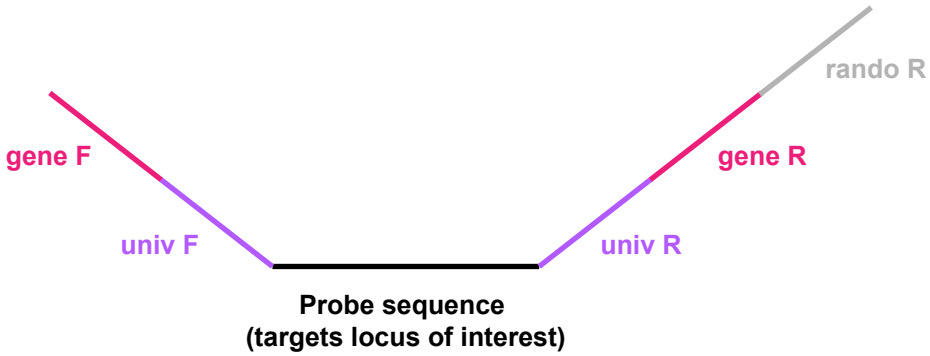
K. *Transgene S-ss*^{def} / *ss*$^{arista\ 2}$ adult eye. *Transgene S-ss*^{def} upregulated Ss(Rh4) expression from the *ss*$^{arista\ 2}$ allele into 100% of R7s, indicating that *Transgene S* performed transvection with *ss*$^{arista\ 2}$ in the eye. Red: Rh4, blue: Rh3.

L. *Transgene S-ss*^{def} / *ss*$^{arista\ 2}$ arista. *Transgene S-ss*^{def} / *ss*$^{arista\ 2}$ had aristapedia, indicating that *Transgene S* did not perform transvection to rescue *ss*$^{arista\ 2}$ expression in the arista. Scale bar= 50 μm. White arrow indicates arista.

A



B



Supplemental Fig. 20

Supplemental Figure 20: Barcoding primer scheme for DNA Oligopaints FISH probes.

A. Schematic of barcoding primer scheme for Oligopaints probe libraries containing sublibraries. univ: universal primer, sub: sublibrary primer.

B. Schematic of barcoding primer scheme for Oligopaints probe libraries without sublibraries. univ: universal primer, rando: random primer.

Chapter 4: Activating and repressing stochastic gene expression between chromosomes

4.1: Abstract

DNA elements act across long genomic distances to regulate gene expression in processes including enhancer-promoter interactions and genetic imprinting. In the *Drosophila* gene-regulatory phenomenon of transvection, DNA elements on one allele of a gene act between chromosomes on another allele of the gene to increase or decrease expression. Despite the discovery of transvection over 60 years ago, little is known about its biological role. Furthermore, it is unclear whether separable DNA elements contribute to activating or repressing transvection at a single locus. Here, we use the *spineless* (*ss*) locus as a paradigm to study gene activation and repression between chromosomes. We find a biological role for transvection in controlling the stochastic expression of naturally occurring *ss* alleles in the fly retina. Additionally, we use CRISPR deletions of sequences across the *ss* locus to determine the DNA elements required for activating and repressing transvection. We find that separable mechanisms control activation and repression of *ss* between chromosomes: activation requires an intact enhancer and promoter in *cis*, while repression requires two Polycomb Response Elements and an insulator element. Finally, we find that *ss silencer 1* can repress *ss* between chromosomes even when translocated to other regions of the genome. The translocated *silencer 1* may require high-affinity button sequences to bring it in to proximity with the *ss* locus. Our findings provide mechanistic insight into how

DNA elements regulate stochastic gene expression between chromosomes, and suggest a model in which button regions “reconstitute” genes to maintain proper expression in the presence of chromosome rearrangements.

4.2: Introduction

Long-distance interactions between DNA regulatory elements are an essential mechanism controlling gene expression. One well-studied example of DNA element action at a distance involves enhancer-promoter interactions, which in cases such as the *sonic hedgehog*, *Hoxb*, and *Hoxd* loci in mammals can occur at distances of up to 1 Mb (*183-185*). In addition to long-range interactions along a single chromosome, DNA elements can act between chromosomes. For example, olfactory receptor enhancers in mice interact between chromosomes to form a super-enhancer that interacts with an active olfactory receptor gene (*33*). Furthermore, mammalian imprinting control regions act between chromosomes to ensure proper expression of maternal or paternal alleles (*75, 186*). Aberrant action of DNA elements between chromosomes is observed in cases of lymphoma and myeloma, where translocations of the *IgH* enhancer to chromosome 14 incorrectly activate the *CCND1* locus on the homologous chromosome (*87*). However, little is known about the mechanisms controlling the action of DNA elements between chromosomes.

The most well-studied example of interchromosomal DNA element interactions occurs in the *Drosophila melanogaster* phenomenon of transvection, where the regulatory elements of a gene on one chromosome control expression

of the same gene on a homologous chromosome. Transvection was originally discovered over 60 years ago by Ed Lewis at the *ubx* locus (*55*), and has subsequently been described for a large number of *Drosophila* genes (*90, 97*). Transvection is facilitated by the colocalization, or "pairing," of homologous chromosomes, which occurs in *Drosophila* in nearly all somatic cells throughout development (*54*). Homologous chromosome pairing is driven by "button" elements, regions of high pairing affinity that are interspersed along chromosome arms (*57-59, 72*). With the exceptions of the *Abd-b* and *spineless* (*ss*) loci, pairing and transvection are typically disrupted by chromosome rearrangements (*55, 59, 72, 90, 91, 96, 97*).

In most cases, transvection involves interactions between two paired mutant alleles. DNA elements on each allele act between chromosomes to rescue gene expression. Often, the enhancer of one mutant allele acts on the promoter of the other mutant allele to activate gene expression (*90*). Enhancer action between chromosomes is often more efficient in the absence of a promoter in *cis (90-95)*. Insulator elements, which are involved in DNA looping interactions (*187*), have also been linked to activating transvection (*72, 102, 104, 147*).

Transvection-related phenomena can also repress genes between chromosomes, in processes including *brown* dominant silencing, *zeste-white* silencing, and pairing-sensitive silencing (*105, 106, 108, 124, 133, 188*). DNA elements such as insulators and Polycomb Response Elements (PREs) have been linked to repressing transvection at many loci across the genome (*67, 69,*

108, 111-115, 145, 146). However, very little work has focused on separating the DNA elements that specifically contribute to activating and repressing transvection to control gene expression at a single locus. The advent of CRISPR provides the unique opportunity to examine the contribution of specific DNA elements to transvection through the precise deletion of regions of interest.

Because transvection has largely been studied in the context of mutant alleles, the biological role of transvection between wild-type alleles is unclear. It has been proposed that transvection ensures proper gene expression by enhancing the level of transcription between paired alleles (*189*). However, a role for transvection between naturally occurring alleles has yet to be experimentally demonstrated.

Here, we use transvection between alleles of *ss* as a model to study activation and repression between chromosomes. We find a biological role for transvection in “averaging” the expression frequencies of naturally occurring *ss* alleles. Using CRISPR to precisely delete DNA elements across the *ss* locus, we dissect the elements required for transvection between *ss* alleles, and separate the activating and repressing functions of *ss* transvection. Activation of *ss* between chromosomes requires an intact enhancer and promoter in *cis*, while repression of *ss* between chromosomes requires an insulator element and two PREs. Finally, we observe *ss* pairing and transvection in the presence of dramatic chromosomal rearrangements, showing that pairing of homologous regions can “reconstitute” genes between chromosomes to maintain proper gene expression.

4.3: Results

Transvection "averages" expression of wild-derived alleles

While transvection between mutant alleles is well-studied, the role of transvection in regulating the expression of functional, naturally occurring alleles is poorly understood. To investigate the biological role of transvection, we used the *ss* locus as a paradigm. In the fly retina, Ss is expressed stochastically in ~67% of R7 photoreceptors to activate expression of the photopigment Rhodopsin 4 (Rh4) and repress expression of Rhodopsin 3 (Rh3)(**Fig. 1A-B**). In R7s lacking Ss, Rh3 is expressed (**Fig. 1A**). Therefore, Rh4 and Rh3 can be used as a readout for Ss expression.

Mutant alleles of *ss* with expression frequencies that vary from the wild-type 67% communicate via transvection (also known as Interchromosomal Communication)(*97*) to determine a final, intermediate Ss^{ON}:Ss^{OFF} ratio. For example, when a "high-frequency" allele of *ss* that is expressed in 100% of R7s is crossed with a protein coding null allele of *ss* that is expressed in 0% of R7s, the resulting progeny express Ss/Rh4 in 90% of R7s, a frequency between that of the high- and low-expressing alleles (**Fig. 1C**)(*97*). Therefore, transvection between *ss* alleles can be measured quantitatively by assessing changes in the Rh3:Rh4 ratio.

Naturally occurring alleles of *ss* from the Drosophila Genome Resource Panel (DGRP) display wide variation in expression frequencies (*100, 101*). To investigate whether these wild-derived *ss* alleles could perform transvection, we examined crosses between DGRP alleles with Ss/Rh4 expression frequencies

ranging from 41% to 74%. If the two DGRP alleles did not perform transvection, we hypothesized that their progeny would exhibit a Ss/Rh4 expression frequency higher than either of the parents due to independent expression of each parental allele. We hypothesized that if the two DGRP alleles performed transvection, then their progeny would exhibit an intermediate Rh3:Rh4 ratio between the ratios of the two parents.

We examined four categories of crosses. In Group 1, the Rh3:Rh4 expression ratios of both parents and their progeny were not statistically different (**Fig. 1D**). In Groups 2-4, the parents had statistically different Rh3:Rh4 expression ratios (**Fig. 1E-G**). The Rh3:Rh4 ratio of Group 2 progeny was not statistically different from either parent (**Fig. 1E**), the Rh3:Rh4 ratio of Group 3 progeny was statistically different from one parent (**Fig. 1F**), and the Rh3:Rh4 ratio of Group 4 progeny was statistically different from both parents (**Fig. 1G**). In all cases, cross progeny had a Ss/Rh4 expression frequency that fell between the frequencies of the parents, rather than an expression frequency higher than either parent (**Fig. 1D-G**). Thus, wild-derived alleles of *ss* performed transvection to “average” the expression frequencies of high and low alleles.

To further investigate transvection between wild-derived alleles, we examined a *ss* allele containing a naturally occurring insertion known as *sin (101).* Fly lines that are homozygous for *sin* display a decreased Ss expression frequency, with only 50% of R7s expressing Ss/Rh4, while wild-type flies lacking *sin* express Ss/Rh4 in 67% of R7s (**Fig. 1H**)(*101*). In a *sin /* + background,

Ss/Rh4 was again expressed at an intermediate expression frequency (60%) between the Rh3:Rh4 ratio of either parent (**Fig. 1H**).

Together, these data suggest a biological role for transvection in patterning the *Drosophila* retina. By roughly "averaging" the expression frequencies of variant alleles, *ss* transvection may ensure that ratios of Rh4- and Rh3-expressing R7s are kept within a certain acceptable range for normal color vision.

An enhancer and promoter in cis are required for gene activation between chromosomes

ss transvection upregulates and downregulates expression between chromosomes to generate intermediate transvection frequencies. While many studies have focused on the DNA elements required for *trans*-activation or *trans*-silencing (*67, 69, 72, 90, 92, 102, 104, 108, 111-115, 145-147*), it remains unclear how activating and repressing elements work cooperatively to regulate the expression of a single gene between chromosomes. To separate the activating and repressing functions of *ss* transvection, we generated CRISPR deletions of DNA elements within the *ss* locus and tested their effects on *ss* transvection.

The *ss* locus contains two enhancers (*early enhancer* and *late enhancer*) and a promoter (*ss promoter*) that regulate its expression (**Fig. 2A**). When hemizygous with a *ss* deficiency (*ss def*), wild-type *ss* is expressed in 56% of R7s (**Fig. 2A**)(*97*). CRISPR deletion of either enhancer or the promoter (*early*

enhΔ, late enhΔ, and *promΔ*) caused a complete loss of Ss/Rh4 expression when placed over a *ss* deficiency (*ss def*)(**Fig. 2A**). A protein coding null of *ss* (*prot null*), which contains intact *cis* regulatory elements, also resulted in 0% Ss/Rh4 expression when placed over *ss def* (**Fig. 2A**).

To test these CRISPR deletions for transvection, we first crossed them with the *prot null*. *early enhΔ / prot null* and *late enhΔ / prot null* both had near-wild type Ss/Rh4 expression frequencies (**Fig. 2B-C**)**,** indicating that the intact enhancers on the *prot null* allele were performing transvection on the *enhΔ* alleles to drive Ss expression. *promΔ / prot null* displayed 0% Ss/Rh4 expression (**Fig. S1A-B**); although the enhancers of the *promΔ* allele could theoretically act on the intact promoter of the *prot null* allele, mutations in the coding regions of the *prot null* allele prevented the production of any functional Ss protein.

To further test the ability of these alleles to perform transvection, we utilized *Transgene A*, which contains the first 46 kb of the *ss* locus, including the *early enhancer, late enhancer,* and *ss promoter* (**Fig. S2A**). Because *Transgene A* lacks the last five exons of *ss*, it does not produce functional Ss protein. However, it drives GFP expression in 100% of R7s (*72*). Additionally, *Transgene A* performs transvection on wild-type alleles of *ss* to upregulate Ss/Rh4 into 100% of R7s (*72, 97*). We recombined *Transgene A* onto the *ss def* allele to test its effects on the *early enhΔ* and *late enhΔ* alleles (**Fig. S2A**). We hypothesized that if *Transgene A* could perform transvection with the *enhΔ* alleles, then Ss/Rh4 expression frequency would increase to nearly 100%, similar to the effects of *Transgene A* on a wild type control (**Fig. S2C**). If *Transgene A* could not perform

transvection with the *enhΔ* alleles, we expected to observe a 0% Ss/Rh4 expression frequency, similar to *early enhΔ* / *ss def* and *late enhΔ* / *ss def* (**Fig. S2B**). *Transgene A* upregulated Ss expression of *early enhΔ* and *late enhΔ* into nearly 100% of R7s (**Fig. S2D-E**), supporting the conclusion that the intact *early enhancer* and *late enhancer* can act between chromosomes on the promoters of the *early enhΔ* and *late enhΔ* alleles to drive Ss expression.

In many cases, enhancer action between chromosomes is more efficient in the absence of a promoter in *cis* (*9, 17-21*). We therefore hypothesized that if we crossed *early enhΔ* and *late enhΔ* with *promΔ*, the intact enhancer of *promΔ* would act on the intact promoter of *early* or *late enhΔ* to restore Ss/Rh4 expression frequency to wild type. Surprisingly, *early enhΔ / promΔ* exhibited only partial rescue of Ss/Rh4 frequency (Ss/Rh4=27%; **Fig. 2D**), while *late enhΔ / promΔ* had no rescue (Ss/Rh4 = 2%; **Fig. 2E**), suggesting that *early enhancer* and *late enhancer* require a promoter in *cis* to work efficiently between chromosomes.

To test the hypothesis that the *ss* enhancers require an intact promoter in *cis* to perform transvection, we examined flies that were transheterozygous for *early enhΔ / late enhΔ*, so that each allele contained an intact promoter and one intact enhancer. These flies expressed Ss/Rh4 in 53% of R7s (**Fig. 2F**). Therefore, activating transvection at the *ss* locus requires an intact enhancer and promoter in *cis* to drive expression between chromosomes.

Two PREs and an insulator element are required for gene repression between chromosomes

To investigate the elements required for repressive transvection between *ss* alleles, we examined a silencer element at the 5' end of the *ss* locus (*silencer 1*). This silencer contains an insulator element (*72*) and two putative Polycomb Response Elements (PREs), which bind the repressive Polycomb Group Complex (**Fig. 3A**). In flies homozygous for a complete *silencer 1* deletion (*silencer 1Δ*), Ss/Rh4 expression frequency increases to 99% of R7s (**Fig. 3B**)(*97*). Over *ss def*, an allele of *ss* in which both PREs are deleted (*PRE12Δ*) causes upregulation of Ss/Rh4 into 85% of R7s (**Fig. 3C**)(*72*).

To test whether the PREs within *silencer 1* could repress Ss/Rh4 expression frequency, we crossed *PRE12Δ* with *prot null*, which has an intact *silencer 1* but mutated protein coding region. We predicted that if the PREs can act between chromosomes to repress expression, then the Ss/Rh4 expression frequency of *PRE12Δ / prot null* would decrease compared to the Ss/Rh4 expression frequency of *PRE12Δ / ss def*. *PRE12Δ / prot null* flies had a 62% Ss/Rh4 expression frequency (**Fig. 3D**), indicating that the PREs on the *prot null* allele could act between chromosomes to repress *PRE12Δ* expression.

To further test the ability of the *ss* PREs to repress expression between chromosomes, we next examined *silencer 1Δ / prot null*. We hypothesized that the PREs on the *prot null* allele would act between chromosomes to repress expression from the *silencer 1Δ* allele. Surprisingly, *silencer 1Δ / prot null* remained at a 99% Ss/Rh4 expression frequency (**Fig. 3E**)(*97*), indicating that

the PREs on *prot null* could not repress expression of the *silencer 1Δ* allele. An insulator element is present in *PRE12Δ* but absent in *silencer 1Δ* (**Fig. 3B-C**), suggesting that an intact insulator on both *ss* alleles is required for the PREs to act between chromosomes and repress Ss expression.

Because both *ss* enhancers require the *ss promoter* in *cis* to perform transvection, we next tested whether the *ss* PREs also require the *ss promoter* to act between chromosomes by examining *PRE12Δ / promΔ*. We hypothesized that if the PREs require the *ss promoter* to act between chromosomes, then a loss of PRE repression in *PRE12Δ / promΔ* would cause Ss/Rh4 expression in ~85% of R7s, similar to *PRE12Δ / ss def*. If the PREs do not require the *ss promoter* to act between chromosomes, we predicted that Ss/Rh4 expression in *PRE12Δ / promΔ* would be similar to wild type expression frequency, because the PREs would act effectively between chromosomes to decrease Ss expression. We observed 57% Ss/Rh4 expression frequency in the *PRE12Δ / promΔ* background (**Fig. S3A-B**), indicating that the *ss* PREs do not require the *ss promoter* to repress *ss* expression between chromosomes.

Silencer 1 performs transvection between rearranged chromosomes

Classical transvection is disrupted by chromosome rearrangements (*55*). However, chromosome rearrangements do not disrupt *ss* transvection (*72, 97*). To further investigate the mechanisms that facilitate transvection between rearranged copies of *ss*, we examined the $ss^{translocation}$ and $ss^{inversion}$ alleles. The $ss^{translocation}$ allele contains a translocation between the X chromosome and

chromosome 3, with a breakpoint directly downstream of *silencer 1* (**Fig. 4A-E**). The $ss^{inversion}$ allele has a breakpoint immediately downstream of *silencer 1*, which moves the coding region of *ss* 12 Mb away from *silencer 1* (**Fig. S4A-E**). $ss^{translocation}$ */ ss def* caused an increase in *ss* expression frequency to 100% Ss/Rh4 (**Fig. 4A**), indicating that the translocated *silencer 1* was not acting between chromosomes on the *ss* coding regions to repress Ss expression. However, $ss^{translocation}$ */ +* and $ss^{translocation}$ */ prot null* returned Ss/Rh4 expression frequency to near-wild-type levels (Ss/Rh4=73%; **Fig. 4B-C**), providing additional support for the conclusion that 2 insulator copies are required for the *ss* PREs to repress between chromosomes.

To further test the hypothesis that an insulator on both alleles is required for repressing transvection between copies of *ss*, we crossed $ss^{translocation}$ with *PRE12Δ*, which contains the insulator, and *silencer 1Δ*, which lacks the insulator. $ss^{translocation}$ / *PRE12Δ* decreased Ss/Rh4 expression frequency to 73% (**Fig. 4D**), while $ss^{translocation}$ */ silencer 1Δ* Ss/Rh4 expression frequency remained at 99% (**Fig. 4E**). Together, these data indicated that the *ss* PREs require the presence of an insulator on both alleles to repress *ss* expression between chromosomes.

In the $ss^{translocation}$ */ +* background, either *silencer 1* on $ss^{translocation}$ or *silencer 1* on the wild type *ss* allele could be repressing Ss expression between chromosomes. However, in a $ss^{translocation}$ */ PRE12Δ* background, the only intact copy of *silencer 1* is on the $ss^{translocation}$ allele (**Fig. 4D**). Since Ss/Rh4 expression frequency is 73% rather than 100% in this background (**Fig. 4D**), we concluded

that *silencer 1* on the $ss^{translocation}$ allele was responsible for repressive transvection.

To further investigate the mechanisms that facilitate PRE-driven repression in trans, we examined $ss^{inversion}$ (**Fig. S4A-E**). Similar to $ss^{translocation}$ / *ss def*, $ss^{inversion}$ / *ss def*, where only one copy of the insulator was present, increased Ss/Rh4 expression frequency to 100% (**Fig. S4A**)(*97*). In $ss^{inversion}$ / + and $ss^{inversion}$ / *prot null* backgrounds, where two copies of the insulator were present, Ss/Rh4 expression frequency decreased to 84% and 90%, respectively (**Fig. S4B-C**)(*97*), These data again suggested that the presence of an insulator on each allele allows the *ss* PREs to act across long distances and repress Ss.

Together, these data suggest that *silencer 1* can repress expression between chromosomes in the presence of significant chromosomal aberrations.

A template ss copy is not required for silencer 1 colocalization with the remainder of ss

The PREs on $ss^{translocation}$ and $ss^{inversion}$ do not repress *ss* between chromosomes when crossed with *ss def* (**Fig. 4A; Fig. S4A**), but they can repress *ss* when crossed with wild type *ss, prot null*, or *PRE12Δ* (**Fig. 4B-D; Fig. S4B-C**). We hypothesized that a two-step mechanism might facilitate interactions between the translocated PREs and the intact *ss* locus. In step 1, the additional copy of *ss* might act as a “template” for homologous pairing, bringing the PREs of $ss^{translocation}$ and $ss^{inversion}$ into close proximity with the *ss* coding region (**Fig. 5A**). In step 2, the two copies of the insulator might allow the PREs to repress *ss*

across long distances (**Fig. 5B**). We predicted that if interactions between the *ss* PREs and the *ss* coding region require a template, then *silencer 1* of $ss^{translocation}$ and $ss^{inversion}$ would be in closer physical proximity to the remainder of the *ss* locus over a wild type allele than over *ss def*.

To test this hypothesis, we used DNA Oligopaints FISH (*149, 170*) to visualize the localization of *silencer 1* and the rest of the *ss* locus for the $ss^{inversion}$ allele. We targeted the region immediately downstream of the proximal inversion breakpoint with green probes and the region immediately upstream of the distal inversion breakpoint with red probes (**Fig. 5C-D**). We hypothesized that if *silencer 1* of $ss^{inversion}$ was interacting with the rest of the *ss* locus, then we would observe a green puncta close to a red puncta in the nucleus (**Fig. 5F**), whereas if *silencer 1* of $ss^{inversion}$ was not interacting with the rest of the *ss* locus, then the green and red puncta would be far apart in the nucleus (**Fig. 5G**), similar to a wild-type control (**Fig. 5E**). In a wild-type control, the red and green punctae were far apart in the nucleus (**Fig. 5H, K**). As we hypothesized, the red and green punctae were significantly closer together in the $ss^{inversion}$ / + background (**Fig. 5I, K**), indicating that *silencer 1* of $ss^{inversion}$ was interacting with the remainder of the *ss* locus. Intriguingly, despite the absence of a template copy of *ss*, the red and green punctae were also close together in the $ss^{inversion}$ / *ss def* background (**Fig. 5J-K**). Together, these data suggest that a template copy of *ss* is not required for *silencer 1* to colocalize with the remainder of the *ss* locus.

We therefore propose that other sequences near the *ss* locus mediate interactions between *silencer 1* and the *ss* protein coding region. *Drosophila*

chromosomes contain a large number of high-affinity “button” sequences, which bring homologous sequences together across the genome. Two buttons (*Button 1* and *Button 2*), located 20 kb and 60 kb downstream of endogenous *ss*, are present in $ss^{inversion}$ and *ss def* (*72*). We hypothesize that these buttons facilitate interactions between the different portions of the *ss* locus, allowing “reconstitution” of the gene and maintenance of proper expression in the presence of chromosome rearrangements.

4.4: Discussion

Despite the discovery of transvection over 60 years ago (*55*), the biological role of transvection has remained elusive. We find that transvection occurs between naturally derived, wild-type alleles of the *ss* gene. When flies homozygous for a high-frequency *ss* allele are crossed with flies homozygous for a low-frequency *ss* allele, their progeny have an intermediate Ss/Rh4 expression frequency between that of the higher and lower alleles (**Fig. 1C-H**). Our data suggest a biological role for transvection in preventing extremely high or low Rh4 expression frequency. The “averaging” of *ss* alleles may keep Rh4:Rh3 expression ratios within a certain range required for proper color vision. Supporting this idea, the naturally occurring insertion *sin* causes a decreased Rh4:Rh3 expression ratio and shifts innate color preference from green to blue (*101*), indicating that maintenance of proper *ss* expression frequency is important for normal vision.

While there is a large body of work examining the DNA elements involved in gene activation between chromosomes *(72, 90-95, 102, 104, 147)* or gene repression between chromosomes (*67, 69, 108, 111-115, 145, 146*), few studies have elucidated the complete networks of activating and repressing elements that work together between different alleles of the same gene to control overall gene expression. Using precise CRISPR deletions of DNA elements across the *ss* locus, we have identified a complex combination of DNA regulatory elements that governs *ss* transvection. Two *ss* enhancers and the *ss* promoter activate *ss* expression between chromosomes (**Fig. 2B-F**), while two PREs and an insulator repress *ss* expression between chromosomes (**Fig. 3; Fig. 4; Fig. S4**). Thus, our findings indicate that activating and repressing transvection are separable mechanisms.

The loss of a promoter in *cis* can increase the ability of an enhancer to act between chromosomes *(90-95)*. However, the *ss* enhancers require a promoter in *cis* in order to activate expression between chromosomes (**Fig. 2D-F**). This finding raises the possibility there are two distinct classes of transvection-competent genes, one in which the transvecting enhancer works more efficiently in the absence of a promoter in *cis,* and one in which the transvecting enhancer requires a promoter in *cis.* As promoter requirements have only been investigated for a subset of transvection-competent genes, including *yellow*, *apterous*, *Abd-b*, and *Ubx (90-95),* it is possible that many less well-characterized genes fall into the *ss* transvection category.

While *ss* represents a potential new class of transvection-competent genes that require an enhancer and promoter in *cis*, it is unclear how these enhancer-promoter interactions facilitate activation between chromosomes. We propose three models: 1. The enhancer simultaneously loops to the *cis* and *trans* promoter to activate gene expression from both alleles (simultaneous looping; **Fig. 6A**) 2. The enhancer loops to the promoter in *cis*, and the resulting transcription initiation complex allows the enhancer-promoter complex to activate the copy of the promoter on the other chromosome (enhancer-promoter complex; **Fig. 6B**) 3. Because the two alleles of *ss* are paired, enhancer looping to the promoter in *cis* causes simultaneous looping of the other *ss* allele, resulting in an open conformation that allows gene expression (looping to active conformation; **Fig. 6C**).

While chromosome rearrangements often disrupt transvection (*55, 90*), *ss* transvection occurs in the presence of rearrangements, even when rearrangement breakpoints lie within the gene locus itself (**Fig. 4; Fig. S4**). The robustness of *ss* transvection in the presence of chromosome rearrangements is likely due to the fact that the *ss* locus encompasses a “button” element, which is sufficient to drive pairing between non-homologous chromosomes (*72*). Two additional button elements lie immediately downstream of the *ss* locus (*72*) and may facilitate interactions between copies of *ss* in cases where rearrangement breakpoints lie within the *ss* locus. Supporting the conclusion that button elements allow transvection in the presence of chromosome rearrangements,

transvection at the *Abd-B* locus, which overlaps a button (*72*), is not disrupted by rearrangements (*59, 91, 96*).

In the presence of major chromosome rearrangements such as the the $ss^{inversion}$ allele, *ss silencer 1* still interacts with the remainder of the *ss* locus. In a $ss^{inversion}$ / *ss def* background, *silencer 1* remains in close physical proximity with the protein-coding regions of *ss* (**Fig. 5J-K**). This physical proximity is potentially mediated by two button elements immediately downstream of the *ss* locus (*72*), which are present on both the $ss^{inversion}$ and *ss def* alleles. Despite being in close physical proximity with the remainder of the *ss* locus, *silencer 1* does not repress *ss* expression in a a $ss^{inversion}$ / *ss def* background (**Fig. S4A**), indicating that physical interaction between *silencer 1* and the *ss* protein coding region is not sufficient for *silencer 1*-mediated repression. Rather, an insulator on both *ss* alleles is required for *silencer 1* repression between chromosomes (**Fig. 3; Fig. 4; Fig. S4**).

The ability of buttons to “reconstitute” *ss* suggests a genome-wide mechanism that maintains proper gene expression in the presence of chromosome aberrations. As this mechanism allows the genome to be robust to rearrangements, it may also have implications for the evolution of long-range gene regulation. In cases where inversions and translocations separate an enhancer from its promoter by distances of 1 Mb or more, nearby buttons might allow a distant enhancer to loop back to its promoter and drive normal gene expression. Over time, this novel enhancer location might evolve to be the endogenous enhancer site. Requirements for long-distance DNA element action

observed at the *ss* locus, such as a promoter in *cis* or two copies of an insulator, might be lost over time, allowing efficient enhancer action at a distance independent of other DNA elements. Together, our findings suggest an essential role for DNA element action between chromosomes in maintaining proper genome organization and gene expression.

4.5: Materials and Methods

Antibodies

Antibodies were used at the following dilutions: mouse anti-Lamin B (DSHB ADL67.10 and ADL84.12), 1:100; mouse anti-Rh3 (1:100) (gift from S. Britt, University of Colorado), rabbit anti-Rh4 (1:100) (gift from C. Zuker, Columbia University), and Alexa 488 Phalloidin (1:80) (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). All secondary antibodies were Alexa Fluor-conjugated (1:400) and made in donkey (Molecular Probes).

Antibody Staining

Adult were dissected as described (*169*) and fixed for 15 min with 4% formaldehyde at room temperature. Retinas were washed three times in PBS plus 0.3% Triton X-100 (PBX), then incubated with primary antibodies diluted in PBX overnight at room temperature. Next, retinas were washed three times in PBX and incubated in PBX for >4 hr. Retinas were then incubated with secondary antibodies diluted in PBX overnight at room temperature. Finally, retinas were rinsed three times in PBX and incubated in PBX for >2 hr. Retinas

were mounted in SlowFade Gold Antifade Reagent (Invitrogen). Images were acquired using a Zeiss LSM 700 confocal microscope.

Quantification of Rh3:Rh4 expression frequencies

Frequency of Rh3 (Ss^{OFF}) and Rh4 (Ss^{ON}) expression in R7s was scored in adults. Six or more retinas were scored for each genotype (N). 100 or more R7s were scored for each retina (n). Frequency was assessed using custom semi-automated software (see below) or manually.

Image processing

Image analysis was performed as described in (*101*). Briefly, we employed a custom algorithm to identify the positions of individual R7 photoreceptors within an image of the fly retina. First, individual fluorescence images from each wavelength channel were denoised using a homomorphic filter(*190*) and Gaussian blur. Next, R7 boundaries were located using the Canny edge detection method (*191*). Cells were then roughly segmented using the convex hull algorithm (*192*). Active contouring (*193*) was used to refine the segments to fit the R7s more closely. Finally, a watershed transform was applied to the image, dividing it into regions that each contain a single R7. Regions were excluded by size or distance from the center to prevent artifacts due to the curvature of the fly retina. For the remaining regions, normalized intensities from the Rh3 and Rh4 channels were compared in order to assign each region a label, indicating that its R7# is stained with Rh3 or Rh4. A MATLAB (The MathWorks, Inc.) script that

implements our algorithm is available at https://app.assembla.com/spaces/roberts-lab-public/wiki/Fly_Retina_Analysis.

Generation of CRISPR lines

CRISPR lines were generated as described in references (*101, 172-174*). $ss^{late\ enh\ del}$ and $ss^{PRE12\ deletion}$ were generated in reference (*72*) using methods and oligos detailed in references (*72, 101, 172-174*). For each CRISPR deletion, sense and antisense DNA oligos for the forward and reverse strands of four gRNAs were designed to generate BbsI restriction site overhangs. The oligos were annealed and cloned into the pCFD3 cloning vector (Addgene, Cambridge, MA). A single-stranded DNA homology bridge was generated with 60-bp homologous regions flanking each side of the predicted cleavage site and an EcoRI restriction site to aid in genotyping. The gRNA constructs (125 ng/μl) and homologous bridge oligo (100 ng/μl) were injected into *Drosophila* embryos (BestGene, Inc., Chino Hills, CA). Single males were crossed with a balancer stock (*yw; +; TM2/TM6B*), and F1 female progeny were screened for the insertion via PCR, restriction digest, and sequencing. Single F1 males whose siblings were positive for the deletion were crossed to the balancer stock (*yw; +; TM2/TM6B*), and the F2 progeny were screened for the deletion via PCR, restriction digest, and sequencing. Deletion-positive flies from multiple founders were used to establish independent stable stocks.

The following oligos were used for the $ss^{early\ enh\ del}$ CRISPR:

Table 1: Oligos used for *ss*$^{early\ enh\ del}$ CRISPR

Oligo name	Sequence
Homologous bridge	AAATAGATGGCTATTAACTATAGTTATTGAAGTGGGTCCATCAATCCATCTTCTCCATTAGAATTCCT CATAAGAATCTTAAAAATGATAAAGATATTTTCTTCGATAGCTTTATTTCAGATATCA
gRNA 1F	GTCGCAAAACAATATGAGGTCTAA
gRNA 1R	AAACTTAGACCTCATATTGTTTTGC
gRNA 2F	GTCGAAGTGGCCTGGGCTTATCTC
gRNA 2R	AAACGAGATAAGCCCAGGCCACTT
gRNA 3F	GTCGTTTAAGATTCTTATGAGATG
gRNA 3R	AAACCATCTCATAAGAATCTTAAA
gRNA 4F	GTCGTATAATTGCGGCTACTTATC
gRNA 4R	AAACGATAAGTAGCCGCAATTATA
genotype F	GCTCCGACTAATTGCAGACAG
genotype R	CCCTTACAAACGAAAAGATCA

The following oligos were used for the *ss*$^{prom\ del}$ CRISPR:

Table 2: Oligos used for *ss*$^{prom\ del}$ CRISPR

Oligo name	Sequence
Homologous bridge	CTTCCACTTTGCTACACTTCACTCCACTCCACTCGACTCAGCTCACTTATTAGTGCCACCGAATTC GATAAACCAGAGCCCACGAGCAACAACACTACCAACACAAACGGCAAAAGTGCAAGTGA
gRNA 1F	GTCGAAATCTCGCTGCGTAAACTA
gRNA 1R	AAACTAGTTTACGCAGCGAGATTT
gRNA 2F	GTCGAGGCAGCGCATGTCAATCGG
gRNA 2R	AAACCCGATTGACATGCGCTGCCT
gRNA 3F	GTCGCAAAAGCGCTAACGCAAGAA
gRNA 3R	AAACTTCTTGCGTTAGCGCTTTTG
gRNA 4F	GTCGCGTGGGCTCTGGTTTATCAA
gRNA 4R	AAACTTGATAAACCAGAGCCCACG
genotype F	CAAGAAGAGGAAAGGCAGTGC
genotype R	TTT GAT TCG AGT GGC TGC GAT

Drosophila lines

Flies were raised on standard cornmeal-molasses-agar medium and grown at 25° C.

Table 3: Genotypes of *Drosophila* lines

Fly line	Full genotype	Source	Figures
Wild type	*yw; +;+ or yw; sp/CyO;+*	N/A	1B, H; 3A; 5A, D, G
DGRP Cross 1 Parent 1	*yw; Sp/CyO; DGRP57*	Bloomington, (*100*)	1D
DGRP Cross 1 Parent 2	*yw; Sp/CyO; DGRP228*	Bloomington, (*100*)	1D
DGRP Cross 1 progeny	*yw; Sp/CyO; DGRP57/DGRP228*	Bloomington, (*100*)	1D
DGRP Cross 2 Parent 1	*yw; Sp/CyO; DGRP57*	Bloomington, (*100*)	1D
DGRP Cross 2 Parent 2	*yw; Sp/CyO; DGRP352*	Bloomington, (*100*)	1D
DGRP Cross 2 progeny	*yw; Sp/CyO; DGRP57/DGRP352*	Bloomington, (*100*)	1D
DGRP Cross 3 Parent 1	*yw; Sp/CyO; DGRP338*	Bloomington, (*100*)	1D
DGRP Cross 3 Parent 2	*yw; Sp/CyO; DGRP379*	Bloomington, (*100*)	1D
DGRP Cross 3 progeny	*yw; Sp/CyO; DGRP338/DGRP379*	Bloomington, (*100*)	1D
DGRP Cross 4 Parent 1	*yw; Sp/CyO; DGRP379*	Bloomington, (*100*)	1D
DGRP Cross 4 Parent 2	*yw; Sp/CyO; DGRP802*	Bloomington, (*100*)	1D
DGRP Cross 4 progeny	*yw; Sp/CyO; DGRP379/DGRP802*	Bloomington, (*100*)	1D
DGRP Cross 5 Parent 1	*yw; Sp/CyO; DGRP338*	Bloomington, (*100*)	1D
DGRP Cross 5 Parent 2	*yw; Sp/CyO; DGRP802*	Bloomington, (*100*)	1D
DGRP Cross 5 progeny	*yw; Sp/CyO; DGRP338/DGRP802*	Bloomington, (*100*)	1D
DGRP Cross 6 Parent 1	*yw; Sp/CyO; DGRP40*	Bloomington, (*100*)	1E
DGRP Cross 6 Parent 2	*yw; Sp/CyO; DGRP228*	Bloomington, (*100*)	1E

DGRP Cross 6 progeny	*yw; Sp/CyO; DGRP40/DGRP228*	Bloomington, (*100*)	1E
DGRP Cross 7 Parent 1	*yw; Sp/CyO; DGRP40*	Bloomington, (*100*)	1E
DGRP Cross 7 Parent 2	*yw; Sp/CyO; DGRP352*	Bloomington, (*100*)	1E
DGRP Cross 7 progeny	*yw; Sp/CyO; DGRP40/DGRP352*	Bloomington, (*100*)	1E
DGRP Cross 8 Parent 1	*yw; Sp/CyO; DGRP352*	Bloomington, (*100*)	1E
DGRP Cross 8 Parent 2	*yw; Sp/CyO; DGRP379*	Bloomington, (*100*)	1E
DGRP Cross 8 progeny	*yw; Sp/CyO; DGRP352/DGRP379*	Bloomington, (*100*)	1E
DGRP Cross 9 Parent 1	*yw; Sp/CyO; DGRP352*	Bloomington, (*100*)	1E
DGRP Cross 9 Parent 2	*yw; Sp/CyO; DGRP338*	Bloomington, (*100*)	1E
DGRP Cross 9 progeny	*yw; Sp/CyO; DGRP352/DGRP338*	Bloomington, (*100*)	1E
DGRP Cross 10 Parent 1	*yw; Sp/CyO; DGRP40*	Bloomington, (*100*)	1F
DGRP Cross 10 Parent 2	*yw; Sp/CyO; DGRP379*	Bloomington, (*100*)	1F
DGRP Cross 10 progeny	*yw; Sp/CyO; DGRP40/DGRP379*	Bloomington, (*100*)	1F
DGRP Cross 11 Parent 1	*yw; Sp/CyO; DGRP57*	Bloomington, (*100*)	1F
DGRP Cross 11 Parent 2	*yw; Sp/CyO; DGRP379*	Bloomington, (*100*)	1F
DGRP Cross 11 progeny	*yw; Sp/CyO; DGRP57/DGRP379*	Bloomington, (*100*)	1F
DGRP Cross 12 Parent 1	*yw; Sp/CyO; DGRP57*	Bloomington, (*100*)	1F
DGRP Cross 12 Parent 2	*yw; Sp/CyO; DGRP802*	Bloomington, (*100*)	1F
DGRP Cross 12 progeny	*yw; Sp/CyO; DGRP57/DGRP802*	Bloomington, (*100*)	1F
DGRP Cross 13 Parent 1	*yw; Sp/CyO; DGRP228*	Bloomington, (*100*)	1F

DGRP Cross 13 Parent 2	*yw; Sp/CyO; DGRP802*	Bloomington, (*100*)	1F
DGRP Cross 13 progeny	*yw; Sp/CyO; DGRP228/DGRP802*	Bloomington, (*100*)	1F
DGRP Cross 14 Parent 1	*yw; Sp/CyO; DGRP379*	Bloomington, (*100*)	1F
DGRP Cross 14 Parent 2	*yw; Sp/CyO; DGRP195*	Bloomington, (*100*)	1F
DGRP Cross 14 progeny	*yw; Sp/CyO; DGRP379/DGRP195*	Bloomington, (*100*)	1F
DGRP Cross 15 Parent 1	*yw; Sp/CyO; DGRP40*	Bloomington, (*100*)	1G
DGRP Cross 15 Parent 2	*yw; Sp/CyO; DGRP195*	Bloomington, (*100*)	1G
DGRP Cross 15 progeny	*yw; Sp/CyO; DGRP40/DGRP195*	Bloomington, (*100*)	1G
DGRP Cross 16 Parent 1	*yw; Sp/CyO; DGRP57*	Bloomington, (*100*)	1G
DGRP Cross 16 Parent 2	*yw; Sp/CyO; DGRP338*	Bloomington, (*100*)	1G
DGRP Cross 16 progeny	*yw; Sp/CyO; DGRP57/DGRP338*	Bloomington, (*100*)	1G
DGRP Cross 17 Parent 1	*yw; Sp/CyO; DGRP57*	Bloomington, (*100*)	1G
DGRP Cross 17 Parent 2	*yw; Sp/CyO; DGRP195*	Bloomington, (*100*)	1G
DGRP Cross 17 progeny	*yw; Sp/CyO; DGRP57/DGRP195*	Bloomington, (*100*)	1G
DGRP Cross 18 Parent 1	*yw; Sp/CyO; DGRP352*	Bloomington, (*100*)	1G
DGRP Cross 18 Parent 2	*yw; Sp/CyO; DGRP195*	Bloomington, (*100*)	1G
DGRP Cross 18 progeny	*yw; Sp/CyO; DGRP352/DGRP195*	Bloomington, (*100*)	1G
DGRP Cross 19 Parent 1	*yw; Sp/CyO; DGRP228*	Bloomington, (*100*)	1G
DGRP Cross 19 Parent 2	*yw; Sp/CyO; DGRP338*	Bloomington, (*100*)	1G
DGRP Cross 19 progeny	*yw; Sp/CyO; DGRP228/DGRP338*	Bloomington, (*100*)	1G

ss^{sin} / ss^{sin}	*yw; +; ss^{sin} / ss^{sin}*	(*101*)	1H
ss^{sin} / wild type	*yw; +; ss^{sin} / +*	(*101*)	1H
wild type / ss def	*yw; +; + / Df(3R)Exel7330*	Bloomington, (*163*)	2A; S1B
protein null / ss def	*yw; +; $ss^{D115.7}$ / Df(3R)Exel7330*	Bloomington, (*163, 164*)	2A
promoter Δ / ss def	*yw; +; $ss^{promoter\ \Delta}$ / Df(3R)Exel7330*	Bloomington, (*163*)	2A; S1B
early enhancer Δ / ss def	*yw; +; $ss^{early\ enhancer\ \Delta}$ / Df(3R)Exel7330*	Bloomington, (*163*)	2A; S1B
late enhancer Δ / ss def	*yw; +; $ss^{enhancer\ deletion}$ / Df(3R)Exel7330*	Bloomington, (*72, 163*)	2A; S1B
early enhancer Δ / protein null	*yw; +; $ss^{early\ enhancer\ \Delta}$ / $ss^{D115.7}$*	(*164*)	2B
late enhancer Δ / protein null	*yw; +; $ss^{enhancer\ deletion}$ / $ss^{D115.7}$*	(*72, 164*)	2C
early enhancer Δ / promoter Δ	*yw; +; $ss^{early\ enhancer\ \Delta}$ / $ss^{promoter\ \Delta}$*	N/A	2D
late enhancer Δ / promoter Δ	*yw; +; $ss^{enhancer\ deletion}$ / $ss^{promoter\ \Delta}$*	(*72*)	2E
early enhancer Δ / late enhancer Δ	*yw; +; $ss^{early\ enhancer\ \Delta}$ / $ss^{enhancer\ deletion}$*	(*72*)	2F
PRE12Δ / ss def	*yw; +; $ss^{PRE12\ deletion}$ / Df(3R)Exel6269*	Bloomington, (*72, 163*)	3B
sil1Δ / sil1Δ	*yw; +; $ss^{RS6279exc44}$*	*(167)*	3C
PRE12Δ / protein null	*yw; +; $ss^{PRE12\ deletion}$ / Df(3R)Exel7330*	Bloomington, (*72, 163*)	3D
sil1Δ / protein null	*yw; +; $ss^{RS6279exc44}$ / $ss^{D115.7}$*	(*164, 167*)	3E
$ss^{translocation}$ / PRE12Δ	T(1;3)$ss^{D114.3}$/ +; +; *$ss^{PRE12\ deletion}$* / T(1;3)$ss^{D114.3}$	(*72, 164*)	3F
$ss^{translocation}$ / sil1Δ	T(1;3)$ss^{D114.3}$/ +; +; T(1;3)$ss^{D114.3}$ / *$ss^{RS6279exc44}$*	(*164, 167*)	3G

PRE12Δ / promoter Δ	*yw; +; ss$^{PRE12\ deletion}$ / ss$^{promoter\ Δ}$*	(*72*)	3H
ss^{translocation} / ss def	T(1;3)*ss$^{D114.3}$*/ +; +; T(1;3)*ss$^{D114.3}$* / *Df(3R)Exel6269*	Bloomington, (*163, 164*)	4A
ss^{translocation} / wild type	T(1;3)*ss$^{D114.3}$*/ +; +; T(1;3)*ss$^{D114.3}$* / +	(*164*)	4B
ss^{translocation} / protein null	T(1;3)*ss$^{D114.3}$*/ +; +; T(1;3)*ss$^{D114.3}$* / *ss$^{D115.7}$*	(*164*)	4C
ss^{inversion} / ss def	*yw; +; ss^{GS2553exc80} / Df(3R)Exel7330*	Bloomington, (*163, 167*)	4D; 5C, F-H
ss^{inversion} / wild type	*yw; +; ss^{GS2553exc80} / +*	(*167*)	4E; 5B, E, G, I
ss^{inversion} / protein null	*yw; +; ss^{GS2553exc80} / ss$^{D115.7}$*	(*164, 167*)	4F
wild type/ Transgene A	*yw; +/ CyO; +/ pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330*	Bloomington, (*97, 163*)	S1C
early enhancer Δ / Transgene A	*yw; +/ CyO; ss$^{early\ enhancer\ Δ}$; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330*	Bloomington, (*97, 163*)	S1D
late enhancer Δ / Transgene A	*yw; +/ CyO; ss$^{enhancer\ deletion}$; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330*	Bloomington, (*72, 97, 163*)	S1E

Oligopaints probe libraries

Table 4: Genome coordinates targeted by Oligopaints probe libraries

Probe set	**Oligopaints library name**	**Genome coordinates targeted**	**Conjugated fluorophore**	**Figures**
ss^{inversion} Probe 1	high freq library 1	3R: 4,183,722-4,243,722	Cy5	5D-F
ss^{inversion} Probe 2	high freq library 2	3R: 16,404,001-16,464,001	Cy3	5D-F

FISH distance quantifications

Quantifications were performed as described in (*72*). All quantifications were performed in 3D on z-stacks with a slice thickness of 0.2 μm. Quantifications were performed manually using Fiji (*175, 176*). To chart the z position of each

FISH dot, a line was drawn through the dot and the Plot Profile tool was used to assess the stack in which the dot was brightest. To determine the x-y distance between the two FISH dots, a line was drawn from the center of one dot to the center of the other dot and the length of the line was measured with the Plot Profile tool. The distance between the FISH dots was then calculated in 3D. A total of 50 nuclei from three eye discs were quantified for each genotype (i.e. N=3, n=50).

Oligopaints probe design

Probes for DNA FISH were designed using the Oligopaints technique (*149, 170*), as described in (*72*). Target sequences were run through the bioinformatics pipeline available at http://genetics.med.harvard.edu/oligopaints/ to identify sets of 50-bp optimized probe sequences (i.e. “libraries”) tiled across the DNA sequence of interest. Five 19-bp barcoding primers, gene F and R, universal (univ) F and R, and random (rando) R, were appended to the 5’ and 3’ ends of each probe sequence. The gene F and R primers allowed PCR amplification of a probe library of interest out of the total oligo pool, and the univ F and R primers allowed conjugation of fluorophores, generation of single-stranded DNA probes, and PCR addition of secondary sequences to amplify probe signal. The rando R primer was appended at the 3’ end to maintain a constant sequence length between all probes that were synthesized in the custom oligo pool.

Barcoding primer sequences were taken from a set of 240,000 randomly generated, orthogonal 25-bp sequences (*171*) and run through a custom script to select 19-bp sequences with ≤15-bp homology to the *Drosophila* genome. Primers were appended to probe sequences using the orderFile.py script available at http://genetics.med.harvard.edu/oligopaints/. Completed probe libraries were synthesized as custom oligo pools by Custom Array, Inc. (Bothell, WA).

DNA FISH

DNA FISH was performed as described in (*72*). 20-50 eye-antennal discs attached to mouth hooks from third instar larvae were collected on ice and fixed in 129 μL ultrapure water, 20 μL 10X PBS, 1 μL Tergitol NP-40, 600 μL heptane, and 50 μL fresh 16% formaldehyde. Tubes containing the fixative and eye discs were shaken vigorously by hand, then fixed for 10 minutes at room temperature with nutation. Eye discs were then given three quick washes in 1X PBX, followed by three five-minute washes in PBX at room temperature with nutation. Eye discs were then removed from the mouth hooks and blocked for 1 hour in 1X PBX+1% BSA at room temperature with nutation. They were then incubated in primary antibody diluted in 1X PBX overnight at 4°C with nutation. Next, eye discs were washed three times in 1X PBX for 20 minutes and incubated in secondary antibody diluted in 1X PBX for two hours at room temperature with nutation. Eye discs were then washed two times for 20 minutes in 1X PBX, followed by a 20-minute wash in 1X PBS. Next, discs were given one 10-minute wash in 20%

formamide+2X SSCT (2X SSC+.001% Tween-20), one 10-minute wash in 40% formamide+2X SSCT, and two 10-minute washes in 50% formamide+2X SSCT. Discs were then predenatured by incubating for four hours at 37°C, three minutes at 92°C, and 20 minutes at 60°C. Primary probes were added in 45 μL hybridization buffer consisting of 50% formamide+2X SSCT+2% dextran sulfate (w/v), + 1 μL RNAse A. All probes were added at a concentration of ≥5 pmol fluorophore/μL. For FISH experiments in which a single probe was used, 4 μL of probe was added. For FISH experiments in which two probes were used, 2 μL of each probe was added. After addition of probes, eye discs were incubated at 91°C for three minutes and at 37°C for 16-20 hours with shaking. Eye discs were then washed for 1 hour at 37°C with shaking in 50% formamide+2X SSCT. 1 μL of each secondary probe was added at a concentration of 100 pmol/μL in 50 μL of 50% formamide+2X SSCT. Secondary probes were hybridized for 1 hour at 37°C with shaking. Eye discs were then washed twice for 30 minutes in 50% formamide+2X SSCT at 37°C with shaking, followed by three 10-minute washes at room temperature in 20% formamide+2X SSCT, 2X SSCT, and 2X SSC with nutation. Discs were mounted in SlowFade Gold immediately after the final 2X SSC wash, and imaged using a Zeiss LSM700 confocal microscope.

Statistical analysis for FISH

FISH datasets were tested for a Gaussian distribution using a D'Agostino and Pearson omnibus normality test and a Shapiro-Wilk normality test. These tests indicated a non-Gaussian distribution for the wild type control and for $ss^{inversion}$ /

WT, so datasets were tested for statistical significance using a one-way ANOVA on ranks with Dunn's multiple comparisons test.

4.6: Author Contributions

The author list for this article was as follows: Kayla Viets, Chaim Chernoff, Elizabeth Urban, Jeong Han, Adrienne Chen, Caitlin Anderson, Sang Tran, Daniel Konzman, and Robert J. Johnston Jr. I was a co-first author on this publication with Chaim Chernoff and Elizabeth Urban. I was responsible for conceiving experiments, guiding other authors in completion of experiments, analyzing data, writing and editing the text, and performing the experiments in Figure 1D-G. Chaim Chernoff was responsible for conceiving experiments, analyzing data, editing the text, and performing the experiments in Figures 3-5 and Supplemental Figures 1-3. Elizabeth Urban was responsible for conceiving experiments, guiding other authors in completion of experiments, analyzing data, editing the text, and performing the experiments in Figure 2. Jeong Han was responsible for performing the experiments in Supplemental Figure 4. Adrienne Chen assisted with fly crosses and data analysis. Caitlin Anderson generated CRISPR alleles and performed the experiments for Figure 1H. Sang Tran generated CRISPR alleles. Daniel Konzman performed preliminary experiments examining nuclear localization of the $ss^{translocation}$ allele. Robert J. Johnston was the corresponding author and was responsible for conceiving experiments and editing the text.

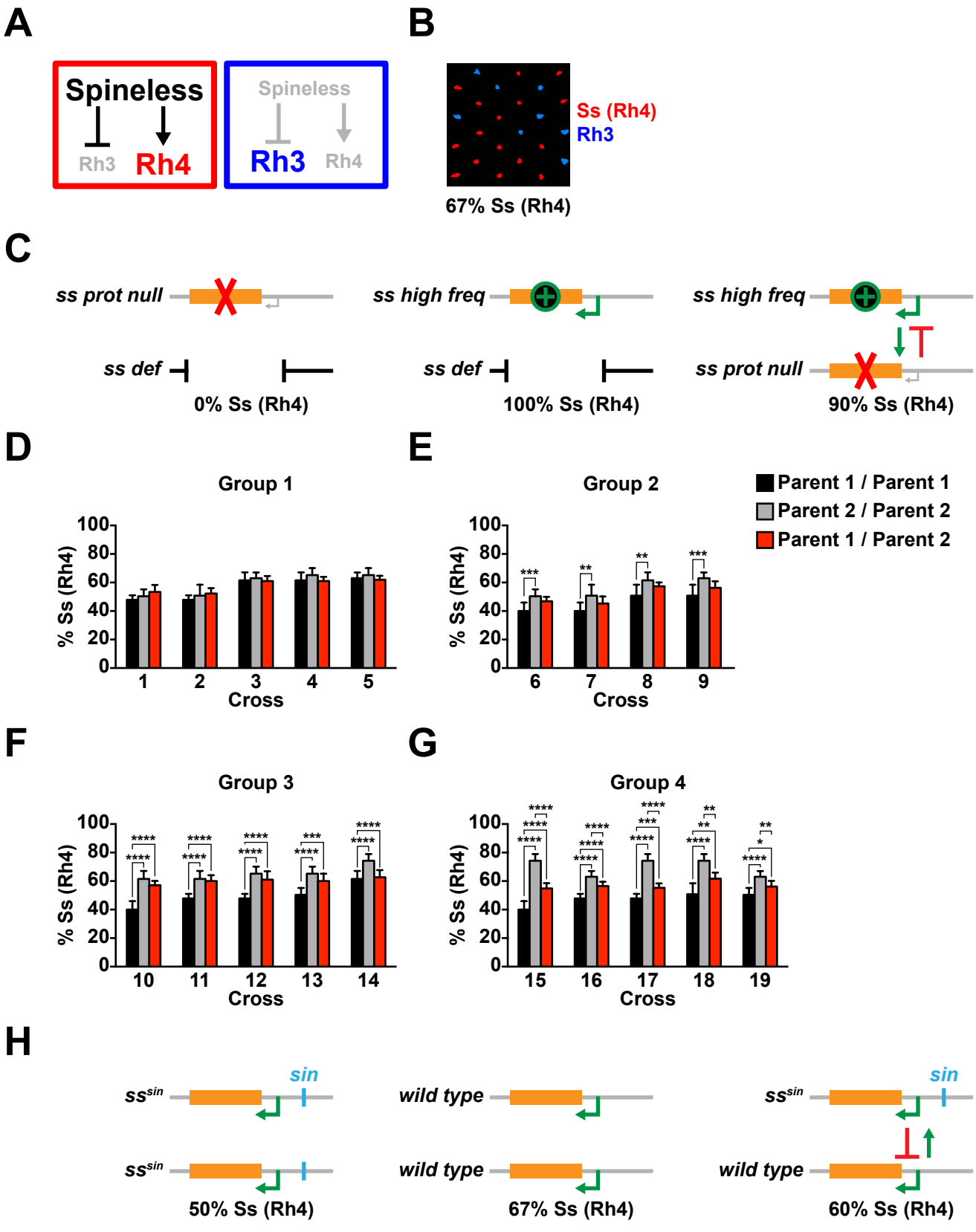

A
Spineless
Rh3
Rh4
Spineless
Rh3
Rh4
B
Ss (Rh4)
Rh3
67% Ss (Rh4)
C
ss prot null
ss def
0% Ss (Rh4)
ss high freq
ss def
100% Ss (Rh4)
ss high freq
ss prot null
90% Ss (Rh4)
D
Group 1
% Ss (Rh4)
Cross
E
Group 2
% Ss (Rh4)
Cross
Parent 1 / Parent 1
Parent 2 / Parent 2
Parent 1 / Parent 2
F
Group 3
% Ss (Rh4)
Cross
G
Group 4
% Ss (Rh4)
Cross
H
ss^sin
sin
ss^sin
50% Ss (Rh4)
wild type
wild type
67% Ss (Rh4)
ss^sin
sin
wild type
60% Ss (Rh4)

Figure 1

Figure 1: *ss* transvection "averages" the expression frequencies of naturally derived alleles.

A. Spineless (Ss) activates Rh4 and represses Rh3.

B. Ss is expressed in ~67% of R7s. Red: Ss (Rh4), blue: Rh3.

C. A cross between flies containing the low-frequency *ss prot null* allele and flies containing the high-frequency *ss high freq* allele results in progeny with an intermediate Ss (Rh4) expression frequency. Red "X" indicates a mutation in exon 4 of *ss prot null* that prevents the production of a functional protein product. Green "+" indicates a mutation that causes upregulation of *ss high freq*. Green and gray arrows indicate transcription start site. Orange box indicates *ss* protein coding region.

D-G. Graphs of Ss (Rh4) expression for DGRP cross Groups 1-4. ****=p<0.0001, ***=p<0.001, **=p<0.005, *=p<0.05, unpaired t-test.

D. Group 1: Ss (Rh4) frequencies of parents and progeny not statistically different.

E. Group 2: Ss (Rh4) frequencies of parents statistically different, Ss (Rh4) frequency of progeny not statistically different from either parent.

F. Group 3: Ss (Rh4) frequencies of parents statistically different, Ss (Rh4) frequency of progeny statistically different from one parent.

G. Group 4: Ss (Rh4) frequencies of parents statistically different, Ss (Rh4) frequency of progeny statistically different from both parents.

H. A cross between flies homozygous for ss^{sin} and flies homozygous for wild type *ss* results in progeny with an intermediate Ss (Rh4) expression frequency. Green

arrow indicates transcription start site. Orange box indicates *ss* protein coding region.

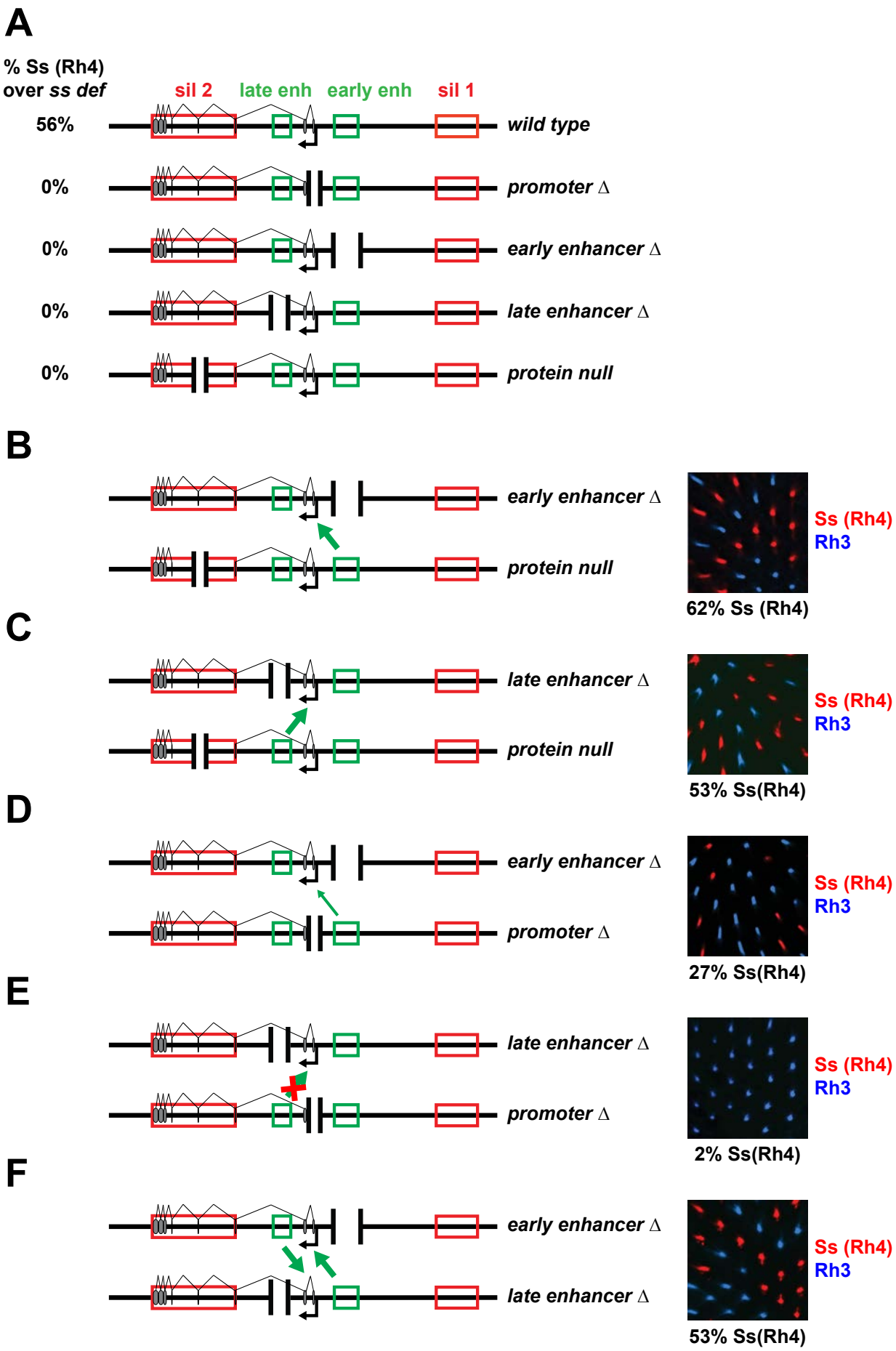

A
% Ss (Rh4)
over ss def
sil 2
late enh
early enh
sil 1
56%
wild type
0%
promoter Δ
0%
early enhancer Δ
0%
late enhancer Δ
0%
protein null
B
early enhancer Δ
protein null
Ss (Rh4)
Rh3
62% Ss (Rh4)
C
late enhancer Δ
protein null
Ss (Rh4)
Rh3
53% Ss(Rh4)
D
early enhancer Δ
promoter Δ
Ss (Rh4)
Rh3
27% Ss(Rh4)
E
late enhancer Δ
promoter Δ
Ss (Rh4)
Rh3
2% Ss(Rh4)
F
early enhancer Δ
late enhancer Δ
Ss (Rh4)
Rh3
53% Ss(Rh4)

Figure 2

Figure 2: Activating *ss* transvection requires an enhancer and promoter in *cis*.

A. Schematics of the wild type, *promoter Δ*, *early enhancer Δ*, *late enhancer Δ*, and *protein null* alleles. Ss (Rh4) percentages indicate the expression frequency for each allele over *ss def.* sil 1: silencer 1, early enh: early enhancer, late enh: late enhancer, sil 2: silencer. Gray ovals: exons, black arrow: promoter.

B-F. Schematics and representative images of Ss (Rh4) expression frequency for *early enhancer Δ / protein null*, *late enhancer Δ / protein null, early enhancer Δ / promoter Δ*, *late enhancer Δ / promoter Δ,* and *early enhancer Δ / late enhancer Δ*.

B-C, F. Large green arrow indicates that activating transvection is occurring between *ss* alleles.

D. Small green arrow indicates that activating transvection is occurring between *ss* alleles, but at a weaker level than in wild-type flies.

E. Green arrow with red “X” indicates that activating transvection is not occurring between *ss* alleles.

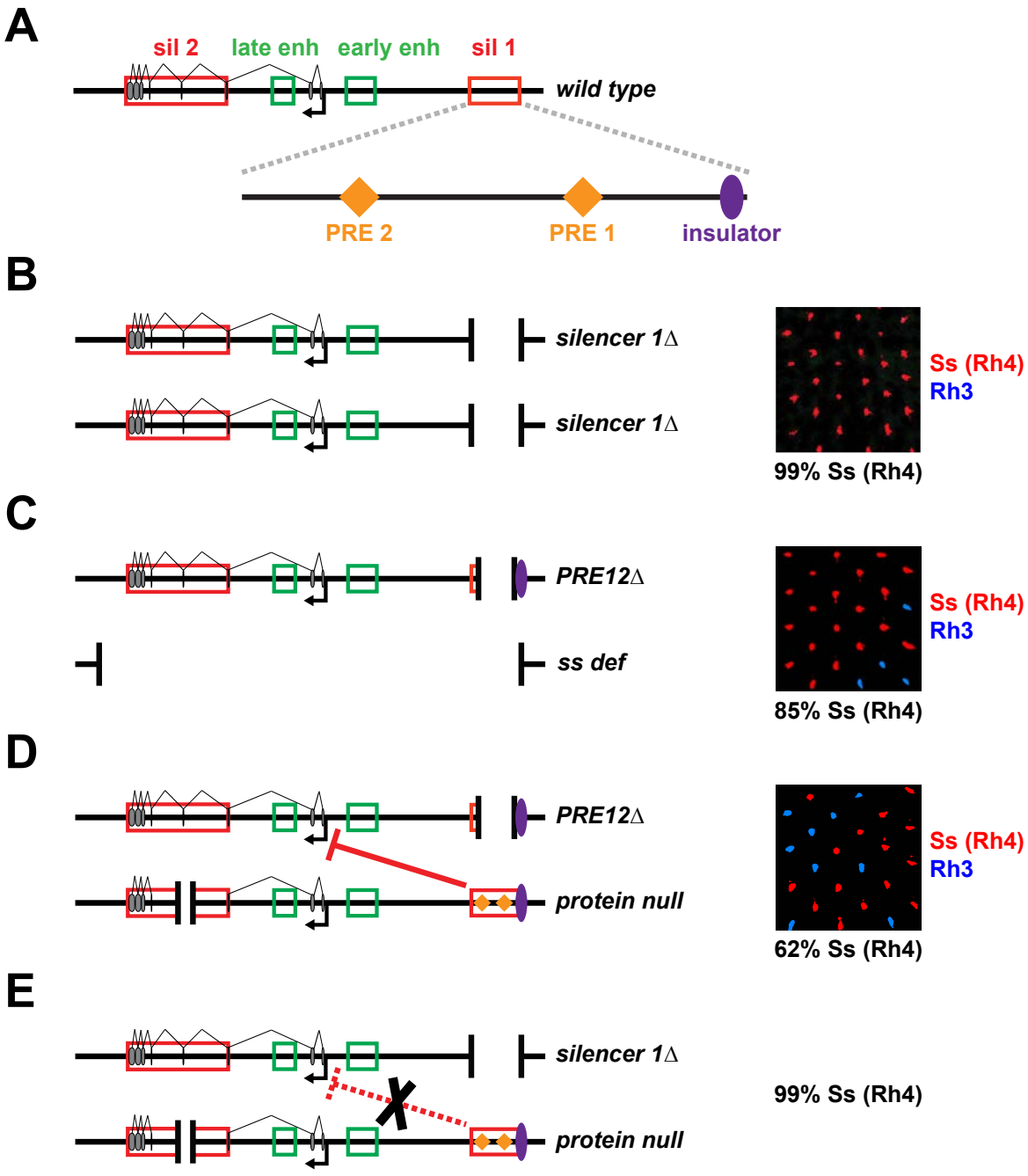

A
sil 2
late enh
early enh
sil 1
wild type
PRE 2
PRE 1
insulator
B
silencer 1Δ
silencer 1Δ
Ss (Rh4)
Rh3
99% Ss (Rh4)
C
PRE12Δ
ss def
Ss (Rh4)
Rh3
85% Ss (Rh4)
D
PRE12Δ
protein null
Ss (Rh4)
Rh3
62% Ss (Rh4)
E
silencer 1Δ
protein null
99% Ss (Rh4)

Figure 3

Figure 3: Repressing *ss* transvection requires two PREs and an insulator element.

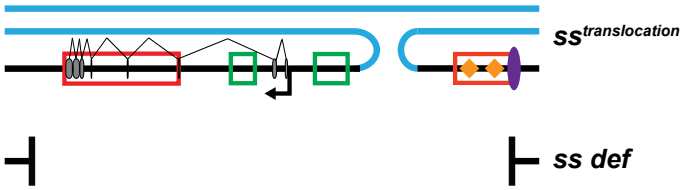
A. Schematic of the wild-type *ss* locus with an expanded view of *silencer 1*, which contains two PREs and an insulator. sil 1: silencer 1, early enh: early enhancer, late enh: late enhancer, sil 2: silencer. Gray ovals: exons, black arrow: promoter.

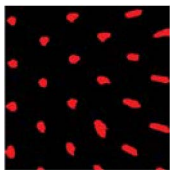
B-E. Schematics and representative images of Ss (Rh4) expression frequency for *silencer1Δ / silencer1Δ, PRE12Δ / ss def*, *PRE12Δ / protein null,* and *silencer 1Δ / protein null.* Purple oval: insulator, orange diamonds: PREs.

D. Red flat arrow indicates that *silencer 1* on *protein null* is performing repressing transvection on *PRE12Δ.*

E. Dotted red flat arrow with black "X" indicates that *silencer 1* on *protein null* is not performing repressing transvection on *silencer 1Δ.*

A

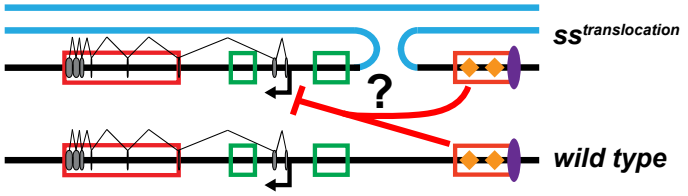


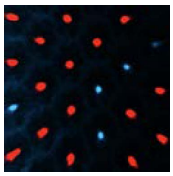


Ss (Rh4)
Rh3

100% Ss (Rh4)

B

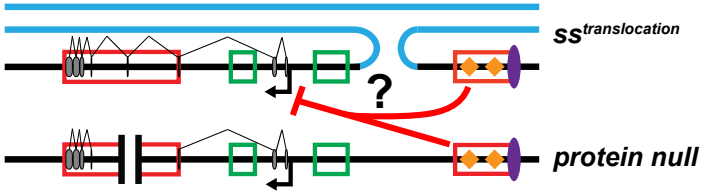


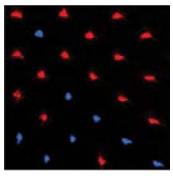


Ss (Rh4)
Rh3

73% Ss (Rh4)

C

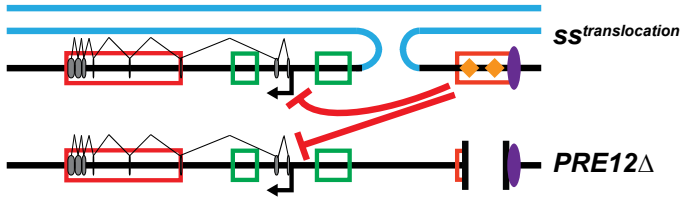


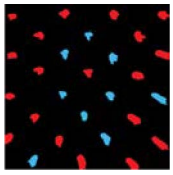


Ss (Rh4)
Rh3

73% Ss (Rh4)

D

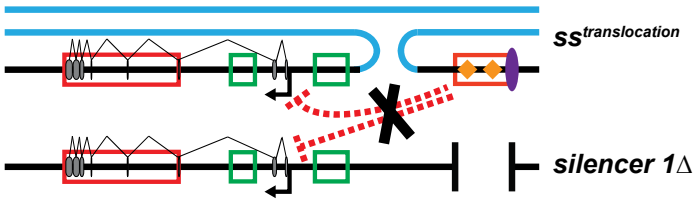




Ss (Rh4)
Rh3

73% Ss (Rh4)

E



Ss (Rh4)
Rh3

99% Ss (Rh4)

Figure 4

Figure 4: Silencer 1 repressing transvection on *ss^{translocation}* requires two copies of an insulator.

A-E: Schematics and representative images of Ss (Rh4) expression frequency for *ss^{translocation} / ss def*, *ss^{translocation} /* wild type, *ss^{translocation} / protein null*, *ss^{translocation} / PRE12Δ*, and *ss^{translocation} / silencer 1Δ.* Red boxes: *silencers 1* and *2*, green boxes: *early enhancer* and *late enhancer*, gray ovals: exons, black arrow: promoter, blue lines: X chromosome, black lines: Chr 3.

B. Red flat arrows with a "?" indicate that it is unclear which copy of *silencer 1* is repressing Ss (Rh4) expression frequency in the *ss^{translocation} /* wild type background.

C. Red flat arrows with a "?" indicate that it is unclear which copy of *silencer 1* is repressing Ss (Rh4) expression frequency in the *ss^{translocation} / protein null* background.

D. Red flat arrows indicate that *silencer 1* on *ss^{translocation}* is performing repressing transvection on *PRE12Δ* and on the protein coding region of *ss^{translocation}.*

E. Dotted red flat arrows with black "X" indicate that *silencer 1* on *ss^{translocation}* is not performing repressing transvection on *silencer 1Δ* or on the protein coding region of *ss^{translocation}.*

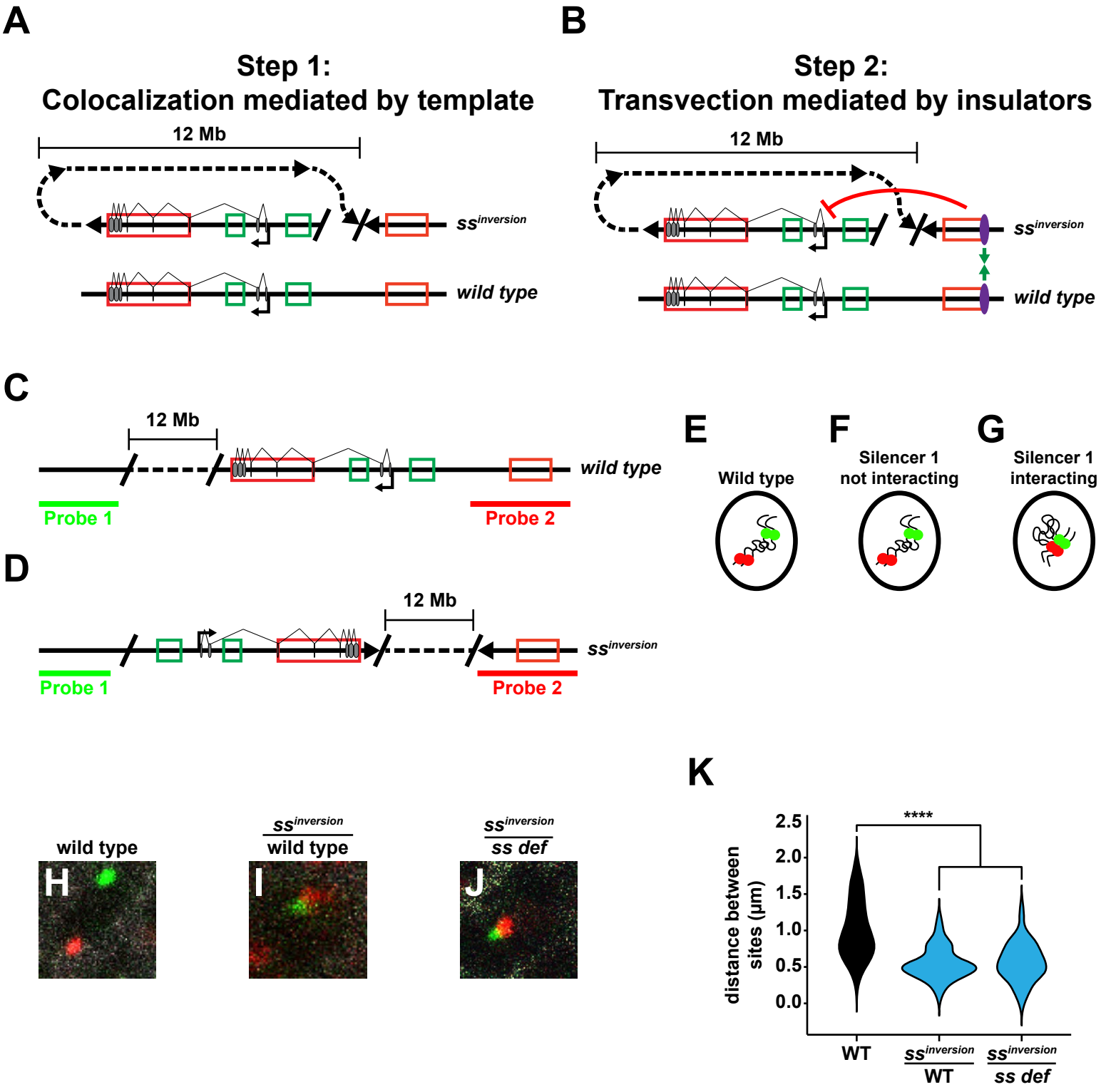

A
Step 1:
Colocalization mediated by template
12 Mb
ss inversion
wild type
B
Step 2:
Transvection mediated by insulators
12 Mb
ss inversion
wild type
C
12 Mb
wild type
Probe 1
Probe 2
D
12 Mb
ss inversion
Probe 1
Probe 2
E
Wild type
F
Silencer 1
not interacting
G
Silencer 1
interacting
wild type
H
ss inversion
wild type
I
ss inversion
ss def
J
K
****
distance between sites (µm)
2.5
2.0
1.5
1.0
0.5
0.0
WT
ss inversion
WT
ss inversion
ss def

Figure 5

Figure 5: The inverted *ss silencer 1* colocalizes with the *ss* protein coding region independent of a template allele.

A. Schematic showing a possible chromatin conformation for step 1 of our hypothesized repressive transvection mechanism. The wild type allele of *ss* acts as a template for homologous pairing, allowing the protein coding region of $ss^{inversion}$ to loop back and interact with *silencer 1* of $ss^{inversion}$. Red boxes: *silencers 1* and *2*, green boxes: *early enhancer* and *late enhancer*, gray ovals: exons, black arrow: promoter. Dotted lines indicate the 12-Mb region that separates *silencer 1* and the *ss* protein coding region on the $ss^{inversion}$ allele.

B. Schematic of step 2 of our hypothesized repressive transvection mechanism. Two copies of the insulator allow *silencer 1* of $ss^{inversion}$ to repress the protein coding region of $ss^{inversion}$. Green arrows indicate an interaction between the two copies of the insulator. Red flat arrow indicates that *silencer 1* of $ss^{inversion}$ is performing repressive transvection on the protein coding region of $ss^{inversion}$. Red boxes: *silencers 1* and *2*, green boxes: *early enhancer* and *late enhancer*, gray ovals: exons, black arrow: promoter, purple ovals: insulator. Dotted lines indicate the 12-Mb region that separates *silencer 1* and the *ss* protein coding region on the $ss^{inversion}$ allele.

C-D. Schematics indicating the locations on a wild-type *ss* allele and on $ss^{inversion}$ targeted by probe sets 1 and 2. Red boxes: *silencers 1* and *2*, green boxes: *early enhancer* and *late enhancer*, gray ovals: exons, black arrow: promoter.

E-G. Schematics of nuclear localization of probe sets 1 and 2 in a wild type control, a nucleus where *silencer 1* of $ss^{inversion}$ is not interacting with the protein

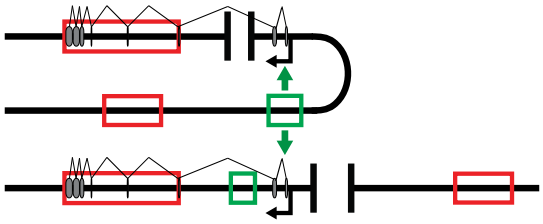
coding region of *ss*^{inversion}, and a nucleus where *silencer 1* of *ss*^{inversion} is interacting with the protein coding region of *ss*^{inversion}.

H-J. Representative DNA FISH images of wild type control, *ss*^{inversion} / wild type, and *ss*^{inversion} / *ss def*. Green: probe set 1, red: probe set 2, white: Lamin B (nuclear marker). Scale bars=1 μm.

K. Quantification of **5H-J**. Black: control, blue: experimental lines. ****=p<0.0001, one-way ANOVA on ranks with Dunn's multiple comparisons test.

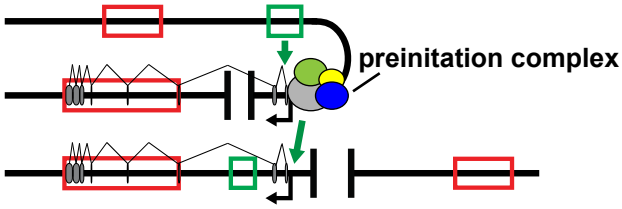
A

Simultaneous looping



B

Enhancer-promoter complex



C

Looping to active conformation

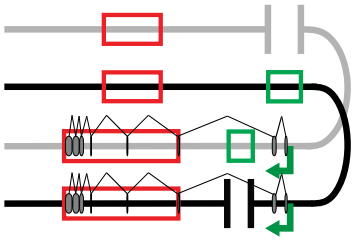


Figure 6

Figure 6: Models for the mechanism of *ss* activating transvection.

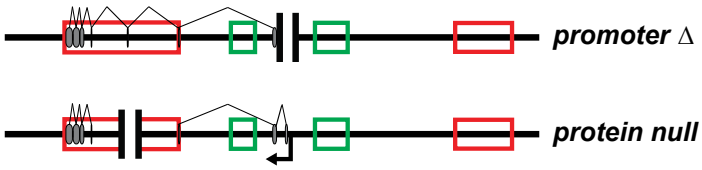
A-C. Red boxes: *silencers 1* and *2*, green boxes: *early enhancer* and *late enhancer*, gray ovals: exons.

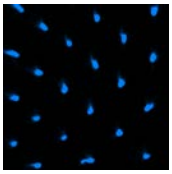
A. Simultaneous looping. Green arrows indicate that the enhancer is activating expression from both *ss* alleles. Black arrows: promoter.

B. Enhancer-promoter complex. Green arrows indicate that the enhancer is activating the promoter in *cis*, and that together the enhancer and promoter in *cis* activate the promoter in *trans*. Black arrows: promoter.

C. Looping to active conformation. Green arrows indicate that both alleles are expressed due to an activating looping conformation.

A





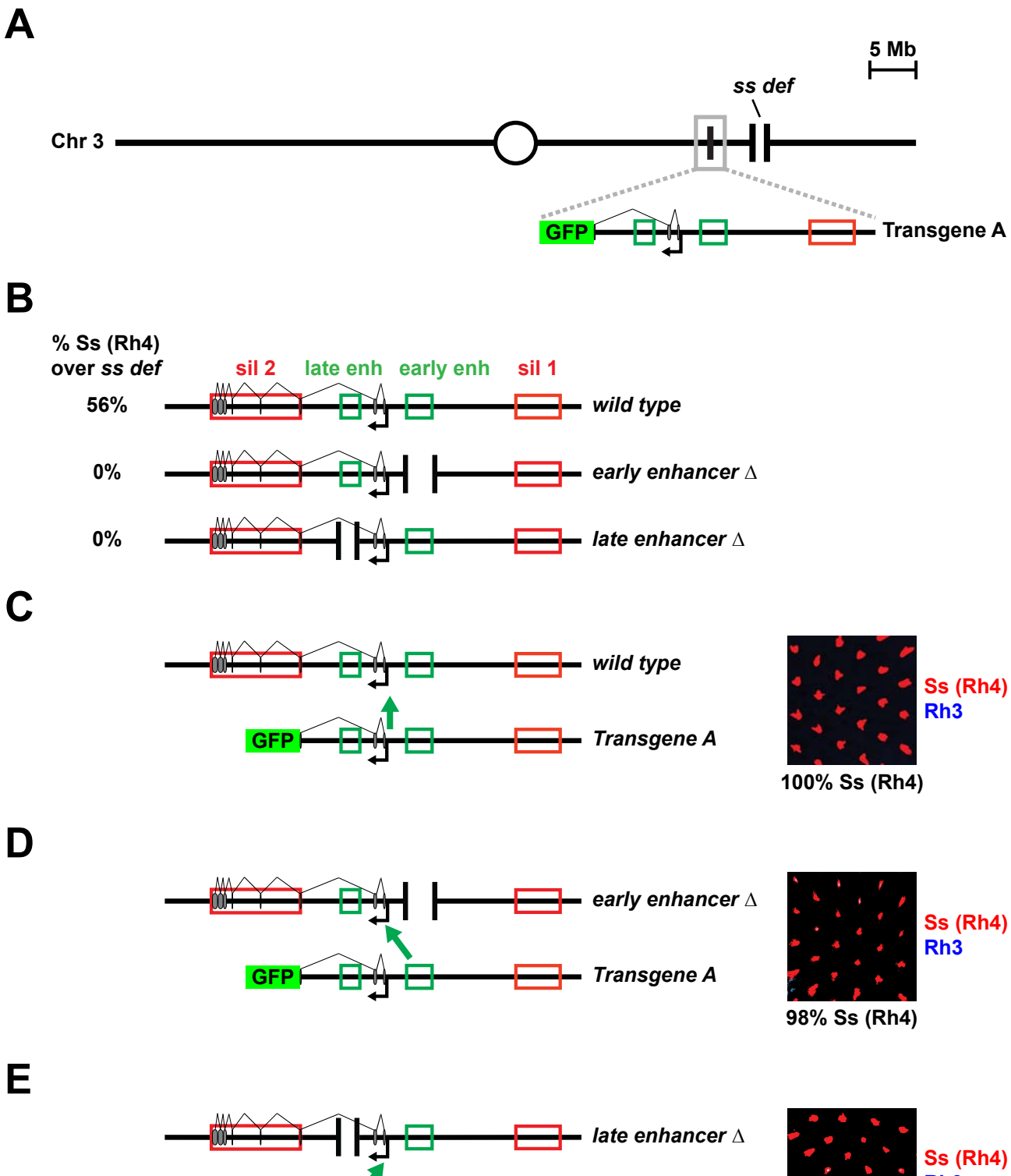
0% Ss(Rh4)

Supplemental Fig 1

Supplemental Figure 1: *promoter* Δ does not perform transvection with *protein null.*

A. Schematic of *promoter* Δ */ protein null*. Red boxes: *silencers 1* and *2*, green boxes: *early enhancer* and *late enhancer*, black arrow: promoter, gray ovals: exons.

B. Representative image of *promoter* Δ */ protein null*.



Supplemental Fig 2

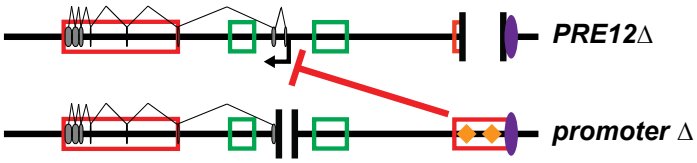
Supplemental Figure 2: *late enhancer Δ* and *early enhancer Δ* are competent to perform transvection with Transgene A.

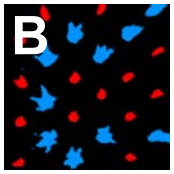
A. Schematic of the *Transgene A* allele, in which *Transgene A* is recombined with *ss def*, with an expanded view of *Transgene A,* which contains the *cis* regulatory regions and first three exons of *ss* and is marked with a GFP tag. Red box: *silencer 1*, green boxes: *early enhancer* and *late enhancer*, black arrow: promoter, gray ovals: exons.

B. Schematics of the wild type, *early enhancer Δ*, and *late enhancer Δ* alleles. Ss (Rh4) percentages indicate the expression frequency for each allele over *ss def*. sil 1: silencer 1, early enh: early enhancer, late enh: late enhancer, sil 2: silencer. Gray ovals: exons, black arrow: promoter.

C-E. Schematics and representative images of Ss (Rh4) expression frequency for wild type / *Transgene A*, *early enhancer Δ / Transgene A,* and *late enhancer Δ / Transgene A*. Green arrows indicate that *Transgene A* is performing activating transvection.

A





Ss (Rh4)
Rh3

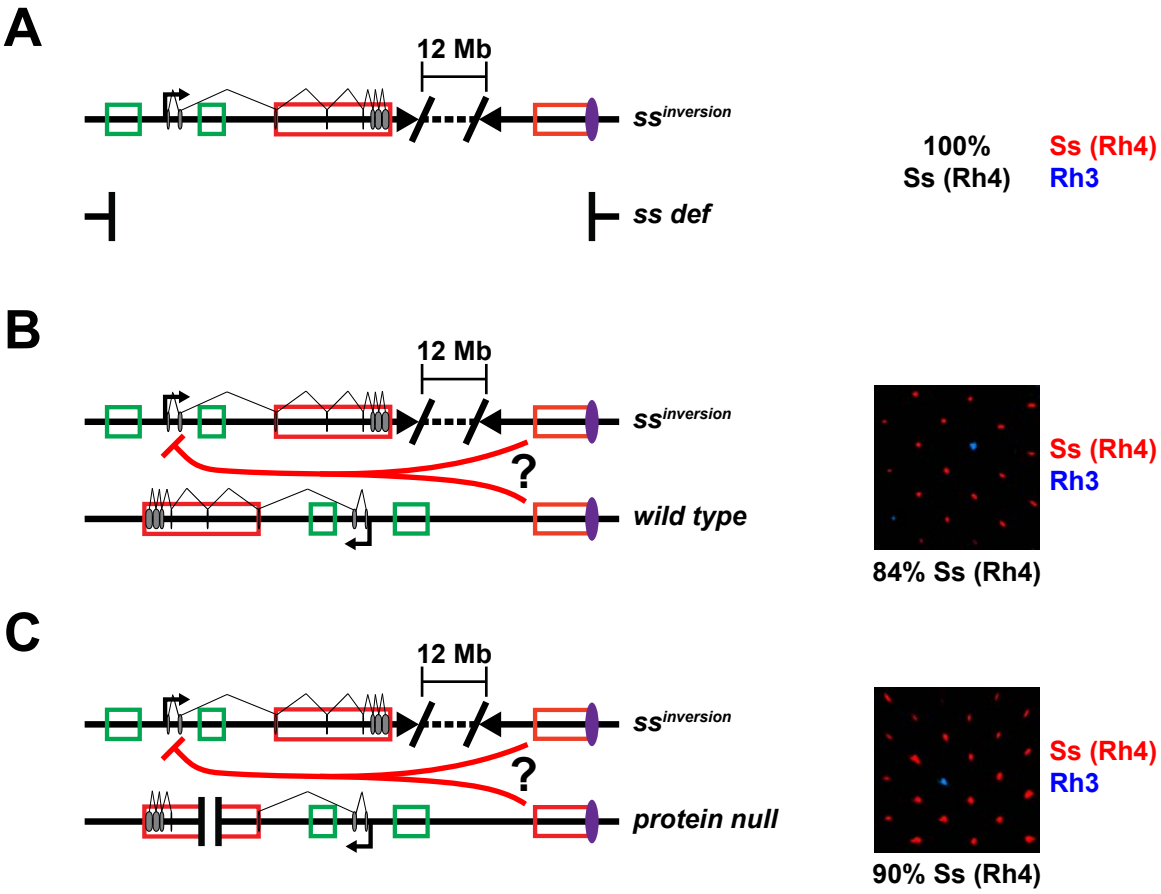
57% Ss (Rh4)

Supplemental Fig. 3

Supplemental Figure 3: The PREs of *promoter Δ are competent to perform repressing transvection on PRE12Δ*.

A. Schematic of *PRE12Δ / promoter Δ*. Red boxes: *silencers 1* and *2*, green boxes: *early enhancer* and *late enhancer*, black arrow: promoter, gray ovals: exons, orange diamonds: PREs, purple ovals: insulators. Red flat arrow indicates that *silencer 1* on *promoter Δ* is performing repressing transvection on *PRE12Δ*.

B. Representative image of *PRE12Δ / promoter Δ*.



Supplemental Fig. 4

Supplemental Figure 4: Silencer 1 repressing transvection on *ss^{inversion}* requires two copies of an insulator.

A-C. Schematics and representative images of *ss^{inversion} / ss def, ss^{inversion} / wild type, ss^{inversion} / protein null, ss^{inversion} / PRE12Δ,* and *ss^{inversion} / silencer 1Δ*. Red boxes: *silencers 1* and *2*, green boxes: *early enhancer* and *late enhancer*, black arrow: promoter, gray ovals: exons, purple ovals: insulators.

B. Red flat arrows with a "?" indicate that it is unclear which copy of *silencer 1* is repressing Ss (Rh4) expression frequency in the *ss^{inversion} /* wild type background.

C. Red flat arrows with a "?" indicate that it is unclear which copy of *silencer 1* is repressing Ss (Rh4) expression frequency in the *ss^{inversion} / protein null* background.

Chapter 5: Unpublished Data

5.1: Button pairing is position-independent

Transgene E, a “pairer” transgene that spans a button on chromosome 3R, drives pairing with endogenous *ss* when inserted on chromosome 2L and on chromosome 3L (*72*), suggesting that buttons drive pairing with their homologous sequences independent of their location in the genome. To test this hypothesis, we examined an additional 3R button transgene, *Transgene A*, which drives pairing when inserted on chromosome 2L (*72*). We inserted *Transgene A* onto chromosome 3L and tested whether it drove pairing with its endogenous locus on chromosome 3R using DNA Oligopaints FISH (**Fig. 1A**)(*149, 170*). As the endogenous and transgenic sequences were identical, we distinguished between them using a two-color FISH strategy, labeling the sequence neighboring the endogenous locus with red probes and the sequence neighboring the transgene insertion site with green probes (**Fig. 1A**). Similar to *Transgene E*, *Transgene A* drove pairing with its endogenous locus when inserted onto chromosome 3L (**Fig. 1B-C, E**), indicating that buttons drive pairing regardless of their genomic location.

Transgenes that do not encompass entire buttons do not drive pairing with their endogenous loci (“non-pairers”)(*72*). To test whether this inability to drive pairing was affected by transgene insertion site, we examined *Transgene L*, which does not drive pairing with its endogenous locus on 3R when inserted onto chromosome 2L (*72*). We inserted *Transgene L* onto chromosome 3L and tested its ability to drive pairing using a two-color FISH strategy (**Fig. 1A**). *Transgene L*

did not drive pairing with its endogenous locus from chromosome 3L (**Fig. 1B, D-E**), suggesting that non-pairer transgenes cannot drive pairing from any location in the genome.

5.2: Identification of additional buttons across the *Drosophila* genome

We previously identified button sequences located on chromosomes X, 2L, 2R, and 3R (*72*). TADs are a feature of buttons: out of all transgenes tested, 80% of pairers span a complete TAD, while only 12% of non-pairers span a complete TAD (*72*). To further test the hypothesis that TADs drive pairing, we selected two transgenes encompassing entire TADs on chromosomes 3L and 3R (*Transgenes EE* and *GG*; **Fig. 2E, K**) and one transgene that did not encompass an entire TAD, taken from chromosome 3L (*Transgene FF*; **Fig. 2F**). We inserted each transgene at the same site on chromosome 2L and tested whether they could drive pairing with their endogenous locus using a one-color FISH strategy, in which the identical transgene and endogenous sequences were labeled with the same red fluorescent probes (**Fig. 2A, H**). With this 1-color strategy, FISH punctae $\leq$0.4 μm apart could not be distinguished as separate and were assigned a distance of 0.4 μm apart. *Transgenes EE* and *GG* drove pairing (**Fig. 2B-C, G, I-J, L**), while *Transgene FF* did not drive pairing (**Fig. 2B, D, G**), further indicating that TADs drive pairing.

We have identified button TADs on chromosomes X, 2L, 2R, 3L, and 3R, strongly supporting the conclusion that specialized TADs interspersed across the *Drosophila* genome are responsible for bringing homologous chromosomes

together. *Drosophila* chromosome 4 also pairs, but its TADs are less well-defined than the other *Drosophila* chromosomes (*11*). We hypothesized that weaker TADs on chromosome 4 might drive the homologous pairing of this chromosome. Therefore, we adjusted our TAD-calling parameters to identify TADs on chromosome 4. Unlike for the other chromosomes, where we used a directionality index score of 0.8 or higher to call a TAD, we called a TAD on chromosome 4 if there was any positive directionality index peak followed by a negative directionality index peak. Using these parameters, we identified a TAD near the *dati* locus on chromosome 4 (**Fig. 2P**). We inserted *Transgene HH*, which spans this TAD, onto chromosome 2L and tested whether it could drive pairing with its endogenous locus on chromosome 4 (**Fig. 2M**). Similar to transgenes spanning TADs taken from other chromosomes, *Transgene HH* drove pairing with its endogenous locus (**Fig. 2N-O, Q**). Together, our data suggest that TADs drive homologous chromosome pairing between all of the *Drosophila* chromosomes.

5.3: Polycomb mutations affect pairing between copies of *ss*

TADs are an important contributor to homologous chromosome pairing, but we were curious whether *trans* factors are also involved in driving pairing between transgenes and their homologous endogenous loci. Previous studies have identified the Polycomb Group Complex as an important factor in both homologous pairing and interchromosomal clustering of heterologous sequences (*67, 70, 71*). We therefore tested the effects of the Polycomb mutant alleles Pc^{4}

and Pc^{xt} on pairing between *Transgene E* on chromosome 2L and its endogenous locus on chromosome 3R (**Fig. 3A, F**). Both Polycomb mutant alleles caused a reduction in pairing compared to a wild type control (**Fig. 3A-J**), indicating that Polycomb contributes to homologous chromosome pairing.

5.4: Investigating the effects of condensin mutations on pairing between *ss* copies

The condensin II protein complex has been previously identified as an "anti-pairer"; overexpression of condensin proteins causes a loss of pairing, while knockdown of condensin proteins increases pairing (*60, 66*). We used alleles of CapH2, a component of the condensin complex, to investigate the role of condensin II in pairing of *Transgene E* on chromosome 2L with its endogenous site on chromosome 3R (**Fig. 3A**). We hypothesized that CapH2 overexpression would cause decreased pairing, while CapH2 mutants would cause increased pairing. Using a one-color FISH strategy, we found that overexpressing CapH2 in our cells of interest (larval photoreceptors) using an *eyeless* driver did not affect *Transgene E* pairing (**Fig. 3K, O**). Surprisingly, one of the three CapH2 mutants (Mutant 1) we examined caused a decrease in pairing between *Transgene E* and its endogenous site (**Fig. 3L, O**). Mutants 2 and 3 had no effect on *Transgene E* pairing (**Fig. 3M-O**). Thus, further investigation with additional condensin II mutants and with stronger drivers of CapH2 overexpression is required to determine the role of condensin in transgene-driven pairing.

5.5: Individual DNA element deletions do not affect *ss* pairing

CRISPR deletions of DNA elements within the *spineless* (*ss*) locus, including the promoter and PREs (**Fig. 4A**), reduce transvection between alleles of *ss (98)*. To test whether this loss of transvection is due to a loss of homologous pairing between *ss* copies, we examined the effects of CRISPR deletions on pairing between endogenous *ss* and *Transgene E*, which contains the *ss* locus. We used a one-color FISH strategy to test whether *Transgene E* could drive pairing with alleles of *ss* containing CRISPR deletions of the promoter (*prom∆*) or PREs (*PRE12∆*)(**Fig. 4A-B**). *Transgene E* drove pairing with both *prom∆* and *PRE12∆*, indicating that deletions of small DNA elements within the *ss* locus do not affect homologous pairing between copies of *ss* (**Fig. 4C-F, H**). These data further support our earlier conclusion that homologous chromosome pairing is necessary but not sufficient for transvection: while these CRISPR alleles are still competent to pair, they cannot perform transvection.

5.6: Deletion of the interior of the *ss* TAD reduces pairing between copies of *ss*

TADs drive homologous pairing, but what features of TADs allow them to find their homologues within the nucleus? Studies using haplotype-resolved HiC suggest that TAD boundaries may be responsible for pairing (*62, 63*). Supporting this hypothesis, insulators have been previously linked to pairing and are enriched at TAD boundaries (*11, 69*). To examine whether TAD boundaries alone are sufficient to drive pairing, we generated a CRISPR allele in which the

entire *ss* locus is deleted, removing a large portion of the interior of the TAD that contains *ss* (*ss full*Δ; **Fig. 4A**). We then tested *Transgene E*, which contains the *ss* locus, for pairing with this allele using a one-color FISH strategy (**Fig. 4B**). *Transgene E* did not drive pairing with *ss full* Δ (**Fig. 4C, G-H**), supporting a model in which TAD interiors work with TAD boundaries to drive homologous chromosome pairing.

5.7: *ss* nuclear localization changes with *ss* expression state

The localization of a relative to various nuclear subcompartments can have major effects on its expression (*14-19*). For example, active genes are often targeted to the interior of the nucleus (**Fig. 5A**), while inactive genes are often localized near the repressive nuclear lamina at the nuclear periphery (**Fig. 5B**)(*14-19*). To investigate the relationship between *ss* nuclear localization and *ss* expression state, we examined *ss* localization in three cell types: 1. R7 larval photoreceptors, where Ss is expressed in 67% of cells; 2. The larval eye disc peripodial membrane, where Ss is not expressed; and 3. The larval antennal disc, where Ss is expressed in 100% of cells.

We used DNA Oligopaints FISH probes to fluorescently label the *ss* locus and visualize is location within the nucleus. To quantify *ss* nuclear localization in each cell type, we measured the shortest distance from the *ss* FISH puncta to the nuclear lamina in 3D. To account for variation in nuclear size, we normalized all distance measurements by dividing them by the diameter of the nucleus. We then sorted the distance measurements into five bins and determined the

percentage of nuclei for each cell type that fell into each bin (**Fig. 5C**). We hypothesized that in the peripodial membrane, where Ss is off, *ss* would localize near the nuclear lamina. In the antennal disc, where Ss is on, we hypothesized that *ss* would localize to the nuclear interior. In R7s, where Ss is expressed in a subset of cells, we hypothesized that *ss* localization would vary between the nuclear lamina and nuclear interior.

As we hypothesized, *ss* localization was biased towards the interior of the nucleus in the antennal disc, where Ss is expressed (**Fig. 5C**). Unexpectedly, *ss* localization was biased towards the nuclear periphery in R7s, where Ss is expressed in 67% of cells (**Fig. 5C**). In the peripodial membrane, where Ss is not expressed, *ss* localization was broadly distributed between the nuclear periphery and nuclear interior (**Fig. 5C**).

These data are consistent with a model in which *ss* is targeted to different activating or repressing compartments in each cell type. While the nuclear periphery is typically a repressive environment, active genes often localize to the nuclear pore complex (NPC)(*14-16*). In R7s, *ss* may localize to the NPC in the 67% of cells where it is active, and to the nuclear lamina in the 33% of R7s where it is repressed (**Fig. 5D**). A change in localization from the lamina to the NPC upon expression might require less drastic rearrangement of nuclear architecture than a change from the nuclear periphery to the nuclear interior, allowing easier control of *ss* expression in this variable cell type. In the peripodial membrane, where *ss* is repressed, *ss* nuclear localization may be more random because this cell type does not require precise regulation of *ss* for its

development (**Fig. 5E**). Additional regulatory mechanisms, such as epigenetic state and repressive *trans* factors, may contribute to *ss* repression in this cell type. In the antennal disc, where *ss* is ubiquitously expressed, *ss* may target to the nuclear interior to ensure its robust and consistent expression (**Fig. 5F**).

5.8: Materials and Methods

Generation of CRISPR lines

promΔ and *PRE12*Δ CRISPR lines were generated in references (*72, 98*) using methods and oligos detailed in references (*72, 98, 101, 172-174*). *ss full*Δ was generated as described in references (*72, 98, 101, 172-174*). For *ss full*Δ, sense and antisense DNA oligos for the forward and reverse strands of four gRNAs were designed to generate BbsI restriction site overhangs. The oligos were annealed and cloned into the pCFD3 cloning vector (Addgene, Cambridge, MA). A single-stranded DNA homology bridge was generated with 60-bp homologous regions flanking each side of the predicted cleavage site. The gRNA constructs (125 ng/μl) and homologous bridge oligo (100 ng/μl) were injected into *Drosophila* embryos (BestGene, Inc., Chino Hills, CA). Single males were crossed with a balancer stock (*yw; +; TM2/TM6B*), and F1 female progeny were screened for the insertion via PCR, restriction digest, and sequencing. Single F1 males whose siblings were positive for the deletion were crossed to the balancer stock (*yw; +; TM2/TM6B*), and the F2 progeny were screened for the deletion via PCR, restriction digest, and sequencing. Deletion-positive flies from multiple founders were used to establish independent stable stocks.

The following oligos were used for the *ss full∆* CRISPR:

Table 1: Oligos used for *ss full∆* CRISPR

Oligo name	Sequence
Homologous bridge	AGACAAACTGTCTGCAGGTCCTGTTCCTCCTGCTCCTTCTGCTCCTCCTGCAGGCCATAT CTGTGGTCGAGTGTTGTTGGCGCTAAGATTGACCTTAAAATCGGATTTTTTGTTTTTTTT
gRNA 1F	GTCGATCGTCTATCTGGGCGCGTG
gRNA 1R	AAACCACGCGCCCAGATAGACGAT
gRNA 2F	GTCGTCAGCCTCGCAGATTGGATA
gRNA 2R	AAACTATCCAATCTGCGAGGCTGA
gRNA 3F	GTCGTGATTGCGAGTCTTGAGCTG
gRNA 3R	AAACCAGCTCAAGACTCGCAATCA
gRNA 4F	GTCGGATGGCGCGAGCAAGGTTTT
gRNA 4R	AAACAAAACCTTGCTCGCGCCATC
genotype F	GCTGCTTGTGTTGTTGTCGTC
genotype R	CGTGCATCAGTGTGTGAGTT

Drosophila lines

Flies were raised on standard cornmeal-molasses-agar medium and grown at 25° C.

Table 2: Genotypes of *Drosophila* lines

Fly line	Full genotype	Source	Figures
Wild type	*yw; +; +* or *yw; pm181>Gal4, UAS>mcd8GFP/CyO; +*	(*158*)	1B, E; 2B, G, I, L, N, Q; 3A, D-E, M; 4C, H; 5A-C
Transgene A 3L	*yw; +; pBAC{CH321-94A21}VK00033*	(*159*)*	1C, E
Transgene L 3L	*yw; +; pBAC{CH321-65B2}VK00033*	(*159*)*	1D-E
Transgene EE	*yw; pBac{CH321-88G14}VK00037; +*	(*159*)*	2C, E, G
Transgene FF	*yw; pBAC{CH321-92P24}VK00037; +*	(*159*)*	2D, F-G
Transgene GG	*yw; pBAC{CH321-32O17}VK00037, +/SM5-TM6B*	(*159*)*	2J-L
Transgene HH	*yw; pBac{CH321-86F02}VK00037; +*	(*159*)*	2O-Q
WT+E	*yw; pBac{CH321-28L15}VK00037/+; +*	*(97, 159)*	3B, D, F, H, M; 4D, H

Pc^4 + E	*yw; pBac{CH321-28L15}VK00037/CyO; Pc^4/TM6B*	(*159, 194*)	3C-D
Pc^{xt} + E	*yw; pBac{CH321-28L15}/+; Pc^{xt}/+*	(*159, 195*)	3G-H
CapH2 OE + E	*yw; pBac{CH321-28L15}VK00037/CyO; P{ey1x-GAL4.Exel}3/UAS>CapH2*	Bloomington, (*66, 159*)	3I, M
CapH2 Mutant 1 + E	yw; *pBac{CH321-28L15}VK00037/CyO; $CapH2^{eo3210}$/ Df(3R)Exel6159*	(*66, 159, 163*)	3J, M
CapH2 Mutant 2 + E	yw; *pBac{CH321-28L15}VK00037/CyO; $CapH2^{eo3210}$*	(*66, 159*)	3K, M
CapH2 Mutant 3 + E	yw; *pBac{CH321-28L15}VK00037/CyO; $CapH2^{z3-0019}$/ Df(3R)Exel6159*	(*66, 159, 163*)	3L-M
promΔ + E	yw; *pBac{CH321-28L15}VK00037/CyO; ss $^{prom\Delta}$/ Df(3R)Exel6269*	(*98, 159, 163*)	4A, E, H
PRE12Δ + E	yw; *pBac{CH321-28L15}VK00037/CyO; $ss^{PRE12\Delta}$*	(*72, 159*)	4A, F, H
ss fullΔ + E	yw; *pBac{CH321-28L15}VK00037/CyO; $ss^{full\ deletion}$*	(*159*)	4A, G-H

*Constructs were purchased from the CHORI *Drosophila melanogaster* BAC library collection (*159*) and sent to Rainbow Transgenic Flies, Inc. (Camarillo, CA) for injection.

Antibodies

Mouse anti-Lamin B (DSHB ADL67.10 and ADL84.12) antibody was used at a dilution of 1:100. Secondary antibody (Molecular Probes, 488 donkey anti-mouse) was Alexa Fluor-conjugated and used at a dilution of 1:400.

Oligopaints probe libraries

Table 3: Genome coordinates targeted by Oligopaints probe libraries

Probe set	Oligopaints library name	Genome coordinates targeted	Conjugated fluorophore	Figures
Neighboring 3L insertion site	pBJ250>3L(65B2) insertion site	3L: 6,442,676-6,502,676	Cy5	1B-D; 2B

3L-3R control sequence	*spineless* 50-kb extension (left)	3R: 16,320,533-16,370,533	Cy3	1B
Transgene A neighboring endogenous sequence	downstream of 94A21	3R: 16,240,324-16,290,324	Cy3	1C
Transgene L neighboring endogenous sequence	downstream of 60D22	3R: 16,844,756-16,894,756	Cy3	1D
2L-3L control sequence	upstream of clamp DNA	2L: 22,115,720-22,165,720	Cy3	2B
Transgene EE	Nf-YA 25-kb left extension+Nf-YA DNA+Nf-YA 25-kb right extension	3L: 9,414,490-9,468,232	Cy3	2C
Transgene FF	caup DNA	3L: 12,596,018-12,693,639	Cy3	2D
Neighboring 2L insertion site	right of 2L>22A3 transgene insertion site	2L: 1,582,821-1,642,821	Cy5	2I, N; 3A, E-G; 4C
endogenous *ss*	old *ss* 90K library	3R: 16,374,660-16,430,430	Cy3	2I; 3A-C, E, I-L; 4C-G; 5A-B
Transgene GG	downstream of 38G20	3R: 16,263,284-16,313,284	Cy3	2J
Transgene HH	dati DNA	4: 375,555-436,956	Cy3	2N-O
WT+E 2-color probe	*spineless* 50-kb extension (right)	3R: 16,435,681-16,485,681	Cy3	3F-G

Oligopaints probe design

Probes were designed as described in references (*72, 98*).

DNA FISH

FISH was performed using the protocol described in references (*72, 98*).

Pairing quantifications

Quantifications were performed as described in references (*72, 98*). 50 total nuclei across three eye/antennal discs were quantified for each dataset (i.e. N=3, n=50).

HiC mapping and TAD calling

HiC analysis was performed as described in (*72*).

Statistical analysis

All datasets were tested for a Gaussian distribution using a D'Agostino and Pearson omnibus normality test and a Shapiro-Wilk normality test. If either test indicated a non-Gaussian distribution for any of the datasets in an experiment, datasets were tested for statistical significance using a Wilcoxon rank-sum test (for single comparisons) or a one-way ANOVA on ranks with Dunn's multiple comparisons test (for multiple comparisons). If both the D'Agostino and Pearson and the Shapiro-Wilk tests indicated a Gaussian distribution for all datasets in an experiment, datasets were tested for statistical significance using an unpaired t-test with Welch's correction (for single comparisons) or an ordinary one-way ANOVA with Dunnett's multiple comparisons test (for multiple comparisons).

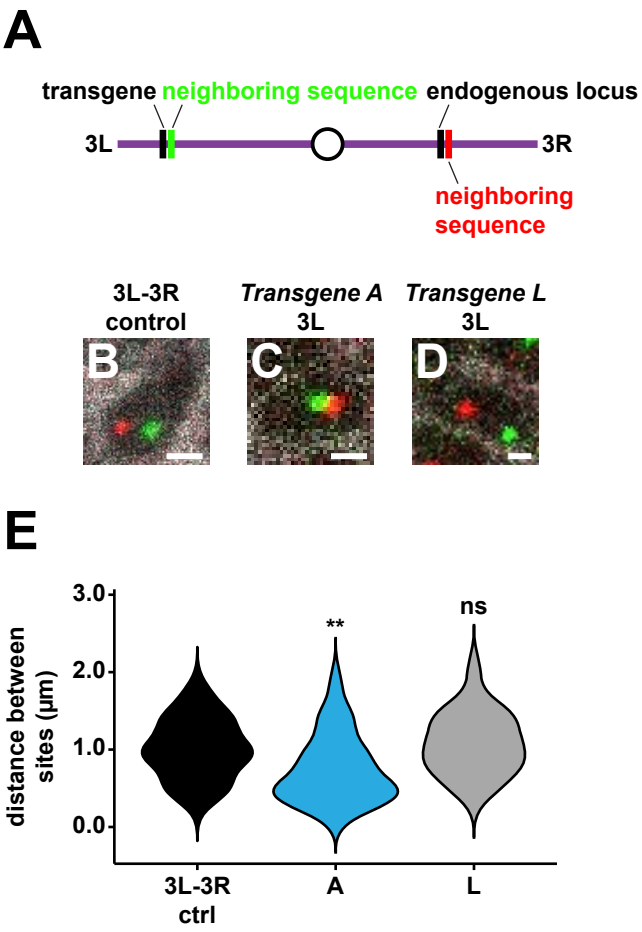

A
transgene
neighboring sequence
endogenous locus
3L
3R
neighboring sequence
3L-3R control
Transgene A 3L
Transgene L 3L
B
C
D
E
distance between sites (µm)
3.0
2.0
1.0
0.0
**
ns
3L-3R ctrl
A
L

Figure 1

Figure 1: Buttons drive pairing independent of their localization in the genome.

A. Schematic indicating the locations targeted by FISH probes in our two-color FISH strategy.

B-D. Representative images for *3L-3R control, Transgene A 3L*, and *Transgene L 3L*. White: Lamin B, green: probes neighboring transgene insertion site, red: probes neighboring endogenous site. Scale bars=1 μm.

E. Quantification of **1B-D.** **=p<0.005, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test.

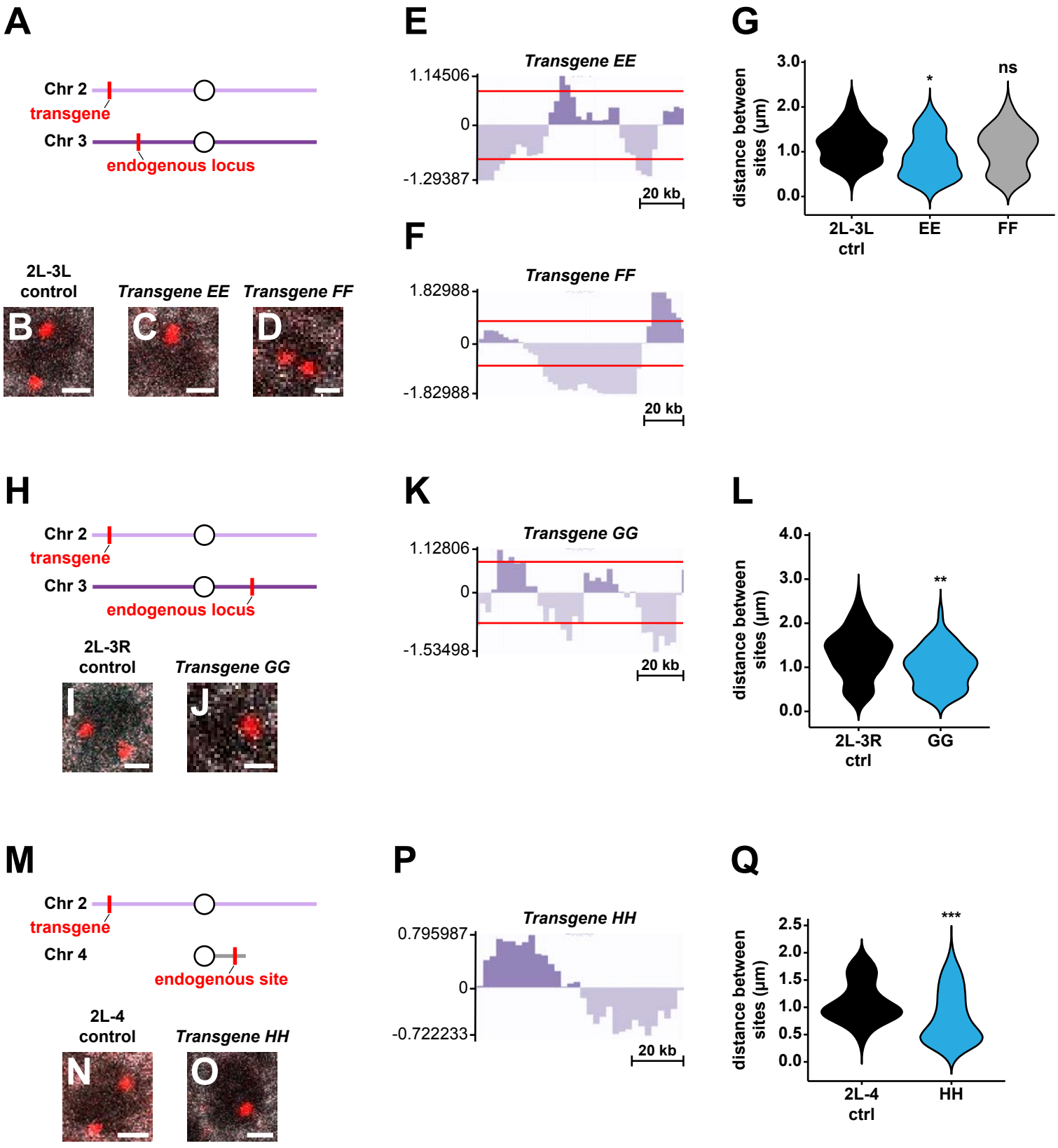

A
Chr 2
transgene
Chr 3
endogenous locus
2L-3L control
Transgene EE
Transgene FF
B
C
D
E
Transgene EE
1.14506
0
-1.29387
20 kb
F
Transgene FF
1.82988
0
-1.82988
20 kb
G
distance between sites (µm)
3.0
2.0
1.0
0.0
*
ns
2L-3L ctrl
EE
FF
H
Chr 2
transgene
Chr 3
endogenous locus
2L-3R control
Transgene GG
I
J
K
Transgene GG
1.12806
0
-1.53498
20 kb
L
distance between sites (µm)
4.0
3.0
2.0
1.0
0.0
**
2L-3R ctrl
GG
M
Chr 2
transgene
Chr 4
endogenous site
2L-4 control
Transgene HH
N
O
P
Transgene HH
0.795987
0
-0.722233
20 kb
Q
distance between sites (µm)
2.5
2.0
1.5
1.0
0.5
0.0
***
2L-4 ctrl
HH

Figure 2

Figure 2: Identification of additional buttons across the *Drosophila* genome.

A, H, M. Schematics indicating the locations targeted by FISH probes in our one-color FISH strategy.

B-D, I-J, N-O. Representative images for *2L-3L control, Transgene EE, Transgene FF, 2L-3R control, Transgene GG, 2L-4 control,* and *Transgene HH*. White: Lamin B, red: probes labeling the transgene and endogenous sequences. Scale bars=1 μm.

E-F, K. Representative directionality indices showing TAD coverage of each transgene, based on consensus TAD calls generated from the analysis of 14 Hi-C datasets (*177-180*). Red lines indicate a directionality index signal of 0.8 or -0.8, the cutoff for a TAD. Representative directionality indices are from NCBI accession numbers GSE63515 (*Transgene EE*) and GSE38468 (*Transgenes FF* and *GG*). Directionality indices are shown for the entire region spanned by each transgene.

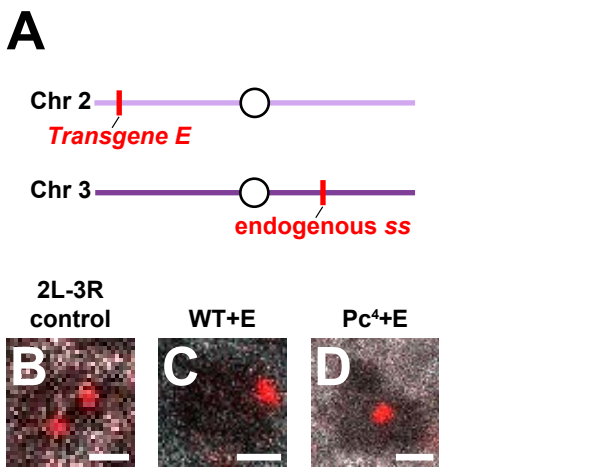
G. Quantification of **2B-D**. *=p<0.05, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test. Black: control, blue: pairer, gray: non-pairer.

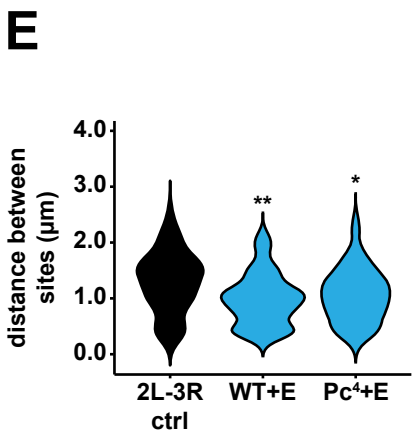
L. Quantification of **2I-J.** **=p<0.005, Wilcoxon rank-sum test. Black: control, blue: pairer.

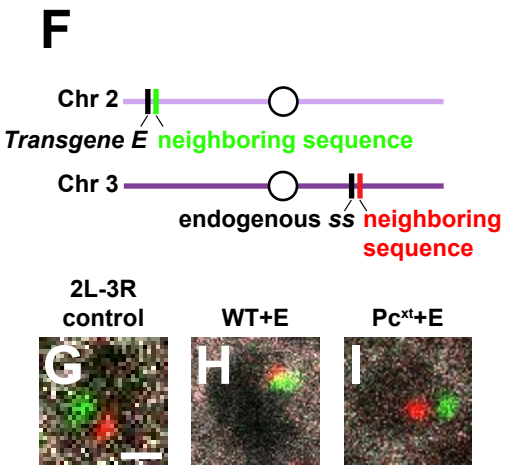
P. Representative directionality index showing TAD coverage of *Transgene HH*, based on consensus TAD calls generated from the analysis of 14 Hi-C datasets (*177-180*). Representative directionality index is from NCBI accession number

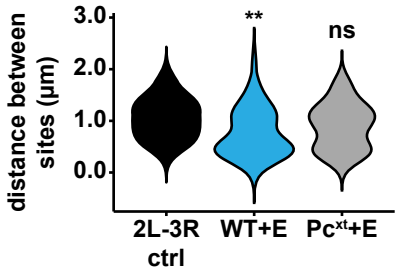
GSE63515. Directionality index is shown for the entire region spanned by *Transgene HH*.

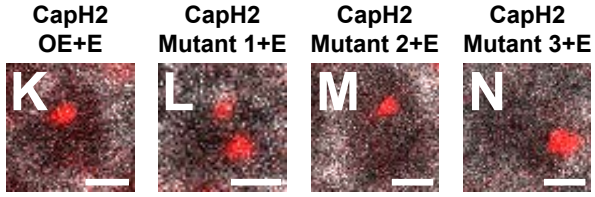
Q. Quantification of **2N-O.** ***=p<0.001, Wilcoxon rank-sum test. Black: control, blue: pairer.


A
Chr 2
Transgene E
Chr 3
endogenous ss
2L-3R control
WT+E
Pc4+E
B
C
D


E
distance between sites (µm)
4.0
3.0
2.0
1.0
0.0
**
*
2L-3R ctrl
WT+E
Pc4+E


F
Chr 2
Transgene E
neighboring sequence
Chr 3
endogenous ss
neighboring sequence
2L-3R control
WT+E
Pcxt+E
G
H
I


J
distance between sites (µm)
3.0
2.0
1.0
0.0
**
ns
2L-3R ctrl
WT+E
Pcxt+E


CapH2 OE+E
CapH2 Mutant 1+E
CapH2 Mutant 2+E
CapH2 Mutant 3+E
K
L
M
N

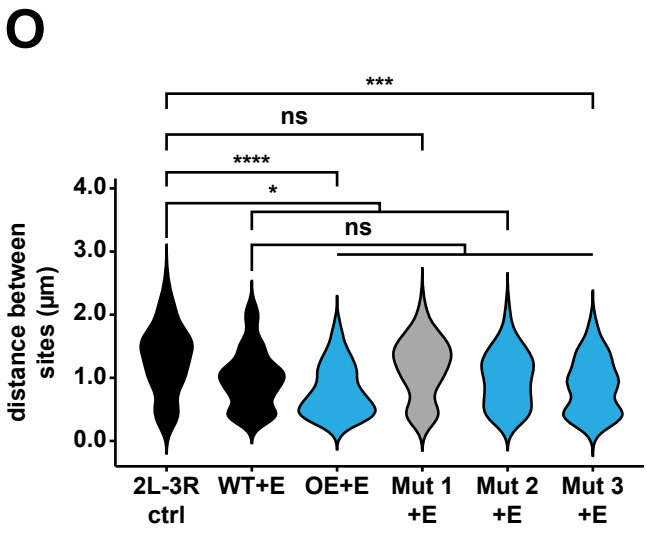

O
***
ns
****
*
ns
distance between sites (µm)
4.0
3.0
2.0
1.0
0.0
2L-3R ctrl
WT+E
OE+E
Mut 1 +E
Mut 2 +E
Mut 3 +E

Figure 3

Figure 3: Examining the effects of *trans* factor mutations on *ss* pairing.

A. Schematic indicating the locations targeted by FISH probes in our one-color FISH strategy.

B-D. Representative images for one-color *2L-3R control*, one-color *WT+E,* and Pc^{4}*+E.* White: Lamin B, red: probes labeling the endogenous *ss* and *Transgene E* sequences. Scale bars=1 μm.

E. Quantification of **3B-D**. **=p<0.005, *=p<0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test. Black: control, blue: pairer.

F. Schematic indicating the locations targeted by FISH probes in our two-color FISH strategy.

G-I. Representative images for two-color *2L-3R control*, two-color *WT+E*, and Pc^{xt}*+E*. White: Lamin B, green: probes neighboring *Transgene E* insertion site, red: probes neighboring endogenous *ss*. Scale bars=1 μm.

J. Quantification of **3G-I.** **=p<0.005, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test. Black: control, blue: pairer, gray: non-pairer.

K-N. Representative images for *CapH2 OE+E, CapH2 Mutant 1+E, CapH2 Mutant 2+E,* and *CapH2 Mutant 3+E*. White: Lamin B, red: probes labeling the endogenous *ss* and *Transgene E* sequences. Scale bars=1 μm.

O. Quantification of **3K-N.** ****=p<0.0001, ***=p<0.001, *=p<0.05, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test. Black: control, blue: pairer, gray: non-pairer.

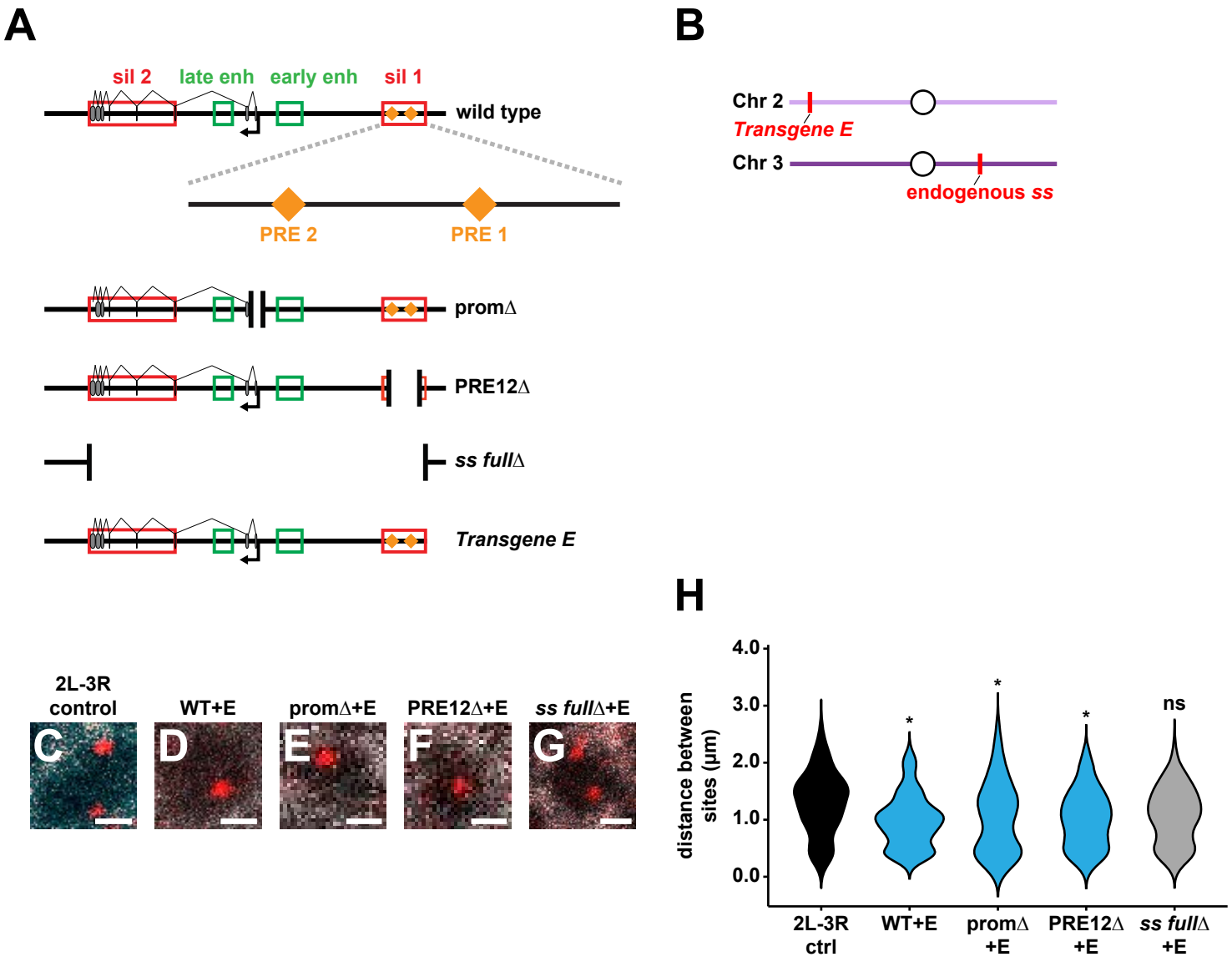

A
sil 2
late enh
early enh
sil 1
wild type
PRE 2
PRE 1
promΔ
PRE12Δ
ss fullΔ
Transgene E
B
Chr 2
Transgene E
Chr 3
endogenous ss
2L-3R control
WT+E
promΔ+E
PRE12Δ+E
ss fullΔ+E
C
D
E
F
G
H
distance between sites (µm)
4.0
3.0
2.0
1.0
0.0
*
*
*
ns
2L-3R ctrl
WT+E
promΔ +E
PRE12Δ +E
ss fullΔ +E

Figure 4

Figure 4: Examining the effects of DNA element CRISPR deletions on *ss* pairing.

A. Schematic of wild type *ss,* the *promΔ, PRE12Δ*, and *ss fullΔ* CRISPR alleles, and *Transgene E*. Sil 1: silencer 1, early enh: early enhancer, late enh: late enhancer, sil 2: silencer 2, gray ovals: exons, black arrow: transcription start site.

B. Schematic indicating the locations targeted by FISH probes in our one-color FISH strategy.

C-G. Representative images for *2L-3R control, WT+E, promΔ+E, PRE12Δ+E*, and *ss fullΔ+E*. White: Lamin B, red: probes labeling the endogenous *ss* and *Transgene E* sequences. Scale bars=1 μm.

H. Quantification of **4C-G.** *=p<0.05, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test. Black: control, blue: pairer, gray: non-pairer.

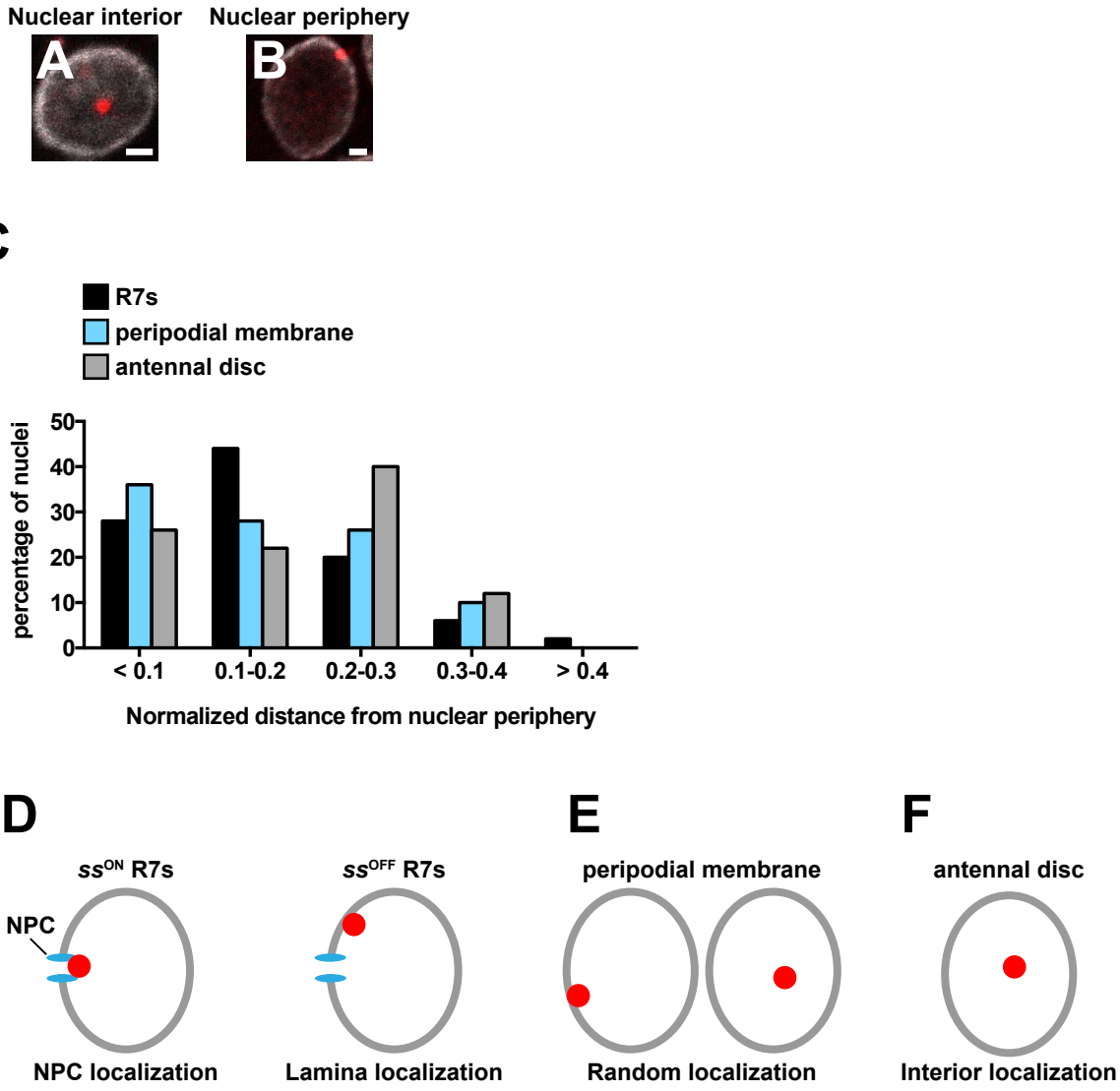

Nuclear interior
Nuclear periphery
A
B
C
R7s
peripodial membrane
antennal disc
percentage of nuclei
50
40
30
20
10
0
< 0.1
0.1-0.2
0.2-0.3
0.3-0.4
> 0.4
Normalized distance from nuclear periphery
D
ssON R7s
NPC
NPC localization
ssOFF R7s
Lamina localization
E
peripodial membrane
Random localization
F
antennal disc
Interior localization

Figure 5

Figure 5: The nuclear localization of *ss* varies with *ss* expression state.

A-B. Representative images of *ss* localization to the nuclear interior vs. the nuclear periphery. Red: FISH probes targeting endogenous *ss*, white: Lamin B. Scale bars=1 μm.

C. Quantifications of the normalized distance of *ss* from the nuclear periphery in R7s, peripodial membrane cells, and antennal disc cells.

D-F. Models for the localization of *ss* in R7s, peripodial membrane cells, and antennal disc cells. Red circle: *ss* locus.

D. NPC=Nuclear pore complex.

Chapter 6: Conclusions and Future Directions

6.1: The study of pairing and transvection provides insights into nuclear organization

Nuclear organization is an important factor controlling gene expression, but many questions remain unanswered in the field. What is the biological role of TADs? What are the *cis* and *trans* factors that drive interactions between chromosomes? How do these interactions between chromosomes affect gene expression? How reproducible are these interactions between cell types?

Using *Drosophila* homologous chromosome pairing and transvection as a paradigm, this work begins to answer some of these fundamental questions. We find that specialized TADs drive pairing between transgenes and their homologous endogenous sequences, providing the first evidence of a general feature that contributes to homologous chromosome pairing genome-wide. Furthermore, our data suggest a biological role for TADs in driving interchromosomal interactions.

In addition to TADs, we find that insulators and Polycomb Group Complex proteins may also contribute to homologous pairing. Clusters of insulators, Polycomb proteins, and additional chromatin and architectural proteins may bind to TAD regions at specific intervals and in specific combinations, providing each TAD with a unique “code” that allows it to pair with its homologue. Thus, we propose a model in which specialized TADs bound by complex combinations of proteins act as the buttons that bring homologous chromosomes together. These buttons appear to play an essential role in the structural integrity of the genome:

when chromosome rearrangement breakpoints lie within a gene, buttons "reconstitute" the gene by bringing regulatory elements back into close physical proximity with the correct protein coding region, ensuring that proper gene expression is maintained. While chromosome-wide pairing is not conserved in mammals, buttons may still serve a similar purpose by driving pairing in processes including genomic imprinting and X-inactivation.

Homologous chromosome pairing facilitates gene-regulatory interactions between chromosomes through the phenomenon of transvection. Our studies provide deeper insight into the mechanisms of transvection, its biological role, and the DNA elements that are involved in this process, contributing the following significant findings to the field:

1) Homologous chromosome pairing is necessary but not sufficient for transvection.
2) Pairing and transvection are cell-type-specific, suggesting that the level of pairing in a given cell type affects the efficiency of transvection in that cell type.
3) Transvection controls photoreceptor patterning by averaging the expression frequencies of naturally-occurring *ss* alleles. This is one of the first examples of a biological role for transvection between wild-type alleles.
4) Trans-activation and trans-repression are separable mechanisms that act simultaneously at the same locus to control gene expression between chromosomes. Trans-activation at the *ss* locus

requires a promoter in *cis* to the enhancer, while trans-repression at the *ss* locus requires an insulator and two PREs.

Our data have broad implications for our understanding of interchromosomal gene regulation across organisms. Determining the role of interactions between chromosomes and the DNA elements that are involved can provide insight into mammalian mechanisms such as long-distance enhancer action, nuclear compartmentalization, and chromosome kissing.

6.2: Perspectives and future work

This work suggests a number of exciting potential avenues for future research. Our identification of the features of buttons, including TAD content and enrichment for certain insulators, provides the opportunity to predict additional button elements based on HiC and insulator ChIP-seq data and test them for their ability to drive pairing. As complementary work suggests a role for active transcription in pairing (*62, 63*), RNA-seq data could be used alongside these analyses to assist in button prediction, with the ultimate goal of identifying all of the buttons that initiate pairing across the fly genome. Once an effective method for predicting buttons is developed in *Drosophila*, a similar approach could be applied to find regions that interact between chromosomes in mammals. Such an approach could allow identification of regions whose interactions change in disease states, providing new candidates for targeted disease treatments.

Our finding that TADs drive interactions between chromosomes suggests that TADs play a structural role in organizing the nucleus beyond their role in

controlling gene expression. In mammals, knockdown of cohesin causes a loss of TADs but has surprisingly minimal effects on gene expression (*27-29*). It is possible that the loss of TADs affects interchromosomal interactions but not gene expression in mammalian cells. Performing similar experiments in *Drosophila* to examine the effects of TAD dissolution on homologous chromosome pairing will provide important insight into the role of TADs in organizing the nucleus.

The mechanisms that allow TADs to interact between chromosomes remain unclear. While two recent studies suggest that TAD boundaries are more tightly paired than TAD interiors (*62, 63*), our data suggest that TAD boundaries alone are not sufficient to drive pairing. *Transgene E*, which contains the *ss* locus, does not pair with a CRISPR allele in which the interior of the *ss* TAD is deleted (**Chapter 5, Fig. 4G-H**). Generating CRISPR knockouts of other TAD interiors and testing their effects on pairing will clarify the roles of TAD interiors and boundaries in this process. It is also possible that other mechanisms contribute to TAD interactions. We find that insulators and the Polycomb Group Complex may be involved in pairing, suggesting that each TAD might bind a unique combination of proteins, allowing it to find its homologue in the nucleus. A screen for factors that affect transgene pairing with its endogenous site could identify additional proteins involved in this process. Alternatively, the specific folding confirmation of a TAD may allow it to pair with its unique partner.

A transgene pairing with its endogenous locus represents a novel interchromosomal interaction that does not occur in wild type nuclei. Such an interaction may have major effects on chromosome territory organization, as it

involves physical colocalization between two sequences on heterologous chromosomes. The disruption of chromosome territories and resulting changes in gene expression may explain why novel interactions between homologous sequences are often linked to disease in mammals (*22, 23, 26, 41, 52*).Thus, understanding how aberrant homologous pairing affects chromosome organization could have major implications for the treatment of disease. Future studies could perform DNA FISH against whole chromosomes to investigate chromosome territory organization in the presence and absence of transgene pairing. Any identified changes in nuclear architecture could be confirmed using HiC, and RNA-seq on cells with or without the transgene would provide information on any changes in gene expression resulting from the disruption of chromosome territories.

Finally, our observation that pairing and transvection occur in the larval eye disc but not the antennal disc suggests a model in which the level of pairing in a tissue controls the efficiency of transvection in that tissue. One prediction of this model is that forcing pairing in a weak-pairing cell type would improve transvection efficiency. To test this prediction, the same forced pairing system we developed in the eye disc (**Chapter 3, Fig. 5M-T**) could be used to identify regions along chromosome 3R that naturally loop to endogenous *ss* in the antennal disc. We could insert the transvection-competent *Transgene S* (**Chapter 3, Fig. 5E, O-S**) at these looped sites to test its ability to perform transvection with endogenous *ss* in the antennal disc, providing insights into the link between pairing and transvection in diverse cell types.

Disruption of nuclear architecture is linked to a large number of human diseases (*13, 20-25*). Further work investigating the role of TADs and *trans* factors in driving physical interactions between chromosomes and identifying the DNA elements responsible for interchromosomal gene regulation will provide insight into the mechanisms that contribute to these diseases. Ultimately, a fuller understanding of nuclear organization and its effects on gene expression over the course of development will enable improved diagnosis and treatment of diseases including limb malformations, Angelman and Prader Willi syndromes, and cancer.

Appendix 1: The *mir-279/996* cluster represses receptor tyrosine kinase signaling to determine cell fates in the *Drosophila* eye

© 2018. Published by The Company of Biologists Ltd | Development (2018) 145, dev159053. doi:10.1242/dev.159053

RESEARCH ARTICLE

The *mir-279/996* cluster represses receptor tyrosine kinase signaling to determine cell fates in the *Drosophila* eye

Hong Duan[1,*], Luis F. de Navas[1,*], Fuqu Hu[1], Kailiang Sun[1,2], Yannis E. Mavromatakis[3], Kayla Viets[4], Cyrus Zhou[4], Joshua Kavaler[5], Robert J. Johnston[4], Andrew Tomlinson[3] and Eric C. Lai[1,‡]

ABSTRACT

Photoreceptors in the crystalline *Drosophila* eye are recruited by receptor tyrosine kinase (RTK)/Ras signaling mediated by Epidermal growth factor receptor (EGFR) and the Sevenless (Sev) receptor. Analyses of an allelic deletion series of the *mir-279/996* locus, along with a panel of modified genomic rescue transgenes, show that *Drosophila* eye patterning depends on both miRNAs. Transcriptional reporter and activity sensor transgenes reveal expression and function of miR-279/996 in non-neural cells of the developing eye. Moreover, *mir-279/996* mutants exhibit substantial numbers of ectopic photoreceptors, particularly of R7, and cone cell loss. These miRNAs restrict RTK signaling in the eye, since *mir-279/996* nulls are dominantly suppressed by positive components of the EGFR pathway and enhanced by heterozygosity for an EGFR repressor. miR-279/996 limit photoreceptor recruitment by targeting multiple positive RTK/Ras signaling components that promote photoreceptor/R7 specification. Strikingly, deletion of *mir-279/996* sufficiently derepresses RTK/Ras signaling so as to rescue a population of R7 cells in R7-specific RTK null mutants *boss* and *sev*, which otherwise completely lack this cell fate. Altogether, we reveal a rare setting of developmental cell specification that involves substantial miRNA control.

KEY WORDS: *Drosophila*, R7 photoreceptor, RTK signaling, MicroRNA

INTRODUCTION

The *Drosophila* eye is a choice model system for studying cell fate specification owing to its highly stereotyped array of pattern elements. Each eye consists of ∼800 ommatidial units, each of which contains eight photoreceptors of distinct identities, four cone cells, and about eight pigment cells; a mechanosensory bristle organ develops at alternate ommatidial vertices. The orderly acquisition of cell fates during eye development is coordinated by multiple signaling pathways and transcription factors (Kumar, 2012).

Initially, a proneural zone defined by the basic helix-loop-helix activator Atonal is resolved into single R8 photoreceptors by Notch pathway signaling. Each R8 nucleates a developing ommatidium, and a stepwise set of events mediated by Epidermal growth factor receptor (EGFR) and receptor tyrosine kinase (RTK) signaling progressively recruit the R2/5, R3/R4, R1/6 and R7 photoreceptors to each ommatidial cluster (Freeman, 1996). A specialized RTK signal transduced by the Sevenless (Sev) receptor specifies the final photoreceptor, R7. In parallel to EGFR and Sev signaling, Notch signaling defines photoreceptor subtypes (Cagan and Ready, 1989). Further non-sensory cell fates are subsequently recruited to each ommatidial cluster, including cone cells followed by primary and secondary pigment cells.

The existence of extensive regulatory networks mediated by microRNAs (miRNAs) suggests broad possibilities for their requirement during development or physiology (Flynt and Lai, 2008; Sun and Lai, 2013). As is true for most tissues, loss of core miRNA biogenesis factors such as Dicer-1 or Pasha causes substantial defects in the developing *Drosophila* eye (Lee et al., 2004; Smibert et al., 2011). Beyond the general requirement for miRNA biogenesis in this tissue, some individual miRNAs and miRNA sites influence eye development. For example, studies of the hypermorphic *Enhancer of split m8*^{D} [*E(spl)m8*^{D}] allele revealed its post-transcriptional repression by K box motifs (Lai et al., 1998) – indeed long before these were recognized as binding sites for K box family miRNAs (Lai, 2002; Lai et al., 2005). Specific mutation of K boxes from an *E(spl)m8* genomic transgene sensitizes the *Notch*^{split} background, yielding a synthetic, smaller rough eye (Lai et al., 1998). The bantam miRNA is required for the growth and proliferation of all imaginal discs; thus, loss of bantam reduces eye tissue and increases apoptosis (Brennecke et al., 2003; Hipfner et al., 2002). The *mir-263a/b* loci are essential for development of eye interommatidial bristles, and protect the shaft cells of these sensory organs from apoptosis (Hardiman et al., 2002; Hilgers et al., 2010).

By contrast, many other *Drosophila* miRNAs connected to eye development lack substantial defects when mutated on their own, but are sensitive to genetic background or environmental stress. For example, miR-7 positively regulates photoreceptor specification by repressing the neural inhibitor *yan* (*aop*), a transcriptional repressor in the EGFR pathway (Li and Carthew, 2005). Although deletion of *mir-7* alone has only minor effects on eye development, its deletion sensitizes the eye to alteration in EGFR signaling (Li and Carthew, 2005) or temperature fluctuation (Li et al., 2009). Similarly, deletion of *mir-11*, located in the intron of *E2f1*, does not have substantial effects by itself, but this condition renders the eye sensitive to E2F1-dependent, DNA damage-induced apoptosis (Truscott et al., 2011). These and other examples have led to the notion that miRNAs are primarily important for robustness, but are individually mostly dispensable for normal developmental programs.

In this study, we elucidate crucial roles for the *mir-279/996* locus during *Drosophila* eye development. These seed-related miRNAs

[1]Department of Developmental Biology, Sloan-Kettering Institute, 1275 York Ave, Box 252, New York, NY 10065, USA. [2]Program in Neuroscience, Weill Cornell Medical College, New York, NY 10065, USA. [3]Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, 701 West 168th Street, New York, NY 10032, USA. [4]Department of Biology, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA. [5]Department of Biology, Colby College, Waterville, ME 04901, USA.
*These authors contributed equally to this work

‡Author for correspondence (laie@mskcc.org)

E.C.L., 0000-0002-8432-5851

Received 5 September 2017; Accepted 28 February 2018

Development (2018) 145, dev159053. doi:10.1242/dev.159053

are expressed from an operon and are functionally equivalent in several neural settings (Sun et al., 2015), including during suppression of CO_2 neurons (Cayirlioglu et al., 2008; Hartl et al., 2011), control of circadian behavior (Luo and Sehgal, 2012), and control of mechanosensory organ development (Kavaler et al., 2018). We now show that these miRNAs are deployed in non-neuronal cells of the developing eye, and their deletion strongly alters eye cell fates, yielding ectopic photoreceptors and loss of cone cells. Focusing on ectopic R7 photoreceptors, we use genetic interactions to demonstrate that miR-279/996 restrict RTK/Ras signaling, which normally promotes R7 specification. This is attributable to their direct repression of multiple positive components of RTK signaling pathways. Strikingly, the efficacy of endogenous *mir-279/996* in restricting RTK/Ras signaling is substantial enough that deletion of these miRNAs can rescue a population of R7 photoreceptors in the absence of the Boss ligand or the Sev receptor.

These findings highlight how a single miRNA locus can exert phenotypically substantial, and not merely fine-tuning, roles in multiple biological settings. Moreover, these miRNAs achieve similar functional roles (neural repression) through mechanistically distinct strategies (i.e. by repressing RTK/Ras components in the eye, by repressing a Notch inhibitor in mechanosensory organs, or by repressing transcription factors in the olfactory system).

RESULTS

The *mir-279/996* locus is essential for normal eye development

The seed-related *mir-279* and *mir-996* were previously considered to be expressed from independent transcription units, with *mir-279* being solely required in various developmental settings (Cayirlioglu et al., 2008; Luo and Sehgal, 2012; Yoon et al., 2011). However, we recently clarified that these miRNAs are functionally overlapping and co-expressed as an operon, and that '*mir-279*-specific' deletions also impair the expression of miR-996 (Sun et al., 2015). Our key genetic reagents include hypomorphic and null alleles of the *mir-279/996* operon, and wild-type and modified genomic transgenes expressing only miR-279 or miR-996 (Fig. 1A).

In our search for novel *mir-279/996* functions, we observed that our mutants exhibit defects in the normally crystalline adult eye (Fig. 1B). Taking advantage of our allelic series, we found that the severity of eye roughening is exacerbated by decreasing dosage of *mir-279/996* (Fig. 1C-F). The phenotypes of *mir-279/996[15C/15C]* double-deletion homozygotes were rescued by a 16.6 kb *mir-279/996* genomic transgene, demonstrating that the eye defects are due to loss of this miRNA locus (Fig. 1G). Moreover, sole expression of either miR-279 or miR-996 from the same genomic context restored normal adult eyes to *[15C/15C]* null animals (Fig. 1H,I). Thus, these miRNAs exhibit substantial functional redundancy across diverse *in vivo* settings (Sun et al., 2015).

To gain initial insight into underlying defects caused by lack of miR-279/996, we sectioned adult eyes of various heteroallelelic mutants. Normal eyes exhibit seven of the eight photoreceptor rhabdomeres in a given section (Fig. 1J), arranged as six large outer photoreceptors (R1-6) surrounding a small inner photoreceptor (R7/8, depending on the apical-basal position of the section). Notably, all three mutant combinations examined showed ectopic outer and inner photoreceptors (Fig. 1K-M). Based on the position and morphology of the latter, the ectopic inner photoreceptors were preferentially R7. We annotated representative sections according to the key in Fig. 1, with circles of different colors indicating ommatidia with one or more outer R cells, and arrowheads designating ectopic R7 cells. Fig. 1N-R shows magnifications of individual normal and mutant ommatidia, and highlights the variety of abnormal photoreceptor combinations present in *mir-279/996* heteroallelic eyes. Quantification revealed that the strong mutant combination *ex36/15C* exhibited stronger defects than either hypomorphic combination examined, and >21% of ommatidia in *ex36/15C* exhibited ectopic photoreceptors (Fig. 1S). Notably, although there are eight photoreceptor subtypes, these quantitative analyses indicate that a majority of mutant ommatidia in all three transheterozygous backgrounds examined exhibit ectopic R7 cells. Thus, specification of the R7 photoreceptor appears particularly sensitive to the miRNAs.

Overall, these data reveal that defective allocation of cell fates contributes to eye roughening in *mir-279/996* mutants, and that both miRNAs are required for normal eye development. We note that although the eye is sensitive in revealing developmental abnormalities, it is actually rare for miRNA mutants to exhibit retinal phenotypes. Indeed, in a recent survey of new deletion alleles covering ∼100 *Drosophila* miRNAs, including most operons analyzed as cluster knockouts, only one novel locus (*mir-92a*) affected eye morphology and barely any others discernibly affected external development (Chen et al., 2014). Therefore, these eye specification defects in *mir-279/996* mutants represent some of the most overt developmental phenotypes detected among miRNA mutants.

Transcriptional activity of *mir-279/996* in the pupal eye

We examined the spatial expression of miR-279/996 in the developing eye. We generated a transcriptional reporter in which cytoplasmic GFP was knocked into a 16.6 kb genomic fragment that extends into the flanking upstream and downstream protein-coding genes (Fig. 2A). This genomic fragment provides full rescue of *mir-279/996* function in multiple settings (Sun et al., 2015), including during eye development (Fig. 1G-I), indicating that it contains all relevant cis-regulatory information.

In eyes staged at 45 h after puparium formation (APF), the differentiation of all ommatidial cell types has occurred, and they have adopted their appropriate relative positions in the mature unit eye. The nuclei of the different cell types adopt characteristic locations along this axis, and can be divided into four layers (Fig. 2B). Most apically lie the quartet of cone cell nuclei, below which are the nuclei of the paired primary pigment cells. The third level is made from the eight photoreceptor nuclei, and most basally lie the nuclei of the secondary/tertiary pigment cells and the bristle cells. The *mir-279/996-GFP* reporter is expressed throughout the cone and primary pigment cell layer [as marked by expression of nuclear DPax2 (Shaven), Fig. 2C], as well as in the accessory pigment and bristle cells, but is clearly excluded from photoreceptors (as marked by nuclear Elav, Fig. 2D,E). Therefore, this miRNA locus is deployed in multiple non-neuronal eye cell types, but is excluded from photoreceptors.

Endogenous activity of miR-279/996 in non-neural cells of the pupal eye

As some miRNAs are regulated post-transcriptionally, patterned expression of the *mir-279/996-GFP* reporter does not necessarily equate to their spatial activity. We therefore generated a miRNA activity sensor, composed of a ubiquitously expressed GFP reporter bearing two sequences complementary to miR-279 (Fig. 2A). In theory, this *tub-GFP-miR-279* transgene should be repressed by both miR-279 (via RNAi) and miR-996 (via miRNA seed matching). Interestingly, expression of the *tub-GFP-miR-279*

Development (2018) 145, dev159053. doi:10.1242/dev.159053

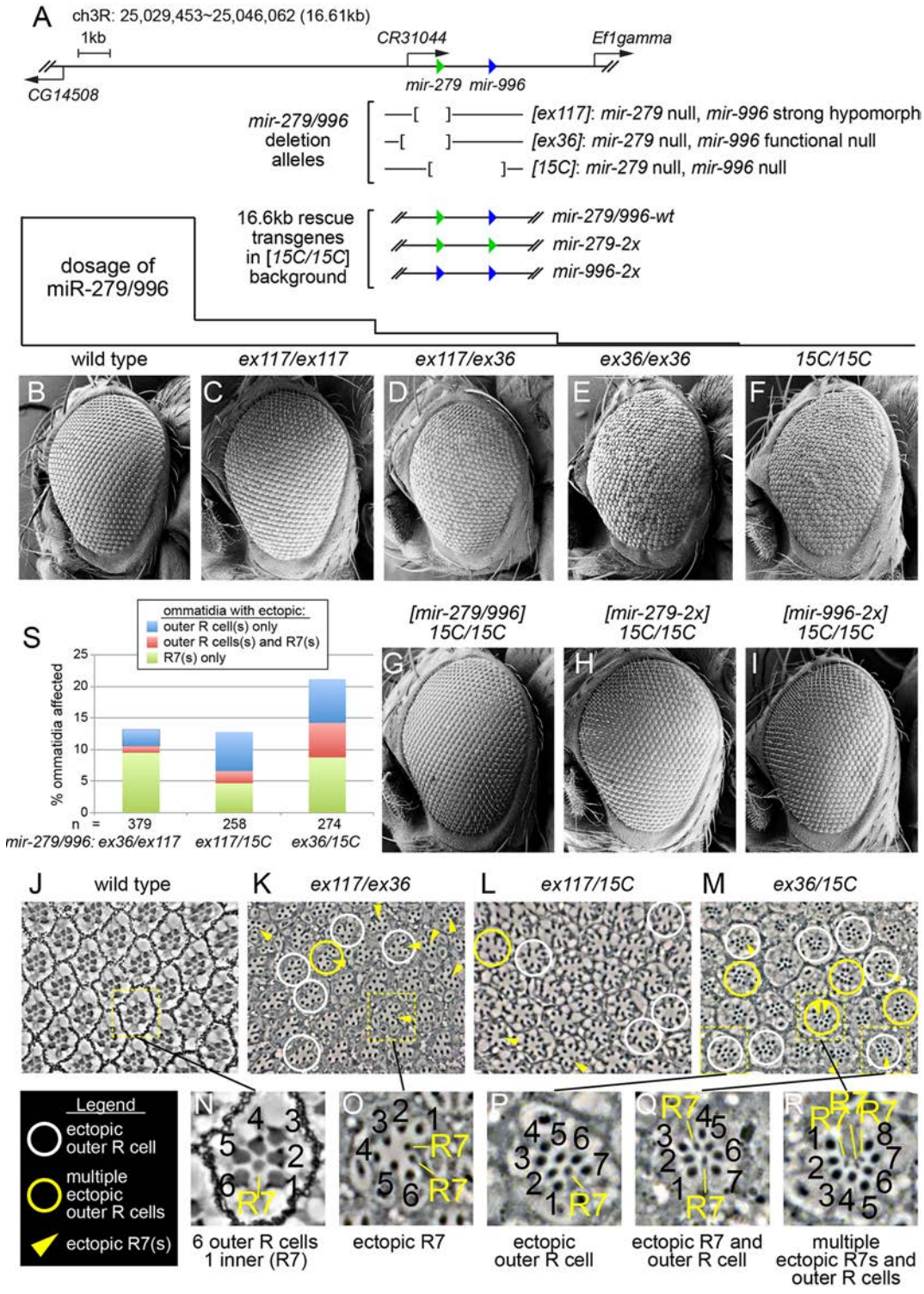


Fig. 1. *mir-279/996* alleles and corresponding adult *Drosophila* eye phenotypes. (A) The *Drosophila mir-279/996* genomic region, along with three deletion alleles and three rescue transgenes built into a 16.6 kb genomic backbone. (B-I) Adult *Drosophila* eyes analyzed by scanning electron microscopy. (B) The wild-type eye exhibits a regular, crystalline organization. (C) A hypomorphic condition that is deleted for *mir-279* and is impaired for miR-996 biogenesis appears externally normal, whereas genotypes that progressively remove miR-279/996 activity exhibit overt roughening and ommatidial disorganization (D-F). (G-I) The exterior eye phenotype caused by deletion of both *mir-279* and *mir-996* (in *[15C]* homozygotes) can be rescued by supplying both miRNAs (G), only *mir-279* (H) or only *mir-996* (I). (J-M) Plastic sections through adult eyes of wild type (J) and three heteroallelic *mir-279/996* combinations (K-M). Normal ommatidia exhibit six large outer photoreceptor rhabdomeres (R1-6) surrounding a smaller inner photoreceptor rhabdomere (R7 or R8, depending on the apical-basal position). All of the transheterozygous *mir-279/996* mutants exhibit populations of ommatidia with ectopic outer and/or inner photoreceptors, as annotated in the key to the left. Based on their position and morphology, the ectopic inner photoreceptors are predominantly R7. Some ommatidia can contain both ectopic outer and R7 photoreceptors, and the frequency of mutant ommatidia is noticeably higher in the strongest mutant transheterozygote *ex36/15C* (M). (N-R) Magnifications of individual ommatidia highlighting normal and different combinations of mutant photoreceptor identities. (S) Quantification of photoreceptor subtypes in the three heteroallelic *mir-279/996* mutants examined. R7 is the predominant photoreceptor subtype affected in all genotypes.

Development (2018) 145, dev159053. doi:10.1242/dev.159053

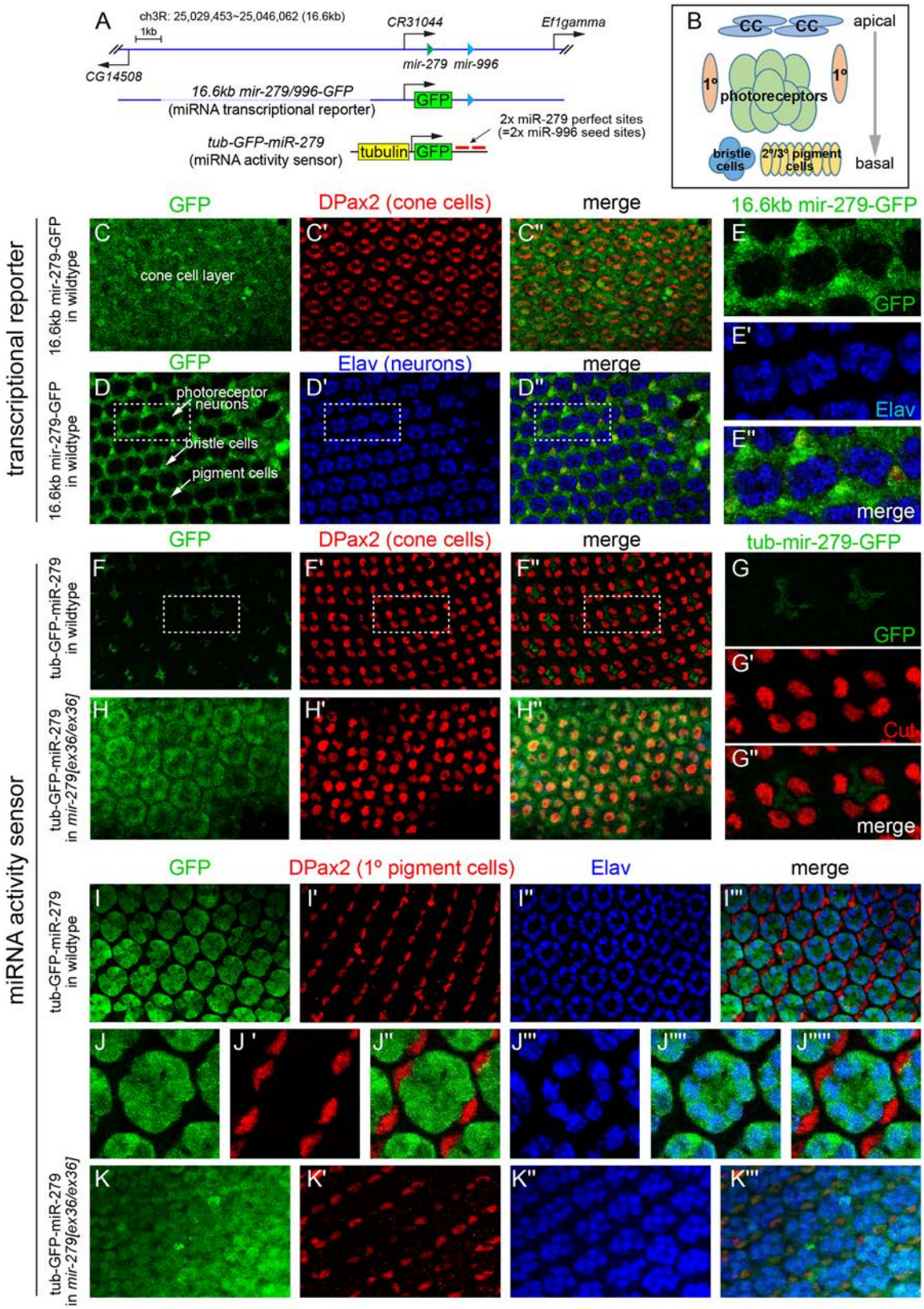


Fig. 2. See next page for legend.

sensor was complementary to that of the transcriptional reporter; sensor GFP expression was robust in photoreceptor cells, but absent in cone cells, pigment cells and the non-neuronal cells of interommatidial bristle organs (Fig. 2F,G,I). The non-neuronal activity of miR-279/996 was particularly apparent at higher magnification of photoreceptor and primary pigment cells, since

Development (2018) 145, dev159053. doi:10.1242/dev.159053

Fig. 2. Expression and activity of *mir-279/996* in the developing pupal eye. (A) Transgenes to detect transcriptional activity or functional repression by *mir-279/996*; the former is a positive readout of miRNA expression, whereas the latter is a negative sensor of miRNA activity. (B) Schematic of ommatidial nuclei in the fly eye. Cone cell (CC) nuclei are located most apically, primary pigment cells and photoreceptors reside medially, and secondary and tertiary pigment cells and interommatidial bristle nuclei are located basally. (C-E″) Expression of the 16.6 kb *mir-279/996-GFP* transgene is readily detected throughout the cone cell layer as marked by DPax2 (C) and in the pigment cells and bristle cells (D), but is excluded from Elav$^+$ photoreceptor neurons. (E) Magnification of the boxed region in D, highlighting exclusion of *mir-279/996-GFP* activity from Elav$^+$ cells. (F-K‴) Functional repression detected by the *tub-GFP-miR-279* sensor. Note that this is a cytoplasmic sensor, whereas the cell-specific markers are nuclear. (F) The *tub-GFP-miR-279* sensor is largely excluded from the cone cell layer. (G) Magnified view of the boxed region in F. (H) The *tub-GFP-miR-279* sensor is reactivated in cone cells when placed in the *mir-279/996[ex36/ex36]* background. (I) The *tub-GFP-miR-279* sensor is coincident with Elav$^+$ photoreceptors but is excluded from pigment cells. (J-J⁗) Higher magnification view emphasizing the on and off spatial pattern of the miR-279 sensor in adjacent Elav$^+$ and DPax2$^+$ cells, respectively. (K-K‴) The *tub-GFP-miR-279* sensor is reactivated in non-neuronal cells within the photoreceptor plane of *mir-279/996[ex36/ex36]* mutants.

GFP was specifically absent from DPax2$^+$ primary pigment cells that are directly adjacent to Elav$^+$ photoreceptors (Fig. 2J).

If the non-uniform expression of the *tub-GFP-miR-279* sensor is truly imposed by these miRNAs, its spatial activity should become equalized in *mir-279/996* mutants. Indeed, expression of the *tub-GFP-miR-279* sensor became uniform throughout the cone cells (Fig. 2H) and the pigment and bristle cells (Fig. 2K) of *mir-279/996[15C/15C]* eyes. We conclude that the *mir-279/996* locus is not only transcriptionally active, but also mediates strong functional repression within diverse non-neuronal cells of the developing eye.

miR-279/996 are required for normal specification of ommatidial cell fates

We characterized the cellular bases of adult eye patterning defects in *mir-279/996* mutants (Fig. 1), which were already becoming evident during the course of our expression pattern studies (Fig. 2). We initially used antibodies to the adherens junctions component Armadillo (Arm). Arm stains throughout the apical surface of photoreceptor cells at 36 h APF, but becomes restricted to zonula adherens junctions by 45 h APF (Zelhof and Hardy, 2004). At the latter time point, Arm staining at the *z*-level in which the photoreceptor apices are well separated reveals seven structures, as R8 resides basally to the others (Fig. 3A). Deletion of *mir-279/996* increased the number of Arm$^+$ photoreceptors observed in single confocal sections (Fig. 3B). These phenotypes were rescued by presence of the wild-type transgene, indicating that they result from the loss of *mir-279/996* (Fig. 3C). Defects were also evident at the cone cell layer, where Arm staining normally outlines the four cone cells in each ommatidia. By contrast, *mir-279/996* mutant ommatidia frequently contained only three cone cells, a phenotype that was also rescued by the wild-type transgene (Fig. 3D-F).

Although the distribution of abnormal ommatidia could be stronger in certain areas of a given eye, there did not appear to be a reproducible spatial preference (e.g. dorsal/ventral or central/peripheral) of phenotypes among eyes. Therefore, to quantify these defects, we utilized all available tissue from all pupal eye preparations. We note that occasional ommatidia exhibited six photoreceptors in a section; however, this did not prove to genuinely reflect photoreceptor loss. Careful examination in confocal stacks revealed these to be due to disturbed cell arrangements, since other photoreceptor apices could be found at other *z*-levels (Fig. S1). Overall, ∼20% of ommatidia have one or two extra photoreceptors in *mir-279/996[15C/15C]* (Fig. 3G), which corresponds to the quantifications from adult sections (Fig. 1). In addition, ∼25% of ommatidia were missing one or two cone cells (Fig. 3H), with a very minor frequency (0.68%) exhibiting five cone cells. The wild-type genomic transgene fully rescued both the ectopic photoreceptor and missing cone cell phenotypes.

We confirmed the cone cell and photoreceptor defects using other markers. We used Discs large (Dlg, or Dlg1) to label cell membranes and DPax2 to label cone cells, and confirmed a substantial population of three-cone ommatidia; these defects were rescuable (Fig. 3I-K). We then analyzed the general neuronal marker Elav to assess over-recruitment of R cells. Although quantification of Elav is challenging due to its diffuse immunoreactivity and the fact that photoreceptor nuclei do not all reside at the same *z*-level at mid-pupal stages, we clearly observed ommatidia with ectopic Elav$^+$ cells in *mir-279/996* mutants (Fig. 3L-N). By carefully examining cells along the *z*-axis in confocal stacks, we confirmed that wild-type ommatidia are associated with eight DAPI-stained Elav$^+$ photoreceptor nuclei, while all ommatidia exhibiting supernumerary photoreceptors with Arm staining carried a correlating increase of DAPI-stained Elav$^+$ nuclei (Fig. S1). This excludes the possibility that ectopic Arm$^+$ structures simply reflect mispositioning of R8, and demonstrates that the miRNA mutants develop ectopic photoreceptors.

Finally, we examined the effects of ectopically expressing *mir-279/996* in the photoreceptors. We first used *GMR-Gal4* to drive expression of *UAS-DsRed-mir-279* in all cells of the developing retina. This treatment strongly disrupted the retina, making it challenging to monitor any cell fate changes. We therefore used the more restricted driver *sev-Gal4*, which is active in R3/R4, R1/R6, R7 and cone cells. Adult eyes of *sev-Gal4; UAS-DsRed-mir-279* showed a selective effect in the R3/4 photoreceptors, which were often degenerate with missing or vestigial rhabdomeres (Fig. S2). Although it is unclear why R3/4 are preferentially sensitive to ectopic *mir-279*, this result suggests that elevated miR-279 can interfere with photoreceptor formation.

miR-279/996 are predominantly required to restrict R7 photoreceptor fate

To gain insight into the identity of ectopic photoreceptors in *mir-279/996* null mutants, we examined a panel of cell-specific markers. At 45 h APF, we observed two classes of supernumerary photoreceptors. Staining for Sens (R8) and BarH1 (R1/6) revealed a low frequency of photoreceptors of heterogeneous fate (Fig. S3). By contrast, staining for the R7 marker Prospero (Pros) revealed a substantial population of ommatidia with two, and sometimes three, Pros$^+$ cells; these were fully suppressed by the genomic *mir-279/996* transgene (Fig. 3O-Q). These cells colabeled with Elav, indicating that they are photoreceptors. Some retinal regions showed much higher densities of the ectopic Pros$^+$ cells than others (Fig. 3R). However, as we could not assign any specific area of the retina that was consistently more affected, our quantifications utilized all available tissue across samples. This revealed that ∼10% of 45 h APF ommatidia bear supernumerary Pros$^+$ cells (Fig. 3S), confirming that R7 comprises a dominant subclass of the ectopic photoreceptors induced by the loss of *mir-279/996*. This supports our morphological classifications and quantifications from adult sections (Fig. 1J-S).

To determine whether mid-pupal cells that ectopically express Pros indeed differentiate as R7s, we stained for the R7-specific rhodopsins

Development (2018) 145, dev159053. doi:10.1242/dev.159053

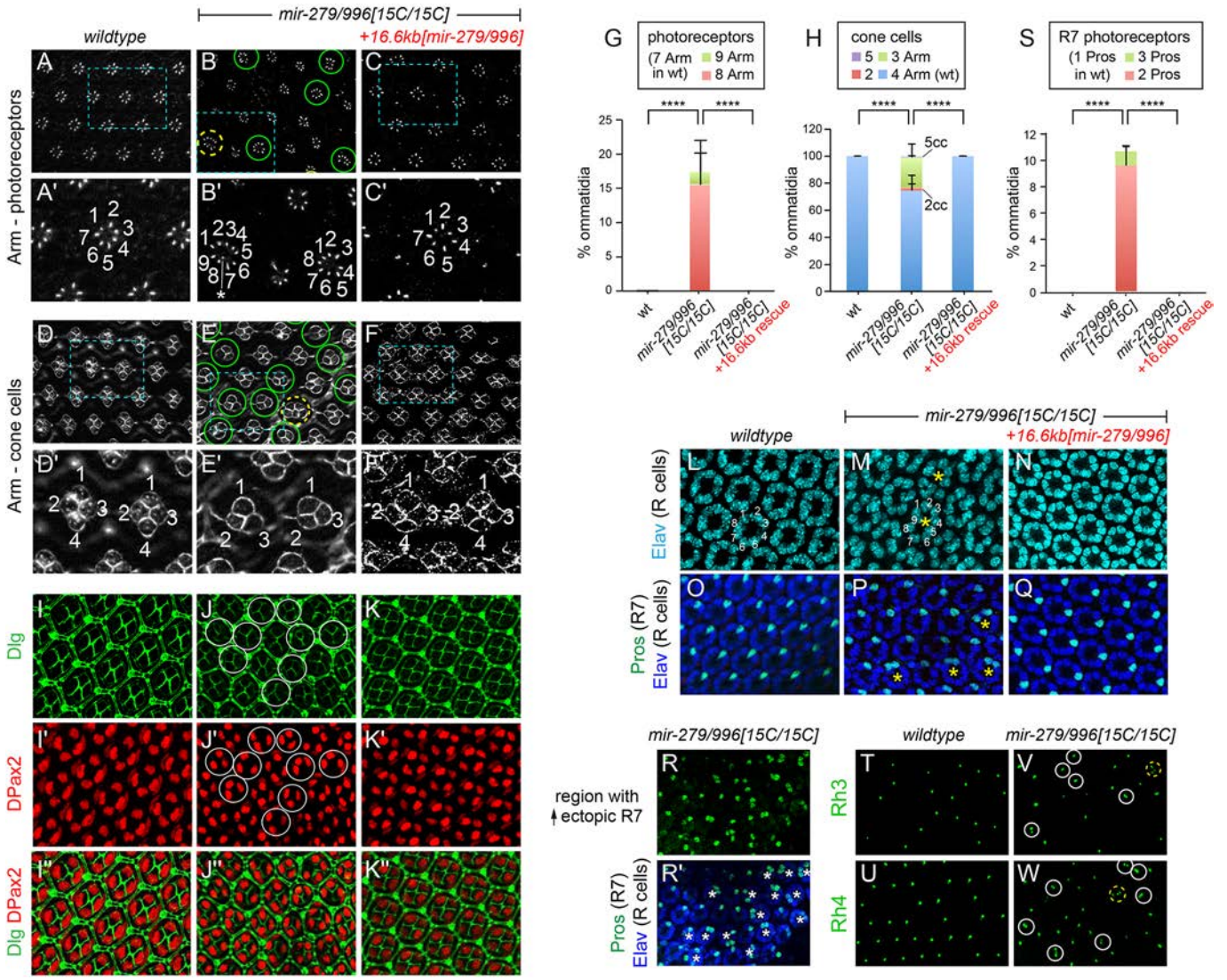


Fig. 3. Cell specification defects in *mir-279/996* mutant eyes. Stainings and quantifications in A-S were from ∼45 h APF pupal eyes, while stainings in T-W were from adult eyes. (A-F) Arm labels the zonula adherens of apically constricted photoreceptors. (A) Seven of the eight neurons are labeled by Arm in a single optical section at the *z*-level in which the photoreceptor apices are well separated. (B) *15C* homozygotes frequently have supernumerary photoreceptors, with ommatidia bearing eight Arm^{+} photoreceptors in a single optical section (green circles), and occasionally more (the yellow circle indicates an ommatidium with nine or possibly ten photoreceptors). (C) These defects were suppressed by a 16.6 kb *mir-279/996* genomic transgene. (A′-C′) Higher magnifications of individual ommatidial groups (boxed regions in A-C) with photoreceptors labeled. A potential tenth photoreceptor in B′ is indicated with an asterisk. (D) In the cone cell layer, Arm labels groups of four cones in each wild-type ommatidium. (E) *mir-279/996[15C/15C]* mutants frequently have only three (or even two) cone cells (circles); very rarely, five cone cell clusters are seen (yellow circle). (F) Cone cell defects are fully rescued by the genomic transgene. (D′-F′) Higher magnifications of individual ommatidial groups (boxed regions in D-F) with cone cells labeled. (G,H) Quantification of ommatidia with aberrant photoreceptor number (G) or cone cell number (H) in *mir-279/996* mutants and rescues. Error bars indicate s.d. ****P<0.0001, one-way ANOVA with Tukey's HSD post-hoc test. (I-K″) Co-staining for Dlg (green, cell membranes) and DPax2 (red, cone cell nuclei). (I) The regular pattern of four cone cells per ommatidium is seen in wild type (wt). (J) *mir-279/996[15C/15C]* mutants frequently exhibit three cone cells per ommatidium, which is rescued by a genomic transgene (K). (L-N) Staining for the neuronal marker Elav demonstrates mutant ommatidia with ectopic photoreceptors (asterisks). (O-Q) Co-staining for Elav and Pros shows that many ectopic photoreceptors are R7 cells. (R,R′) Region of *mir-279/996* mutant eye that shows an especially high frequency of ectopic Pros photoreceptors (asterisks). (S) Quantification of ommatidia with ectopic Pros^{+} R7 cells. Error bars indicate s.d. ****P<0.0001, one-way ANOVA with Tukey's HSD post-hoc test. (T-W) Staining for rhodopsins selectively expressed in terminally differentiated R7 cells: Rh3 (T,V) and Rh4 (U,W). *mir-279/996* mutant ommatidia are labeled; white circles indicate two R7 cells and yellow circles indicate three R7 cells.

Rh3 and Rh4 in adult eyes. Their expression is stochastic, but mutually exclusive, within mature R7 neurons and thus defines two functional subclasses (Fig. 3T,U). We observed ∼12.7% of ommatidia with ectopic Rh3 and/or Rh4 staining (Fig. 3V,W), concordant with the results of ectopic Pros reactivity at 45 h APF. We further analyzed Pros expression in the genetic combinations *mir-279/996[15C/ex36]* and *[ex117/ex117]* (Fig. 1A). These exhibited 8.8% and 1.7% of ommatidia with ectopic Pros^{+} cells, respectively, consistent with the fact that these different *mir-279* deletions express lower and higher levels of miR-996, respectively (Sun et al., 2015). Finally, rescue experiments with *mir-279*-only and *mir-996*-only genomic transgenes showed that both could completely rescue ommatidial cellular organization in the *15C* null homozygotes, including full suppression of ectopic Pros^{+} cells (Fig. S4).

Overall, these cytological tests validate observations from the adult that defective mutant eye phenotypes are a direct consequence

Development (2018) 145, dev159053. doi:10.1242/dev.159053

of miRNA loss, and that both miR-279 and miR-996 can direct normal eye patterning and neuronal suppression.

miR-279/996 repress multiple positive RTK/Ras signaling factors

Supernumerary photoreceptors typically result from ectopic activation of RTK/Ras signaling. As the photoreceptor subtype most sensitive to endogenous *mir-279/996* was R7, we focused our efforts on understanding its specification. Two RTKs are active in *Drosophila* eye development: EGFR and the Sev receptor (Hafen et al., 1987; Schejter and Shilo, 1989). Ectopic activation of either RTK, or their downstream pathways, will trigger cone cell precursors to adopt the R7 fate (Basler et al., 1991). Strikingly, multiple positive core components of the EGFR and Sev signaling pathways bear highly conserved miR-279/996 seed matches in their 3′ UTRs (Fig. 4A). These include *rhomboid* (*rho*) and *roughoid* (*ru*, also known as *rho3*), which encode two serine-type endopeptidases that generate active EGFR ligands, and *bride of sevenless* (*boss*), which encodes the ligand for the Sev receptor. The sites in *ru* and *boss* are optimal ‘8mer’ sites, whereas the site in *rho* is a 7mer-1A site (Lewis et al., 2005). All of these sites are conserved in all sequenced drosophilids, indicating that they are under strong selective constraint (Fig. 4A,B).

We used luciferase sensor assays in cultured cells (Sun et al., 2015) to demonstrate that the 3′ UTRs of RTK pathway factors can be repressed by miR-279 and miR-996. We first tested a validated sensor for the key miR-279 target *nerfin-1* (Cayirlioglu et al., 2008) and observed nearly 90% reduction. We show here that miR-996 has similar activity to miR-279 (Fig. 4C). *nerfin-1* is an unusually strong miRNA target due to five conserved matches to the same miRNA seed, a number that few fly mRNAs bear. By comparison, we observed reduction by half to two-thirds in the levels of the *rho*, *ru* and *boss* 3′ UTR sensors (Fig. 4C), which is comparable to more ‘typical’ miRNA targets. To demonstrate direct regulation, we mutated the cognate binding sites. This abrogated the response of the *rho*, *ru* and *boss* 3′ UTR sensors, indicating that regulation is mediated directly via individual miR-279/996 seed matches (Fig. 4C).

We next generated *tubulin-GFP* transgenes to the full 3′ UTRs of these RTK signaling genes and examined their regulation in the fly. When a control transgene was crossed into a background expressing *UAS-DsRed-mir-279* using *ptc-Gal4*, the expression of the miRNA can be marked in DsRed^{+} cells, which do not alter GFP expression (Fig. 4D). However, when the three RTK pathway sensor transgenes were assayed, all three were cell-autonomously suppressed in the miRNA-expressing domain (Fig. 4E-G). Therefore, *rho*, *ru* and

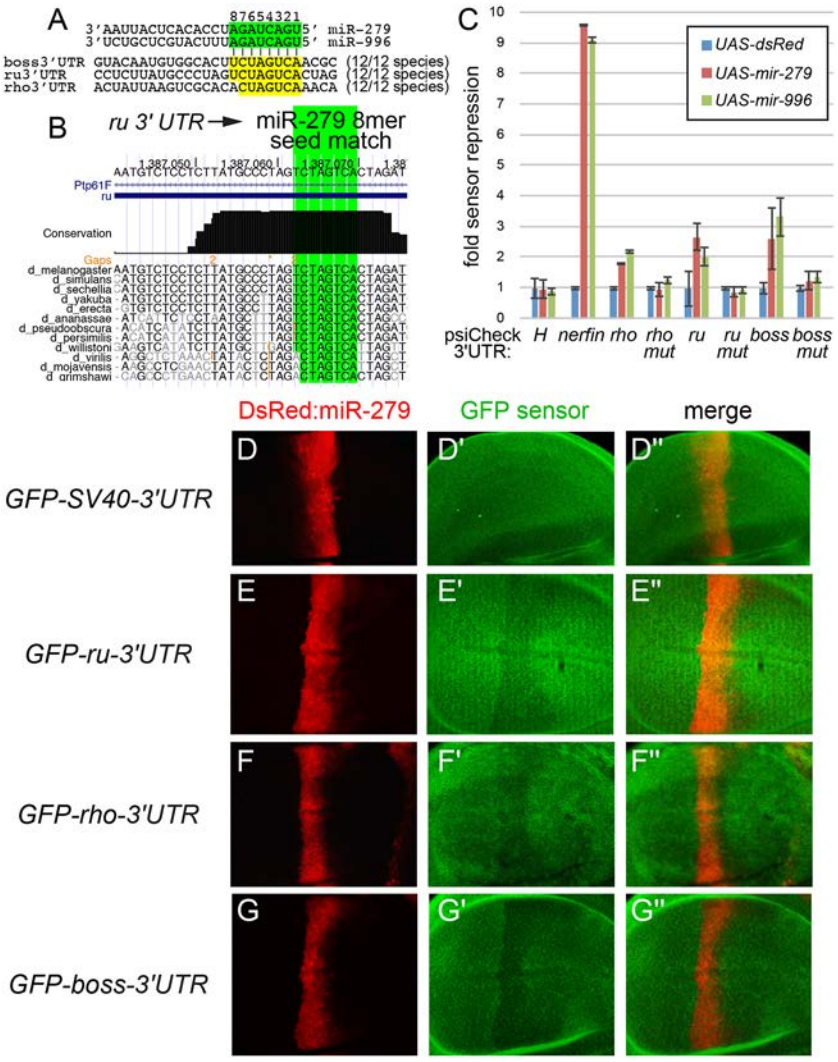


Fig. 4. miR-279/996 directly repress multiple positive components of RTK/Ras pathways. (A) Among 3′ UTRs bearing target sites for the shared miR-279/996 seed sequence, three (*ru*, *rho*, *boss*) are positive components of RTK/Ras that promote photoreceptor/R7 specification. *rho* bears a 7mer-1A site, whereas *boss* and *ru* both contain high-affinity 8mer sites. (B) All of these sites are conserved across the 12 sequenced *Drosophila* genomes, as exemplified for *ru*. (C) Luciferase sensor assays demonstrate target 3′ UTR repression by miR-279 and miR-996. *Hairless* (*H*) is used as a negative control, while *nerfin-1* is a positive control that is exceptionally well repressed by both miRNAs. *rho*, *ru* and *boss* are repressed 2- to 3-fold by ectopic miR-279/996 in a seed-dependent manner. Error bars indicate s.d. (D-G″) Evidence that miR-279 represses Ras pathway 3′ UTRs *in vivo*. Shown are the central portions of wing imaginal discs that express GFP ubiquitously (*tub-GFP-3′ UTR* sensors), in a genetic background in which miR-279 is ectopically expressed in a central stripe labeled by DsRed. (D) miR-279 does not affect a control GFP sensor, but induces cell-autonomous repression of GFP sensors linked to *ru* (E), *rho* (F) and *boss* (G) 3′ UTRs.

Development (2018) 145, dev159053. doi:10.1242/dev.159053

boss are bona fide targets of these miRNAs. These results provide *in vitro* and *in vivo* evidence for post-transcriptional regulation of multiple positive components of RTK/Ras signaling by miR-279/996.

miR-279/996 represses RTK/Ras signaling during eye development

We sought to connect the action of the miRNAs on RTK signaling to the phenotypes observed in *mir-279/996* mutant eyes by assaying for dominant genetic interactions. Even though miRNAs typically have large cohorts of conserved targets (at least 130 *Drosophila* genes bear conserved 3′ UTR seed matches for miR-279/996, http://www.targetscan.org), at least some miRNA mutant phenotypes are highly dose sensitive on the level of individual targets. This implies that particular genes or pathways, out of the entire target network, may drive a particular miRNA mutant phenotype (Dai et al., 2012).

Both Nerfin-1 and Escargot (Esg) are key miR-279 targets, and their heterozygosity can suppress the specification of ectopic CO_2-sensing neurons in the miRNA mutants (Cayirlioglu et al., 2008; Hartl et al., 2011). However, heterozygosity for neither *nerfin-1* nor *esg* modulated the eye roughening of *mir-279/996[15C]* homozygous eyes; in fact, double heterozygosity for these target genes did not suppress eye roughening of the miRNA deletion (Fig. 5A-C). This suggests that the miR-279/996 target cohort that is crucial in olfactory system development is not substantially required during eye development.

Since the ectopic R7s likely resulted from ectopic activation of the RTK pathway, and many RTK-associated genes are functionally targeted by miR-279/996, we next examined genetic interactions with genes associated with RTK activation. Notably, heterozygosity for *Egfr*, *ru*, *rho*, and the nuclear effector *phyllopod* (*phyl*), each yielded detectable dominant suppression of overall adult eye roughness (Fig. 5D-F). The strongest individual suppression was observed with heterozygosity for *phyl*, which was not predicted as a direct miR-279/996 target. However, as a nuclear component of the RTK/Ras pathway, Phyl integrates signaling from both EGFR and Sev signaling, and was previously isolated by virtue of its strong dominant suppression of an activated Ras phenotype (Chang et al., 1995; Dickson et al., 1995). Nevertheless, to test whether combined loss of direct targets could produce greater suppression, we generated a *ru, rho, 15C* recombinant chromosome and backcrossed this to *mir-279/996[15C]*. Double heterozygosity for *ru* and *rho* strongly rescued the regularity of *[15C/15C]* external eyes (Fig. 5G).

We examined the cellular basis of these morphological rescues. Arm staining at 45 h APF served as a convenient readout of photoreceptor numbers. Indeed, multiple components of the EGFR/Ras pathway dominantly suppressed the frequency of ommatidia with ectopic photoreceptors (Fig. 5H-L). We extended these interaction tests by examining the secreted EGFR inhibitor Argos (Freeman et al., 1992; Schweitzer et al., 1995). Whereas *argos* heterozygotes (*argos/+*) were normal in an otherwise wild-type background, *argos/+* enhanced ectopic photoreceptors in *mir-279/996[15C]* mutants (Fig. 5M).

We recapitulated these findings by specific analysis of R7 cells. In particular, double heterozygosity of *nerfin-1* and *esg* did not modify the number of $Pros^+$ cells (Fig. 5N), further indicating that dominant phenotypic suppression is not a trivial genetic outcome, even among demonstrated 'crucial' miRNA targets. On the other hand, heterozygosity for multiple positive EGFR/Ras pathway components (e.g. *ru*, *rho* and *phyl*) each reduced R7 cell numbers (Fig. 5O-R), whereas heterozygosity for *argos* strongly increased R7 cells (Fig. 5S).

We quantified the nature of these genetic interactions at the level of cell fate readouts. Analyses of photoreceptor numbers (Arm^+ cells, Fig. 5T) and R7 photoreceptors ($Pros^+/Elav^+$ cells, Fig. 5U) validated the magnitude and directionality of these phenotypic modifications, and provide further evidence that miR-279/996 endogenously suppress Ras/RTK signaling during eye development. Importantly, assignment of miRNA control of R7 specification extended to terminal differentiation. Supernumerary R7 photoreceptors expressing Rh3 or Rh4 were reduced to <2% in *ru, rho* double heterozygotes, and to ∼0.5% in *phyl* heterozygotes. Reciprocally, whereas *argos* heterozygotes did not exhibit ectopic R7 cells (Fig. 5U), *argos* dominantly enhanced their presence in 15C homozygotes (Fig. 5V). Even with the caveat that the morphological aberration of some rhabdomeres interfered with assignment of photoreceptor subtype, at least 31.6% of ommatidia in *argos/+, mir-279/996[15C/15C]* eyes carried an unambiguously ectopic R7 cell.

Finally, we examined interactions with *boss*. Although the R7 phenotypes we observed could relate to the activities of either EGFR or Sev, and *boss* is a functional miR-279/996 target, we did not detect substantial genetic interactions of *boss* heterozygotes with *mir-279/996[15C]* homozygotes (Fig. 5U). Failure to observe a dominant effect does not rule out the involvement of Boss/Sev signaling. Nevertheless, the strong genetic interactions of *mir-279/996* mutants with positive and negative Ras pathway factors suggest that the phenotype preferentially relates to secretion and antagonism of the EGFR ligand.

Deletion of *mir-279/996* partially bypasses Boss and Sev for R7 specification

The R7 precursor is unusual in that it requires potent activation of the RTK/Ras pathway, an activation level that EGFR alone cannot supply (Tomlinson et al., 2011). Instead, activation of both EGFR and Sev is required, and if Sev alone is removed all ommatidia lack R7s, as this cell transforms into an equatorial cone cell (Tomlinson and Ready, 1986) (Fig. 6A,B). We therefore extended the previous heterozygote experiments by investigating the effect of full *mir-279/996* loss-of-function in the *sev* null background. In particular, we were interested whether EGFR pathway activation caused by lack of *mir-279/996* might restore R7 differentiation in the absence of Sev.

We sectioned adult eyes of *mir-279/996[15C/ex36]* transheterozygotes in a *sev[d2]* background. Whereas *sev[d2]* does not differentiate any R7s, concomitant loss of *mir-279/996* yielded a striking population of rescued R7s (Fig. 6C). We note that in *sev* mutant ommatidia the R8 cell can migrate into the apical regions and present a central rhabdomere that can be mistaken for an R7. To ensure that the rescued cells were bona fide R7s, we examined the basal regions of the R7-bearing ommatidia in adjacent sections, and confirmed the presence of the endogenous R8 (Fig. 6C′). Since the rescued R7s reside in the appropriate position, the simplest scenario is that the equatorial cone cell has been transformed into R7, as opposed to one of the other three cone cell types.

We next examined *sev* and *sev; mir-279/996* double-mutant eyes in 45 h pupae. As expected, *sev* mutants completely lack $Pros^+$ R7 cells (Fig. 6D). We were able to confirm that the antibody staining was successful since the $Pros^+$ sheath cells of the interommatidial bristle sensory organs were labeled. Sheath cells do not express Elav and are thus distinct from the $Elav^+$ neurons of these bristle organs (Fig. 6D′,D″). By contrast, *sev; mir-279/996* double mutants differentiated a population of $Pros^+$ R7 cells, which were confirmed as photoreceptors since they colabeled with Elav (Fig. 6E).

Development (2018) 145, dev159053. doi:10.1242/dev.159053

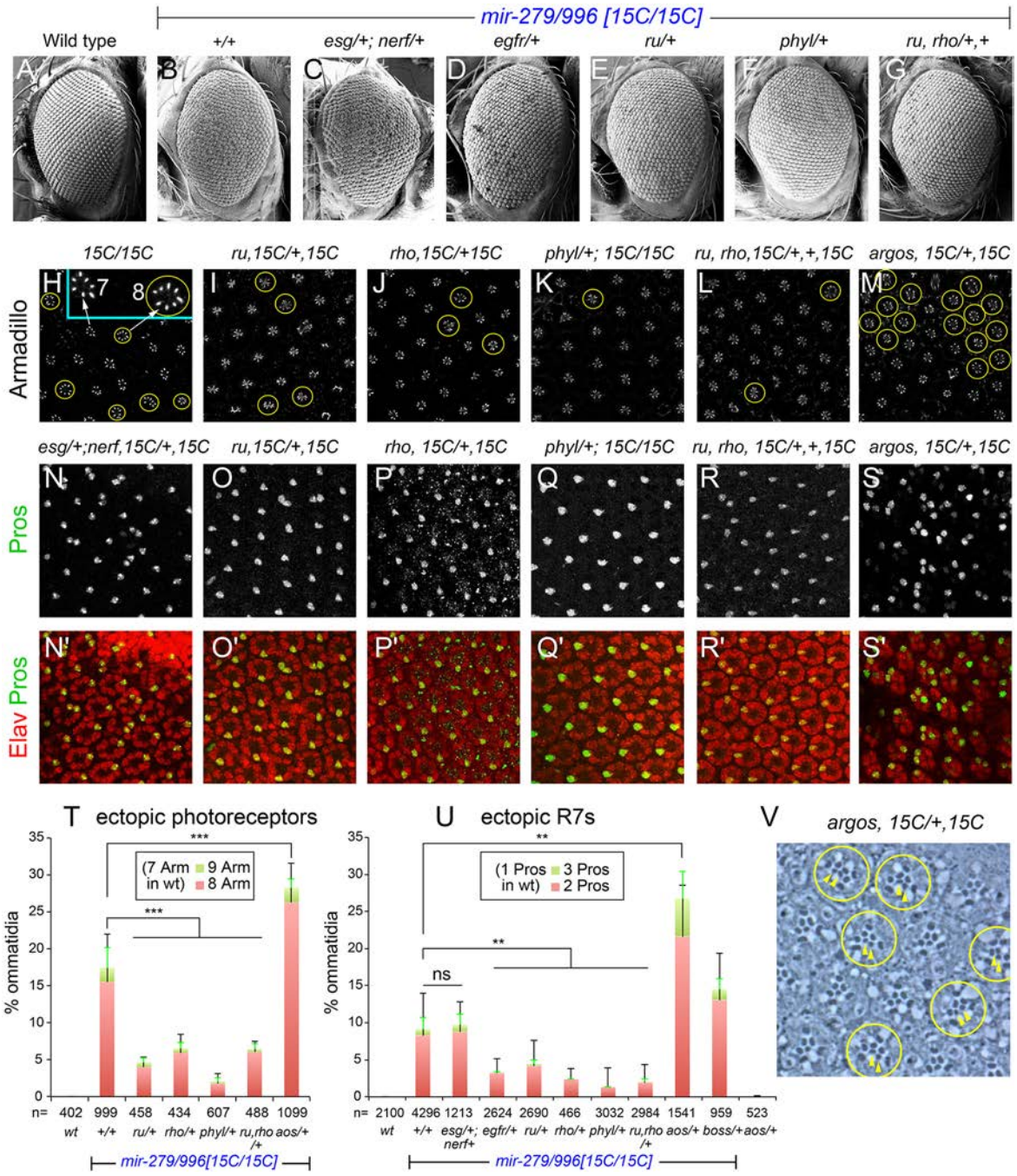


Fig. 5. *mir-279/996* mutant phenotypes are due to elevated Ras pathway activity. Except for the wild-type eye (A), all other adult and pupal eye samples are homozygous for *mir-279/996[15C]*, with other heterozygous mutations as indicated. (A-G) Scanning electron microscopy of adult eyes. (A) Normal regular arrangement of wild-type ommatidia. The rough eye of the *mir-279/996[15C]* homozygote (B) is not modified by double heterozygosity for *nerfin-1* and *esg* (C), but is partially suppressed by *Egfr*/+ (D) and *ru*/+ (E), and strongly suppressed by *phyl*/+ (F) and *ru, rho*/+, + (G). (H-S′) 45 h APF eyes stained for the indicated markers. (H-M) Arm staining focused on photoreceptor apices. Mutant ommatidia with ectopic photoreceptors (eight in one optical section) are circled; the inset (H) shows examples of ommatidia with seven and eight photoreceptors at higher magnification. The phenotype of *[15C/15C]* (H) is dominantly suppressed by heterozygosity for positive Ras pathway components (I-L) and dominantly enhanced by *argos*, a Ras pathway inhibitor (M). (N-S) Pros (green) and Elav (red) staining to detect R7 cells. The phenotype of *[15C/15C]* is not substantially modified by double heterozygosity for *nerfin-1* and *esg* (N), but is dominantly suppressed by heterozygosity for positive Ras pathway components (O-R) and dominantly enhanced by *argos* (S). (T,U) Quantification of ectopic photoreceptor (T) and ectopic R7 (U) phenotypes in various genotypes. Error bars indicate s.d. $^{**}P<0.01$, $^{***}P<0.001$, one-way ANOVA with Tukey's post-hoc test. (V) Retinal section of *argos, 15C/15C* adult eye illustrates a high frequency of ectopic R7 cells (arrowheads), which can be identified based on their central position within the circled ommatidia.

Similar to *sev* mutants, loss of *boss* function ablates R7 specification (Fig. 6F). By contrast, *boss[1], mir-279/996[15C]* double homozygotes displayed clear rescue of a population of Pros^{+} photoreceptors in R7 positions (Fig. 6G). We quantified the restoration of Pros^{+} R7 cells to *sev* and *boss* mutants by concomitant loss of *mir-279/996* (Fig. 6H). Overall, these data provide striking evidence for endogenous restriction of EGFR pathway activity by these miRNAs, which constitute novel crucial players in one of the classic paradigms of cell fate specification.

Development (2018) 145, dev159053. doi:10.1242/dev.159053

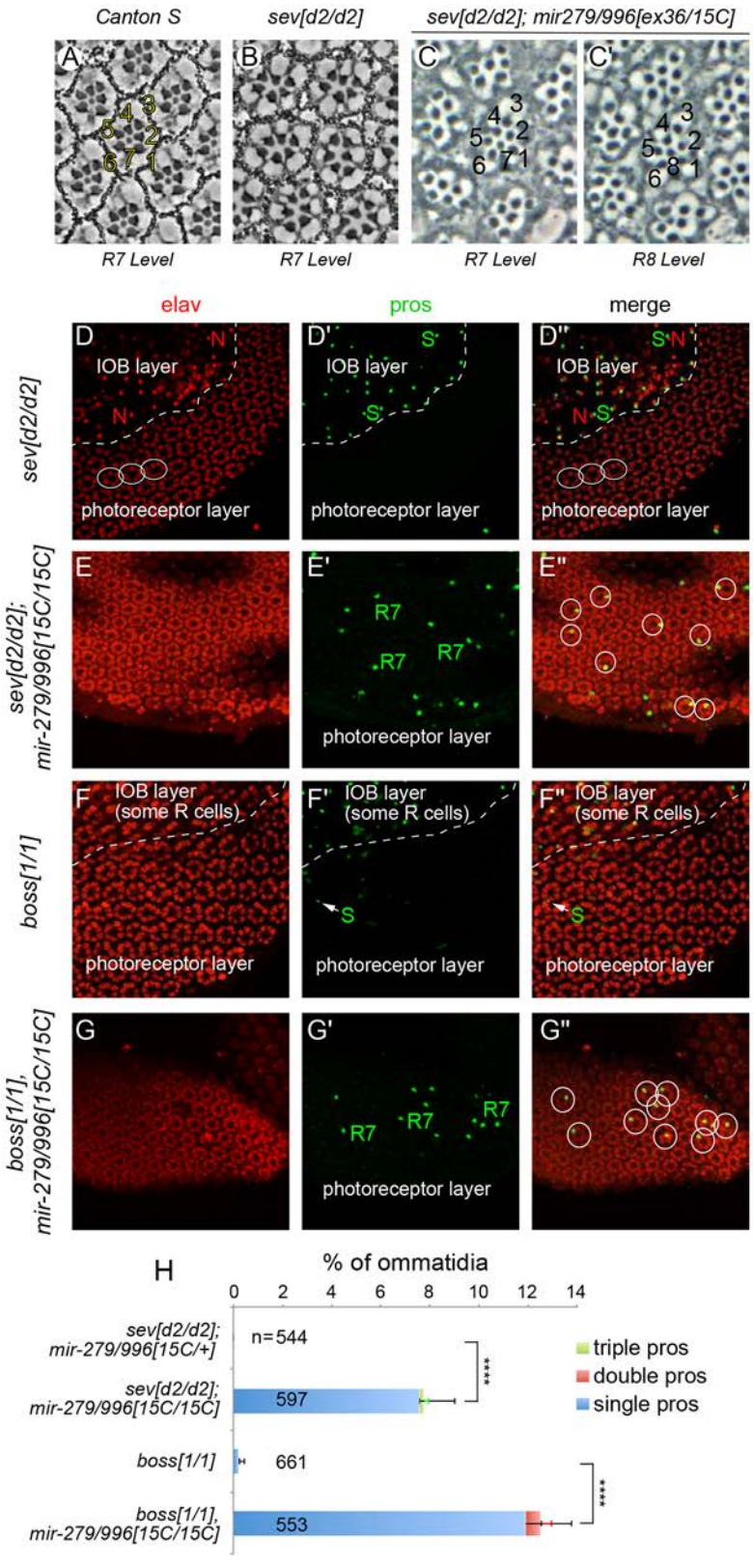


Fig. 6. Deletion of *mir-279/996* partially rescues R7 cells in *sev* and *boss* mutants. (A-C) Plastic sections of adult eyes. (A) Section of Canton S at the R7 cell level shows the characteristic trapezoidal arrangement of seven photoreceptor rhabdomeres in each ommatidia. (B) Section of *sev[d2]* mutant shows only six photoreceptors, with the centrally located R7 absent. (C) Example of rescued R7 cell in *sev[d2]; mir-279/996[ex36/15C]* eye; an adjacent section from the R8 level confirms the presence of R8 (C′), demonstrating that the cell assigned as R7 is not a misplaced R8. (D-G″) Double staining for Pros (green) and Elav (red) at 45 h APF. (D) *sev* mutant eye lacks Pros$^+$ photoreceptors. Proper staining is confirmed by including a section that overlaps the interommatidial bristle (IOB) layer, where adjacent pairs of Pros$^+$ sheath (S) and Elav$^+$ neuronal (N) cells are present in each IOB organ. (E) *sev; mir-279/996* double-mutant eye shows presence of Pros$^+$ (R7) cells; their identity as photoreceptors is evidenced by colabeling for Elav (circled in E″). (F) *boss* mutant eye lacks Pros$^+$ photoreceptors. (G) *boss, mir-279/996* double-mutant eye exhibits rescue of a population of R7 cells (circled in G″). (H) Quantification of R7 rescue in *sev* and *boss* mutants by concomitant deletion of *mir-279/996*. Error bars indicate s.d. ****P<0.0001, one-way ANOVA with Tukey's post-hoc test.

DISCUSSION

miR-279/996 and eye development

During *Drosophila* eye development, the level of RTK pathway activation determines whether a cell can adopt the photoreceptor fate. Multiple mechanisms, including activation of Notch (Tomlinson et al., 2011) and release of Argos (Freeman et al., 1992), ensure that the pathway is activated only in presumptive photoreceptors. Here, we presented molecular and genetic evidence indicating that miR-279/996 suppress RTK activation in non-photoreceptor cells during eye development. Of note, the fact that

Development (2018) 145, dev159053. doi:10.1242/dev.159053

mir-279/996 mutants exhibit a rich set of reciprocal genetic interactions with positive and negative Ras pathway factors suggests that this locus might have unknowingly been hit in previous large-scale dominant modifier screens in the *Drosophila* eye used to isolate components of the RTK pathway (Karim et al., 1996; Simon et al., 1991).

Supernumerary photoreceptors typically result from ectopic activation of RTK/Ras signaling in either of two groups of cells that are not normally destined to become photoreceptors (Basler et al., 1991). Mystery cells are constituents of the early ommatidia that are subsequently lost from the structure (Tomlinson and Ready, 1987). However, if their RTK/Ras pathway is activated, they can generate supernumerary photoreceptors of the R1-6 class (with large rhabdomeres). Cone cells are added to developing ommatidia after the photoreceptors. If RTK/Ras signaling is inappropriately activated in these cells, they generate photoreceptors of the R7 class (with small rhabdomeres) (Basler et al., 1991). As we observe substantial populations of excess photoreceptors with large as well as small rhabdomeres in *mir-279/996* mutants, our data are consistent with the occurrence of both mystery cell transformations and cone cell transformations, although we have focused on the latter as the inferred source of ectopic R7 cells. Further studies are needed to characterize the basis and subtypes of other photoreceptor transformations in *mir-279/996* mutants.

The two RTKs active in *Drosophila* eye development, Sev and EGFR, share intracellular transduction pathways but differ markedly in their ligands. Boss is the ligand for Sev and is an integral plasma membrane protein, whereas Spitz, the ligand for EGFR, is a diffusible peptide released by a subset of differentiating photoreceptors. Although we identify *boss* as a conserved and functional miR-279/996 target, our genetic experiments do not yet implicate miR-279/996 as restricting Boss/Sev signaling. Instead, the data are consistent with the scenario that miR-279/996 may regulate the presentation of ligands to EGFR. In particular, genes that positively regulate EGFR ligands (*ru* and *rho*) are both conserved targets and dominant suppressors of *mir-279/996* mutant eye phenotypes.

One of the most striking phenotypic readouts in *mir-279/996* mutants is the fact that a subset of R7 cells can be rescued by deletion of the miRNA locus in *boss* or *sev* mutants. It is well documented that directed misexpression of activated Ras pathway components can rescue R7 cells in Sev pathway mutants (Basler et al., 1991; Fortini et al., 1992). However, there are very few loss-of-function mutants that can rescue R7 cells in *boss* or *sev* mutants. Mutations of the Ras target gene inhibitor Yan (Lai and Rubin, 1992) and the neural inhibitor Ttk (Lai et al., 1996; Xiong and Montell, 1993) can induce R7 photoreceptors in *sev* mutants, but mutation of the Ras pathway inhibitor Argos does not (Freeman et al., 1992). Of note, the H214 enhancer trap (Mlodzik et al., 1992) is selectively active in R7 precursors and not other photoreceptors. In the absence of *sev*, the R7 precursor transforms into a cone cell precursor and yet retains H214 expression. This was the first evidence that the R7 precursor could receive positional information independently of Sev. Whether H214 expression relates to miR-279/996 activity, and whether it responds to the Notch signals that the R7 precursor receives, remains to be investigated.

Overall, the endogenous impact of miR-279/996 in the eye is more profound than the typical view of 'fine-tuning' regulation attributed to most miRNAs. Indeed, the eye phenotypes of *mir-279/996* mutants place it among a small cohort of miRNA mutants in any species that, when deleted, have overt consequences on external morphology and assignment of cell fate. Although the miR-279 family is not found in chordates (Mohammed et al., 2014), its genetic attributes might have relevance to cancer mechanisms. The setting-specific nature of *mir-279/996* mutant phenotypes provides precedent that potential human miRNA deletions that unleash Ras signaling in disease and cancer might await genetic discovery.

Multiple developmental roles for miR-279/996

A striking dichotomy has emerged from the genetic analysis of miRNA biology (Lai, 2015). The founding miRNAs and even miRNA target sites were identified from genetic aberrations of miRNA loci and 3′ UTRs that yielded highly penetrant morphological defects. Alongside bioinformatic and molecular evidence that miRNAs target a majority of animal transcriptomes, it is widely considered that miRNAs have broad impacts on gene expression. On the other hand, systematic collections of miRNA deletions demonstrate that the vast majority do not cause seemingly overt phenotypes (Chen et al., 2014; Miska et al., 2007). Thus, miRNA loci were probably historically undersampled not on account of the small mutational target size of mature miRNAs, but rather because they only rarely exhibit notable phenotypes that would have permitted their isolation from forward genetic screens.

This dichotomy is intensified by the fact that among the minority of miRNAs with overt phenotypic impacts, several are involved in distinct biological processes, often via different target outputs (Smibert and Lai, 2010). In this regard, *mir-279/996* is an exemplar of a highly pleiotropic miRNA locus. Forward genetic screening originally identified a crucial role for miR-279, later extended to miR-996, in preventing CO_2-sensing olfactory neuron development from maxillary palps (Cayirlioglu et al., 2008; Sun et al., 2015). In this setting, the miRNAs repress neural transcription factors encoded by *nerfin-1* and *esg* to restrict CO_2-responsive neurons (Cayirlioglu et al., 2008; Hartl et al., 2011). Additional studies revealed that these miRNAs control different aspects of JAK-STAT signaling. In the CNS, repression of the JAK-STAT ligand Unpaired is important for circadian behavior (Luo and Sehgal, 2012), whereas regulation of the STAT transcription factor influences ovarian border cell dynamics (Monahan and Starz-Gaiano, 2013; Yoon et al., 2011). Most recently, we found that miR-279/996 suppress neuronal specification in the mechanosensory lineage by repressing the Notch inhibitor Insensible (Kavaler et al., 2018). Here, we extend the role of miR-279/996 to suppressing photoreceptor specification by limiting EGFR/Ras signaling. Notably, the ectopic neuron phenotypes in the olfactory lineage, mechanosensory lineage, and the eye are genetically dose sensitive to distinct target cohorts, and might thus represent convergent 'anti-neural' outputs for this miRNA operon.

MATERIALS AND METHODS

Drosophila strains

We used our published *mir-279/996* alleles *[ex117]*, *[ex36]* and *[15C]*, the genomic 16.6 kb transgenes of *mir-279/996*, *mir-279-2x* and *mir-996-2x* (Sun et al., 2015), and *UAS-DsRed-mir-279* (Bejarano et al., 2012). Other alleles and stocks utilized were: *ru[1]* (BDSC #575), *rho[ve-1]* (BDSC #628), *su(ve)[1] ru[1] rho[ve-1] h[1] th[1]* (BDSC #617), *argos[rlt]* (BDSC #7336), *phyl[2245]-SR2-3* (G. Rubin, Janelia Farm, VA, USA), *Egfr[f24]* (A. Simcox, Ohio State University, OH, USA), *boss[1]* (H. Kramer, UT Southwestern, TX, USA), *nerfin-1[54]* (W. Odenwald, NIH, Washington DC, USA) and *sev[d2]* (A. Tomlinson, Columbia University, NY, USA). Appropriate recombinants were constructed to assay genetic interactions with *mir-279/996* alleles.

To generate the *mir-279-GFP* transgene, we used recombineering to retrieve genomic fragments of 11.8 kb upstream and of 4.7 kb downstream of the *mir-279* hairpin from the BAC CH322-35G11 (BACPAC Resources)

Development (2018) 145, dev159053. doi:10.1242/dev.159053

and cloned them between the *Asc*I and *Not*I sites of attB-P[acman]-Amp^R (H. Bellen, Baylor College of Medicine, TX, USA). *hsp70-GFP-SV40* was PCR cloned from the pEGFP vector (Brennecke et al., 2003) and inserted into the 5′ end of the 4.7 kb genomic fragment, then the resultant EGFP +4.7 kb piece was digested out and ligated with the 11.8 kb upstream sequence to generate the 16.6 kb *mir-279-GFP* construct. Transgenic flies were generated using the PhiC31 system (BestGene).

We generated transgenic sensors by cloning two complementary sequences to miR-279, or the 3′ UTRs of *rho*, *ru*, *boss* and *neur*, downstream of *tub-GFP*. The 3′ UTRs were amplified from genomic DNA using the primers listed in Table S1, and cloned into the *Xba*I/*Xho*I sites of *tub-GFP*. Transgenic animals were generated by co-injection with Δ2-3 helper plasmid (gift of G. Rubin).

Immunohistochemistry

We used a standard protocol for immunostaining of imaginal discs (Lai and Rubin, 2001). Primary antibodies were rat anti-Elav (1:30, 7E8A10, DSHB), mouse anti-Cut (1:25, 2B10, DSHB), mouse anti-Prospero (1:20, MR1A, DSHB), mouse anti-Dlg (1:100, 4F3, DSHB), rabbit anti-DPax2 [1:2000; gift of J. Kavaler (Kavaler et al., 2018)], rabbit anti-GFP (1:1250, A-11122, Invitrogen), chicken anti-GFP (1:1000, ab13970, Abcam), mouse anti-Rh3 (1:10; gift of S. Britt, University of Colorado), rabbit anti-Rh4 (1:100; gift of C. Zuker, Columbia University), rabbit anti-BarH1 (1:500; gift of Kwang-Wook Choi, KAIST South Korea) and guinea pig anti-Senseless (1:2500; gift of Hugo Bellen, Baylor College of Medicine). We used appropriate secondary antibodies conjugated to Alexa 488, 568 and 647 (1:500, Molecular Probes).

Luciferase sensor assays

We used previously described luciferase 3′ UTR sensors in the psiCheck2 backbone for *nerfin-1*, *ru*, *rho* and *boss* (Sun et al., 2015). Point mutations in miR-279/996 seed matches were introduced by site-directed mutagenesis with the oligonucleotides listed in Table S1 and confirmed by sequencing.

S2 cells were plated in 96-well plates and transfected with 12.5 ng/well either UAS-DsRed empty vector or UAS-DsRed-miRNA constructs, 12.5 ng/well psiCheck2 plasmid and 6.25 ng/well ub-Gal4. Transfections were performed using Effectene Transfection Reagent (Qiagen) according to the manufacturer's instructions. Luciferase values were measured 3 days after transfection. We normalized transfection efficiencies using control firefly luciferase carried within psiCheck2, and fold repression was normalized against empty UAS-DsRed and empty psiCheck2 plasmid. We present representative data from quadruplicate sensor assays, for which each set was performed at least three times and found to yield qualitatively similar results. The S2 cells were recently authenticated as being male cells based on expression of the male *Sxl* transcript isoform but were not tested for other contamination.

Acknowledgements

We thank the Bloomington *Drosophila* Stock Center (BDSC), the Developmental Studies Hybridoma Bank (DSHB), and numerous colleagues (cited in the Materials and Methods) for providing fly stocks and antibodies.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: H.D., E.C.L.; Methodology: H.D., R.J.J., A.T., E.C.L.; Validation: H.D., L.F.d.N.; Formal analysis: H.D., L.F.d.N., F.H., Y.E.M., K.V., C.Z., J.K., A.T.; Investigation: H.D., L.F.d.N., F.H., K.S., Y.E.M., K.V., C.Z., A.T.; Resources: H.D., F.H., K.S.; Data curation: H.D., L.F.d.N., K.V., C.Z., J.K., E.C.L.; Writing - original draft: E.C.L.; Writing - review & editing: R.J.J., A.T., E.C.L.; Visualization: H.D., L.F.d.N., E.C.L.; Supervision: R.J.J., A.T., E.C.L.; Project administration: E.C.L.; Funding acquisition: E.C.L.

Funding

Work in the R.J.J. group is supported by National Institutes of Health (NIH) R01-EY025598. Work in the A.T. group was supported by NIH R01-EY026217. Work in the E.C.L. group was supported by the Burroughs Wellcome Fund and the NIH (R01-NS074037, R01-NS083833 and R01-GM083300) and Memorial Sloan-Kettering Cancer Center core grant P30-CA008748. Deposited in PMC for release after 12 months.

Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.159053.supplemental

References

Basler, K., Christen, B. and Hafen, E. (1991). Ligand-independent activation of the sevenless receptor tyrosine kinase changes the fate of cells in the developing Drosophila eye. *Cell* **64**, 1069-1081.

Bejarano, F., Bortolamiol-Becet, D., Dai, Q., Sun, K., Saj, A., Chou, Y.-T., Raleigh, D. R., Kim, K., Ni, J.-Q., Duan, H. et al. (2012). A genome-wide transgenic resource for conditional expression of Drosophila microRNAs. *Development* **139**, 2821-2831.

Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. and Cohen, S. M. (2003). *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25-36.

Cagan, R. L. and Ready, D. F. (1989). *Notch* is required for successive cell decisions in the developing *Drosophila* retina. *Genes Dev.* **3**, 1099-1112.

Cayirlioglu, P., Kadow, I. G., Zhan, X., Okamura, K., Suh, G. S., Gunning, D., Lai, E. C. and Zipursky, S. L. (2008). Hybrid neurons in a microRNA mutant are putative evolutionary intermediates in insect CO_2 sensory systems. *Science* **319**, 1256-1260.

Chang, H. C., Solomon, N. M., Wassarman, D. A., Karim, F. D., Therrien, M., Rubin, G. M. and Wolff, T. (1995). phyllopod functions in the fate determination of a subset of photoreceptors in Drosophila. *Cell* **80**, 463-472.

Chen, Y.-W., Song, S., Weng, R., Verma, P., Kugler, J.-M., Buescher, M., Rouam, S. and Cohen, S. M. (2014). Systematic study of Drosophila microRNA functions using a collection of targeted knockout mutations. *Dev. Cell* **31**, 784-800.

Dai, Q., Smibert, P. and Lai, E. C. (2012). Exploiting Drosophila genetics to understand microRNA function and regulation. *Curr. Top. Dev. Biol.* **99**, 201-235.

Dickson, B. J., Domínguez, M., van der Straten, A. and Hafen, E. (1995). Control of Drosophila photoreceptor cell fates by phyllopod, a novel nuclear protein acting downstream of the Raf kinase. *Cell* **80**, 453-462.

Flynt, A. S. and Lai, E. C. (2008). Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat. Rev. Genet.* **9**, 831-842.

Fortini, M. E., Simon, M. A. and Rubin, G. M. (1992). Signalling by the sevenless protein tyrosine kinase is mimicked by Ras1 activation. *Nature* **355**, 559-561.

Freeman, M. (1996). Reiterative use of the EGF receptor triggers differentiation of all cell types in the Drosophila eye. *Cell* **87**, 651-660.

Freeman, M., Klämbt, C., Goodman, C. S. and Rubin, G. M. (1992). The argos gene encodes a diffusible factor that regulates cell fate decisions in the Drosophila eye. *Cell* **69**, 963-975.

Hafen, E., Basler, K., Edstroem, J. E. and Rubin, G. M. (1987). Sevenless, a cell-specific homeotic gene of Drosophila, encodes a putative transmembrane receptor with a tyrosine kinase domain. *Science* **236**, 55-63.

Hardiman, K. E., Brewster, R., Khan, S. M., Deo, M. and Bodmer, R. (2002). The bereft gene, a potential target of the neural selector gene cut, contributes to bristle morphogenesis. *Genetics* **161**, 231-247.

Hartl, M., Loschek, L. F., Stephan, D., Siju, K. P., Knappmeyer, C. and Kadow, I. C. (2011). A new Prospero and microRNA-279 pathway restricts CO_2 receptor neuron formation. *J. Neurosci.* **31**, 15660-15673.

Hilgers, V., Bushati, N. and Cohen, S. M. (2010). Drosophila microRNAs 263a/b confer robustness during development by protecting nascent sense organs from apoptosis. *PLoS Biol.* **8**, e1000396.

Hipfner, D. R., Weigmann, K. and Cohen, S. M. (2002). The bantam gene regulates Drosophila growth. *Genetics* **161**, 1527-1537.

Karim, F. D., Chang, H. C., Therrien, M., Wassarman, D. A., Laverty, T. and Rubin, G. M. (1996). A screen for genes that function downstream of Ras1 during Drosophila eye development. *Genetics* **143**, 315-329.

Kavaler, J., Duan, H., Aradhya, R., de Navas, L. F., Joseph, B., Shklyar, B. and Lai, E. C. (2018). miRNA suppression of a Notch repressor directs non-neuronal fate in Drosophila mechanosensory organs. *J. Cell Biol.* **217**, 571-583.

Kumar, J. P. (2012). Building an ommatidium one cell at a time. *Dev. Dyn.* **241**, 136-149.

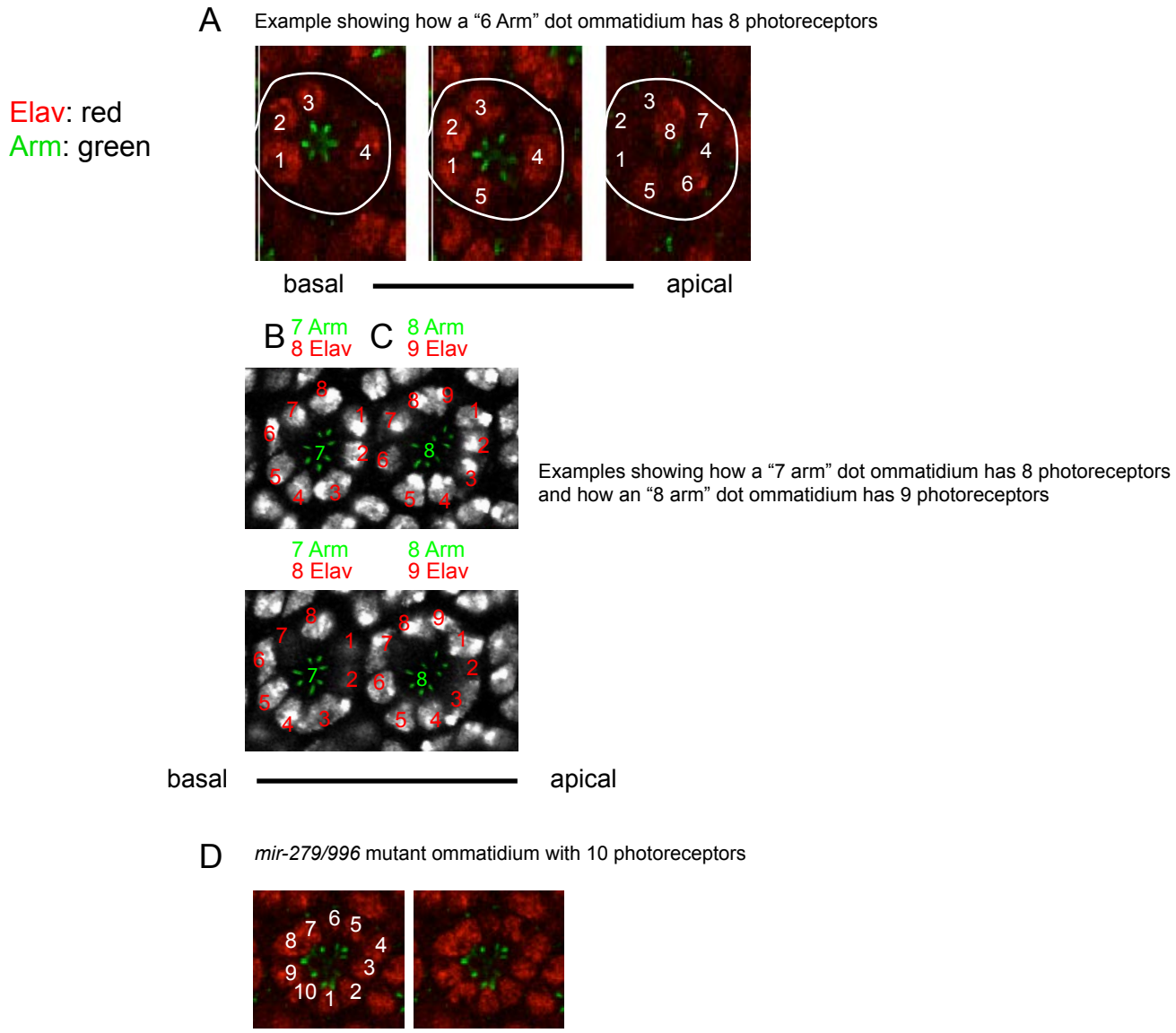
Lai, E. C. (2002). microRNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30**, 363-364.

Lai, E. C. (2015). Two decades of miRNA biology: lessons and challenges. *RNA* **21**, 675-677.

Lai, Z.-C. and Rubin, G. M. (1992). Negative control of photoreceptor development in *Drosophila* by the product of the *yan* gene, an ETS domain protein. *Cell* **70**, 609-620.

Lai, E. C. and Rubin, G. M. (2001). *neuralized* functions cell-autonomously to regulate a subset of Notch-dependent processes during adult *Drosophila* development. *Dev. Biol.* **231**, 217-233.

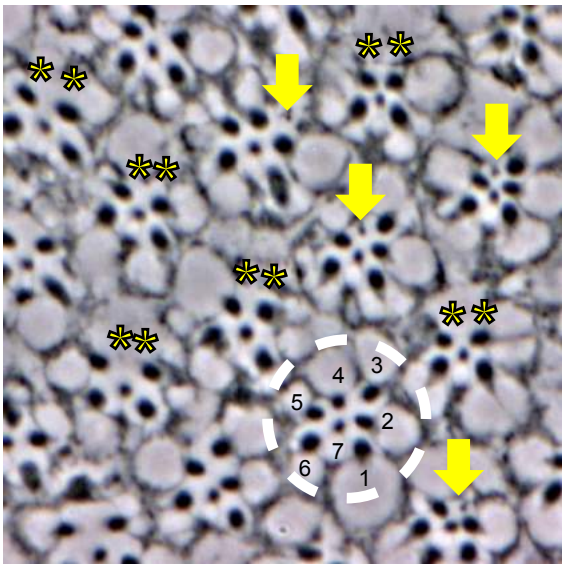
Lai, Z. C., Harrison, S. D., Karim, F., Li, Y. and Rubin, G. M. (1996). Loss of tramtrack gene activity results in ectopic R7 cell formation, even in a sina mutant background. *Proc. Natl. Acad. Sci. USA* **93**, 5025-5030.
Lai, E. C., Burks, C. and Posakony, J. W. (1998). The K box, a conserved 3′ UTR sequence motif, negatively regulates accumulation of *Enhancer of split* Complex transcripts. *Development* **125**, 4077-4088.
Lai, E. C., Tam, B. and Rubin, G. M. (2005). Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev.* **19**, 1067-1080.
Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J. and Carthew, R. W. (2004). Distinct roles for drosophila dicer-1 and dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69-81.
Lewis, B. P., Burge, C. B. and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20.
Li, X. and Carthew, R. W. (2005). A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye. *Cell* **123**, 1267-1277.
Li, X., Cassidy, J. J., Reinke, C. A., Fischboeck, S. and Carthew, R. W. (2009). A microRNA imparts robustness against environmental fluctuation during development. *Cell* **137**, 273-282.
Luo, W. and Sehgal, A. (2012). Regulation of circadian behavioral output via a microRNA-JAK/STAT circuit. *Cell* **148**, 765-779.
Miska, E. A., Alvarez-Saavedra, E., Abbott, A. L., Lau, N. C., Hellman, A. B., McGonagle, S. M., Bartel, D. P., Ambros, V. R. and Horvitz, H. R. (2007). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. *PLoS Genet.* **3**, e215.
Mlodzik, M., Hiromi, Y., Goodman, C. S. and Rubin, G. M. (1992). The presumptive R7 cell of the developing Drosophila eye receives positional information independent of sevenless, boss and sina. *Mech. Dev.* **37**, 37-42.
Mohammed, J., Siepel, A. and Lai, E. C. (2014). Diverse modes of evolutionary emergence and flux of conserved microRNA clusters. *RNA* **20**, 1850-1863.
Monahan, A. J. and Starz-Gaiano, M. (2013). Socs36E attenuates STAT signaling to optimize motile cell specification in the Drosophila ovary. *Dev. Biol.* **379**, 152-166.
Schejter, E. D. and Shilo, B.-Z. (1989). The Drosophila EGF receptor homolog (DER) gene is allelic to faint little ball, a locus essential for embryonic development. *Cell* **56**, 1093-1104.
Schweitzer, R., Howes, R., Smith, R., Shilo, B.-Z. and Freeman, M. (1995). Inhibition of Drosophila EGF receptor activation by the secreted protein Argos. *Nature* **376**, 699-702.
Simon, M. A., Bowtell, D. D. L., Dodson, G. S., Laverty, T. R. and Rubin, G. M. (1991). Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase. *Cell* **67**, 701-716.
Smibert, P. and Lai, E. C. (2010). A view from Drosophila: multiple biological functions for individual microRNAs. *Semin. Cell Dev. Biol.* **21**, 745-753.
Smibert, P., Bejarano, F., Wang, D., Garaulet, D. L., Yang, J.-S., Martin, R., Bortolamiol-Becet, D., Robine, N., Hiesinger, P. R. and Lai, E. C. (2011). A Drosophila genetic screen yields allelic series of core microRNA biogenesis factors and reveals post-developmental roles for microRNAs. *RNA* **17**, 1997-2010.
Sun, K. and Lai, E. C. (2013). Adult-specific functions of animal microRNAs. *Nat. Rev. Genet.* **14**, 535-548.
Sun, K., Jee, D., de Navas, L. F., Duan, H. and Lai, E. C. (2015). Multiple in vivo biological processes are mediated by functionally redundant activities of *Drosophila mir-279* and *mir-996*. *PLoS Genet.* **11**, e1005245.
Tomlinson, A. and Ready, D. F. (1986). Sevenless: a cell-specific homeotic mutation of the Drosophila eye. *Science* **231**, 400-402.
Tomlinson, A. and Ready, D. F. (1987). Neuronal differentiation in Drosophila ommatidium. *Dev. Biol.* **120**, 366-376.
Tomlinson, A., Mavromatakis, Y. E. and Struhl, G. (2011). Three distinct roles for notch in Drosophila R7 photoreceptor specification. *PLoS Biol.* **9**, e1001132.
Truscott, M., Islam, A. B. M. M. K., Lopez-Bigas, N. and Frolov, M. V. (2011). mir-11 limits the proapoptotic function of its host gene, dE2f1. *Genes Dev.* **25**, 1820-1834.
Xiong, W. C. and Montell, C. (1993). tramtrack is a transcriptional repressor required for cell fate determination in the Drosophila eye. *Genes Dev.* **7**, 1085-1096.
Yoon, W. H., Meinhardt, H. and Montell, D. J. (2011). miRNA-mediated feedback inhibition of JAK/STAT morphogen signalling establishes a cell fate threshold. *Nat. Cell Biol.* **13**, 1062-1069.
Zelhof, A. C. and Hardy, R. W. (2004). WASp is required for the correct temporal morphogenesis of rhabdomere microvilli. *J. Cell Biol.* **164**, 417-426.



Supplementary Figure 1. Correspondence of Arm+ constrictions with Elav+ photoreceptors.

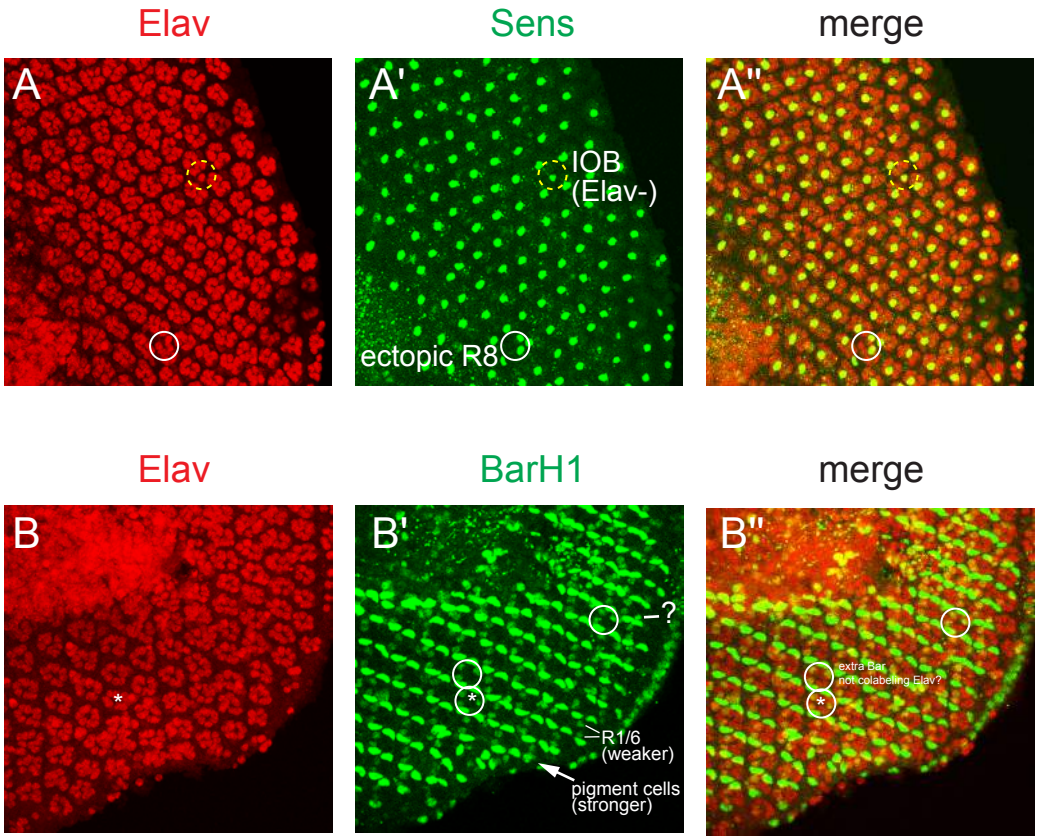
Shown are illustrative ommatidia from *mir-279/996[15C/15C]* eyes, with Arm stained in green and Elav in red. There are normally eight photoreceptors in an ommatidium, but their nuclei do not usually lie in a single plane, and the region where the apical zonula adherens labeled by Arm usually present only 7 constrictions. Therefore, the correspondence of Arm and Elav staining usually requires examination of staining along the Z-axis. (A) Example showing how a 6 dot ommatidia has 8 photoreceptors. (B) Example showing how a 7 dot ommatidia has 8 photoreceptors. (C) Example showing how an 8 dot ommatidia has 9 photoreceptors. (D) Example of an ommatidia with 10 Elav+ photoreceptors, all located in the same plane.

sev-Gal4>UAS-DsRed-mir-279



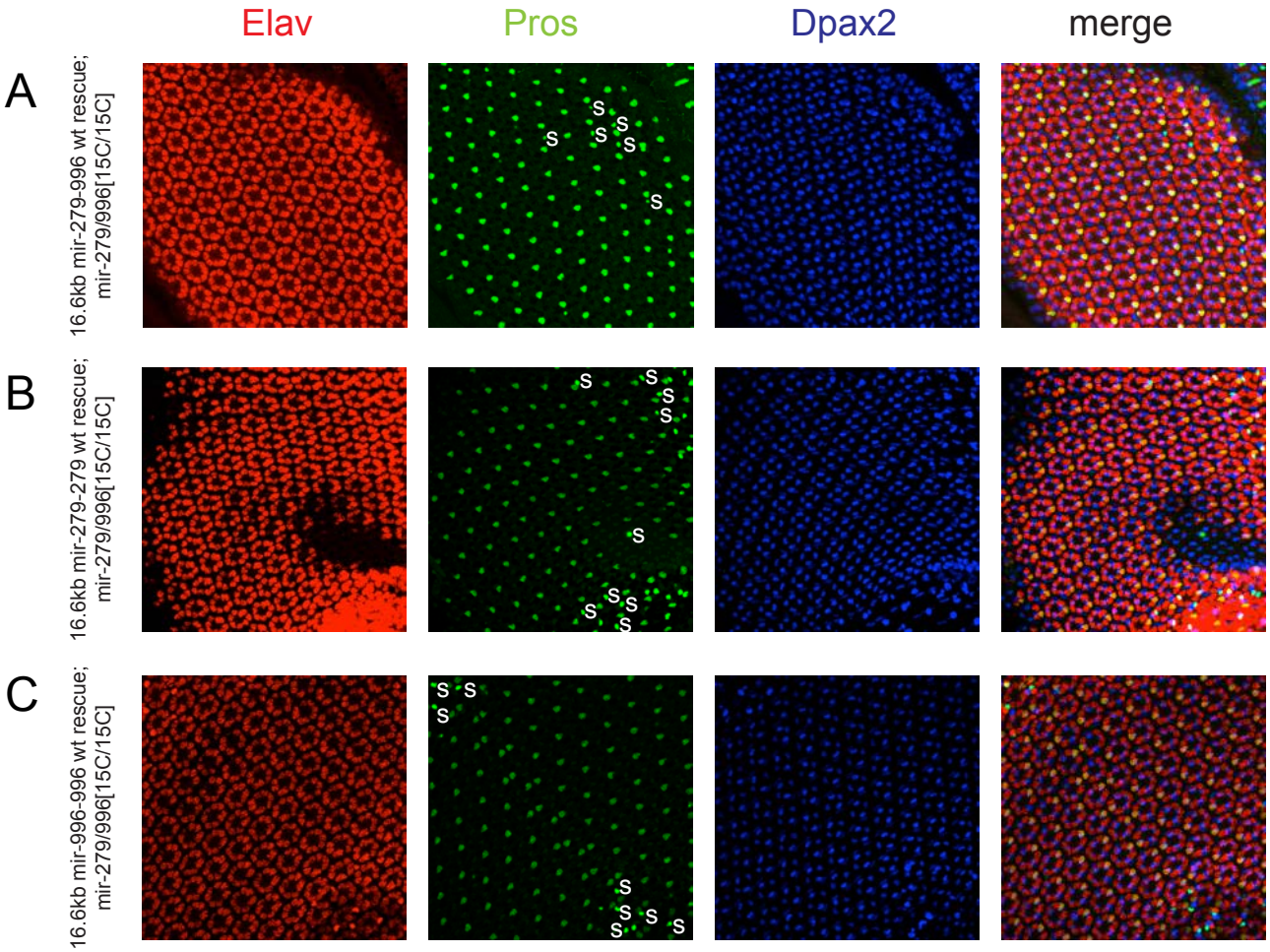
Supplementary Figure 2. Adult retinal sections in mir-279/996 gain-of-function conditions.

sev-Gal4>UAS-DsRed-mir-279 eye shows severe disruption of normal ommatidial patterning. *sev-Gal4* is expressed in R3/R4, R1/R6, R7, and cone cells. Dotted white circle highlights an ommatidia with the normal 7 rhabdomeres. Asterisks indicate absent R3/4 rhabdomeres, while yellow arrows indicate rudimentary/degenerated rhabdomeres.



Supplementary Figure 3. Analysis of photoreceptor subtype-specific markers in mir-279/996 mutants.

Staining for Sens (R8) and BarH1 (R1/6) (in green) reveal sporadic ommatidia with ectopic photoreceptors of these subtypes. Identity of cells as photoreceptors is confirmed by co-staining with Elav (in red). Note that non-neuronal cells express Sens (e.g., within interommatidial bristles, IOB) and BarH1 (e.g. within pigment cells).



Supplementary Figure 4. Rescue of *mir-279/996* deletion mutants by genomic transgenes expressing only individual miRNAs.

Shown are 45 hr pupal eyes stained for the general photoreceptor (R) marker Elav (in red), the R7 marker Prospero (in green), and the cone cell marker D-Pax2 (in blue). Note that these markers also label some additional cell types, which may be variably observed due to the 3D architecture and curved structure of the eye. For example, Prospero is also expressed in the sheath cell of interommatidial bristles that are located at alternate vertices of ommatidia. Sheath cell nuclei are distinguished for R7 cells by their location and by their absence of Elav expression; a subset of sheath cells are labeled (S) for clarity. All of the tissues shown are homozygous for the *mir-279/996[15C]* deletion, which normally exhibits highly aberrant eye development (see main Figures). Each genotype also bears a 16.6 kb genomic transgene of the mir-279/996 locus (panels A-A'''), either fully wildtype or modified so that the mir-279 hairpin replaces the mir-996 hairpin (16.6 kb 2x-mir-279, panels B-B''') or vice versa (16.6 kb 2x-mir-996, panels C-C'''). Regular organization and appropriate cell fate patterning of the eye is observed in *mir-279/996[15C/15C]* deletion mutants bearing any of these genomic rescue transgenes.

Supplementary Table 1	
Primers for amplifying 3'UTRs for wt luciferase and tub-GFP sensors	
rho1-3'UTR-for-xba	GCTCTAGAgagatcgagagacagagagt
rho1-3'UTR-rev-xho	CCGCTCGAGtgtaggggattacgatgctc
ru-3'UTR-for-xba	GCTCTAGAgcagcatctgatgaatgacc
ru-3'UTR-rev-xho	CCGCTCGAGtaagcctaagtaggcaactc
Boss-3'UTR-for-xba	GCTCTAGAatccacaacgtcctccacct
Boss-3'UTR-rev-xho	CCGCTCGAGtcttggttcgtcagtgatt
neur-3'UTR-for-xba	GCTCTAGActacaccacgtagaaaggtc
neur-3'UTR-rev-xho	CCGCTCGAGtccgctttgcttgtccttca
Primers for mutating miR-279/996 on sensors	
ru miR-279 site mutation Fwd	TGTCTCCTCTTATGCCCTAG AGATCAGT CTAGATTATAAGACTTTTCC
ru miR-279 site mutation Rev	GGAAAAGTCTTATAATCTAG ACTGATCT CTAGGGCATAAGAGGAGACA
rho miR-279 site mutation Fwd	GTAAATACTATTAAGTCGCACA GATCAGT AACAACAACAACAGCAACAG
rho miR-279 site mutation Rev	CTGTTGCTGTTGTTGTTGTT ACTGATC TGTGCGACTTAATAGTATTTAC
boss miR-279 site mutation Fwd	CGAGTGTACAATGTGGCACT AGATCAGT ACGCTCCCAGTACGTGAATTC
boss miR-279 site mutation Rev	GAATTCACGTACTGGGAGCGT ACTGATCT AGTGCCACATTGTACACTCG

Appendix 2: Regulatory logic driving stable levels of *defective proventriculus* expression during terminal photoreceptor specification in flies

© 2017. Published by The Company of Biologists Ltd | Development (2017) 144, 844-855 doi:10.1242/dev.144030

RESEARCH ARTICLE

Regulatory logic driving stable levels of *defective proventriculus* expression during terminal photoreceptor specification in flies

Jenny Yan[1,*], Caitlin Anderson[1], Kayla Viets[1], Sang Tran[1], Gregory Goldberg[2,‡], Stephen Small[2] and Robert J. Johnston, Jr[1,§]

ABSTRACT

How differential levels of gene expression are controlled in post-mitotic neurons is poorly understood. In the *Drosophila* retina, expression of the transcription factor Defective Proventriculus (Dve) at distinct cell type-specific levels is required for terminal differentiation of color- and motion-detecting photoreceptors. Here, we find that the activities of two *cis*-regulatory enhancers are coordinated to drive *dve* expression in the fly eye. Three transcription factors act on these enhancers to determine cell-type specificity. Negative autoregulation by Dve maintains expression from each enhancer at distinct homeostatic levels. One enhancer acts as an inducible backup ('dark' shadow enhancer) that is normally repressed but becomes active in the absence of the other enhancer. Thus, two enhancers integrate combinatorial transcription factor input, feedback and redundancy to generate cell type-specific levels of *dve* expression and stable photoreceptor fate. This regulatory logic may represent a general paradigm for how precise levels of gene expression are established and maintained in post-mitotic neurons.

KEY WORDS: Defective proventriculus, Shadow enhancer, Dark shadow enhancer, Rhodopsin, *Drosophila* retina, Photoreceptor, Spineless, Orthodenticle, Spalt

INTRODUCTION

Genes are expressed at distinct cell type-specific levels at different times during development. Expression is often transient, arising for short periods of time to trigger downstream regulatory pathways. For example, expression driven by the *eve* stripe 2 enhancer, perhaps the best-understood regulatory DNA element, is very short-lived, persisting for only ∼15 min after the mature stripe is fully formed during embryonic development in flies (Bothma et al., 2014). By contrast, gene expression in post-mitotic neurons must be maintained on long timescales, often for the lifetime of the organism. Establishing and maintaining distinct levels of transcription factors is particularly important for neuronal fate and function across species. For example, in worms, low levels of the transcription factor MEC-3 specify the elaborate dendritic patterning of PVD pain-sensing neurons, whereas high MEC-3 determines the simple morphology of AVM and PVM touch neurons (Smith et al., 2013). Similarly, flies use differences in levels of the homeodomain transcription factor Cut to control dendritic branching complexity in sensory neuron subtypes (Grueber et al., 2003). In mice, the Hox accessory factor FoxP1 acts as a dose-dependent determinant of motor neuron subtype identity (Dasen et al., 2008). Beyond these cases, there are numerous examples of differential transcription factor expression in neuronal subtypes, such as the unique expression levels of Brn3b in ipRGC subtypes (Chen et al., 2011).

Establishing and maintaining distinct levels of gene expression for the lifetime of a neuron presents specific challenges. Regulatory mechanisms must ensure that expression levels remain within a narrow range for days and even years while providing robustness against acute perturbations caused by activity and environment. In some cases, the transcription factors that dictate cell type-specific expression levels have been identified (Corty et al., 2016), but how these regulatory inputs are interpreted by DNA elements has not been characterized. Furthermore, it is unclear how transcription factor feedback and *cis*-regulatory redundancy contribute to ensuring proper expression levels in neurons.

Expression of transcription factors at cell type-specific levels is required for the terminal specification of motion- and color-detecting photoreceptors in the *Drosophila* retina. The *Drosophila* compound eye consists of approximately 800 ommatidia, or unit eyes, each containing eight photoreceptors (PRs) (Wolff and Ready, 1993) (Fig. 1E). The outer PRs (R1-R6) express the broad spectrum-sensitive Rhodopsin 1 (Rh1) and detect motion (Hardie, 1985), whereas the inner PRs (R7 and R8) express color-sensitive Rhodopsin proteins (Rh3-Rh6) (Gao et al., 2008; Yamaguchi et al., 2010). Two ommatidial subtypes, pale (p) and yellow (y), are randomly distributed in the retina at a ratio of 35:65 (Bell et al., 2007; Franceschini et al., 1981) (Fig. 1A-D). The p subtype contains UV-sensitive Rh3 in pR7 and blue-sensitive Rh5 in pR8, whereas the y subtype contains UV-sensitive Rh4 in yR7 and green-sensitive Rh6 in yR8 (Fig. 1A-C) (Chou et al., 1996; Fortini and Rubin, 1990; Johnston and Desplan, 2010). The specification of these photoreceptor subtypes is controlled by a complex network of transcription factors and other regulators (Hsiao et al., 2013; Jukam and Desplan, 2011; Jukam et al., 2013, 2016; Mikeladze-Dvali et al., 2005; Viets et al., 2016; Wernet et al., 2006).

Differential expression of the K50 homeodomain transcription factor Defective proventriculus (Dve) is crucial for terminal specification of photoreceptors in the fly eye (Johnston et al., 2011; Thanawala et al., 2013). Dve is expressed in a unique pattern, with high levels in the outer PRs, low levels in yR7s and no expression in pR7s or R8s (Fig. 1K). High Dve in motion-detecting outer PRs represses expression of color-detecting Rh3, Rh5 and Rh6. Low levels of Dve in yR7s repress Rh3 to maintain exclusive expression of Rh4 in yR7 subtypes in the main ventral region of the retina (Fig. 1L,

[1]Department of Biology, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218-2685, USA. [2]Center for Developmental Genetics, Department of Biology, New York University, 100 Washington Square East, New York, NY 10003-6688, USA.
*Present address: Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA. ‡Present address: Laboratory of Bacteriology, The Rockefeller University, New York, NY 10065, USA.

§Author for correspondence (robertjohnston@jhu.edu)

R.J.J., 0000-0002-5775-6218

Received 24 August 2016; Accepted 2 January 2017

Development (2017) 144, 844-855 doi:10.1242/dev.144030

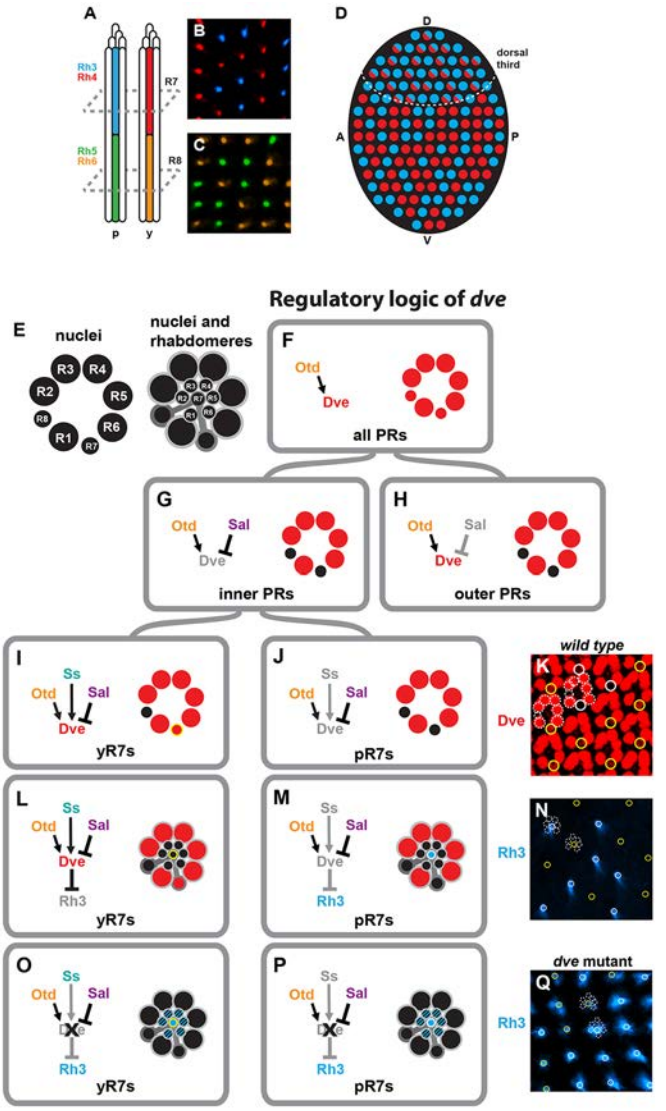


Fig. 1. The regulatory logic controlling Dve. (A) Rh3 (blue) and Rh4 (red) expression in pR7s and yR7s coordinates with Rh5 (green) and Rh6 (orange) expression in pR8s and yR8s in adults. (B,C) Rh3, Rh4, Rh5 and Rh6 in cross-sectional view at the levels depicted by the gray dashed lines in A. Images were taken in adult flies. (D) In the adult fly eye, two ommatidial subtypes, the Rh3-expressing pR7s (blue circles) and the Rh4-expressing yR7s (red circles), are randomly distributed in the retina at a ratio of 35:65. This mutual exclusivity in expression breaks down in the dorsal third region, where Rh3 and Rh4 are co-expressed in the y subtype (half red/half blue circles). A, anterior; P, posterior; D, dorsal; V, ventral. (E) Nuclei and rhabdomeres of the 8 PRs (R1-8) that make up the fly ommatidium. Large outer black circles represent nuclei; smaller inner circles represent rhabdomeres. (F-J,L,M,O,P) Regulatory logic governing *dve*. (Left) Gene network. (Right) Dve expression pattern. Solid color represents consistent expression. Hatched colors indicate variable derepressed expression. (F) Otd activates Dve in all PRs. (G) Sal represses Dve in inner PRs. (H) The absence of Sal allows Dve expression in outer PRs. (I) Ss activates Dve in yR7s. (J) The absence of Ss prevents Dve expression in pR7s. (K) The interactions in F-J yield the expression pattern of Dve: high expression in outer PRs; low expression in yR7s; no expression in pR7s and R8s in pupae. Yellow circles indicate yR7 nuclei with *dve* on; solid white circles indicate pR7 nuclei with *dve* off; dashed white circles are nuclei of outer PRs and R8s. (L) Dve represses Rh3 in yR7s. (M) The absence of Dve allows Rh3 expression in pR7s. (N) Rh3 is expressed only in pR7s where Dve is absent in adults. (O-Q) In *dve* mutants, Rh3 is expressed in all R7s and variably derepressed in outer PRs in adults. (N,Q) Yellow circles indicate yR7 rhabdomeres. Solid white circles indicate pR7 rhabdomeres. Dashed white circles are rhabdomeres of outer PRs.

N). In the dorsal third, Dve levels are lowered further in yR7s to allow co-expression of Rh3 in Rh4-expressing cells (Fig. 1D). The absence of Dve expression allows expression of Rh3 in pR7s (Fig. 1M-N) and Rh5 and Rh6 in R8s (Johnston et al., 2011).

Changes in levels of Dve expression have a dramatic impact on Rhodopsin expression and photoreceptor fate. In *dve* null mutants, Rh3 is derepressed in all R7s, and Rh3, Rh5 and Rh6 are variably expressed in outer PRs (Fig. 1O-Q) (Johnston et al., 2011; Sood et al., 2012). In *dve* hypomorphic mutants, where levels of Dve are lowered but not completely lost, Rh3 is still derepressed in all R7s, but only Rh6 is expressed in outer PRs (Johnston et al., 2011). When Dve levels are subtly lowered upon mutation of upstream regulators, the dorsal region of Rh3 and Rh4 co-expression is expanded from one-third of the retina to the entire dorsal half (Thanawala et al., 2013). The misexpression of Rhs in *dve* mutants causes defects in low-intensity light discrimination (Johnston et al., 2011). Deleterious effects are also seen when Dve levels are increased: raising levels of Dve in yR7s causes loss of Rh3/Rh4 co-expression in the specialized dorsal third region (Mazzoni et al., 2008; Thanawala et al., 2013), whereas overexpression in R8s represses Rh5 and Rh6 completely (Johnston et al., 2011). Thus, the differential expression of *dve* in photoreceptors is important for proper Rh expression and visual function.

Cell type-specific levels of Dve are achieved through regulation by the K50 homeodomain transcription factor Orthodenticle (Otd), the zinc-finger transcription factors Spalt major and Spalt related [referred to collectively as Spalt (Sal)], and the PAS-bHLH transcription factor Spineless (Ss). Otd activates Dve in all PRs (Fig. 1F), Sal represses Dve in the inner PRs (Fig. 1G-H) and Ss re-activates Dve in yR7s (Fig. 1I,J) (Johnston, 2013; Johnston et al., 2011).

To determine how these transcription factors dictate cell type-specific levels of Dve expression, we analyzed the *cis*-regulatory logic controlling *dve* and identified two enhancers, *yR7 enh* and *outer enh*, that together induced expression recapitulating endogenous Dve expression. *yR7 enh* is activated by Ss, Sal and Otd in yR7 cells, whereas *outer enh* is activated by Otd in all PRs and repressed by Sal in inner PRs. Negative feedback by Dve onto both enhancers maintains proper levels of Dve expression. This autoregulation is particularly important for *yR7 enh*, which is dramatically upregulated in yR7s when Dve feedback is ablated. Interestingly, we also observed derepression of *yR7 enh* in outer PRs in *dve* mutants, suggesting that *yR7 enh* serves as an inducible backup or 'dark' shadow enhancer in these cells. Shadow enhancers are DNA elements that drive redundant expression patterns and ensure robust gene expression in cases of genetic and environmental perturbation (Bothma et al., 2015; Frankel et al., 2010; Hong et al., 2008; Miller et al., 2014; Nolte et al., 2013; Perry et al., 2010; Wunderlich et al., 2015). *yR7 enh* represents an unusual 'dark' shadow enhancer as it is normally repressed and only becomes active when Dve driven by the primary *outer enh* is compromised. Together, the *yR7 enh* and *outer enh* integrate combinatorial transcription factor input, negative feedback and redundancy to ensure distinct cell type-specific levels of *dve* expression required for stable photoreceptor specification.

RESULTS

Two enhancers determine yR7- and outer PR-specific expression of Dve

The *dve* gene locus is ~65 kb with two alternative transcriptional starts driven by the *dve-A* promoter or *dve-B* promoter (Fig. 2A). Deletion of the *dve-A* promoter caused derepression of Rh3 in yR7s in the dorsal half of the retina (Fig. S1A,B), while Rh5 and Rh6 expression were unaffected (Fig. S1C). This incomplete *dve*

Development (2017) 144, 844-855 doi:10.1242/dev.144030

phenotype is consistent with a decrease in Dve levels in yR7s (Thanawala et al., 2013), suggesting that the *dve-A* promoter is required for normal Dve expression. To test the role of the *dve-B* promoter, we employed a CRISPR strategy to delete a ~1.5 kb region encompassing the *dve-B* promoter and first exon. Deletion of the *dve-B* promoter did not alter Dve-regulated Rh expression (Fig. S1D-F), suggesting that the *dve-B* promoter is not required for Dve expression. As the *dve-A* promoter is required for normal Dve expression, we used this promoter as the minimal promoter in enhancer reporters.

To identify *cis*-regulatory elements controlling *dve* expression, we generated transgenes containing 3-6 kb DNA fragments from the *dve* locus and the *dve-A* promoter driving nuclear GFP (Fig. 2A, *dve enh>GFP*). The *dve-A* promoter alone drove extremely weak GFP expression in pigment cells and R4 PRs, and therefore did not recapitulate normal Dve expression in all outer PRs and yR7s (Fig. S1H).

Two constructs drove GFP expression that together recapitulated endogenous Dve expression in midpupation [i.e. ~48 h after puparium formation (APF)]. *outer enh* drove expression in outer PRs (Fig. 2A,E), and *yR7 enh* drove expression specifically in a subset of R7s (Fig. 2A,B). This subset corresponded to yR7 fate, as 68% of R7s had strong GFP expression and perfectly co-expressed Ss (i.e. yR7s), whereas 32% had weak or no GFP and lacked Ss (i.e. pR7s) (Fig. 2B-D).

Additionally, *weak yR7 enh* drove weak expression in yR7s (Fig. 2A, Fig. S1M-O, described further below), and *dorsal R7 enh* drove expression in dorsal posterior R7s (Fig. 2A, Fig. S1P-Q). Four enhancers drove weak expression in all PRs (*all PRs enh 1-4*) (Fig. 2A, Fig. S1I-L).

Janelia Research Campus and the Vienna *Drosophila* Resource Center (VDRC) both generated lines that express Gal4 driven by fragments of the *dve* locus (Fig. 2A). Expression driven by these fragments was consistent with results from our *dve enh* reporter constructs. GMR40E08, a ~3 kb fragment that overlaps with *outer enh*, drove strong GFP expression in outer PRs, whereas other constructs that either did not overlap or only partially overlapped with *outer enh* or *yR7 enh* did not show significant expression (Fig. 2A).

As *yR7 enh* and *outer enh* recapitulated endogenous Dve expression, we further characterized the temporal dynamics of these two enhancers. At midpupation, Dve protein is expressed strongly in outer PRs and weakly in yR7s (Johnston et al., 2011), similar to GFP expression driven by *yR7 enh* and *outer enh* (Fig. S2B,F,J). In third instar larvae, analysis of Dve protein expression was obscured by non-specific antibody staining (Fig. S2A) (Johnston et al., 2011). Although *outer enh* was not expressed, *yR7 enh* was expressed in a subset of R7s (Fig. S2E,I), suggesting that Dve is expressed in larval yR7s. In adults, Dve protein is expressed in yR7s and outer PRs (Fig. S2C,D). Similarly, *outer enh* drove GFP expression in outer PRs in adults (Fig. S2K,L). *yR7 enh* drove expression in all R7s in adults (Fig. S2G,H), suggesting that additional activators present only in the adult stage induce *yR7 enh* expression in all R7s, and that this enhancer is missing DNA elements that prevent ectopic Dve expression in adults.

Together, the spatiotemporal dynamics of these enhancers are consistent with endogenous *dve* expression. Next, we tested how upstream transcription factors control expression of these two enhancers.

yR7 enh is activated by Ss, Sal and Otd

yR7 enh drives expression in yR7 cells (Fig. 3A). Dve is expressed at lower levels in yR7s in the dorsal third, allowing IroC-induced activation of Rh3 and co-expression of Rh3 and Rh4 (Johnston et al., 2011). Similar to endogenous Dve expression, *yR7 enh* is expressed at lower levels in dorsal third (DT) yR7s when compared with the rest of the retina (Fig. 3D).

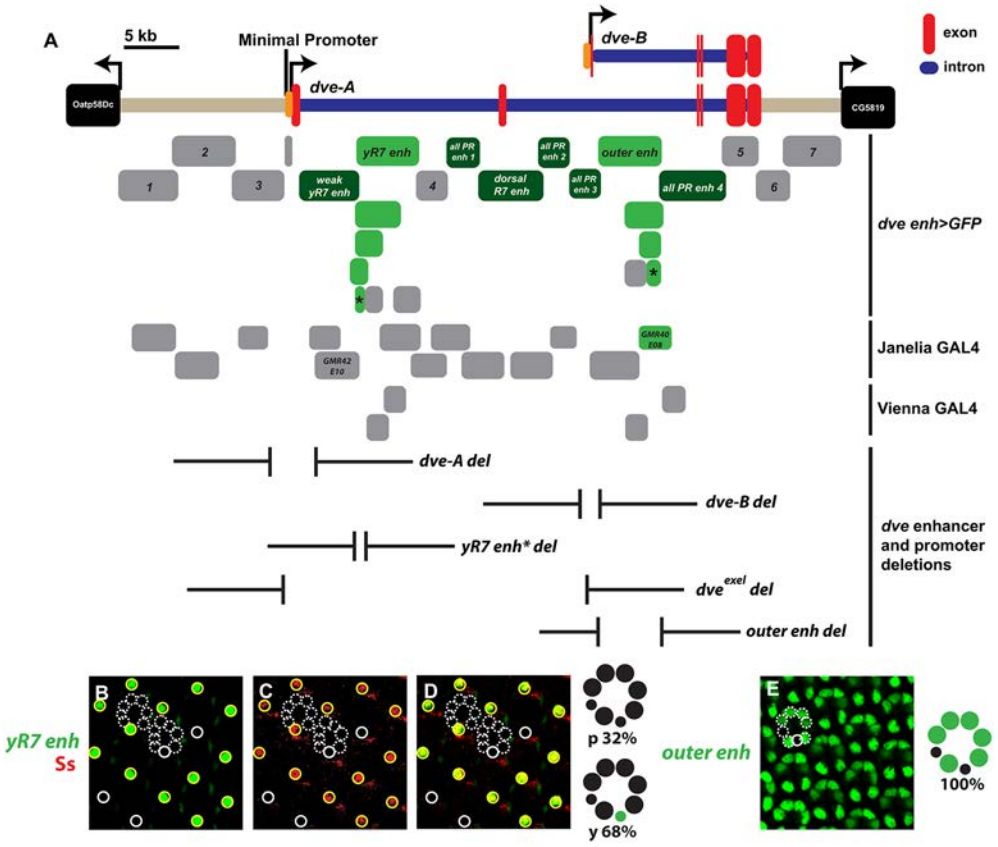


Fig. 2. Two enhancers recapitulate Dve expression. (A) The *dve* locus, reporter constructs and deletions. Reporter constructs (*dve enh>GFP*) consist of fragments of the *dve* locus and the *dve-A* promoter driving nuclear GFP. Smaller fragments, denoted with an asterisk, represent the shortest truncations generated that recapitulate the expression level of the original *dve enh>GFP* constructs. Janelia Research Campus and VDRC stocks contain fragments of the *dve* locus driving Gal4. Light green fragments drive strong expression. Dark green fragments drive weak expression. Gray fragments did not drive expression. (B-D) *yR7 enh* drives expression in yR7s at mid-pupation; Ss is a marker for yR7s. Yellow circles indicate yR7 cells; solid white circles indicate pR7 cells; dashed white circles are outer PRs and R8s. In schematics, black circles indicate no GFP expression and green circles indicate GFP expression. (E) *outer enh* constructs drive GFP expression in outer PRs at mid-pupation. Dashed white circles indicate outer PRs and R8s; solid white circle indicates R7.

Development (2017) 144, 844-855 doi:10.1242/dev.144030

Otd is required for Dve expression in yR7s (Johnston et al., 2011). *yR7 enh* failed to induce expression in yR7s in *otd* mutants, suggesting that Otd is required for activation of this enhancer (Fig. 3B).

Ss induces expression of Dve in yR7s (Johnston et al., 2011). Expression of *yR7 enh* was lost in *ss* mutants (Fig. 3C). Ectopic expression of Ss in all PRs induced strong *yR7 enh* expression in all R7s and weak expression in all other PRs (Fig. 3E), suggesting that another factor acts with Ss to activate strong *yR7 enh* expression.

As Sal is important for R7 fate (Mollereau et al., 2001), we posited that Sal may work with Ss to activate *yR7 enh*. Expression of *yR7 enh* was completely lost in *sal* mutants (Fig. 3F), whereas ectopic expression of Sal in all PRs induced *yR7 enh* expression in a random subset of R1 and R6 outer PRs (Fig. 3G). We showed previously that ectopic Sal induced Ss in a random subset of R1 and R6 outer PRs (Johnston and Desplan, 2014). These data suggest that Ss and Sal function together to activate expression of *yR7 enh*.

Supporting our hypothesis, ectopic expression of both Ss and Sal induced strong *yR7 enh* expression in all PRs (Fig. 3H), suggesting that Ss and Sal both activate expression of *yR7 enh*. As Sal induces expression of Ss, and Ss together with Sal induces *yR7 enh*, Ss, Sal and *yR7 enh* form a coherent feed-forward loop (Fig. 3I).

To further elucidate these combinatorial regulatory interactions, we truncated *yR7 enh* to a 0.8 kb fragment (*yR7 enh**) that recapitulated *yR7* expression driven by the entire *yR7 enh* fragment (Fig. 2A; Fig. 3J,L, Fig. S3A). Three other truncations that encompass the 0.8 kb region also recapitulated yR7 expression, whereas two truncations and four GAL4 lines generated by Janelia Research Campus and VDRC that excluded *yR7 enh** failed to drive GFP expression, consistent with the role of *yR7 enh** in driving yR7 specific expression (Fig. 3J). *yR7 enh** contains three conserved Ss binding sites (called Xenobiotic Response Elements/XREs) (Fig. 3K), consistent with regulation by Ss.

weak yR7 enh drove weak GFP expression in yR7s, colocalizing with Ss expression (Fig. S1M-O). *weak yR7 enh* and *yR7 enh** share a ∼250 bp overlap that contains one of the three Ss XRE binding sites (Fig. 3J,K), suggesting that while the shared XRE site can drive GFP in yR7s, strong expression requires the presence of additional XRE sites. The Janelia enhancer GMR42E10 shares a ∼75 bp overlap with *yR7 enh* but does not contain any Ss XRE binding sites (Fig. 3J,K). This construct failed to drive GFP expression, suggesting that at least one Ss XRE binding site is required for yR7-specific expression.

To further test the roles of Ss XRE binding sites, we generated a *yR7 enh** construct that replaces all GCGTG Ss XRE binding sites with AAAAA. This construct showed a near complete loss of yR7 GFP expression, indicating the importance of these sites for Ss activation (Fig. 3M). Very low-level expression of this reporter suggests the presence of additional cryptic Ss sites within *yR7 enh** (Fig. 3M). Searching *yR7 enh** for low-affinity Ss binding motifs (Zhu et al., 2011), we identified two putative sites (GTCTGA and GTGTGA), one of which is conserved (GTCTGA), suggesting that these cryptic/low-affinity sites may drive very low level expression in the absence of core conserved (GCGTG) sites. Together, these data suggest that Ss directly binds the XRE sites in *yR7 enh** to regulate expression. However, we cannot rule out possible indirect mechanisms.

Although *yR7 enh** has three Ss XRE sites, this enhancer contains no predicted Sal sites (Barrio et al., 1996; Sanchez et al., 2011), suggesting that Sal regulates *yR7 enh** either directly via binding to cryptic sites or indirectly through regulation of other intermediary factors. The longer *yR7 enh* contains a Sal binding site, which may contribute to regulation. Genetic epistasis analysis supports an indirect mode of regulation by Sal (Fig. S4; see below).

*yR7 enh** is required for expression of endogenous Dve in yR7s, as CRISPR-generated deletion of *yR7 enh** caused a loss of Dve expression specifically in R7s (Figs 2A, 3N) and a corresponding upregulation of Rh3 in all PRs (Fig. 3O). Similarly, the larger dve^{exel} deletion, covering *yR7 enh* and the *dve-A* promoter, also resulted in Rh3 upregulation in R7s (Fig. 2A, Fig. S1G). Together, these results suggest that yR7-specfic expression of Dve requires *yR7 enh*, which is activated by Ss, Sal and Otd.

Negative feedback onto *yR7 enh* determines homeostatic levels

Expression levels of Dve are precisely controlled to determine region-specific activation or repression of Rh3 in yR7s (Thanawala et al., 2013) (Fig. 1D). Negative feedback is a mechanism that ensures precise, homeostatic levels of gene expression. As Dve is a transcriptional repressor, we hypothesized that Dve feeds back onto *yR7 enh* to control expression levels. To test Dve for negative regulation of *yR7 enh*, Dve was expressed in all PRs at high levels causing a complete loss of *yR7 enh* expression (Fig. 4A). *yR7 enh* was expressed at higher levels in yR7s in *dve* mutant clones compared with wild-type clones (Fig. 4B,C,F), suggesting that Dve driven by *yR7 enh* feeds back to control levels of expression in yR7s (Fig. 4H).

yR7 enh is a 'dark' shadow enhancer for outer PR expression

In addition to yR7s, expression of *yR7 enh* occurred in outer PRs in *dve* mutant clones (Fig. 4D,E,G), suggesting that *outer enh* induces Dve expression to completely repress *yR7 enh* in outer PRs in normal conditions (Fig. 4I,J). As *yR7 enh* was never expressed in pR7s or R8s in wild type or in *dve* mutants (Fig. 4B-E), *yR7 enh* is only competent to drive expression in yR7s and outer PRs, where Dve is normally expressed.

As *outer enh* drives expression in outer PRs in normal conditions and *yR7 enh* drives expression in outer PRs in *dve* mutants, we predicted that deleting *outer enh* would cause *yR7 enh* to drive expression of endogenous *dve* in outer PRs (Fig. 4I,J). Flies with a CRISPR-mediated deletion of *outer enh* displayed expression of Dve in outer PRs (Fig. 4K) and repression of Rh3, Rh5 and Rh6 (i.e. Dve target genes) in outer PRs in 1-week-old adults (Fig. 4L,N), suggesting that *yR7 enh* drives expression in the absence of functional *outer enh*. Although Rh3 expression remained unchanged (Fig. 4M), variable derepression of Rh5 and Rh6 occurred in 4-week-old adults (Fig. 4O), suggesting that expression driven by *yR7 enh* is not sufficient to completely rescue Dve expression due to differences in levels or timing.

As *yR7 enh* can drive expression in outer PRs, *yR7 enh* is a shadow enhancer (i.e. redundant regulatory DNA element) for *outer enh*, the primary enhancer for outer PR expression. Unlike typical shadow enhancers, the *yR7 enh* shadow enhancer is repressed ('dark') in outer PRs under normal conditions due to negative feedback from the primary enhancer (Fig. 4I). We therefore define *yR7 enh* as a 'dark' shadow enhancer, as its expression in outer PRs only occurs when *outer enh* function is lost (Fig. 4J).

Otd/Dve sites play context-dependent roles in *yR7 enh*

As Otd activates and Dve represses *yR7 enh*, we next tested the regulatory roles of canonical Otd/Dve binding sites (also called K50 sites; TAATCC). *yR7 enh** contains two Otd/Dve sites, which are perfectly conserved across at least five out of six *Drosophila* species (Fig. 3K). Replacing these two sites with AAAAAA caused increased levels of GFP expression in yR7s (Fig. 4P), suggesting that these sites mediate repression by Dve but not activation by Otd in yR7s. As Otd is required for expression of *yR7 enh*, the expression of GFP in yR7s in the absence of optimal Otd binding

Development (2017) 144, 844-855 doi:10.1242/dev.144030

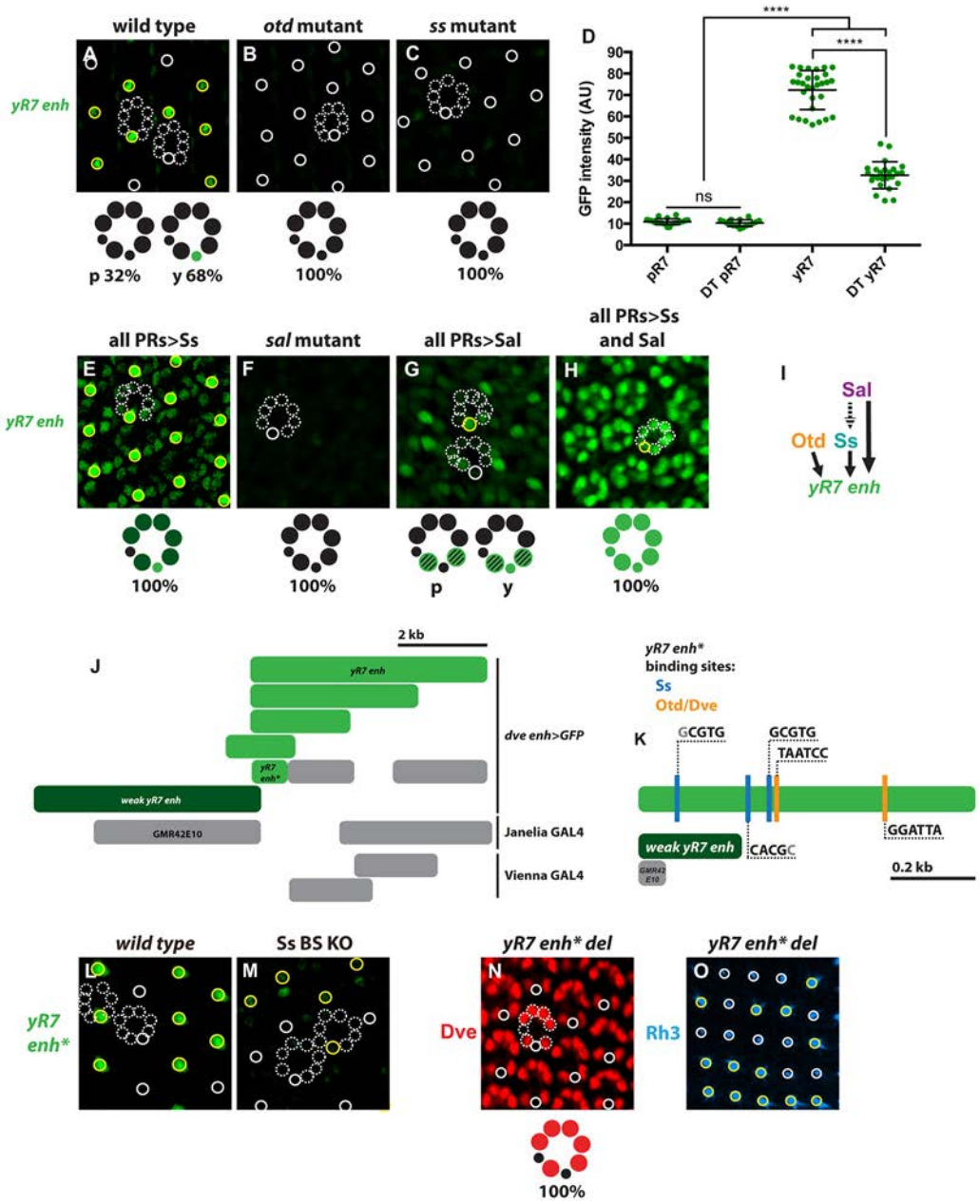


Fig. 3. *yR7 enh* is activated by Otd, Sal and Ss. (A-C,E-H,L-M) Yellow circles indicate yR7s; solid white circles indicate pR7s. Dashed white circles indicate outer PRs and R8s. Light green in ommatidium schematics indicates strong GFP expression; dark green indicates weak expression; crosshatch indicates variable expression; black indicates lack of expression. Images were acquired at mid-pupation. (A) *yR7 enh* is expressed in yR7s. (B) Expression of *yR7 enh* is lost in *otd* mutants. (C) Expression of *yR7 enh* is lost in *ss* mutants. (D) Quantification of GFP intensity in R7 cells shows three distinct intensity levels corresponding to pR7 [including pR7 and dorsal third (DT) pR7], yR7 and DT yR7 expression. Data are mean±s.d., *n*=22 for pR7s, 16 for DT pR7s, 31 for yR7s and 31 for DT yR7s. ****$P<0.0001$, ns indicates $P>0.05$ and not significant (unpaired *t*-test with Welch's correction). All measurements were internally controlled within a single mid-pupal retina. (E) *yR7 enh* is strongly expressed in all R7s and weakly expressed in all PRs when Ss is ectopically expressed in all PRs. (F) Expression of *yR7 enh* is lost in *sal* mutants (white circle indicates presumptive R7). (G) *yR7 enh* is expressed in random R1s and R6s when Sal is ectopically expressed in all PRs. (H) *yR7 enh* is expressed in all PRs when Ss and Sal are ectopically expressed in all PRs (yellow circle indicates presumptive yR7). (I) The regulatory interactions governing *yR7 enh.* Otd, Ss and Sal activate *yR7 enh*, whereas Sal activates stochastic expression of Ss in yR7s (denoted by dashed arrow). (J) A truncated 0.8 kb fragment of *yR7 enh*, indicated by *yR7 enh**, was sufficient to recapitulate GFP expression in yR7 cells. Larger truncations encompassing *yR7 enh** also expressed GFP in yR7 cells, while truncations excluding *yR7 enh** did not drive GFP expression. *weak yR7 enh* shares a ∼250 bp overlap with *yR7 enh**, including one of the three Ss XRE binding sites (Fig. 3K). GMR42E10, a construct generated by Janelia that contains a fragment of *dve* driving Gal4, shares a ∼75 bp overlap with *yR7 enh* that does not contain any Ss XRE binding sites (Fig. 3K). This construct failed to drive GFP expression in yR7 cells. Light-green fragments drive strong GFP expression; dark-green fragments drive weak GFP expression; gray fragments do not drive GFP expression. (K) *yR7 enh** contains three Ss binding sites and two Otd/Dve binding sites. Capitalized black text indicates perfect conservation across six *Drosophila* species. Capitalized gray indicates conservation across five out of the six species. Light-green fragments drive strong GFP expression; dark-green fragments drive weak GFP expression; gray fragments do not drive GFP expression. (L) *yR7 enh** is expressed in yR7s, similar to Dve and *yR7 enh*. (M) Knocking out Ss XRE binding sites in the *yR7 enh** construct resulted in a near complete loss of GFP expression. BS KO, binding site knockout. (N) CRISPR-mediated deletion of *yR7 enh* from the endogenous *dve* locus resulted in loss of Dve specifically in yR7s. Dashed white circles indicate outer PRs and R8s; solid white circles indicate R7s. Red in ommatidium schematic indicates Dve expression. (O) Loss of Dve in yR7s resulted in derepression of Rh3 in adults. Yellow circles indicate yR7s; white circles indicate pR7s; black circles indicate no expression.

Development (2017) 144, 844-855 doi:10.1242/dev.144030

sites suggests that Otd may act through additional Otd-specific cryptic sites or that activation is mediated by another activator downstream of Otd. Mutation of these sites did not cause de-repression in outer PRs, suggesting that these sites mediate both repression by Dve and activation by Otd in outer PRs.

To test whether Dve directly binds the two Otd/Dve sites in *yR7 enh**, we conducted *in vitro* electrophoretic mobility shift assays (EMSAs). Dve bound sequences containing the Otd/Dve sites, and mutation of these sites dramatically decreased binding (Fig. 4Q), suggesting that Dve directly binds the two Otd/Dve sites in *yR7 enh** to repress expression.

As regulation of *yR7 enh** is dependent on Otd/Dve sites, Otd likely directly binds these sites to regulate expression. However, we cannot rule out possible indirect mechanisms.

outer enh is activated by Otd and repressed by Sal

We next characterized *outer enh*, the primary enhancer for Dve expression in outer PRs (Fig. 5A). The dve^{exel} deletion, which removes the first exon of *dve*, the *dve-A* promoter, and *yR7 enh*, showed no derepression of Dve target genes (Rh3, Rh5 and Rh6) in outer PRs (Fig. 2A, Fig. S1G), suggesting that *outer enh* is sufficient to drive Dve expression in outer PRs.

Otd activates Dve expression in all PRs, and Sal represses Dve expression in inner PRs (Johnston et al., 2011). *outer enh* expression was completely lost in *otd* mutants, consistent with a general requirement of Otd for *dve* expression (Fig. 5B). In *sal* mutants, *outer enh* was derepressed in inner PRs (Fig. 5C), suggesting that Sal represses this element in inner PRs. Ectopic expression of Ss in all PRs did not affect *outer enh* expression, consistent with

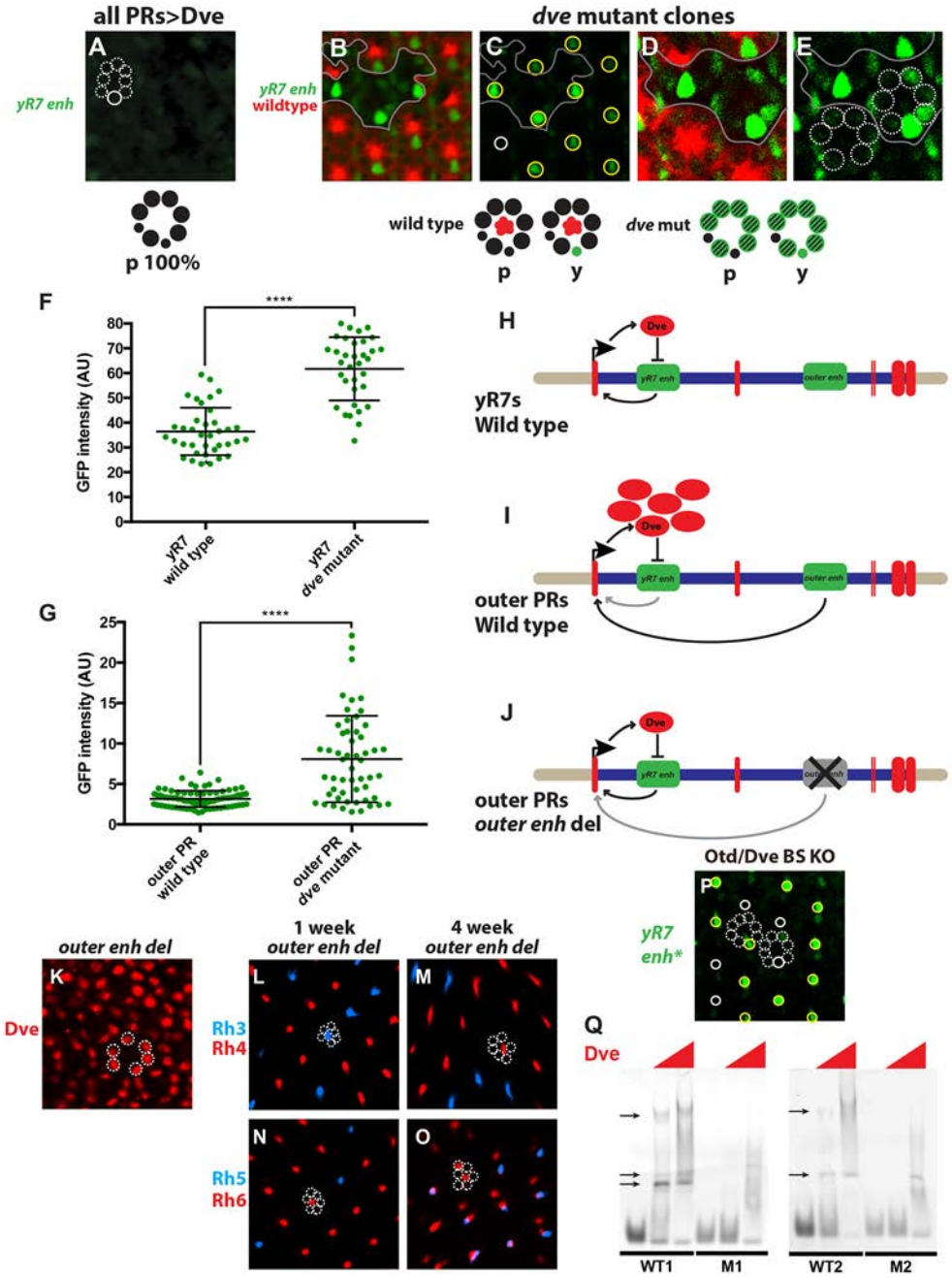


Fig. 4. See next page for legend.

Development (2017) 144, 844-855 doi:10.1242/dev.144030

Fig. 4. Dve feeds back to control *yR7 enh*. (A-E) Expression analysis was conducted on mid-pupal retinas. (A) Expression of *yR7 enh* is lost when Dve is ectopically expressed in all PRs. Dashed white circles indicate outer PRs and R8s; solid white circle indicates R7. In schematic, black circles indicate no GFP expression. (B-E) Yellow circles indicate yR7 cells; white circles indicate pR7 cells. Dashed white circles are outer PRs. Solid gray lines represent the boundary between *dve* mutant clones (indicated by the absence of RFP) and wild-type clones (indicated by the presence of RFP). Green in ommatidium schematic indicates strong GFP expression; crosshatching indicates variable expression; black indicates lack of expression; red spot indicates RFP expression. (B,C) In yR7s, *yR7 enh* is upregulated in *dve* mutant clones compared with wild-type clones. (D,E) In outer PRs, *yR7 enh* is upregulated in *dve* mutant clones compared with wild-type clones. (F) Quantification of yR7 GFP intensity in *dve* mutant and wild-type clones. yR7s in *dve* mutants show greater GFP intensity than in wild-type clones. R7 cells that are GFP positive indicate yR7s. n=37 for wild-type yR7s and n=37 for *dve* mutant yR7s. ****P<0.0001, unpaired *t*-test with Welch's correction. All measurements were internally controlled within a single mid-pupal retina. (G) Quantification of GFP intensity of outer PRs in *dve* mutant and wild-type clones. In wild-type clones, outer PRs are GFP off, whereas *dve* mutant clones show a much greater distribution of GFP expression states. n=84 for wild-type outer PRs and n=54 for *dve* mutant outer PRs. ****P<0.0001, unpaired *t*-test with Welch's correction. All measurements were internally controlled within a single mid-pupal retina. (H) *yR7 enh* induces Dve expression that negatively feeds back onto *yR7 enh* to maintain homeostatic Dve levels in yR7 cells. (I) *outer enh* induces Dve expression that negatively feeds back onto *yR7 enh* to completely repress *yR7* expression in outer PRs. (J) When *outer enh* function is impaired, *yR7 enh* is derepressed in outer PRs. (K) Dve remains expressed in outer PRs upon deletion of *outer enh*. Dashed white circles indicate outer PRs. (L,N) Expression of downstream Dve targets (Rh3, Rh5 and Rh6) is unaffected in *outer enh* deletion mutants in 1-week-old adults. Dashed white circles indicate outer PRs. (M,O) Variable derepression of Rh5 and Rh6 in outer PRs is observed in *outer enh* deletion mutants in 4-week-old adults. Expression of Rh3 is unaffected in *outer enh* deletion mutants in 4-week-old adults. Dashed white circles indicate outer PRs. (P) Knocking out Otd/Dve K50 binding sites resulted in an increased level of GFP in yR7s, suggesting that these sites mediate repression by Dve but not activation by Otd in yR7s. Solid yellow circles indicate yR7s that express GFP; solid white circles indicate pR7s that do not express GFP; dashed white circles indicate outer PRs and R8s. BS KO, binding site knockout. (Q) EMSAs illustrating that the binding of Dve is dependent on K50 Otd/Dve sites in *yR7 enh*. WT, wild-type sequence; M, mutation of K50 Otd/Dve site. Arrows indicate the bands shifted upon Dve binding. Multiple bands are observed likely due to the presence of multiple functional DNA binding domains within Dve (Johnston et al., 2011), yielding higher-order DNA/protein structures.

regulation of this element independent of Ss (Fig. 5D). Thus, combinatorial regulation involving activation by Otd in all PRs and repression by Sal in inner PRs yields the outer PR-specific expression of *outer enh* (Fig. 5E).

We truncated *outer enh* to a 1.3 kb fragment (*outer enh**) that recapitulated the expression of the entire *outer enh* fragment (Figs 2A, 5A,F,H, Fig. S3B). Two larger truncations and a Janelia Gal4 construct (GMR40E08) that encompass this 1.3 kb region also recapitulated expression, whereas fragments that exclude *outer enh** failed to drive GFP, consistent with the role of *outer enh** in driving outer PR-specific expression (Fig. 5F).

*outer enh** has four K50 homeodomain consensus sites (TAATCC) for Otd and Dve (Fig. 5G) (Chaney et al., 2005). *all PR enh 4* shares a 390 bp overlap with *outer enh**, including one of the Otd/Dve binding sites, suggesting that its weak expression in all PRs may be due to the single Otd/Dve binding site functioning independently of the repressive Sal input that regulates the entire *outer enh**.

We generated an *outer enh** construct that removes all TAATCC Otd/Dve binding sites by replacing them with AAAAAA (Fig. 5I). This construct showed a near complete loss of GFP expression in outer PRs, consistent with our model that Otd is required for *outer enh* activation. As regulation of *outer enh** is dependent on Otd/Dve sites, Otd likely directly binds these sites to regulate expression. However, we cannot rule out possible indirect mechanisms.

Although *outer enh** has four Otd/Dve sites, this enhancer contains no predicted Sal sites (Barrio et al., 1996; Sanchez et al., 2011), suggesting that Sal regulates *outer enh** either directly via binding to cryptic sites or indirectly through regulation of other intermediary factors. The longer *outer enh* contains a Sal binding site, which may contribute to regulation.

Feedback onto *outer enh* determines homeostatic levels

As *yR7 enh* is controlled by negative autoregulation, we next tested whether feedback also determines expression levels driven by *outer enh*. As *outer enh* (and Dve) are highly expressed in outer PRs, we expected that *dve* mutants may exhibit subtle increases in expression from *outer enh*. Indeed, in *dve* mutant clones, *outer enh* was expressed at higher levels in outer PRs compared with wild-type clones (Fig. 6B-D). To confirm negative feedback onto *outer enh*, Dve was ectopically expressed in all PRs at high levels (*all PRs>dve*), causing a complete loss of *outer enh* expression (Fig. 6A). Thus, Dve driven by *outer enh* feeds back onto this enhancer to autoregulate and ensure homeostatic levels of expression in outer PRs (Fig. 6E).

To test whether Dve directly binds the four Otd/Dve sites in *outer enh**, we conducted EMSAs. Dve bound sequences containing the Otd/Dve sites, and mutation of these sites dramatically decreased binding (Fig. 6F), suggesting that Dve directly binds the four Otd/Dve sites in *outer enh** to repress expression.

Sal represses *outer enh* to allow Ss-mediated activation of *yR7 enh*

yR7 enh is highly sensitive to levels of Dve feedback, particularly in outer PRs where Dve levels are high. Ss alone is sufficient to induce *yR7 enh* expression at high levels in all R7s but not outer PRs (Fig. 3E). Ss and Sal together are sufficient to induce *yR7 enh* at high levels in outer PRs (Fig. 3H). As Dve driven by *outer enh* feeds back to repress *yR7 enh* in outer PRs (Fig. 4D,E,I,J) and Sal represses Dve expression from *outer enh* (Fig. 5C), Sal may activate *yR7 enh* by repressing *outer enh*.

One prediction of this model is that ectopic Ss should be sufficient to activate *yR7 enh* at high levels in outer PRs in the absence of Dve. Indeed, when Ss is expressed at high levels in all PRs in *otd* mutants that lack Dve (Johnston et al., 2011), *yR7 enh* is activated in all PRs (Fig. S4A).

This result highlights two facets of *yR7 enh* regulation. First, Ss activates *yR7 enh*, whereas Sal represses *outer enh* to allow expression of *yR7 enh*, suggesting that Sal interacts indirectly with *yR7 enh* (Fig. S4B). Second, Ss requires Otd to activate *yR7 enh* in wild-type conditions (Fig. 3B) where Ss levels are low, whereas high levels of Ss are sufficient to override the requirement for Otd (Fig. S4A).

DISCUSSION

Dve is expressed in an intricate pattern with distinct levels in different photoreceptors. The regulation required to achieve this pattern is complex, involving two enhancers controlled by three main mechanisms: combinatorial transcription factor input, negative feedback and enhancer redundancy (Fig. 7). PR-specific Otd, inner PR-specific Sal and yR7-specific Ss work together to induce expression of *yR7 enh* in yR7s (Fig. 7A). By contrast, Otd activates *outer enh* whereas Sal represses this enhancer to yield Dve expression in outer photoreceptors (Fig. 7B).

Development (2017) 144, 844-855 doi:10.1242/dev.144030

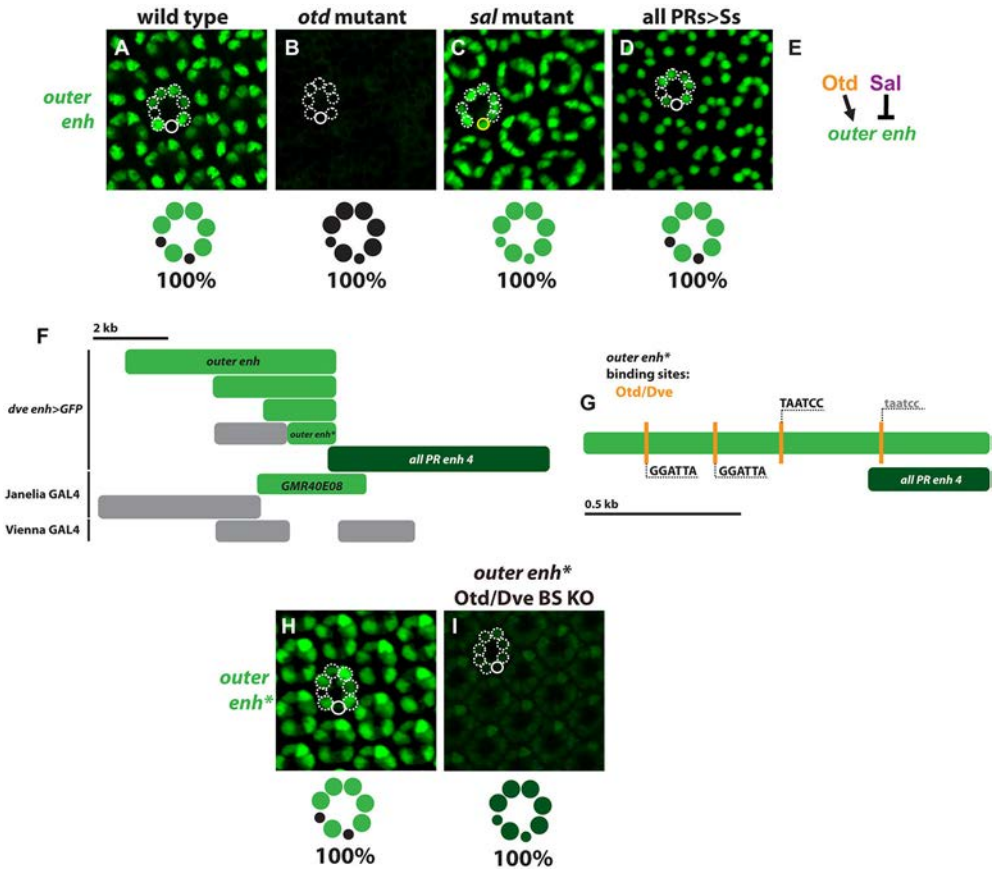


Fig. 5. *outer enh* is regulated by Otd and Sal. (A-D,H,I) Dashed white circles indicate outer PRs and R8s; solid white circles indicate R7s that do not express GFP; solid yellow circles indicate presumptive R7s expressing GFP. In schematics: light green circles indicate strong GFP expression; dark green circles indicate weak GFP expression; black circles indicate no GFP expression. Expression analysis was conducted on mid-pupal retinas. (A) *outer enh* is expressed in outer PRs. (B) Expression of *outer enh* is lost in *otd* mutants. (C) *outer enh* is expressed in all PRs in *sal* mutants. (D) Expression of *outer enh* is unaffected by ectopic expression of Ss in all PRs. (E) Otd activates *outer enh* and Sal represses *outer enh*. (F) A truncated 1.3 kb fragment of *outer enh*, denoted *outer enh**, was sufficient to recapitulate GFP expression in outer PRs. Larger fragments encompassing *outer enh** also expressed GFP in outer PRs, as did the Janelia reporter GMR40E04. Light green fragments drive strong GFP expression; dark green fragments drive weak GFP expression; gray fragments did not drive GFP expression. (G) *outer enh** contains three conserved Otd/Dve binding sites. Capitalized black text indicates perfect conservation across six *Drosophila* species. Lowercase gray text indicates that more than one species shows variation at the site. (H) *outer enh** is expressed in outer PRs, similar to Dve and *outer enh*. (I) Mutating Otd/Dve binding sites in *outer enh** resulted in a loss of expression of GFP. BS KO, binding site knockout.

Once these cell type-specific patterns are set, negative feedback by Dve maintains expression of the two enhancers at distinct levels important for regulation of downstream *rhodopsin* genes (Fig. 7C,D). This negative feedback appears especially crucial for the *yR7 enh*, the expression levels of which determine activation or repression of Rh3 in different regions of the retina. Gene regulatory network motifs involving negative feedback minimize variation in expression levels. With negative feedback, high concentrations of a regulator repress its expression, whereas low levels allow its activation. Negative feedback thus ensures homeostatic levels of expression (Alon, 2007; Becskei and Serrano, 2000; Irvine et al., 1993; Stewart et al., 2013).

As an additional layer of regulation, *outer enh* drives high levels of Dve that repress *yR7 enh* in outer PRs (Fig. 7E). When *outer enh* function is lost, *yR7 enh* becomes active in outer PRs, functioning as a shadow enhancer to provide redundancy and robustness to expression (Fig. 7F). Complex multi-enhancer systems enable genes to integrate multiple regulatory inputs, yielding intricate expression patterns. Although some enhancers account for distinct aspects of regulation, others drive overlapping patterns. Shadow enhancers can compensate for removal of a primary enhancer, resulting in mostly unaltered gene expression (Hong et al., 2008; Miller et al., 2014; Nolte et al., 2013; Perry et al., 2012). These shadow enhancers provide reliability and robustness in pattern formation, allowing crucial patterning genes to be buffered against environmental and genetic variation (Barolo, 2012; Bothma et al., 2015; Frankel et al., 2010; Perry et al., 2010).

We define *yR7 enh* as a dark shadow enhancer, as it is normally repressed in outer PRs but becomes active when the function of the primary enhancer is impaired. We were able to identify the *yR7 enh* dark shadow enhancer because we were characterizing how a complex pattern was controlled by combinatorial transcription factor input and feedback acting on two enhancers. Similar to the generality of shadow enhancers (Cannavo et al., 2016), dark shadow enhancers may be a common mechanism to ensure gene expression. However, they would be challenging to identify as they are active only upon genetic or possibly environmental perturbation.

Dve is a transcriptional repressor (Johnston et al., 2011) that acts directly on *yR7 enh* in outer PRs to repress expression (Fig. 4Q). Generally, transcriptional repressors would likely act directly on dark shadow enhancers to repress them, poising them as backup systems. For transcriptional activators, more complex indirect mechanisms would be required. For example, the primary enhancer could induce the activator to activate expression of a

Development (2017) 144, 844-855 doi:10.1242/dev.144030

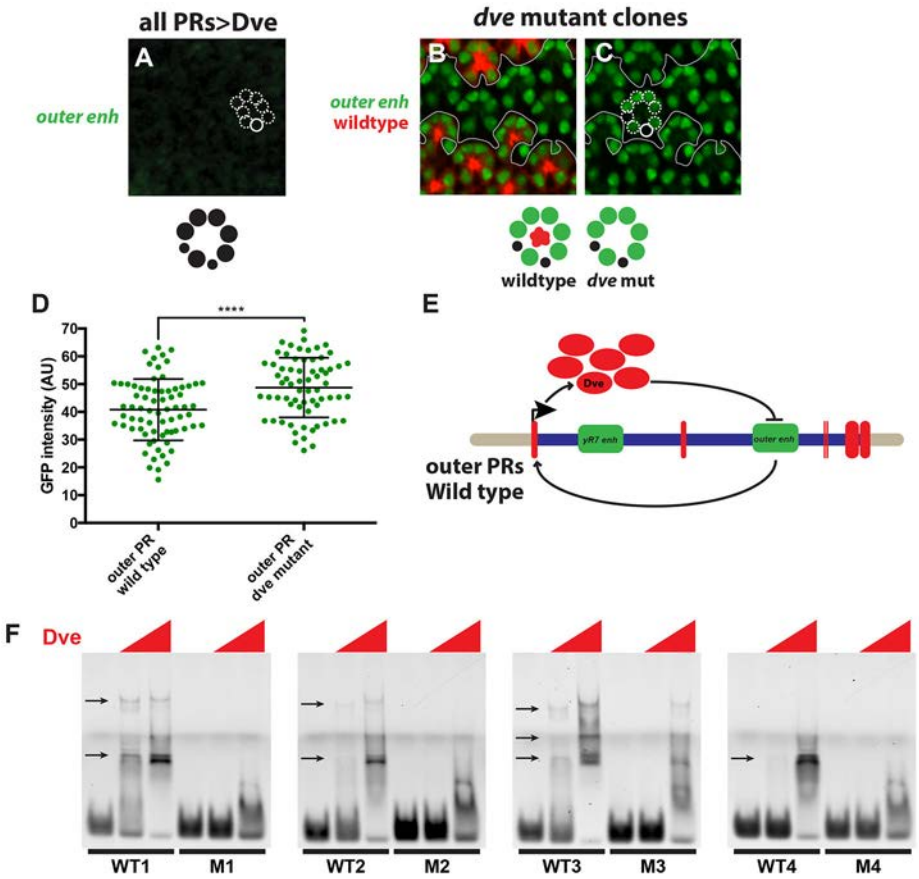


Fig. 6. Dve feeds back to control *outer enh*. (A-C) Dashed white circles indicate outer PRs and R8s; solid white circles indicate R7s. Expression analysis was conducted on mid-pupal retinas. In schematics: green circles indicate GFP expression; black circles indicate no GFP expression; red spots indicate RFP expression. (B,C) Solid gray line represents boundary between *dve* mutant clones (indicated by absence of RFP) and wild-type clones (indicated by presence of RFP). (A) Expression of *outer enh* is lost when Dve is ectopically expressed in all PRs. (B,C) Autoregulatory feedback: in outer PRs, *outer enh* is upregulated in *dve* mutant clones compared with wild-type clones. (D) Quantification of outer PR GFP expression of *outer enh* in *dve* mutant clones compared with wild-type clones. n=72 for wild-type outer PRs; n=67 for *dve* mutant outer PRs. ****P<0.0001, unpaired *t*-test with Welch's correction. All measurements were internally controlled within a single mid-pupal retina. (E) *outer enh* induces Dve expression that negatively feeds back onto *outer enh* to maintain homeostatic levels in outer PRs. (F) EMSAs illustrating that the binding of Dve is dependent on K50 Otd/Dve sites in *outer enh*. WT, wild-type sequence; M, mutation of K50 Otd/Dve site. Arrows indicate the bands shifted upon Dve binding. Multiple bands are observed likely due to the presence of multiple functional DNA binding domains within Dve (Johnston et al., 2011), yielding higher-order DNA/protein structures.

transcriptional repressor, which in turn could repress the dark shadow enhancer. As dark shadow enhancers require feedback, they would likely only be found in genes encoding regulatory factors.

A key aspect of regulation by primary enhancers and dark shadow enhancers is their differential responsiveness to repression. For *outer enh*, normal Dve levels induce a slight decrease in expression. However, for *yR7 enh*, these same levels completely turn off

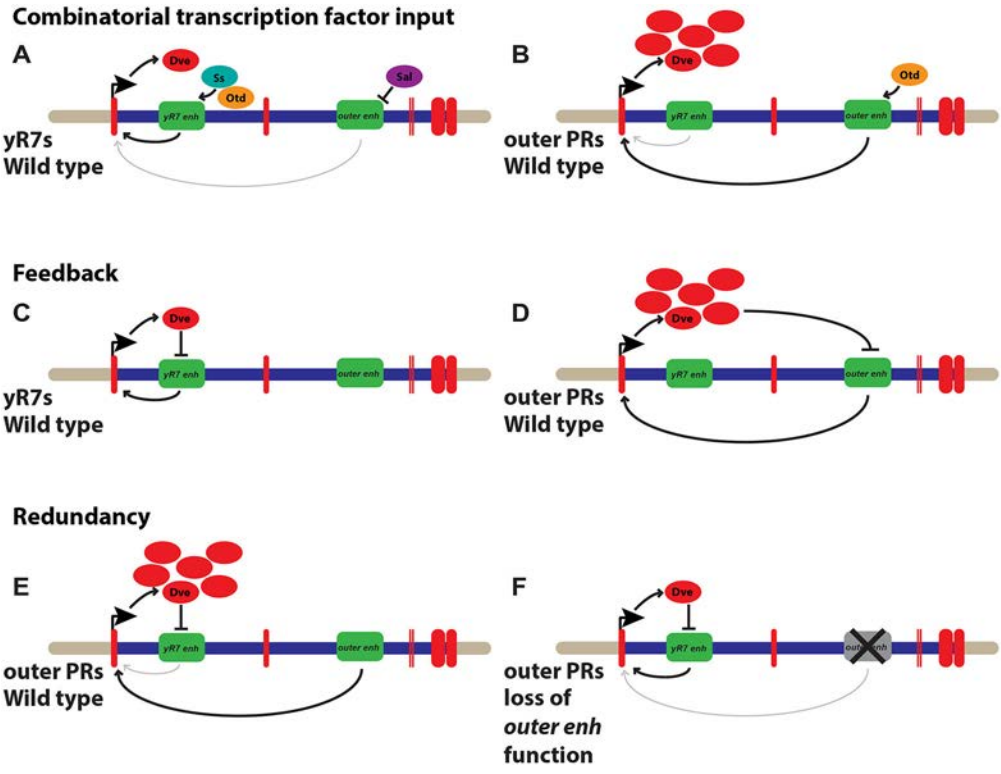


Fig. 7. Combinatorial transcription factor input, feedback and redundancy govern *dve* expression. (A) In yR7 cells, *yR7 enh* is activated by Otd and Ss, while *outer enh* is repressed by Sal. (B) In outer PRs, Otd activates *outer enh*. (C) In yR7 cells, *yR7 enh* induces Dve expression that negatively feeds back onto *yR7 enh* to maintain homeostatic levels. (D) In outer PRs, *outer enh* induces Dve expression that negatively feeds back onto *outer enh* to maintain homeostatic levels. (E) In wild-type outer PRs, *outer enh* induces Dve expression that negatively feeds back onto *yR7 enh* to completely repress expression. (F) Upon loss of *outer enh* function, *yR7 enh* is derepressed and drives expression in outer PRs.

Development (2017) 144, 844-855 doi:10.1242/dev.144030

expression in outer PRs. The difference may lie in activation by Otd: *outer enh* contains four Otd/Dve sites, whereas *yR7 enh* contains two (Figs 3K, 5G). As these sites mediate both activation by Otd and repression by Dve, cooperative action by the four sites in *outer enh* may drive stronger expression and prevent repression. Generally, the primary enhancer is expressed and must be significantly less susceptible to repression than the dark shadow enhancer, which is off.

Expression of Dve in outer PRs is seen in the mosquitos *Anopheles gambiae* and *Aedes aegypti* (Johnston et al., 2011), suggesting a conserved role in Rh regulation that has been maintained over 250 million years of evolution. However, expression of Rhs in R7s of mosquito species is regionalized in contrast to the stochastic pattern in *Drosophila* (Hu and Castelli-Gair, 1999), suggesting that different mechanisms have arisen to regulate Dve and Rh expression in R7s. Dark shadow enhancers may be an ancestral mechanism to ensure gene expression despite evolutionary changes. Furthermore, they may allow the evolution of new functions such as the expression of *yR7 enh* in R7s.

Dark shadow enhancers appear to provide robustness to gene expression and may act as additional mechanisms of canalization (i.e. the ability for individuals in a population to produce similar phenotypes regardless of environmental or genetic perturbation) (Waddington, 1942). Buffering of gene expression occurs at the levels of *cis*-regulatory logic (Dunipace et al., 2013; Frankel et al., 2010; Hong et al., 2008; Staller et al., 2015; Wunderlich et al., 2015) and gene networks (Cassidy et al., 2013; Lott et al., 2007; Manu et al., 2009). Dark shadow enhancers are an interesting integration of these mechanisms, whereby a primary enhancer induces expression of a factor that feeds back to repress a dark shadow enhancer. When expression from the primary enhancer is perturbed, this feedback is broken and the dark shadow enhancer becomes active. Thus, dark shadow enhancers are poised as backup mechanisms for proper gene regulation. As our understanding of complex multi-enhancer systems increases, it will be interesting to see the generality of dark shadow enhancers.

In conclusion, our studies show how two enhancers integrate combinatorial transcription factor input, negative autoregulation and redundancy in *cis*-regulatory elements to determine robust levels of gene expression in photoreceptor neurons. These mechanisms likely play roles in the establishment and maintenance of gene expression levels in other neuronal subtypes.

MATERIALS AND METHODS

Generating *dve enh>GFP* constructs

Fragments (3-6kb; Fig. 2A) were cloned into GFP reporter constructs and injected into flies. Transgenic flies were isolated and stocks were generated (see supplementary Materials and Methods and Table S1).

Drosophila strains

Flies were raised on standard cornmeal medium and grown at room temperature (25°C) (see supplementary Materials and Methods and Tables S2-S4 for complete descriptions of *Drosophila* genotypes).

CRISPR-generated deletions

dve-B promoter, *outer enh* and *yR7 enh* deletions were generated using CRISPR (see supplementary Materials and Methods and Table S5 for further details).

Otd/Dve binding site knockout

Otd/Dve binding site knockouts for *outer enh* and *yR7 enh* were generated using site-directed mutagenesis (see supplementary Materials and Methods for further details).

Electrophoretic mobility shift assay

Binding assays were performed as described previously (Johnston et al., 2011; Li-Kroeger et al., 2008) (see supplementary Materials and Methods for further details).

Antibodies

Antibodies and dilutions used were as follows: mouse anti-prospero (1:10, DSHB), rat anti-Elav (1:50, DSHB), sheep anti-GFP (1:500, Bio-Rad, 4745-1051), mouse anti-Rh3 (1:100; a gift from S. Britt, University of Colorado, Boulder, CO, USA), rabbit anti-Rh4 (1:100; a gift from C. Zuker, Columbia University, New York, USA), mouse anti-Rh5 (1:2000; Tahayato et al., 2003), rabbit anti-Rh6 (1:2000; Tahayato et al., 2003), guinea pig anti-Ss (1:200; a gift from Y. N. Jan, University of California, San Francisco, CA, USA) and rabbit anti-Dve (1:500; Nakagoshi et al., 1998). All secondary antibodies were Alexa-conjugated (1:400; Molecular Probes).

Retina dissection and immunohistochemistry

Retinas were dissected and stained as described previously (Hsiao et al., 2012) (see supplementary Materials and Methods for further details)

Quantification

Fluorescence intensity of nuclear GFP expression of single retinas was quantified using the ImageJ processing program. A small region in the center of each nucleus was selected for fluorescence intensity measurement. Images were taken under subsaturating conditions and comparisons of GFP intensity were drawn between cells of the same retina. Column scatterplots were generated using Graphpad Prism.

Acknowledgements

We are grateful to Steve Britt, Lily Jan, Yuh-Nung Jan, Hideki Nakagoshi, Charles Zuker, the Bloomington Stock Center and the Vienna *Drosophila* Resource Center (VDRC) for generously providing published fly stocks and antibodies. We thank Rejji Kuruvilla, Mike Levine, David Lorberbaum, Richard Mann and Alexandra Neuhaus-Follini for helpful comments on the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: J.Y., G.G., R.J.J.; Methodology: J.Y., S.T., G.G., R.J.J.; Investigation: J.Y., C.A., K.V., S.T., G.G.; Writing – original draft: J.Y., R.J.J.; Writing – review and editing: J.Y., R.J.J., S.S.; Funding acquisition: S.S., R.J.J.; Supervision: R.J.J.

Funding

R.J.J. and S.S. were supported by the National Institute of General Medical Sciences (R01GM106090-01A1/NYU F7784-01). R.J.J. was supported by a Pew Scholar Award from Pew Charitable Trusts (00027373), by a March of Dimes Foundation Basil O'Connor Scholar Award (5-FY15-21) and by the National Eye Institute (R01EY025598). Deposited in PMC for release after 12 months.

Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.144030.supplemental

References

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450-461.

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

Barrio, R., Shea, M. J., Carulli, J., Lipkow, K., Gaul, U., Frommer, G., Schuh, R., Jäckle, H. and Kafatos, F. C. (1996). The spalt-related gene of Drosophila melanogaster is a member of an ancient gene family, defined by the adjacent, region-specific homeotic gene spalt. *Dev. Genes Evol.* **206**, 315-325.

Barrio, R., de Celis, J. F., Bolshakov, S. and Kafatos, F. C. (1999). Identification of regulatory regions driving the expression of the Drosophila spalt complex at different developmental stages. *Dev. Biol.* **215**, 33-47.

Becskei, A. and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593.

Bell, M. L., Earl, J. B. and Britt, S. G. (2007). Two types of Drosophila R7 photoreceptor cells are arranged randomly: a model for stochastic cell-fate determination. *J. Comp. Neurol.* **502**, 75-85.

Development (2017) 144, 844-855 doi:10.1242/dev.144030

Bothma, J. P., Garcia, H. G., Esposito, E., Schlissel, G., Gregor, T. and Levine, M. (2014). Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. *Proc. Natl. Acad. Sci. USA* **111**, 10598-10603.

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

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

Cannavò, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., Ciglar, L., Korbel, J. O. and Furlong, E. E. M. (2016). Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr. Biol.* **26**, 38-51.

Cassidy, J. J., Jha, A. R., Posadas, D. M., Giri, R., Venken, K. J. T., Ji, J., Jiang, H., Bellen, H. J., White, K. P. and Carthew, R. W. (2013). miR-9a minimizes the phenotypic impact of genomic diversity by buffering a transcription factor. *Cell* **155**, 1556-1567.

Chaney, B. A., Clark-Baldwin, K., Dave, V., Ma, J. and Rance, M. (2005). Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome. *Biochemistry* **44**, 7497-7511.

Chen, S.-K., Badea, T. C. and Hattar, S. (2011). Photoentrainment and pupillary light reflex are mediated by distinct populations of ipRGCs. *Nature* **476**, 92-95.

Chou, W.-H., Hall, K. J., Wilson, D. B., Wideman, C. L., Townson, S. M., Chadwell, L. V. and Britt, S. G. (1996). Identification of a novel Drosophila opsin reveals specific patterning of the R7 and R8 photoreceptor cells. *Neuron* **17**, 1101-1115.

Corty, M. M., Tam, J. and Grueber, W. B. (2016). Dendritic diversification through transcription factor-mediated suppression of alternative morphologies. *Development* **143**, 1351-1362.

Dasen, J. S., De Camilli, A., Wang, B., Tucker, P. W. and Jessell, T. M., (2008). Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1. *Cell* **134**, 304-316.

Duncan, D. M., Burgess, E. A. and Duncan, I. (1998). Control of distal antennal identity and tarsal development in Drosophila by spineless-aristapedia, a homolog of the mammalian dioxin receptor. *Genes Dev.* **12**, 1290-1303.

Dunipace, L., Saunders, A., Ashe, H. L. and Stathopoulos, A. (2013). Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. *Dev. Cell* **26**, 536-543.

Fortini, M. E. and Rubin, G. M. (1990). Analysis of cis-acting requirements of the Rh3 and Rh4 genes reveals a bipartite organization to rhodopsin promoters in Drosophila melanogaster. *Genes Dev.* **4**, 444-463.

Franceschini, N., Kirschfeld, K. and Minke, B. (1981). Fluorescence of photoreceptor cells observed in vivo. *Science* **213**, 1264-1267.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F. and Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490-493.

Gao, S., Takemura, S. Y., Ting, C.-Y., Huang, S., Lu, Z., Luan, H., Rister, J., Thum, A. S., Yang, M., Hong, S.-T. et al. (2008). The neural substrate of spectral preference in Drosophila. *Neuron* **60**, 328-342.

Grueber, W. B., Jan, L. Y. and Jan, Y. N. (2003). Different levels of the homeodomain protein cut regulate distinct dendrite branching patterns of Drosophila multidendritic neurons. *Cell* **112**, 805-818.

Hardie, R. C. (1985). Functional organization of the fly retina. In (ed. H. Autrum, D. Ottoson, E. R. Perl, R. F. Schmidt, H. Shimazu, W. D. Willis). *Progress in Sensory Physiology*. Berlin: Springer.

Hong, J.-W., Hendrix, D. A. and Levine, M. S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314.

Hsiao, H. Y., Johnston, R. J., Jukam, D., Vasiliauskas, D., Desplan, C. and Rister, J. (2012). Dissection and immunohistochemistry of larval, pupal and adult Drosophila retinas. *J Vis Exp*, e4347.

Hsiao, H.-Y., Jukam, D., Johnston, R. and Desplan, C. (2013). The neuronal transcription factor erect wing regulates specification and maintenance of Drosophila R8 photoreceptor subtypes. *Dev. Biol.* **381**, 482-490.

Hu, N. and Castelli-Gair, J. (1999). Study of the posterior spiracles of Drosophila as a model to understand the genetic and cellular mechanisms controlling morphogenesis. *Dev. Biol.* **214**, 197-210.

Irvine, K. D., Botas, J., Jha, S., Mann, R. S. and Hogness, D. S. (1993). Negative autoregulation by Ultrabithorax controls the level and pattern of its expression. *Development* **117**, 387-399.

Johnston, R. J., Jr. (2013). Lessons about terminal differentiation from the specification of color-detecting photoreceptors in the Drosophila retina. *Ann. N. Y. Acad. Sci.* **1293**, 33-44.

Johnston, R. J., Jr and Desplan, C., (2010). Stochastic mechanisms of cell fate specification that yield random or robust outcomes. *Annu. Rev. Cell Dev. Biol.* **26**, 689-719.

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

Johnston, R. J., Jr, Otake, Y., Sood, P., Vogt, N., Behnia, R., Vasiliauskas, D., McDonald, E., Xie, B., Koenig, S., Wolf, R. et al. (2011). Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the Drosophila eye. *Cell* **145**, 956-968.

Jukam, D. and Desplan, C. (2011). Binary regulation of Hippo pathway by Merlin/NF2, Kibra, Lgl, and Melted specifies and maintains postmitotic neuronal fate. *Dev. Cell* **21**, 874-887.

Jukam, D., Xie, B., Rister, J., Terrell, D., Charlton-Perkins, M., Pistillo, D., Gebelein, B., Desplan, C. and Cook, T. (2013). Opposite feedbacks in the Hippo pathway for growth control and neural fate. *Science* **342**, 1238016.

Jukam, D., Viets, K., Anderson, C., Zhou, C., DeFord, P., Yan, J., Cao, J. and Johnston, R. J., Jr. (2016). The insulator protein BEAF-32 is required for Hippo pathway activity in the terminal differentiation of neuronal subtypes. *Development* **143**, 2389-2397.

Kuhnlein, R. P. and Schuh, R. (1996). Dual function of the region-specific homeotic gene spalt during Drosophila tracheal system development. *Development* **122**, 2215-2223.

Li-Kroeger, D., Witt, L. M., Grimes, H. L., Cook, T. A. and Gebelein, B. (2008). Hox and senseless antagonism functions as a molecular switch to regulate EGF secretion in the Drosophila PNS. *Dev. Cell* **15**, 298-308.

Lott, S. E., Kreitman, M., Palsson, A., Alekseeva, E. and Ludwig, M. Z. (2007). Canalization of segmentation and its evolution in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10926-10931.

Manu, M., Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M. et al. (2009). Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. *PLoS Biol.* **7**, e1000049.

Mazzoni, E. O., Celik, A., Wernet, M. F., Vasiliauskas, D., Johnston, R. J., Cook, T. A., Pichaud, F. and Desplan, C. (2008). Iroquois complex genes induce co-expression of rhodopsins in Drosophila. *PLoS Biol.* **6**, e97.

Mikeladze-Dvali, T., Wernet, M. F., Pistillo, D., Mazzoni, E. O., Teleman, A. A., Chen, Y.-W., Cohen, S. and Desplan, C. (2005). The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in Drosophila R8 photoreceptors. *Cell* **122**, 775-787.

Miller, S. W., Rebeiz, M., Atanasov, J. E. and Posakony, J. W. (2014). Neural precursor-specific expression of multiple Drosophila genes is driven by dual enhancer modules with overlapping function. *Proc. Natl. Acad. Sci. USA* **111**, 17194-17199.

Mollereau, B., Dominguez, M., Webel, R., Colley, N. J., Keung, B., de Celis, J. F. and Desplan, C. (2001). Two-step process for photoreceptor formation in Drosophila. *Nature* **412**, 911-913.

Nakagawa, Y., Fujiwara-Fukuta, S., Yorimitsu, T., Tanaka, S., Minami, R., Shimooka, L. and Nakagoshi, H. (2011). Spatial and temporal requirement of defective proventriculus activity during Drosophila midgut development. *Mech. Dev.* **128**, 258-267.

Nakagoshi, H., Hoshi, M., Nabeshima, Y.-i. and Matsuzaki, F. (1998). A novel homeobox gene mediates the Dpp signal to establish functional specificity within target cells. *Genes Dev.* **12**, 2724-2734.

Nolte, C., Jinks, T., Wang, X., Martinez Pastor, M. T. and Krumlauf, R., (2013). Shadow enhancers flanking the HoxB cluster direct dynamic Hox expression in early heart and endoderm development. *Dev. Biol.* **383**, 158-173.

O'Gorman, S., Fox, D. T. and Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351-1355.

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

Perry, M. W., Bothma, J. P., Luu, R. D. and Levine, M. (2012). Precision of hunchback expression in the Drosophila embryo. *Curr. Biol.* **22**, 2247-2252.

Sánchez, J., Talamillo, A., González, M., Sánchez-Pulido, L., Jiménez, S., Pirone, L., Sutherland, J. D. and Barrio, R. (2011). Drosophila Sal and Salr are transcriptional repressors. *Biochem. J.* **438**, 437-445.

Smith, C. J., O'Brien, T., Chatzigeorgiou, M., Spencer, W. C., Feingold-Link, E., Husson, S. J., Hori, S., Mitani, S., Gottschalk, A., Schafer, W. R. et al. III, (2013). Sensory neuron fates are distinguished by a transcriptional switch that regulates dendrite branch stabilization. *Neuron* **79**, 266-280.

Sood, P., Johnston, R. J., Jr and Kussell, E. (2012). Stochastic de-repression of Rhodopsins in single photoreceptors of the fly retina. *PLoS Comput. Biol.* **8**, e1002357.

Staller, M. V., Vincent, B. J., Bragdon, M. D. J., Lydiard-Martin, T., Wunderlich, Z., Estrada, J. and DePace, A. H. (2015). Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 785-790.

Stewart, A. J., Seymour, R. M., Pomiankowski, A. and Reuter, M. (2013). Under-dominance constrains the evolution of negative autoregulation in diploids. *PLoS Comput. Biol.* **9**, e1002992.

Tahayato, A., Sonneville, R., Pichaud, F., Wernet, M. F., Papatsenko, D., Beaufils, P., Cook, T. and Desplan, C. (2003). Otd/Crx, a dual regulator for the specification of ommatidia subtypes in the Drosophila retina. *Dev. Cell* **5**, 391-402.

Development (2017) 144, 844-855 doi:10.1242/dev.144030

Terriente, J., Perea, D., Suzanne, M. and Diaz-Benjumea, F. J. (2008). The Drosophila gene zfh2 is required to establish proximal-distal domains in the wing disc. *Dev. Biol.* **320**, 102-112.

Thanawala, S. U., Rister, J., Goldberg, G. W., Zuskov, A., Olesnicky, E. C., Flowers, J. M., Jukam, D., Purugganan, M. D., Gavis, E. R., Desplan, C. et al., Jr. (2013). Regional modulation of a stochastically expressed factor determines photoreceptor subtypes in the Drosophila retina. *Dev. Cell* **25**, 93-105.

Vandendries, E. R., Johnson, D. and Reinke, R. (1996). orthodenticle is required for photoreceptor cell development in the Drosophila eye. *Dev. Biol.* **173**, 243-255.

Viets, K., Eldred, K. C. and Johnston, R. J., Jr. (2016). Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates. *Trends Genet.* **32**, 638-659.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characteristics. *Nature* **150**, 563-565.

Wernet, M. F., Labhart, T., Baumann, F., Mazzoni, E. O., Pichaud, F. and Desplan, C. (2003). Homothorax switches function of Drosophila photoreceptors from color to polarized light sensors. *Cell* **115**, 267-279.

Wernet, M. F., Mazzoni, E. O., Çelik, A., Duncan, D. M., Duncan, I. and Desplan, C. (2006). Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174-180.

Wolff, T. and Ready, D. F. (1993). Pattern formation in the Drosophila retina. In *The Development of Drosophila Melanogaster. II*. (ed. M. Bate and A. Martinez-Arias), pp. 1277-1316. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

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

Yamaguchi, S., Desplan, C. and Heisenberg, M. (2010). Contribution of photoreceptor subtypes to spectral wavelength preference in Drosophila. *Proc. Natl. Acad. Sci. USA* **107**, 5634-5639.

Zhu, L. J., Christensen, R. G., Kazemian, M., Hull, C. J., Enuameh, M. S., Basciotta, M. D., Brasefield, J. A., Zhu, C., Asriyan, Y., Lapointe, D. S. et al. (2011). FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* **39** Suppl. 1, D111- D117.

Supplemental Figures

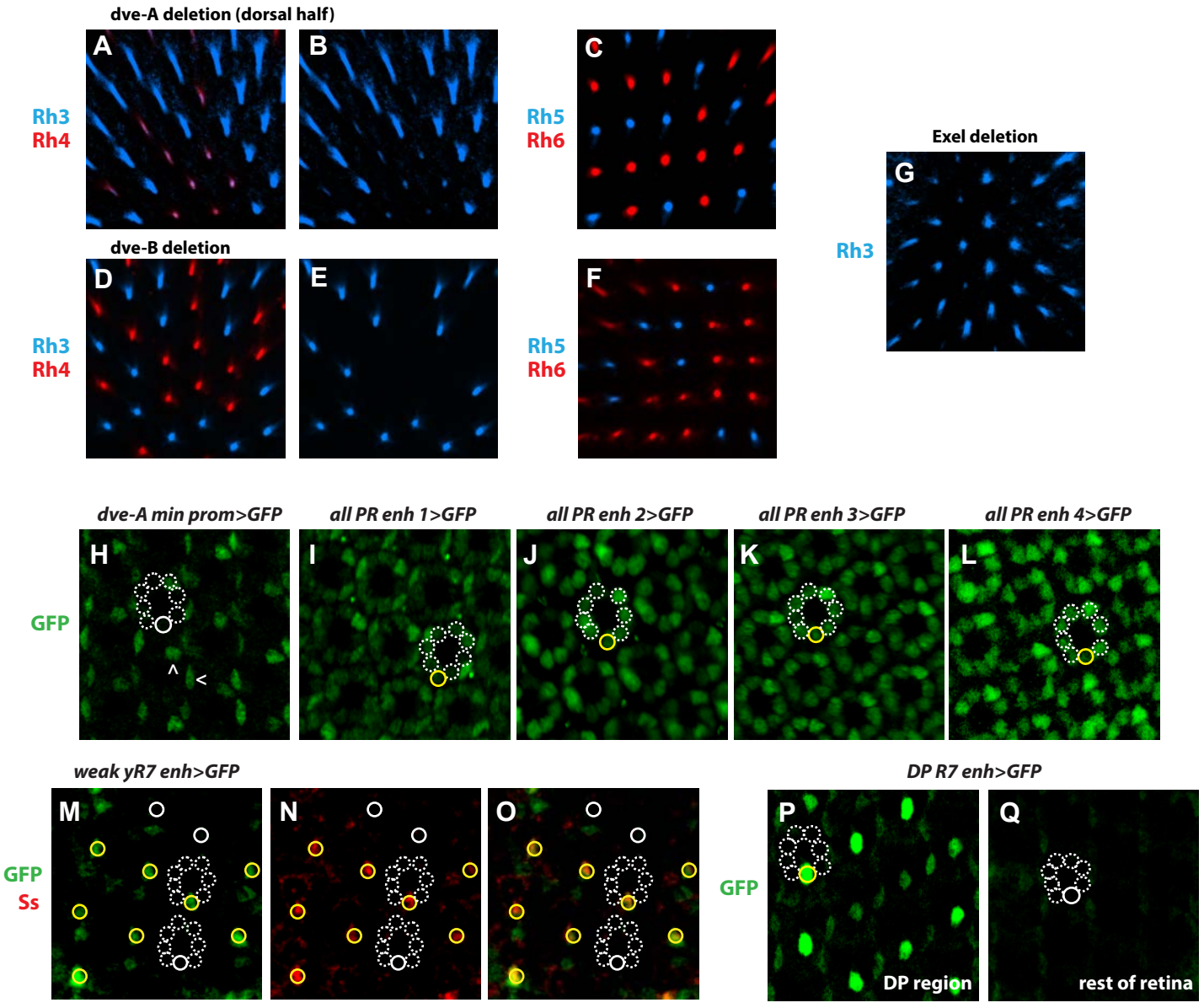


Figure S1. ***dve enh>GFP*** **constructs and *dve-A* and *dve-B* promoter deletions**

A-C. *dve-A* promoter deletion resulted in upregulation of Rh3 in all R7s in the dorsal half of the retina. Dve targets Rh5 and Rh6 displayed wild type expression.

D-F. *dve-B* promoter deletion displayed wild type expression of Dve targets (Rh3, Rh5, and Rh6).

G. dve^{exel} deletion, covering ***y****R7 enh* and the *dve-A* promoter*,* resulted in Rh3 upregulation in R7s.

H. *dve-A min prom* drove weak expression in R4s (denoted by ^) and pigment cells (denoted by <). Dashed white circles represent outer PRs and R8s; solid white circle indicates R7.

I-L. Four enhancers (*all PR enh 1-4*) drove weak GFP expression in all PRs. Dashed white circles represent PRs. Yellow circles denote GFP-expressing R7s.

M-O. *weak* ***y****R7 enh* displayed weak GFP expression in **y**R7s; Ss is a marker for **y**R7s. Dashed white circles represent outer PRs and R8s, solid white circles denote **p**R7s, and solid yellow circles denote **y**R7s.

P-Q. *dorsal R7 enh* drove expression in R7s in the dorsal posterior (DP) region of the retina. Dashed white circles represent PRs. Yellow circles indicate GFP-expressing R7; solid white circle indicates non-expressing R7.

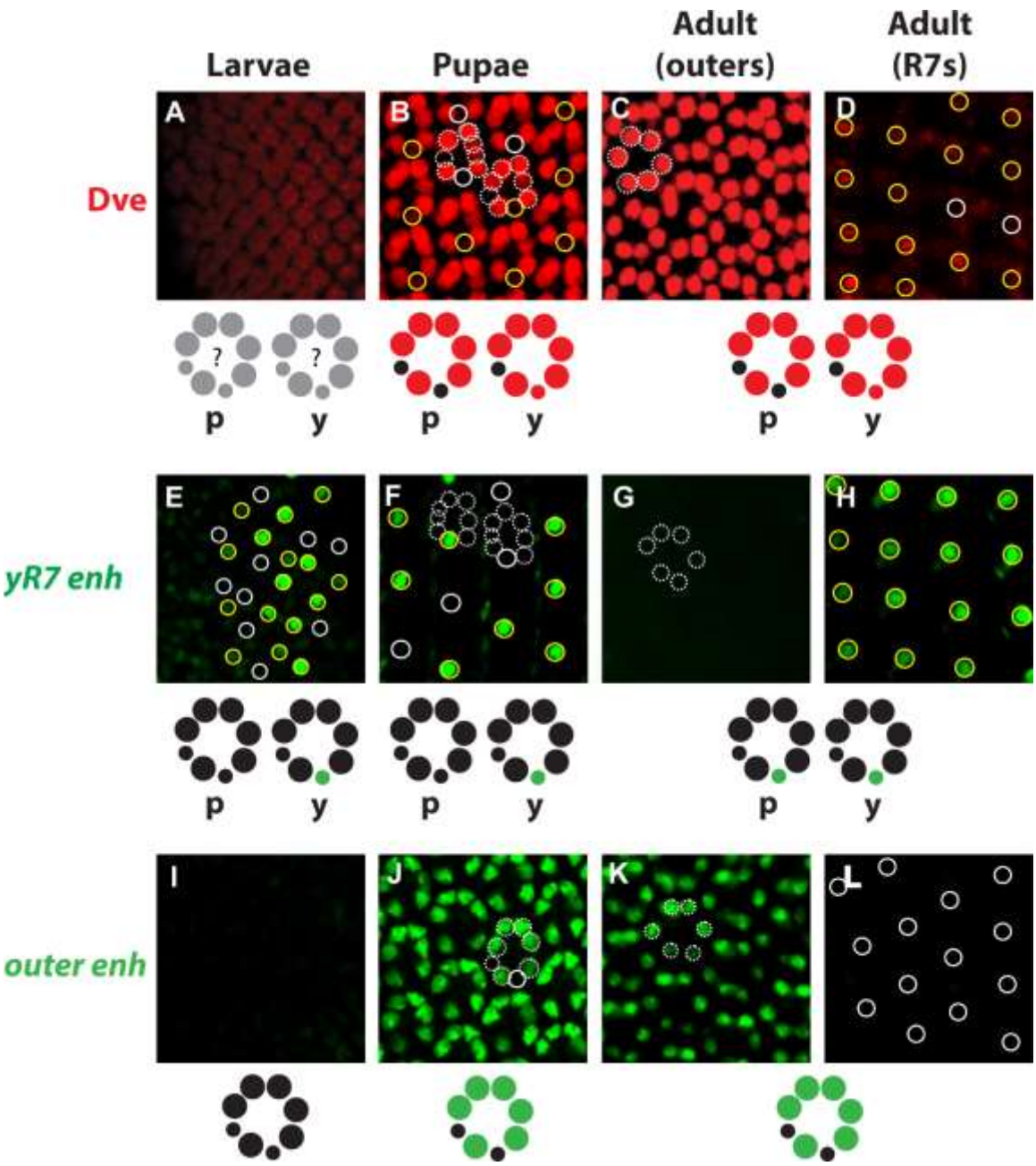


Figure S2. **Differential expression of Dve throughout development**

For A-L, in schematics, red circles indicate endogenous Dve expression, green circles indicate GFP expression, black circles indicate no expression (endogenous Dve or GFP), and gray circles indicate indeterminate expression.

For B-D, yellow circles indicate **y**R7 cells. Solid white circles indicate **p**R7 cells. Dashed white circles are outer PRs and R8s.

For E-H, yellow circles indicate **y**R7 cells. Solid white circles indicate **p**R7 cells. Dashed white circles are outer PRs and R8s.

For J-L, dashed white circles represent outer PRs; solid white circles indicate R7s.

A. Antibody staining for Dve is nonspecific in larvae.

B-D. Dve is highly expressed in outer PRs and weakly expressed in **y**R7s in pupae and adults.

E-F. ***y****R7 enh* is expressed in **y**R7s in larvae and pupae.

G-H. ***y****R7 enh* is expressed in all R7 cells in the adult but is not expressed in outer photoreceptors.

I. *outer enh* is not expressed in larvae.

J-L. *outer enh* drives expression in outer PRs in pupae and adults, but is not expressed in R7s.

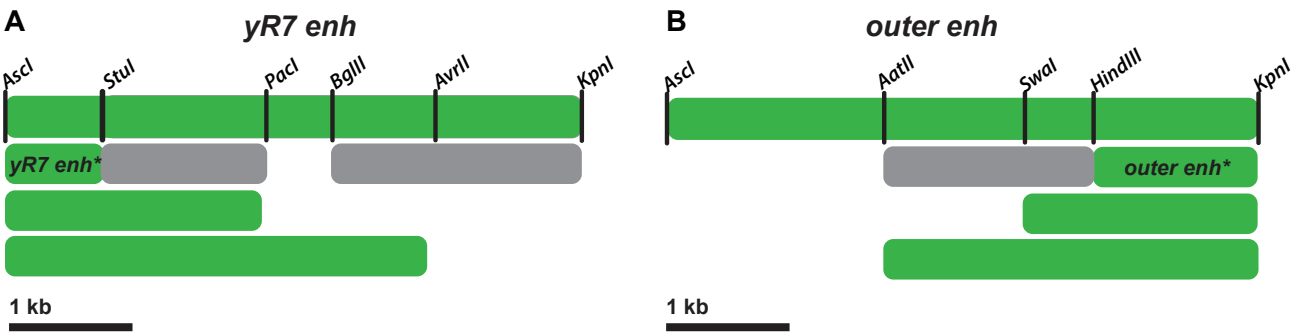


Figure S3. Restriction enzyme sites used to generate *dve enh* truncations

A. Schematic of the ***y****R7 enh* and the restriction enzyme sites used to generate the ***y****R7 enh** and other truncations.

B. Schematic of the outer enh and the restriction enzyme sites used to generate the *outer enh** and other truncations.

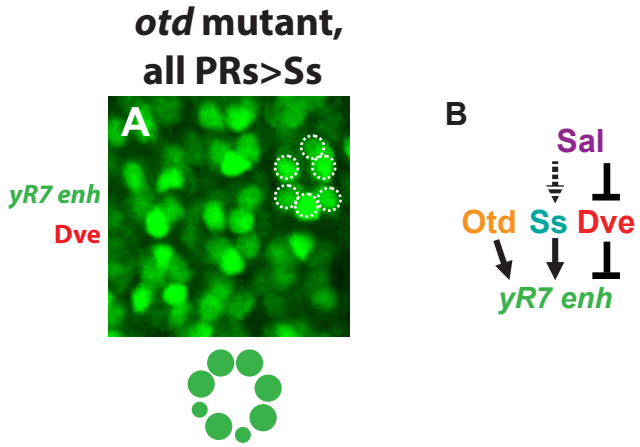


Figure S4. Sal activates yR7 enh by repressing Dve

A. In *otd* mutants, in which Dve is not expressed, ectopic Ss induces ***y****R7 enh* in all PRs.

B. The regulatory interactions governing ***y****R7 enh.* Otd and Ss activate ***y****R7 enh*, while Dve represses ***y****R7 enh*. Sal activates stochastic expression of Ss (denoted by dashed arrow) in **y**R7s and represses Dve.

Supplementary Materials and Methods

Generating *dve enh>GFP* constructs

3-6kb fragments (Fig. 2A) were amplified using DNA isolated from yw^{67} flies and ligated into the pGEM-T easy vector. The 699 bp *dve* minimal promoter was subcloned into a pJR16 nGFPcDNA reporter vector containing a w^{+} marker, generating the pGG14 vector. Other fragments of the *dve* locus were subcloned into pGG14 before microinjection into fly embryos. Constructs were then integrated into the genome via the attP/B system, and injected flies were crossed with a balancer stock with the genotype *y⁻w⁻; +/+; Tm2/Tm6b*. Red-eyed offspring were isolated, and transgenes were balanced over Tm6b. Primers and restriction enzymes used to generate each *dve* enhancer construct are shown in Table S1.

Enhancer constructs ***y**R7 enh>GFP* and *outer enh>GFP* were further truncated using restriction enzymes and blunt end ligation (Fig. S3A,B). Reporter vectors were microinjected into fly embryos, and transgenic lines were established using the same methods as above.

Drosophila strains

Flies were raised on standard cornmeal medium and grown at room temperature (25°C). Transgenic lines used include *dve* enhancer gene constructs generated for this project, as well as the reagents in Table S2.

The GAL4-UAS system was used to ectopically express Sal, Ss, and Dve (Brand and Perrimon, 1993), while the FLP-FRT system was used to create *sal* and *dve* mutant phenotypes (O'Gorman et al., 1991). Shortened and complete genotypes of flies examined are found in Table S3.

Janelia and VDRC Stock Centers generated transgenic lines that express GAL4 driven by flanking non-coding or intronic regions of various genes. GAL4 lines

associated with *dve **y**R7* and *dve outer* enhancer elements were crossed with UAS-nlsGFP. See Table S4.

CRISPR-generated deletions

dve-B promoter, *outer enh*, and ***y**R7 enh* deletions were generated using CRISPR. We designed four gRNAs per deletion, two flanking either side of the deletion. Forward and reverse strands of gRNAs were designed and annealed together to have BbsI restriction site overhangs. gRNAs were then cloned into the pCDF3 cloning vector. Single stranded homologous bridges were generated with 80 bp homologous regions flanking each side of the deletion. An AscI restriction cut site was incorporated into the homologous bridge to facilitate screening. For every deletion, all four gRNAs were injected into *Drosophila* embryos at 125ng/ul each for a total of 500ng, along with 100ng/ul of homologous bridge oligos. Single adult males were then crossed with balancer stocks (yw ; if / cyo ; +), and the progeny were screened for the deletion via PCR. Homologous bridges, gRNAs, and PCR screening primers are shown in Table S5.

Otd/Dve binding site knockout

Otd/Dve binding site knockouts for *outer enh* and ***y**R7 enh* were generated using site-directed mutagenesis, where the K50 homeodomain consensus sites (TAATCC) were replaced with AAAAAA. Constructs were integrated into the genome and *Drosophila* strains were established using the same methods as described above.

Electrophoretic Mobility Shift Assay

Binding assays were performed as previously described (Johnston et al., 2011; Li-Kroeger et al., 2008).

Probes tested were as follows (bold/underline indicates K50 Otd/Dve site or mutated K50 Otd/Dve site):

***y**R7 enh* 1 5’ –CGTGTTAGCCAAACC**TAATCC**AGGCTAAACGAGGG- 3’
***y**R7 enh* 2 5’ –AAATACGCTTATGTC**GGATTA**TCCCATAATTTATG- 3’
***y**R7 enh* mutant 1 5’ –CGTGTTAGCCAAACC**TGCGCC**AGGCTAAACGAGGG- 3’
***y**R7 enh* mutant 2 5’ –AAATACGCTTATGTC**GGCGCA**TCCCATAATTTATG- 3’
outer enh 1 5’ –AGCAAACAACAAAAA**GGATTA**AGTCCAAGACACAC- 3’
outer enh 2 5’ –ATACTTATTTCATTA**GGATTA**TTTTTGACTAACAT - 3’
outer enh 3 5’ –TCACGGCATTAATTA**TAATCC**GCTTAAAAGTTTCA - 3’
outer enh 4 5’ –TCACACAAGGATTCG**TAATCC**TTGCGAGGGACCCA- 3’
outer enh mutant 1 5’ –AGCAAACAACAAAAA**GGCGCA**AGTCCAAGACACAC- 3’
outer enh mutant 2 5’ –ATACTTATTTCATTA**GGCGCA**TTTTTGACTAACAT- 3’
outer enh mutant 3 5’ –TCACGGCATTAATTA**TGCGCC**GCTTAAAAGTTTCA- 3’
outer enh mutant 4 5’ –TCACACAAGGATTCG**TGCGCC**TTGCGAGGGACCCA- 3’

Antibodies

Antibodies and dilutions used were as follows: mouse anti-prospero (1:10)(DSHB), rat anti-Elav (1:50)(DSHB), sheep anti-GFP (1:500), mouse anti-Rh3 (1:100)(gift from S. Britt, University of Colorado), rabbit anti-Rh4 (1:100)(gift from C. Zuker, Columbia University), mouse anti-Rh5 (1:2000)(Tahayato et al., 2003), rabbit anti-Rh6 (1:2000)(Tahayato et al., 2003), guinea pig anti-Ss (1:200)(Gift from Y.N. Jan, University of California, San Francisco), rabbit anti-Dve (1:500)(Nakagoshi et al., 1998). All secondary antibodies were Alexa-conjugated (1:400) (Molecular Probes).

Retina dissection and immunohistochemistry

Retinas were dissected and stained as described previously (Hsiao et al., 2012). Larvae were collected and dissected in ice cold PBS (1x), and retinas were isolated using forceps before fixing for 20 minutes in 4% formaldehyde at RT. Samples were washed three times with PBX and kept in primary antibodies diluted in PBX overnight at 4^{o}C. After three washes with PBX, secondary antibodies diluted in PBX were added, and samples were kept at RT for at least 2 hours. After three more washes, samples were kept in PBX at room temperature overnight, before being mounted flat in Vectashield (Vector Laboratories).

To facilitate pupae collection at the desired mid-pupae stage, flies were raised at 25^{o}C in a 12hr light/12hr dark cycle incubator. Pupae heads were dissected in ice cold PBS (1x) and eye-brain complexes were extracted via pipetting. Fixing, antibody staining and mounting procedures were consistent with those of larvae, but pupal retinas were not isolated from the brain until prior to mounting.

Adult flies were anesthetized on CO_2 pads before their heads were removed using forceps. Fly heads were dissected in ice cold PBS (1x), and retinas were isolated using forceps. Fixing and antibody staining procedures were consistent with those of larvae and pupae, although laminas were not removed until after fixing. Retinas were then mounted using SlowFade Gold Reagent (ThermoFisher Scientific).

For all stages of fly development, samples stained with antibodies were visualized under a Zeiss LSM 700 confocal microscope.

Quantification

Fluorescence intensity of nuclear GFP expression of single retinas was quantified using the ImageJ processing program. A small region in the center of each nucleus was selected for fluorescence intensity measurement. Images were taken under subsaturating conditions and comparisons of GFP intensity were drawn between cells of the same retina. Column scatterplots were generated using Graphpad Prism.

Table S1. *dve enh>GFP* Constructs and Primers

Construct	Primers	Restriction sites
***y**R7 enh*	agtcggcgcgcccacaaccatttcactcctgc agtcggtacccttctcccagtcttcgaatg	*AscI* *KpnI*
outer enh	agtcggcgcgcctcatcctcatccctacctac agtcggtaccacaactgcctttgccttgtg	*AscI* *KpnI*
***y**R7 enh extended to right*	agtcggcgcgccgcctagctaccgtgatcaac agtcggtaccgtttagctcgattacgcttc	*AscI* *KpnI*
dve min promotor	agtcagatcttgatctggctctctggactc agtcggatccgtgggaaagtgttggtaagc	*BglII* *BamHI*
*weak **y**R7 enh*	agtcggcgcgcccggtcagcaggtgagttgag agtcggtacctacgatgacaccgataagcg	*AscI* *KpnI*
dorsal R7 enh	agtcggcgcgcccataatcacaacacgagtcgg agtcggtaccgatggtggcttaactcaatc	*AscI* *KpnI*
all PR enh 1	agtcggcgcgccgcttatctgcggctttgtgg agtcggtaccctcgtcttgtcccgattcca	*AscI* *KpnI*
all PR enh 2	agtcggcgcgccgctagcgcatagagcatagatg agtcggtaccgttgctggcaccaatacacg	*AscI* *KpnI*
all PR enh 3	agtcggcgcgccgtgctgcctacaagtttgga agtcggtaccgccttctgaagactagcac	*AscI* *KpnI*
all PR enh 4	agtcggcgcgcccgaaactcctcgactcacac agtcggtaccccaattcgctgattg	*AscI* *KpnI*
dve enh 1	agtcggcgcgcccactgacatcaattcaccgtc agtcggtaccaggagaaaggagtgagttcg	*AscI* *KpnI*
dve enh 2	agtcggcgcgccccatcccttagagagctttg agtcggtacctgtatctggggaatcggatg	*AscI* *KpnI*
dve enh 3	agtcggcgcgccgccacaatgtcaagcatcaaag agtcggtacccactttcccacagtatcatcttg	*AscI* *KpnI*
dve enh 4	agtcggcgcgcccagagctgaactgaacaatc agtcggtacccgtgtctctgcgctttgtga	*AscI* *KpnI*
dve enh 5	agtcggcgcgccgtcttagtgcagctactgtt agtcggtaccgaaggcttacgaaactaatg	*AscI* *KpnI*
dve enh 6	agtcggcgcgcccagctcgttaagcataagca agtcggtacccgttcccaaattaccccatc	*AscI* *KpnI*
dve enh 7	agtcggcgcgcctggtggtggcgattcatttg agtcggtaccctaccacaaaactagagcaccc	*AscI* *KpnI*

Table S2. *Drosophila* reagent descriptions

Name	Description	Source
otd^{uvi}	hypomorphic allele, fails to produce protein product in the eye	(Vandendries et al., 1996)
FRT40 sal$^{Df(2L)32FP5}$	a deficiency that removes the *sal* gene	(Barrio et al., 1999)
FRT42d dve^{186}	Dve protein null mutant	(Terriente et al., 2008)
FRT40 GMR-hid, cL	eye-specific enhancer driving hid, an activator of apoptosis	Bloomington
FRT42d GMR-RFP	eye-specific enhancer driving RFP	Bloomington
ey-FLP	eye-specific enhancer driving flippase	Bloomington
ss$^{d115.7}$	Ss protein null mutant	(Duncan et al., 1998)
ss$^{Df(3R)Exel7330}$	Deficiency covering the *ss* locus	Bloomington
UAS-Sal	UAS enhancer driving Sal	(Kuhnlein and Schuh, 1996), (Wernet et al., 2003)
UAS-Dve	UAS enhancer driving Dve	(Nakagoshi et al., 1998)
UAS-Ss	UAS enhancer driving Ss	(Duncan et al., 1998)
lGMR-Gal4	eye-specific enhancer driving Gal4	Bloomington
dve^{Exel}	deletion that removes the first exon of *dve, dve-A* promoter, and *dve* ***y****R7* enhancer element	See footnote*
dve^{E181}	deletion allele for the *dve-A* promoter	(Nakagawa et al., 2011)
UAS-nlsGFP	UAS driving nuclear GFP	Bloomington

*We generated the dve^{exel} deletion by using hsFLP-mediated recombination between two FRT sites, inserted by P-elements P(XP)d05100 and P(XP)d08355.

Table S3. *Drosophila* shortened and complete genotypes

Shortened	Complete genotype	Figure
dve enh>GFP	*yw ; + ; dve enh>GFP*	2B-E, 3A, 5A, S1H-Q, S2E-L
all PRs>Ss	*yw ; lGMR>Gal4, UAS>Ss ; dve enh>GFP*	3E, 5D
all PRs>Sal	*yw, UAS>sal ; lGMR>Gal4 ; dve enh>GFP*	3G
all PRs>Dve	*yw ; lGMR>Gal4 ; dve enh>GFP/ UAS>dve*	4A, 6A
sal mutant	*yw ;sal$^{Df(2L)32FP5}$FRT40 GMR>hid FRT40 ; dve enh>GFP/ey>Flp*	3F, 5C
otd mutant	*otd^{uvi} ; + ; dve enh>GFP*	3B, 5B
dve mutant	*ey>FLP ;FRT42d dve^{186}/ FRT42d GMR>RFP ; dve enh>GFP/ +*	4B-E, 6B-C
ss mutant	*yw ; + ; dve enh>GFP, ss^{BL7985def}/ ss$^{d115.7}$*	3C
all PRs>Ss and Sal	*yw, UAS>sal ; lGMR>GAL4, UAS>ss ; dve enh>GFP*	3H
otd mutant, all PRs>Ss	*otd^{uvi} ; lGMR>GAL4, UAS>ss ; dve enh>GFP*	S4A
dve-A del	*yw ; FRT42d* dve^{186} / dve^{E181} *; +*	S1A-C

Table S4. Janelia and VDRC Gal4 stock numbers

Janelia GMR			
49373	49927	48655	46230
45702	50133	45682	46238
45284	46241	50066	48150
41238			
VDRC Stock			
020724	020725	020737	020739

Table S5. Primers for CRISPR

dve-B del	
Homologous bridge	ttttatggatcgcttggcattataatgaacagcggcgtcgccggctggccatgggcgcatggcgc gcccatgggagcaagttggagctgggcaagcccccacatcccatccgcccactgacctaagc c
dveBgRNA1 F	gtcgctggccatgggcgcataat
dveBgRNA1 R	aaacattatgcgcccatggccagc
dveBgRNA2 F	gtcgggataagtacggtgcatgg
dveBgRNA2 R	aaacccatgcaccgtacttatccc
dveBgRNA3 F	gtcgtcatccttccagtgcccat
dveBgRNA3 R	aaacatgggcactggaaggatgac
dveBgRNA4 F	gtcggtgtctgccactgttgaac
dveBgRNA4 R	aaacgttcaacagtggcagacacc
DveBscr F	gctgttgggagattaagttt
DveBscr R	tgccttctgaagactagcac
outer enh del	
Homologous bridge	gctgcctgggcgtccttttctcgggcacttgatagaatttgacaaattgaaaatccttttggcgcgcc gaagcctacttaagtcccttgaaatccttgagatttttgcactggtcaagcaatgataa
outergRNA1 F	gtcggacaaccgctcgccacaaa
outergRNA1 R	aaactttgtggcgagcggttgtcc
outergRNA2 F	gtcgttcaagagtccaggcgacc
outergRNA2 R	aaacggtcgcctggactcttgaac
outergRNA3 F	gtcgaaattaagcaatagtctta
outergRNA3 R	aaactaagactattgcttaatttc
outergRNA4 F	gtcggacttaagtaggcttccca
outergRNA4 R	aaactgggaagcctacttaagtcc
outer-scr F	ccagtgattatgtatggttc
outer-scr R	gagtgatttgggtatttagg
yR7 enh del	
Homologous bridge	acttgctccccgtccgtcgatcgattcaaattaccagcgatttattggcgatcgccagccggcgcg ccgctatggcaatgcaaacaggtgagggtgaattacttgtcctagacaactttgcagtcagc
yR7gRNA1 F	gtcgttgatcacggtagctaggc
yR7gRNA1 R	aaacgcctagctaccgtgatcaac
yR7gRNA2 F	gtcgtgttcgcataacgctggtc
yR7gRNA2 R	aaacgaccagcgttatgcgaacac
yR7gRNA3 F	gtcgtttagctcgattacgcttc
yR7gRNA3 R	aaacgaagcgtaatcgagctaaac
yR7gRNA4 F	gtcgtttgcattgccatagctac
yR7gRNA4 R	aaacgtagctatggcaatgcaaac
yR7-scr F	gatggctaatggcgagagga
yR7-scr R	gcaatcttggcactcccgtt

Appendix 3: The insulator protein BEAF-32 is required for Hippo pathway activity in the terminal differentiation of neuronal subtypes

© 2016. Published by The Company of Biologists Ltd | Development (2016) 143, 2389-2397 doi:10.1242/dev.134700

RESEARCH ARTICLE

The insulator protein BEAF-32 is required for Hippo pathway activity in the terminal differentiation of neuronal subtypes

David Jukam[1,*], Kayla Viets[2], Caitlin Anderson[2], Cyrus Zhou[2], Peter DeFord[2], Jenny Yan[2,‡], Jinshuai Cao[1] and Robert J. Johnston, Jr[2,§]

ABSTRACT

The Hippo pathway is crucial for not only normal growth and apoptosis but also cell fate specification during development. What controls Hippo pathway activity during cell fate specification is incompletely understood. In this article, we identify the insulator protein BEAF-32 as a regulator of Hippo pathway activity in *Drosophila* photoreceptor differentiation. Though morphologically uniform, the fly eye is composed of two subtypes of R8 photoreceptor neurons defined by expression of light-detecting Rhodopsin proteins. In one R8 subtype, active Hippo signaling induces Rhodopsin 6 (Rh6) and represses Rhodopsin 5 (Rh5), whereas in the other subtype, inactive Hippo signaling induces Rh5 and represses Rh6. The activity state of the Hippo pathway in R8 cells is determined by the expression of *warts*, a core pathway kinase, which interacts with the growth regulator *melted* in a double-negative feedback loop. We show that *BEAF-32* is required for expression of *warts* and repression of *melted*. Furthermore, *BEAF-32* plays a second role downstream of Warts to induce Rh6 and prevent Rh5 fate. *BEAF-32* is dispensable for Warts feedback, indicating that *BEAF-32* differentially regulates *warts* and Rhodopsins. Loss of *BEAF-32* does not noticeably impair the functions of the Hippo pathway in eye growth regulation. Our study identifies a context-specific regulator of Hippo pathway activity in post-mitotic neuronal fate, and reveals a developmentally specific role for a broadly expressed insulator protein.

KEY WORDS: Color vision, Photoreceptor, Cell fate, Insulator, *Drosophila* retina, RNAi screen, Hippo pathway, Regulatory networks, Warts tumor suppressor, Rhodopsin

INTRODUCTION

The Hippo signaling pathway is a crucial regulator of growth and apoptosis in organ size control (Irvine and Harvey, 2015; Yu et al., 2015; Zhao et al., 2011). However, proliferation-independent roles for the pathway during animal development have also been discovered. The identification of the Hippo pathway as a regulator of R8 photoreceptor subtype specification in *Drosophila* was among the first examples of a mitosis-independent role for the Hippo pathway in determining cell fate (Mikeladze-Dvali et al., 2005). More recently, the pathway has been shown to regulate dendritic field tiling in neurons (Emoto et al., 2006), cell differentiation in pre-implantation embryos (Cockburn et al., 2013; Nishioka et al., 2009), neuroblast differentiation upon cell cycle exit (Reddy et al., 2010), and hematopoiesis (Milton et al., 2014), among others. Because R8 photoreceptors are post-mitotic neurons and are not competent to divide, they are an excellent system in which to elucidate context-specific mechanisms of Hippo pathway function (Hsiao et al., 2013; Jukam and Desplan, 2011; Jukam et al., 2013). How the Hippo pathway is regulated differently in division and differentiation is incompletely understood. Here, we describe the insulator protein BEAF-32 as a regulator of Hippo pathway activity in cell fate specification in the developing *Drosophila* retina.

The fly eye is composed of ∼800 ommatidia (unit eyes); each ommatidium contains eight photoreceptors named R1-R8 (Hardie, 1985). The outer photoreceptors, R1-R6, express the broad spectrum-detecting Rhodopsin 1 (Rh1; also known as NinaE) and function in motion detection (Heisenberg and Buchner, 1977; Yamaguchi et al., 2008; Wardill et al., 2012). The inner photoreceptors, R7 and R8, are specialized for color vision, with some contribution from R1-R6 (Schnaitmann et al., 2013). Though morphologically uniform, the fly eye is composed of two main types of ommatidia defined by expression of color-sensing Rhodopsins (Rhs) in the inner photoreceptors (Rister et al., 2013). In the 'pale' (p) subtype, pR7s express Rhodopsin 3 (Rh3) and pR8s express Rhodopsin 5 (Rh5) (Fig. 1A). In the 'yellow' (y) subtype, yR7s express Rhodopsin 4 (Rh4) and yR8s express Rhodopsin 6 (Rh6) (Fig. 1B). The ommatidial subtypes are stochastically distributed throughout the eye in a p:y ratio of ∼35:65 (Fig. 1C,F).

The specification of ommatidial subtypes is determined in R7s by the stochastic ON/OFF expression of the PAS-bHLH transcription factor Spineless (Ss) (Johnston and Desplan, 2014; Wernet et al., 2006). The ON/OFF state of Ss determines R8 subtype fate through an inductive signal (Chou et al., 1996; Papatsenko et al., 1997) that results in mutually exclusive R8 expression of the Hippo pathway kinase Warts (Wts) and the growth regulator Melted (Melt). In pR7s lacking Ss, Rh3 is expressed in R7s and a signal from R7s triggers activation of *melt* and repression of *wts*, leading to Rh5 expression in pR8s (Fig. 1A) (Mikeladze-Dvali et al., 2005). In yR7s expressing Ss, Rh4 is expressed and the signal is repressed, causing the default state of *melt* repression and *wts* activation leading to Rh6 expression (Fig. 1B). The double-negative feedback loop between *wts* and *melt* controls the presence or absence of Wts downstream of the constitutively active upstream Hippo pathway (Jukam and Desplan, 2011). Wts negatively regulates Yorkie (Yki), which acts with a network of photoreceptor-specific transcription factors

[1]Center for Developmental Genetics, Department of Biology, New York University, 100 Washington Square East, New York, NY 10003-6688, USA. [2]Department of Biology, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218-2685, USA.
*Present address: Department of Biology, Stanford University, Stanford, CA 94305, USA. ‡Present address: Division of Medical Sciences, Harvard Medical School, 260 Longwood Avenue, Boston, MA 02115, USA.

§Author for correspondence (robertjohnston@jhu.edu)

R.J.J., 0000-0003-4419-0646

Received 30 December 2015; Accepted 11 May 2016

Development (2016) 143, 2389-2397 doi:10.1242/dev.134700

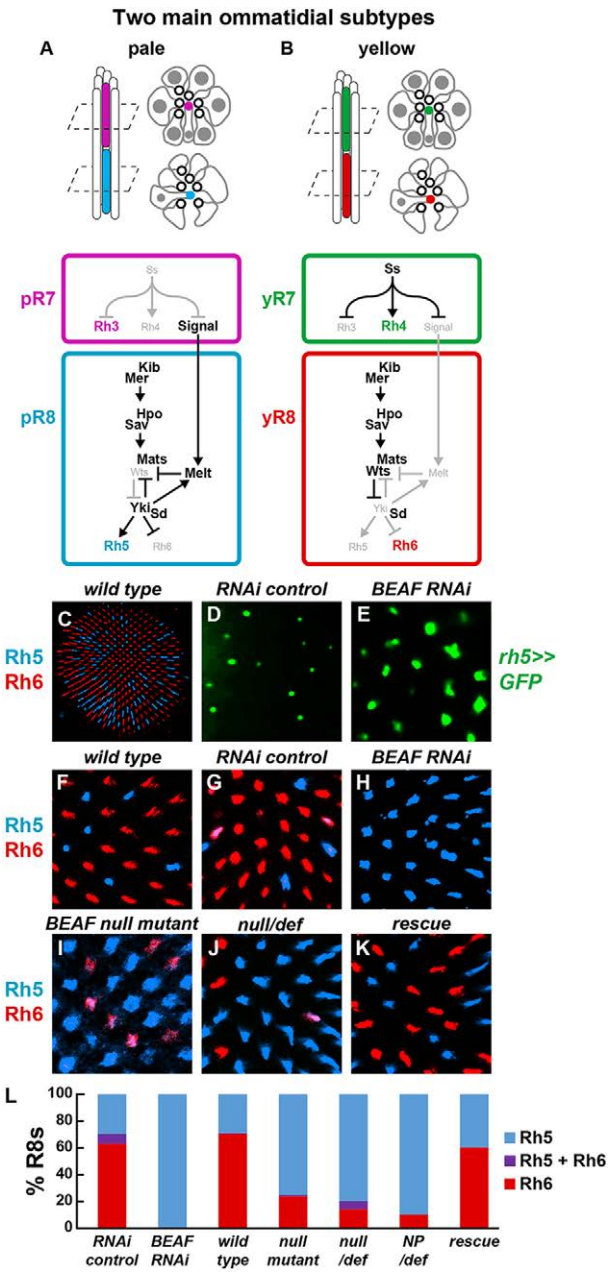


Fig. 1. *BEAF* is required for yR8 subtype specification.
(A,B) Schematic showing eight photoreceptors and a cross-section of their rhabdomeres, the membranous structures containing Rhodopsin (Rh) proteins, within an ommatidium. Gray indicates cell bodies and nuclei. White circles with black outlines indicate outer photoreceptor rhabdomeres. Colored rhabdomeres indicate R7 (top) and R8 (bottom). Below is the regulatory network controlling Rh expression in R7 (top) and R8 (bottom). (A) Pale 'p' ommatidial subtype. (B) Yellow 'y' ommatidial subtype. (C) Retina showing Rh5 and Rh6 expression in stochastic and mutually exclusive R8 subsets. R8 subtypes are visualized by Rh5 (pR8, blue) and Rh6 (yR8, red) antibodies in all panels unless otherwise noted. (D) *rh5≫GFP* was expressed in a subset of R8s in RNAi controls. Visualized by water immersion (see Materials and Methods). (E) *rh5≫GFP* was expressed in most R8s when *BEAF* was knocked down by RNAi. (F,G) Rh5 and Rh6 were expressed in subsets of R8s in wild-type (F) or RNAi Gal4 control (G) retinas. (H-J) Most R8s contained Rh5, and few contained Rh6, in retinas expressing *BEAF* RNAi (H) or homozygous mutant for $BEAF^{AB\text{-}KO}$ (I) or $BEAF^{AB\text{-}KO}$ over a deficiency covering the *BEAF* locus (J). (K) A *BEAF* genomic fragment restored normal Rh5 and Rh6 expression in *BEAF* homozygous null mutants. (L) Quantification of phenotypes.

to transduce Hippo pathway output into expression of Rh5 or Rh6 (Jukam et al., 2013).

Here, we identify the insulator protein BEAF-32 as a regulator of Wts and Hippo pathway activity in R8 subtype specification. *BEAF-32* is required for the expression of *wts* and also functions downstream of Wts to regulate Rhodopsins, but does not noticeably affect growth. Finally, we demonstrate that *BEAF-32* is differentially required for Hippo pathway positive feedback and Rh expression. The role of BEAF-32 in post-mitotic determination of photoreceptor subtypes suggests that insulators have highly specific functions in development.

RESULTS AND DISCUSSION

BEAF-32 is a regulator of the Hippo pathway controlling Rhodopsin expression

To identify transcription factors that regulate the Hippo pathway and R8 subtype specification, we conducted an *in vivo* RNAi screen for genes for which knockdown caused a change in the proportion of R8s expressing a *rh5-LexA, lexAOP-GFP* transcriptional reporter (*rh5≫GFP*) (Vasiliauskas et al., 2011) (i.e. low Hippo pathway activity) (Fig. 1D). We screened 652 lines targeting transcription factor genes, which resulted in 113 lethal phenotypes, 155 eye morphology phenotypes, and one line with a dramatic increase in Rh5.

In the screen, we identified *BEAF-32* as a positive regulator of the Hippo pathway. RNAi knockdown of *BEAF-32* caused a dramatic increase in the proportion of R8s that express *rh5≫GFP* (Fig. 1E). *BEAF-32* RNAi also caused an increase in R8s that express Rh5 protein and a decrease in R8s expressing Rh6 protein (Fig. 1G,H,L).

BEAF-32 (Boundary element-associated factor of 32 kD; hereafter referred to as BEAF) is one of several known *Drosophila* insulator proteins, including CTCF, GAGA factor (also known as Trl), Su(Hw), Zw5 (Dwg), CP190 and Mod(mdg4). BEAF binds preferentially near promoters at several thousand sites in the genome (Emberly et al., 2008; Jiang et al., 2009; Negre et al., 2011; Yang et al., 2012) and generally promotes gene expression. Two BEAF isoforms, BEAF-32A and BEAF-32B, are identical except for the 80 amino acid DNA-binding domain; however, BEAF-32B appears to be the dominant isoform (Jiang et al., 2009; Roy et al., 2007). Though BEAF binds throughout the genome, homozygous *BEAF* mutants, null for both isoforms ($BEAF^{AB\text{-}KO}$), are viable, suggesting that *BEAF* is required for specific developmental processes such as R8 subtype specification.

Similar to *BEAF* RNAi, homozygous null *BEAF* mutants ($BEAF\text{-}32^{AB\text{-}KO}$) (Roy et al., 2007) (Fig. 1I,L) and flies with the *BEAF* null mutant allele over a 105 kb deficiency completely lacking the *BEAF* locus (Fig. 1J,L) displayed an increase in Rh5 and decrease in Rh6 expression. An independent *BEAF* mutant allele caused by a P-element insertion ($BEAF\text{-}32^{NP6377}$) placed over the deficiency showed similar changes in the Rh5:Rh6 ratio (NP/def; Fig. 1L). All three *BEAF* mutant conditions and *BEAF* RNAi displayed significant increases in Rh5 and decreases in Rh6, and any phenotypic variability among these is likely due to differences in genetic background. A genomic fragment containing the *BEAF* gene locus (Roy et al., 2007) rescued the mutant phenotype, restoring the normal Rh5:Rh6 ratio (Fig. 1K,L), indicating that the Rh phenotype is specifically due to loss of *BEAF*. RNAi-mediated knockdown of other insulator genes [*CTCF*, *Cp190*, *mod(mdg4)*, *su(Hw)* and *GAGA factor*] did not significantly increase Rh5 in the retina (Fig. S1). Therefore, the regulation of Rhodopsins in R8s by insulators is likely to be restricted to *BEAF* and not a general property of insulator function.

Development (2016) 143, 2389-2397 doi:10.1242/dev.134700

BEAF acts in R8s downstream of R7 signaling to control Rhodopsins

We next determined the cellular focus of BEAF activity. Consistent with previous reports that *BEAF* is expressed in all cells of the fly (Roy et al., 2007), a 900 bp *BEAF* promoter drove a *BEAF-GFP* transgene in all photoreceptors, including all R8s (Fig. 2A). Photoreceptor-specific expression of a dominant-negative BEAF protein lacking the DNA-binding domain and containing only the protein-binding BED domain (*all PRs>dom neg*) (Gilbert et al., 2006) induced an increase in Rh5 and decrease in Rh6 (Fig. 2B,D). Whole-retina clones of *BEAF* null or P-element insertion mutants displayed changes in Rh expression (Fig. 2C,D) similar to those in viable whole-animal *BEAF* mutants. *BEAF* null mutant clones displayed upregulation of Rh5 and loss of Rh6 compared with wild-type clones (Fig. 2E,F). Thus, *BEAF* is required for proper expression of Rh5 and Rh6 in R8 photoreceptor neurons of the eye.

The most upstream trigger for R8 subtype fate is the stochastic ON/OFF expression of Ss in R7s. Expression of Ss represses an unknown signal to R8s, resulting in Wts expression, active Hippo signaling, and Rh6 expression (Fig. 1B). In the absence of Ss, the signal induces repression of *wts*, leading to inactive Hippo signaling and Rh5 expression (Fig. 1A). Ss was expressed at a similar frequency in *BEAF* null mutant clones as in wild-type clones (Fig. 2G,H), indicating that *BEAF* is not required for Ss expression. Rh3 and Rh4, targets of Ss regulation in R7s (Thanawala et al., 2013; Wernet et al., 2006), were expressed at similar ratios in *BEAF* null mutant and wild-type clones (Fig. 2I,J). Thus, *BEAF* is not required for Ss expression or R7 subtype specification.

We next showed that *BEAF* acts downstream of Ss and the signal to control R8 subtypes. Ectopic expression of Ss in all R7s from a BAC transgene (Johnston and Desplan, 2014) repressed the signal to R8s, causing nearly all R8s to adopt yR8 fate and express Rh6 (Fig. 2K,M,O). Ectopic Ss expression in R7s in *BEAF* null mutants displayed increased Rh5 and decreased Rh6 expression (Fig. 2L,M,P), showing that Ss requires *BEAF* activity to control R8 subtype. Genetic ablation of R7s in *sevenless* (*sev*) mutants removed the signal from R7s to R8s, causing nearly all R8s to acquire yR8 fate and express Rh6 (Fig. 2Q,S,T). *sev*; *BEAF* null double mutant R8s primarily expressed Rh5 (Fig. 2R,S,U), showing that the default Hippo activity ON state of R8s requires *BEAF* to activate Rh6 and repress Rh5. Altogether, these data indicate that *BEAF* acts in R8s downstream of the signal from R7s to control R8 subtype specification (Fig. 2V).

BEAF binds genes encoding the Hippo pathway members and *melt*

To explore how *BEAF* regulates the R8 regulatory network, we examined five independent ChIP datasets (four ChIP-chip, one ChIP-seq) available from the modENCODE consortium (Negre et al., 2011). We identified strong peaks (Fig. 3A, red diamonds) that are likely to be direct binding sites for BEAF and weak peaks (Fig. 3A, unfilled diamonds) that may be the result of DNA looping and insulator-insulator interaction (Liang et al., 2014). Strong BEAF binding peaks were present for all the core members of the Hippo pathway (*hpo*, *sav, mats*, *wts*), upstream regulators known to function in R8 {*kib* (*kibra*), *mer*, *aPKC*, *lgl* [*l(2)gl*]}, and output regulators (*yki*, *sd*) (Fig. 3A). BEAF also bound *melt*, part of the *melt-wts* bistable feedback loop (Fig. 3A). Although BEAF weakly bound *ss*, the fate trigger in R7s, and *otd* (also known as *oc*), a general activator of Rh3 and Rh5 and repressor of Rh6, we did not detect any defects indicative of changes in *ss* (Fig. 2G-J) or *otd* (Fig. 1D-J; Fig. 2E) expression in *BEAF* mutants. Additionally, BEAF does not bind at loci of the other photoreceptor-restricted transcription factors that regulate R8 Rhodopsins (*sens*, *pph13* and *tj*). The absence of the BEAF consensus DNA binding sequence (CGATA) in the *rh5* and *rh6* promoter regions, which are sufficient to induce their subtype specific expression, is consistent with a model wherein BEAF does not regulate Rh5 or Rh6 expression through direct binding. Together, these binding profiles suggest that BEAF could regulate R8 subtype fate by controlling aspects of the Hippo pathway or expression of *melt* and *wts*.

BEAF is required for repression of *melt* and activation of *wts*

Since BEAF bound *melt* and *wts*, which are in a transcriptional double-negative feedback loop crucial for R8 subtype specification, we examined the role of *BEAF* in their regulation. In pR8s, *melt* represses *wts* expression to activate Rh5 and repress Rh6 (Fig. 1A) (Mikeladze-Dvali et al., 2005). In yR8s, *wts* is expressed to repress Rh5 and induce Rh6 (Fig. 1B). *BEAF* null mutants display loss of Rh6 and gain of Rh5 expression, suggesting that *melt* is upregulated and *wts* expression is downregulated. Indeed, *melt* (*melt-lacZ*) is de-repressed (Fig. 3B) and *warts* expression (*warts-lacZ*) is lost (Fig. 3C) in R8s in *BEAF* null mutant clones, indicating that *BEAF* is required for the repression of *melt* and activation of *warts* expression (Fig. 3D).

BEAF is required downstream of Wts and Melt for regulation of Rhodopsins

The Rh phenotype observed in *BEAF* mutants could be caused simply by de-repression of *melt* and loss of *wts* expression (Fig. 3B-D). Alternatively, BEAF could play other roles in the pathway and the *BEAF* mutant phenotype could be due to misregulation of additional downstream genes. In *melt* mutants, *wts* is expressed and the Hippo pathway is active, inducing nearly all R8s to express Rh6 and lose Rh5 (Fig. 4A,C) (Mikeladze-Dvali et al., 2005). Removing *BEAF* in *melt* mutants caused upregulation of Rh5 and downregulation of Rh6 (Fig. 4B,C) compared with *melt* single mutants, suggesting that BEAF acts downstream of or in parallel with *melt* to control Hippo pathway activity and Rh expression state (Fig. 4F).

We next tested whether *wts* and the Hippo pathway require *BEAF* to regulate Rh expression. Misexpression of Wts and Salvador (Sav, an upstream positive regulator of Wts) in wild-type clones (*BEAF*+) induced Rh6 in all R8s (Fig. 4D,E). *BEAF* null mutant clones generated in retinas simultaneously misexpressing Wts and Sav in all photoreceptors resulted in the upregulation of Rh5 and loss of Rh6 (Fig. 4D,E). *BEAF* null mutant clones in retinas with misexpression of Wts alone displayed similar phenotypes (data not shown). Thus, *BEAF* is required for Hippo pathway activity to promote the Rh6 R8 fate. Altogether, our epistasis analysis indicates that BEAF acts downstream of or in parallel to *wts* and *melt* to regulate Rh5 and Rh6 expression (Fig. 4F), as well as upstream of *wts* and *melt* to regulate their expression (Fig. 3D).

Yki but not Sd requires *BEAF* function to regulate Rhodopsins

The transcription factors Yorkie (Yki) and Scalloped (Sd) are heterodimerization partners that regulate Hippo pathway target genes downstream of Wts (Goulev et al., 2008; Wu et al., 2008; Zhang et al., 2008). In pR8s with Hippo pathway OFF, Yki and Sd are active and induce Rh5 and repress Rh6 (Fig. 1A). In yR8s with active Hippo signaling, Yki is inactive and Rh6 is expressed, whereas Rh5 is repressed (Fig. 1B). *yki* null mutant cells are

Development (2016) 143, 2389-2397 doi:10.1242/dev.134700

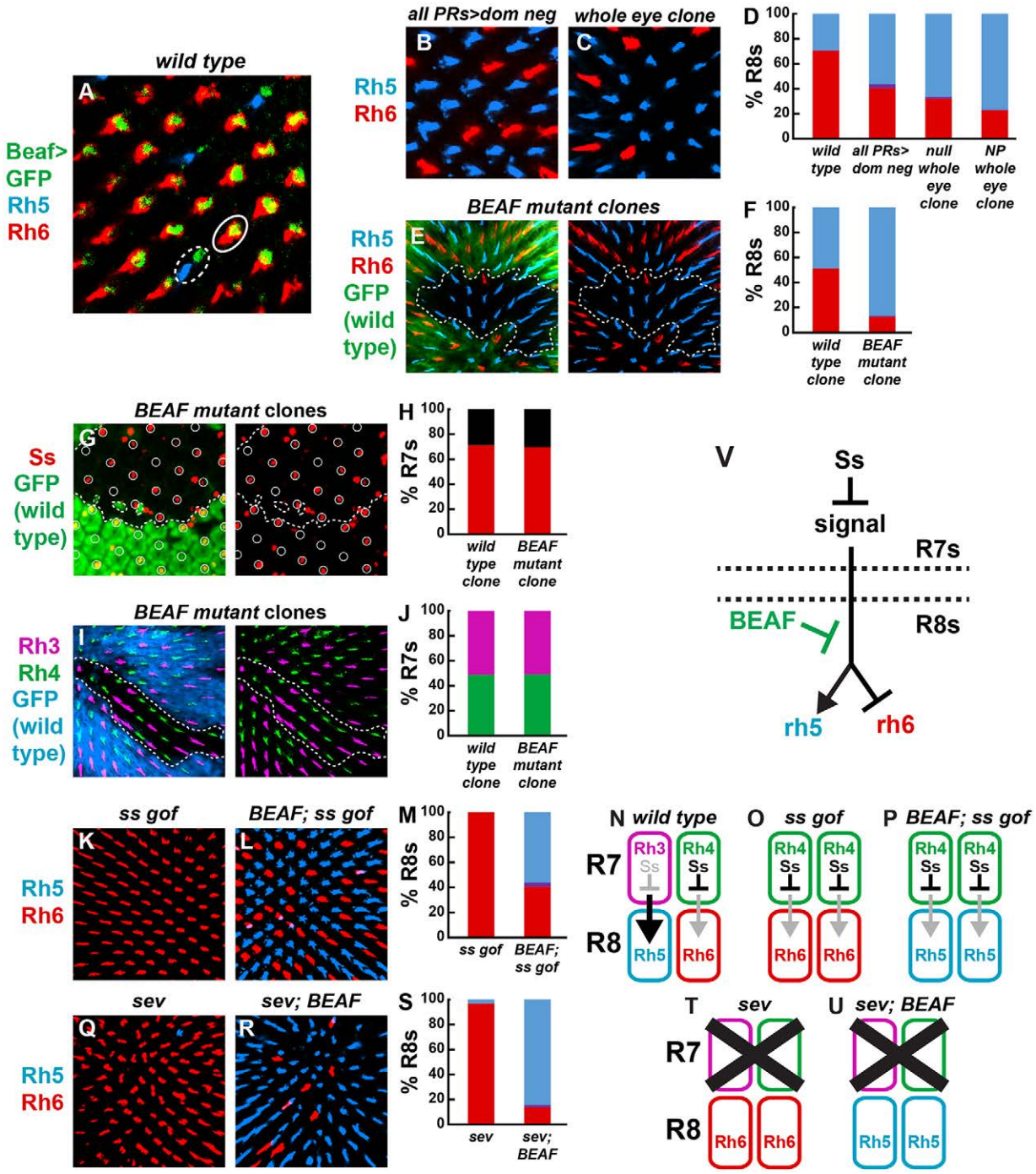


Fig. 2. *BEAF* acts in R8s downstream of R7 signaling to control Rhodopsins. (A) *BEAF-GFP* under control of the *BEAF* promoter was expressed in both R8 subtypes. Rh5-expressing pR8 (dashed oval); Rh6-expressing yR8 (solid oval). (B) Rh5-expressing R8s increased and Rh6-expressing R8s decreased when a BEAF dominant-negative construct was expressed specifically in photoreceptors. (C) A similar phenotype was observed in whole eye *BEAF* null mutant clones. (D) Quantification of the data shown in B,C. (E,F) *BEAF* null mutant clones (GFP–) contained more R8s expressing Rh5 compared with wild-type or heterozygous tissue (GFP+). Dashed lines represent clone boundary in all panels unless otherwise noted. (G,H) Ss was expressed stochastically with similar frequency in *BEAF* null (GFP–) and control (GFP+) tissue in pupal retinas. R7 cells are circled. Red indicates percentage of R7s expressing Ss; black indicates percentage of R7s lacking Ss. (I,J) The Rh3 and Rh4 expression ratio was normal in *BEAF* null mutant clones (GFP–). (K) Ectopic Ss expression in all R7s induced Rh6 and inhibited Rh5 expression in nearly all R8s. (L) Ectopic Ss expression in the absence of BEAF resulted in increased Rh5- and decreased Rh6-expressing R8s. (M) Quantification of the data shown in K,L. (N-P) Schematics depicting wild type, K and L. (Q) Genetic ablation of R7s (and hence the signal to R8) in *sev* mutants caused expression of Rh6 and loss of Rh5 in nearly all R8s. (R) *sev; BEAF* null double mutants displayed upregulation of Rh5 and downregulation of Rh6. (S) Quantification of the data shown in Q,R. (T,U) Schematics describing the observations shown in Q,R. (V) Model for how *BEAF* acts in R8s, downstream of R7 signaling to control Rhodopsins.

eliminated via apoptosis and cannot be examined in adult eyes, but strong expression of RNAi can effectively knockdown *yki* function in the retina (Jukam et al., 2013). RNAi knockdown of *yki* caused a loss of Rh5 and gain of Rh6 in all R8s (Fig. 4G,I). Retinas with simultaneous RNAi knockdown of *yki* and *BEAF* displayed upregulation of Rh5 and loss of Rh6 (Fig. 4H,I) relative to *yki*

Development (2016) 143, 2389-2397 doi:10.1242/dev.134700

RNAi alone. With the caveats inherent to RNAi-based epistasis analysis, we conclude that *BEAF* is required downstream of or in parallel with *yki* to control Rh5 and Rh6 (Fig. 4L).

Sd is present in all R8s and appears to play a permissive role in Rh regulation. *sd* mutants display Rh6 in all R8s and completely lose Rh5 (Jukam et al., 2013). Whereas wild-type clones (*sd*+) in homozygous *BEAF* null mutant tissue upregulated Rh5 and lost Rh6, *sd* mutant clones in *BEAF* mutant tissue expressed Rh6 and lost Rh5 in most R8s (Fig. 4J,K), suggesting that Sd acts downstream of *BEAF* to regulate Rhs. These data are consistent with a model wherein Yki, but not Sd, requires *BEAF* to regulate Rhodopsins, suggesting that Yki and Sd may have separable roles in R8 subtype specification (Fig. 4L). Given the strong but incomplete phenotypic suppression in the above epistasis, however, we cannot exclude more complicated models.

Positive-feedback regulation of *wts* expression is independent of *BEAF*

We next tested *BEAF* for a role in the positive network-level feedback that is a feature of the R8 Hippo pathway, but not the Hippo growth pathway (Jukam et al., 2013). Wts and Yki cross-regulate in a double-negative feedback loop, in which Wts phosphorylates Yki to inactivate it in yR8s and Yki downregulates transcription of *wts* in pR8s (Fig. 1A,B). Thus, Wts activates its expression by inhibiting its repressor Yki (Fig. 4O). In wild-type R8s, *wts* (i.e. *wts-lacZ*) is expressed in yR8s to generate an active Hippo pathway and Rh6 expression. Ectopic expression of *wts* (*GMR-wts*) caused all R8s to express *wts* (Mikeladze-Dvali et al., 2005). Since *BEAF* is required for *wts* expression in otherwise wild-type yR8s, we predicted that *BEAF* would be required for positive-feedback regulation of *wts* when *wts* was ectopically expressed. However, retinas with ectopic *wts* expression in *BEAF* null mutants displayed *wts-lacZ* expression in all R8s (Fig. 4M), suggesting that *BEAF* is not required for the positive feedback regulation of *wts* (Fig. 4O,P). One interpretation of this result is that *BEAF* is required only for initiation of *wts* expression and not its maintenance. Alternatively, the wild-type *wts/yki* feedback loop could be near a threshold that is highly sensitive to *BEAF* regulation, whereas ectopic expression of *wts* biases the regulation strongly towards *wts* expression overcoming the absence of *BEAF*. Consistent with a role for *BEAF* downstream of *wts* and *yki* to regulate opsins, we still observe loss of Rh6 in most *BEAF* mutant cells that express *wts-lacZ* (Fig. 4M,N).

Conclusions

We have shown that the insulator protein BEAF is required for a post-mitotic neuronal fate decision in *Drosophila* photoreceptors. BEAF regulates Hippo pathway activity to control R8 subtype fate and Rhodopsin expression (Fig. 4Q). First, *BEAF* regulates *wts* and *melt* expression by acting upstream. Second, *BEAF* is required for the Hippo pathway to promote Rh6 and repress Rh5. We also demonstrate that *BEAF* acts downstream of or in parallel with Yki for regulation of Rhodopsins. Finally, we show that this regulation of Rhodopsins is independent of *wts* feedback. It appears likely that BEAF regulates cell specification by permissively promoting Hippo pathway activity and *wts* expression to specify the default yR8 fate.

Despite its role in regulating the Hippo pathway in post-mitotic neuronal fate, *BEAF* appears to be dispensable for Hippo growth signaling in the eye because homozygous null mutants are viable, exhibit no gross external morphological defects, and show no dramatic differences in eye clone size or pupal interommatidial cell number compared with wild-type tissue (Fig. 2E,G,I; Fig. 3B,C; Fig. 4D; Fig. S2). Additionally, *BEAF* depletion did not suppress the under-proliferation in *yki*-RNAi eyes despite suppressing

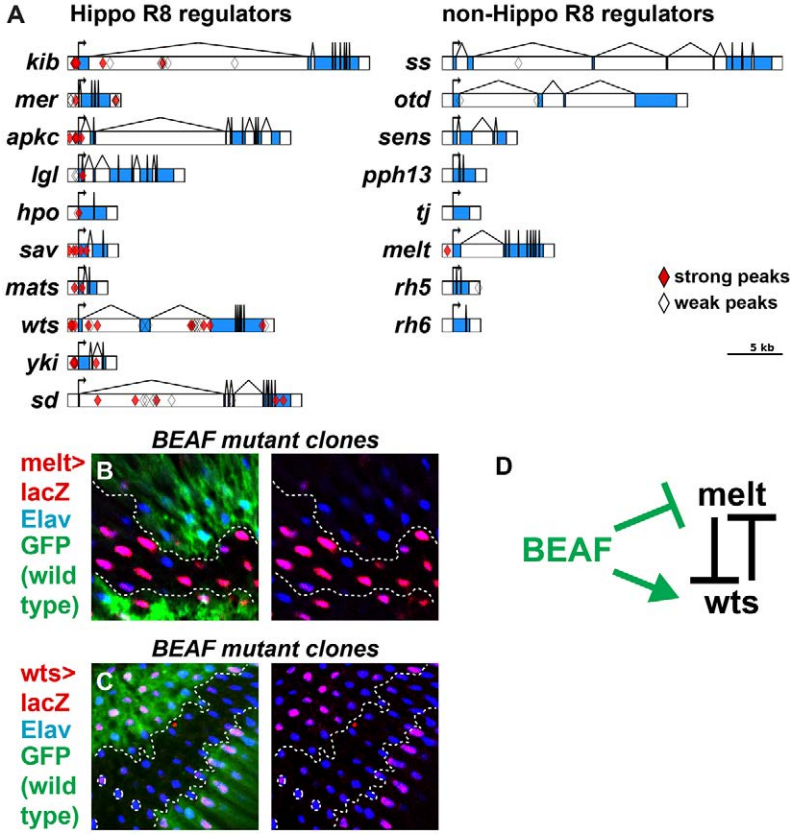


Fig. 3. *BEAF* regulates R8 subtypes by promoting *warts* expression and preventing *melt* expression. (A) Summary of BEAF modEncode ChIP binding data at loci of R8 Hippo pathway genes (left) and non-Hippo R8 subtype regulators (right). Diamonds are ChIP-chip peak centers (red diamonds are strong peaks; unfilled diamonds are weak peaks); exons are blue; non-coding sequence is white. (B) *melt-lacZ* was upregulated in *BEAF* null mutant clones (GFP−) compared with control tissue (GFP+). (C) *wts-lacZ* was lost in *BEAF* null mutant clones (GFP−) compared with control tissue (GFP+). (D) Model for BEAF regulation of *wts* and *melt* expression. Dashed lines represent clone boundary.

Development (2016) 143, 2389-2397 doi:10.1242/dev.134700

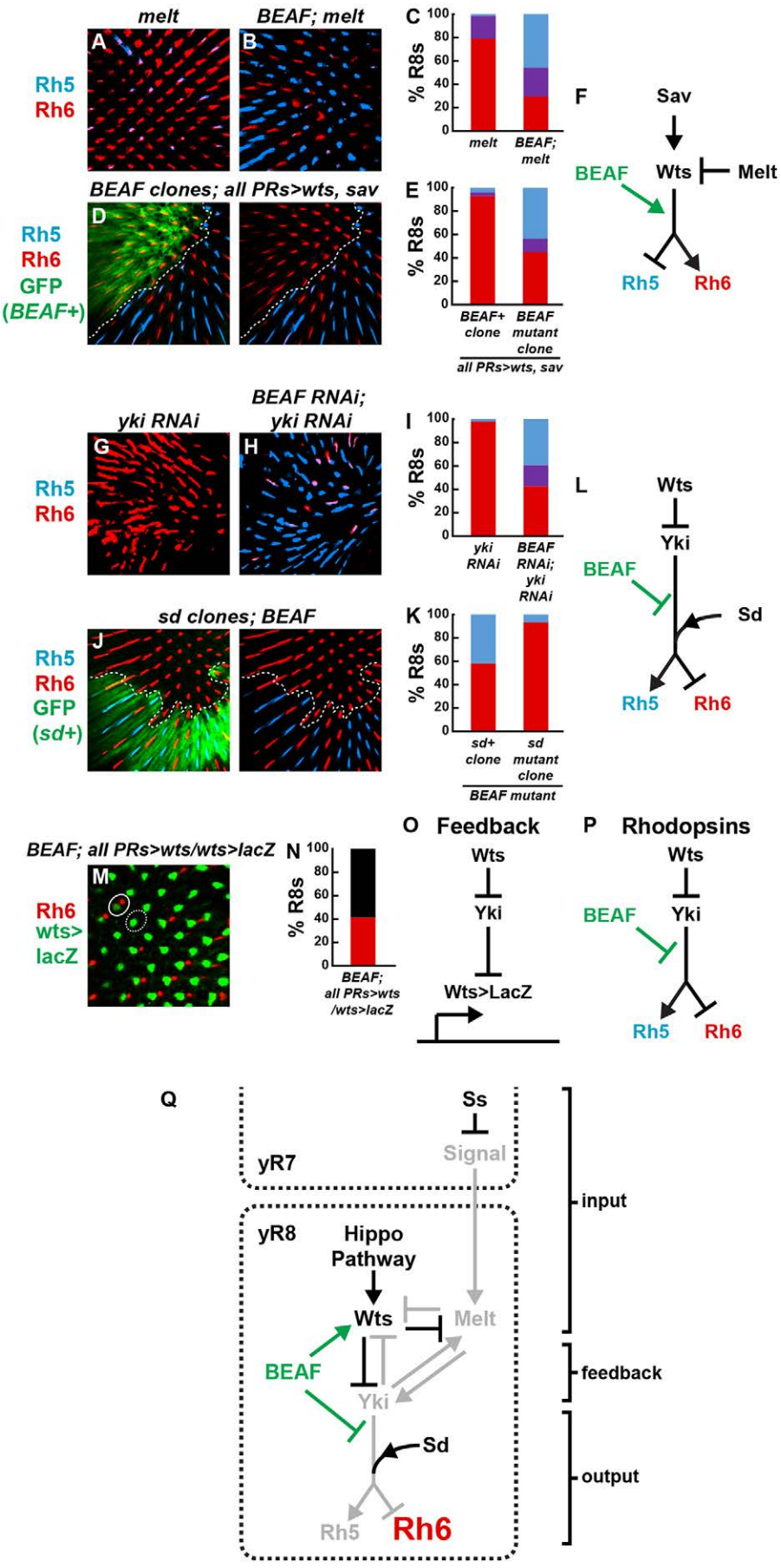


Fig. 4. *BEAF* is essential for Hippo pathway regulation of Rhodopsins, but not positive feedback, in R8 subtype specification. (A) *melt* mutants contained Rh6 in almost all R8s. (B) Rh5 was upregulated in *BEAF; melt* null double mutants. (C) Quantification of the data shown in A,B. (D) 0*GMR-wts+sav* induced Rh6 and inhibited Rh5 expression in otherwise wild-type R8s (GFP+). *BEAF* null mutant clones (GFP–) in *GMR-wts+sav* retinas showed increased Rh5 and decreased Rh6 expression. (E) Quantification of the data shown in D. (F) Schematic showing how *BEAF* acts downstream of *wts* and *melt* to regulate Rhodopsins. (G) *yki-RNAi* retinas displayed Rh6 expression in all R8s. (H) *BEAF-RNAi+yki-RNAi* resulted in upregulation of Rh5 and downregulation of Rh6. (I) Quantification of the data shown in G,H. (J) *sd* mutant clones (GFP–) in whole-eye *BEAF* null mutant background (GFP+) showed Rh6 in almost all R8s. (K) Quantification of the data shown in J. (L) Model for *BEAF* regulation of Rhodopsin output downstream of Yki and upstream of Sd. (M) *wts-lacZ* (green) was expressed in all R8s in *BEAF null*;*GMR-wts* retinas. Circle shows Rh6 and *wts-lacZ* co-expressed; dotted circle shows R8 expressing *wts-lacZ* but not Rh6. Note: R8 nuclei are in a different cell region than Rh-containing rhabdomeres. (N) Quantification of the data shown in M. Red indicates percentage of R8s expressing Rh6; black indicates percentage of R8s lacking Rh6 (presumably expressing Rh5). (O,P) *BEAF* is dispensable for Hippo pathway feedback, but not Hippo pathway regulation of Rhodopsins. (Q) Working model for *BEAF* regulation of R8 subtypes. Arrows indicate genetic regulation.

Development (2016) 143, 2389-2397 doi:10.1242/dev.134700

yki-RNAi Rhodopsin phenotypes (Fig. 4G,H). This differential regulation of the Hippo pathway in R8s compared with growth is consistent with other transcriptional regulators of R8 subtypes (*ewg*, *tj*) having minimal or no proliferation defects (Hsiao et al., 2013; Jukam et al., 2013). It is possible that BEAF regulates the Hippo pathway indirectly in R8s, by acting on yet-to-be-discovered R8-specific Hippo pathway regulators. Alternatively, BEAF may play a larger role in R8s because of Hippo pathway positive feedback, and have less effect in Hippo growth signaling where homeostatic regulation through negative feedback may compensate for the absence of BEAF. The compensation of ectopic Wts on *wts* expression, but not opsin control, in *BEAF* mutants is consistent with such a model.

Non-CTCF insulators appear to be restricted to arthropods, and among several insect species examined (*Anopheles gambiae*, *Apis mellifera* or *Tribolium castaneum*), *BEAF* was present exclusively in the *Drosophila* genus (Heger et al., 2013; Schoborg and Labrador, 2010). We speculate that conserved signaling modules of the Hippo pathway in growth control may be co-opted for cell fate specification by regulatory factors such as BEAF that are unique to dipterans.

Insulators were classically defined as proteins that bind particular DNA sequences to either interfere with promoter-enhancer interactions or prevent chromatin-state position effects from affecting transgenes (Gaszner and Felsenfeld, 2006). This definition has expanded to include proteins that mediate chromosomal interactions to regulate 3D chromatin organization and global gene expression (Bushey et al., 2009; Phillips-Cremins and Corces, 2013; Wood et al., 2011). Despite these studies, surprisingly few roles for insulator proteins in specific biological processes in flies have been characterized, including the regulation of oogenesis (Hsu et al., 2015; Roy et al., 2007; Soshnev et al., 2013) and spermatogenesis (Soltani-Bejnood et al., 2007; Thomas et al., 2005). Our result that BEAF regulates Hippo pathway activity for terminal differentiation of R8 neuronal subtypes, but has no observed effect on general growth control or other photoreceptor fate, indicates that broadly expressed insulators can have exquisitely specific functions in development.

MATERIALS AND METHODS

Drosophila genotypes and stocks

See Table S1 for details of *Drosophila* genotypes and stocks.

Drosophila genetics and transgene descriptions

Flies were raised on corn meal-molasses-agar medium under standard laboratory conditions. y^1, w^{67};+;+ flies were considered 'wild type' and used as a control for Rhodopsin gene expression. All experiments were conducted at 25°C unless otherwise noted.

lGMR-Gal4 (long Glass Multiple Reporter) contains a pentamerized 38 bp Glass binding site and is expressed in all photoreceptors and some other retina cells posterior to the morphogenetic furrow (Wernet et al., 2003). *ey-Gal4* drives transgene expression in eye primordium and eye imaginal discs. *UAS-Dicer2(Dcr2)* is co-expressed to increase RNAi processing efficiency (Dietzl et al., 2007). *warts-lacZ* contains a P-element inserted into the *warts* locus (Justice et al., 1995). *melt-lacZ* contains the first intron of *melted* cloned upstream of *nls:lacZ* (Mikeladze-Dvali et al., 2005).

BEAF-32$^{AB\text{-}KO}$ is semi-viable (Roy et al., 2007). The original *BEAF-32*$^{AB\text{-}KO}$ stock possibly contained a second-site mutation that caused rhabdomere defects, and the chromosome was cleaned during recombination to FRT42D. After recombination, the resulting FRT42D *BEAF-32*$^{AB\text{-}KO}$ flies contained normal rhabdomeres. *BEAF-32*^{NP6377} is a null allele caused by a P-element insertion (Gurudatta et al., 2012). We also recombined this allele onto an FRT42D chromosome, which removed the lethality and growth defects previously described on the chromosome (Gurudatta et al., 2012). The deficiency *Df(2R)BSC429* (Bloomington stock #24933) contains a 105 kb FLP/FRT-derived deletion that completely removes the *BEAF* coding sequence, in addition to several other genes. Placing *BEAF-32*$^{AB\text{-}KO}$ trans-heterozygous over a second, ∼250 kb deficiency (*Df(2R)BSC858*; stock#27928) gave similar results.

Homozygous mutant adult eyes (whole eye mutant clones) were generated using the FLP/FRT system (Xu and Rubin, 1993). FLP recombinase expressed under control of the *eyeless* (*ey*) promoter (*ey-FLP*) (Newsome et al., 2000) induced recombination of FRT chromosomes containing a cell lethal mutation and GMR-hid to remove all non-mutant eye tissue (Stowers et al., 2000). Mutant clones were made using FRT-FLP-mediated recombination between the mutant chromosome and an otherwise wild-type chromosome containing P[*w*+, *ubi-GFP*].

RNAi screen

Transcription factors were defined according to the FlyTF database (Adryan and Teichmann, 2006), which includes manual curation from the literature and computationally generated structure homologies. The data set identifies 1052 candidate DNA-binding proteins, including 753 proposed as transcription factors (∼450 site specific). Another 299 genes did not meet their criteria, but had transcription-related Gene Ontology annotations (Adryan and Teichmann, 2006, 2007).

UAS-RNAi fly lines were obtained from a genome-wide library of *Drosophila* RNAi at the Vienna *Drosophila* RNAi Center (VDRC) (Dietzl et al., 2007). Each transgenic line contained a 300-500 bp inverted hairpin construct under the control of a 10× multimerized UAS promoter. We tested several Gal4 drivers, including *lGMR-Gal4* (strongly expressed in all photoreceptors after the morphogenetic furrow), *sens-Gal4* (strongly expressed in R8 and weakly and variably expressed in other photoreceptors), and *ey-Gal4+lGMR-Gal4* (*ey-Gal4* is strongly expressed in the early eye primordium and disc), with or without co-expression of *UAS-Dicer2* (*Dcr-2*). Dcr-2 is thought to enhance RNAi-processing efficiency in cell types more refractory to UAS-RNAi such as adult neurons (Dietzl et al., 2007).

UAS-Dcr2; *ey-Gal4*, *lGMR-Gal4* was determined to be the optimal Gal4 driver because it induced RNAi phenotypes very similar to phenotypes of a gene's respective loss-of-function mutant for seven out of seven known R7 and R8 subtype regulators: Rh6 was lost with *warts*, *merlin*, *mats* and *sav* RNAi, whereas Rh5 was lost with *melt* RNAi. *spineless* RNAi flies had expansion of Rh3 into all R7 photoreceptors. In addition, *dve* RNAi resulted in an increase in Rh5 expression, in the outer photoreceptors. An independent paper describing a role for Dve in restricting R8 Rhodopsins from R1-R6 was in preparation during the screen and has since been published (Johnston et al., 2011). *otd* RNAi completely removed *rh5* expression, consistent with a requirement for Otd for *rh5* transcription and direct binding to the *rh5* promoter. These positive controls demonstrate that expression of UAS-RNAi under control of *UAS-Dcr2*; *ey-Gal4*, *lGMR-Gal4* is an effective tool to induce loss-of-function developmental phenotypes detectable in adult photoreceptors.

The screen was performed by crossing UAS-RNAi males to virgin female reporter-driver lines. To maximize virgin collection, the female driver stock contained an *hs-hid* transgene on the Y chromosome (Dietzl et al., 2007). We performed two 30 min heat-shocks spaced 8 h apart at 37°C in late larval and early pupal stages to eliminate males.

R8 subtypes were assessed for defects in F1 progeny by examining for a change in the proportion of R8s expressing a *rh5-LexA::VP16*, *lexAOP::GFP* (*rh5≫GFP*) transcriptional reporter. The LexA/lexAOP binary expression system was used to amplify GFP levels while keeping the reporter Gal4 independent (Lai and Lee, 2006; Vasiliauskas et al., 2011).

Water immersion protocol for visualizing *rh5≫GFP*

GFP was visualized in living adult flies by neutralizing the cornea using a water immersion technique (Pichaud and Desplan, 2001). Ten to twelve flies of the appropriate F1 genotype were placed on a streak of clear nail polish (Wet *n* Wild) perpendicular to the straight edge with one retina facing up, in the middle of a 10 cm Petri dish. The dish was placed on the microscope stage and water added to submerge flies. Images of *rh5≫GFP* were taken with a SPOT camera mounted on a fluorescence microscope with a Nikon Plan Fluor 40× objective lens immersed in

Development (2016) 143, 2389-2397 doi:10.1242/dev.134700

water. About 30-40 ommatidia are visible in the same focal plane. If *rh5≫GFP* appeared in >60% of ommatidia of at least two flies, the genotype was later dissected and Rh5 and Rh6 visualized with antibodies. During the pilot screen it was discovered that the ey-Gal4 +lGMR-Gal4 flies exhibited, on occasion, an increase in the Rh6:Rh5 ratio (from 70:30 to 85:15). Therefore we only assayed subtype phenotypes that increased the Rh5 R8 proportion.

BEAF binding analysis

BEAF-32 ChIP data for *Drosophila melanogaster* were obtained from five available studies from modENCODE (http://www.modencode.org). For the purpose of this analysis, the pre-processed peak calls available in the modENCODE's 'dmel-interpreted-1' FTP directory were used. The midpoint of each identified ChIP peak regions was used to annotate the gene diagrams, based on the FlyBase dm3 gene annotation on the UCSC Genome Browser (genome.ucsc.edu). Strong peaks were defined as having >2.5-fold enrichment. Weak peaks were defined as having <2.5-fold enrichment. The scripts used for analysis can be found at https://github.com/pdeford/beaf32-hippo-chip.

Antibodies

Antibodies and dilutions were as follows: mouse anti-Rh3 (1:10; gift from S. Britt, University of Colorado, CO, USA), rabbit anti-Rh4 (1:100; gift from C. Zuker, Columbia University, NY, USA), mouse anti-Rh5 (1:200; Chou et al., 1996), rabbit anti-Rh6 (1:2000; Tahayato et al., 2003), guinea pig anti-Ss 2.21 (1:200; gift from Y.N. Jan, University of California, San Francisco, CA, USA) (Kim et al., 2006), rat anti-ElaV (1:50; Developmental Studies Hybridoma Bank), sheep anti-GFP (1:500; AbD Serotec, 4745-1051), mouse anti-Dlg (1:75; Developmental Studies Hybridoma Bank) and goat anti-β-gal (1:50; Biogenesis, 4600-1409). All secondary antibodies were Alexa Fluor 488-, 555- or 647-conjugated antibodies (1:400) made in donkey (Molecular Probes).

Antibody staining and imaging

Adult or staged pupal retinas were dissected as described (Hsiao et al., 2012). Briefly, retinas were dissected and immediately fixed for 15 min with 4% paraformaldehyde at room temperature. After lamina removal, retinas were rinsed twice in PBX (PBS+0.2% Triton X-100) then washed in PBX for >2 h. Retinas were incubated overnight with primary antibodies diluted in PBX at room temperature, rinsed twice in PBX and then washed in PBX for >4 h. Retinas were incubated 4-6 h or overnight with secondary antibodies diluted in PBX at room temperature, rinsed twice in PBX and then washed in PBX for >2 h. Adult retinas were mounted in SlowFade (Molecular Probes) and pupal retinas in Vectashield (Vector Laboratories) on glass slides with coverslip. Images were acquired using a Leica TCS SP5, Zeiss 710 or Zeiss 780 confocal microscope. Objectives were 10×, 20× or 60×. Images were processed in Photoshop (Adobe) or ImageJ. Brightness or contrast adjustments, if any, were simple linear adjustments made to the entire image, in accordance with journal guidelines. Figures were prepared using Illustrator (Adobe).

Quantification of photoreceptor subtypes

Confocal images were taken and the number of R8 cells that expressed Rh5, Rh6, both, or neither were counted. The percentage of R8s expressing Rh5 (%Rh5) was calculated for each retina, and mean %Rh5 of all retinas within a genotype was used to compare across genotypes. Retinas were scored if there were 75 or more ommatidia present in a single focal plane. Most retinas contained ∼200-300 ommatidia in a single image. For all genotypes, more retinas were observed than quantified to confirm a particular phenotype. Means and standard deviations for all experiments can be found in Table S2.

Acknowledgements

We are grateful to Steve Britt, Steve Cohen, Barry Dickson, Georg Halder, Iswar Hariharan, Craig Hart, Ken Irvine, Jin Jiang, D. J. Pan, Jessica Treisman, Tian Xu, Charles Zuker, the Bloomington Stock Center, the Kyoto Stock Center and the Vienna *Drosophila* RNAi Center (VDRC) for generously providing published fly stocks and antibodies. We thank Pam Geyer and Judy Kassis for helpful comments on the manuscript. We thank Vince Lau for technical assistance.

Competing interests

The authors declare no competing or financial interests.

Author contributions

D.J. and R.J.J. designed the research and RNAi screen. D.J. performed the RNAi screen. D.J., K.V., C.A., C.Z., J.Y., J.C., and R.J.J. performed genetic experiments and analyzed the data. P.D. analyzed the BEAF modEncode ChIP data. D.J. and R.J.J. wrote the manuscript with input from all authors.

Funding

This work was supported by a Pew Scholar Award from Pew Charitable Trusts [00027373 to R.J.J.]; a Basil O'Connor Scholar Award from the March of Dimes Foundation [5-FY15-21 to R.J.J.]; and a grant from the National Institutes of Health [R01EY025598 to R.J.J.]. Deposited in PMC for release after 12 months.

Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.134700.supplemental

References

Adryan, B. and Teichmann, S. A. (2006). FlyTF: a systematic review of site-specific transcription factors in the fruit fly Drosophila melanogaster. *Bioinformatics* **22**, 1532-1533.

Adryan, B. and Teichmann, S. A. (2007). Computational identification of site-specific transcription factors in Drosophila. *Fly* **1**, 142-145.

Bushey, A. M., Ramos, E. and Corces, V. G. (2009). Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. *Genes Dev.* **23**, 1338-1350.

Chou, W.-H., Hall, K. J., Wilson, D. B., Wideman, C. L., Townson, S. M., Chadwell, L. V. and Britt, S. G. (1996). Identification of a novel Drosophila opsin reveals specific patterning of the R7 and R8 photoreceptor cells. *Neuron* **17**, 1101-1115.

Cockburn, K., Biechele, S., Garner, J. and Rossant, J. (2013). The Hippo pathway member Nf2 is required for inner cell mass specification. *Curr. Biol.* **23**, 1195-1201.

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

Emberly, E., Blattes, R., Schuettengruber, B., Hennion, M., Jiang, N., Hart, C. M., Käs, E. and Cuvier, O. (2008). BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites. *PLoS Biol.* **6**, e327.

Emoto, K., Parrish, J. Z., Jan, L. Y. and Jan, Y.-N. (2006). The tumour suppressor Hippo acts with the NDR kinases in dendritic tiling and maintenance. *Nature* **443**, 210-213.

Gaszner, M. and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7**, 703-713.

Gilbert, M. K., Tan, Y. Y. and Hart, C. M. (2006). The Drosophila boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. *Genetics* **173**, 1365-1375.

Goulev, Y., Fauny, J. D., Gonzalez-Marti, B., Flagiello, D., Silber, J. and Zider, A. (2008). SCALLOPED interacts with YORKIE, the nuclear effector of the hippo tumor-suppressor pathway in Drosophila. *Curr. Biol.* **18**, 435-441.

Gurudatta, B. V., Ramos, E. and Corces, V. G. (2012). The BEAF insulator regulates genes involved in cell polarity and neoplastic growth. *Dev. Biol.* **369**, 124-132.

Hardie, R. C. (1985). Functional organization of the fly retina. In *Sensory Physiology* (ed. D. Ottoson), pp. 1-79. New York, NY: Springer-Verlag.

Heger, P., George, R. and Wiehe, T. (2013). Successive gain of insulator proteins in arthropod evolution. *Evolution* **67**, 2945-2956.

Heisenberg, M. and Buchner, E. (1977). The role of retinula cell types in visual behavior of *Drosophila melanogaster*. *J. Comp. Physiol.* **117**, 127-162.

Hsiao, H. Y., Johnston, R. J., Jukam, D., Vasiliauskas, D., Desplan, C. and Rister, J. (2012). Dissection and immunohistochemistry of larval, pupal and adult Drosophila retinas. *J. Vis. Exp.* **69**, e4347.

Hsiao, H.-Y., Jukam, D., Johnston, R. and Desplan, C. (2013). The neuronal transcription factor erect wing regulates specification and maintenance of Drosophila R8 photoreceptor subtypes. *Dev. Biol.* **381**, 482-490.

Hsu, S.-J., Plata, M. P., Ernest, B., Asgarifar, S. and Labrador, M. (2015). The insulator protein Suppressor of Hairy wing is required for proper ring canal development during oogenesis in Drosophila. *Dev. Biol.* **403**, 57-68.

Irvine, K. D. and Harvey, K. F. (2015). Control of organ growth by patterning and hippo signaling in Drosophila. *Cold Spring Harb. Perspect Biol.* **7**, a019224.

Jiang, N., Emberly, E., Cuvier, O. and Hart, C. M. (2009). Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. *Mol. Cell. Biol.* **29**, 3556-3568.

Johnston, R. J., Jr and Desplan, C. (2014). Interchromosomal communication coordinates intrinsically stochastic expression between alleles. *Science* **343**, 661-665.
Johnston, R. J., Jr, Otake, Y., Sood, P., Vogt, N., Behnia, R., Vasiliauskas, D., McDonald, E., Xie, B., Koenig, S., Wolf, R. et al. (2011). Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the Drosophila eye. *Cell* **145**, 956-968.
Jukam, D. and Desplan, C. (2011). Binary regulation of Hippo pathway by Merlin/NF2, Kibra, Lgl, and Melted specifies and maintains postmitotic neuronal fate. *Dev. Cell* **21**, 874-887.
Jukam, D., Xie, B., Rister, J., Terrell, D., Charlton-Perkins, M., Pistillo, D., Gebelein, B., Desplan, C. and Cook, T. (2013). Opposite feedbacks in the Hippo pathway for growth control and neural fate. *Science* **342**, 1238016.
Justice, R. W., Zilian, O., Woods, D. F., Noll, M. and Bryant, P. J. (1995). The Drosophila tumor suppressor gene warts encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. *Genes Dev.* **9**, 534-546.
Kim, M. D., Jan, L. Y. and Jan, Y. N. (2006). The bHLH-PAS protein Spineless is necessary for the diversification of dendrite morphology of Drosophila dendritic arborization neurons. *Genes Dev.* **20**, 2806-2819.
Lai, S.-L. and Lee, T. (2006). Genetic mosaic with dual binary transcriptional systems in Drosophila. *Nat. Neurosci.* **9**, 703-709.
Liang, J., Lacroix, L., Gamot, A., Cuddapah, S., Queille, S., Lhoumaud, P., Lepetit, P., Martin, P. G. P., Vogelmann, J., Court, F. et al. (2014). Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. *Mol. Cell* **53**, 672-681.
Mikeladze-Dvali, T., Wernet, M. F., Pistillo, D., Mazzoni, E. O., Teleman, A. A., Chen, Y.-W., Cohen, S. and Desplan, C. (2005). The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in Drosophila R8 photoreceptors. *Cell* **122**, 775-787.
Milton, C. C., Grusche, F. A., Degoutin, J. L., Yu, E., Dai, Q., Lai, E. C. and Harvey, K. F. (2014). The Hippo pathway regulates hematopoiesis in Drosophila melanogaster. *Curr. Biol.* **24**, 2673-2680.
Nègre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., Kheradpour, P., Eaton, M. L., Loriaux, P., Sealfon, R. et al. (2011). A cis-regulatory map of the Drosophila genome. *Nature* **471**, 527-531.
Newsome, T. P., Asling, B. and Dickson, B. J. (2000). Analysis of Drosophila photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851-860.
Nishioka, N., Inoue, K., Adachi, K.-I., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R. O., Ogonuki, N. et al. (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev. Cell* **16**, 398-410.
Papatsenko, D., Sheng, G. and Desplan, C. (1997). A new rhodopsin in R8 photoreceptors of Drosophila: evidence for coordinate expression with Rh3 in R7 cells. *Development* **124**, 1665-1673.
Phillips-Cremins, J. E. and Corces, V. G. (2013). Chromatin insulators: linking genome organization to cellular function. *Mol. Cell* **50**, 461-474.
Pichaud, F. and Desplan, C. (2001). A new visualization approach for identifying mutations that affect differentiation and organization of the Drosophila ommatidia. *Development* **128**, 815-826.
Reddy, B. V. V. G., Rauskolb, C. and Irvine, K. D. (2010). Influence of fat-hippo and notch signaling on the proliferation and differentiation of Drosophila optic neuroepithelia. *Development* **137**, 2397-2408.
Rister, J., Desplan, C. and Vasiliauskas, D. (2013). Establishing and maintaining gene expression patterns: insights from sensory receptor patterning. *Development* **140**, 493-503.
Roy, S., Gilbert, M. K. and Hart, C. M. (2007). Characterization of BEAF mutations isolated by homologous recombination in Drosophila. *Genetics* **176**, 801-813.
Schnaitmann, C., Garbers, C., Wachtler, T. and Tanimoto, H. (2013). Color discrimination with broadband photoreceptors. *Curr. Biol.* **23**, 2375-2382.
Schoborg, T. A. and Labrador, M. (2010). The phylogenetic distribution of non-CTCF insulator proteins is limited to insects and reveals that BEAF-32 is Drosophila lineage specific. *J. Mol. Evol.* **70**, 74-84.
Soltani-Bejnood, M., Thomas, S. E., Villeneuve, L., Schwartz, K., Hong, C.-S. and McKee, B. D. (2007). Role of the mod(mdg4) common region in homolog segregation in Drosophila male meiosis. *Genetics* **176**, 161-180.
Soshnev, A. A., Baxley, R. M., Manak, J. R., Tan, K. and Geyer, P. K. (2013). The insulator protein Suppressor of Hairy-wing is an essential transcriptional repressor in the Drosophila ovary. *Development* **140**, 3613-3623.
Stowers, R. S., Garza, D., Rascle, A. and Hogness, D. S. (2000). The L63 gene is necessary for the ecdysone-induced 63E late puff and encodes CDK proteins required for Drosophila development. *Dev. Biol.* **221**, 23-40.
Tahayato, A., Sonneville, R., Pichaud, F., Wernet, M. F., Papatsenko, D., Beaufils, P., Cook, T. and Desplan, C. (2003). Otd/Crx, a dual regulator for the specification of ommatidia subtypes in the Drosophila retina. *Dev. Cell* **5**, 391-402.
Tapon, N., Harvey, K. F., Bell, D. W., Wahrer, D. C., Schiripo, T. A., Haber, D. and Hariharan, I. K. (2002). salvador Promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines. *Cell* **110**, 467-478.
Thanawala, S. U., Rister, J., Goldberg, G. W., Zuskov, A., Olesnicky, E. C., Flowers, J. M., Jukam, D., Purugganan, M. D., Gavis, E. R., Desplan, C. et al. (2013). Regional modulation of a stochastically expressed factor determines photoreceptor subtypes in the Drosophila retina. *Dev. Cell* **25**, 93-105.
Thomas, S. E., Soltani-Bejnood, M., Roth, P., Dorn, R., Logsdon, J. M. and McKee, B. D. (2005). Identification of two proteins required for conjunction and regular segregation of achiasmate homologs in Drosophila male meiosis. *Cell* **123**, 555-568.
Vasiliauskas, D., Mazzoni, E. O., Sprecher, S. G., Brodetskiy, K., Johnston, R. J., Jr, Lidder, P., Vogt, N., Celik, A. and Desplan, C. (2011). Feedback from rhodopsin controls rhodopsin exclusion in Drosophila photoreceptors. *Nature* **479**, 108-112.
Wardill, T. J., List, O., Li, X., Dongre, S., McCulloch, M., Ting, C.-Y., O'Kane, C. J., Tang, S., Lee, C.-H., Hardie, R. C. et al. (2012). Multiple spectral inputs improve motion discrimination in the Drosophila visual system. *Science* **336**, 925-931.
Wernet, M. F., Labhart, T., Baumann, F., Mazzoni, E. O., Pichaud, F. and Desplan, C. (2003). Homothorax switches function of Drosophila photoreceptors from color to polarized light sensors. *Cell* **115**, 267-279.
Wernet, M. F., Mazzoni, E. O., Çelik, A., Duncan, D. M., Duncan, I. and Desplan, C. (2006). Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174-180.
Wood, A. M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B. C., Jones, K. C. and Corces, V. G. (2011). Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. *Mol. Cell* **44**, 29-38.
Wu, S., Liu, Y., Zheng, Y., Dong, J. and Pan, D. (2008). The TEAD/TEF family protein Scalloped mediates transcriptional output of the Hippo growth-regulatory pathway. *Dev. Cell* **14**, 388-398.
Xu, T. and Rubin, G. M. (1993). Analysis of genetic mosaics in developing and adult Drosophila tissues. *Development* **117**, 1223-1237.
Xu, T., Wang, W., Zhang, S., Stewart, R. A. and Yu, W. (1995). Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase. *Development* **121**, 1053-1063.
Yamaguchi, S., Wolf, R., Desplan, C. and Heisenberg, M. (2008). Motion vision is independent of color in Drosophila. *Proc. Natl Acad. Sci. USA* **105**, 4910-4915.
Yang, J., Ramos, E. and Corces, V. G. (2012). The BEAF-32 insulator coordinates genome organization and function during the evolution of Drosophila species. *Genome Res.* **22**, 2199-2207.
Yu, F.-X., Zhao, B. and Guan, K.-L. (2015). Hippo pathway in organ size control, tissue homeostasis, and cancer. *Cell* **163**, 811-828.
Zhang, L., Ren, F., Zhang, Q., Chen, Y., Wang, B. and Jiang, J. (2008). The TEAD/TEF family of transcription factor Scalloped mediates Hippo signaling in organ size control. *Dev. Cell* **14**, 377-387.
Zhao, B., Tumaneng, K. and Guan, K.-L. (2011). The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal. *Nat. Cell Biol.* **13**, 877-883.

Table S1. Table of *Drosophila* Genotypes and Stocks

Click here to Download Table S1

Table S2. Table of means and standard deviations

Click here to Download Table S2

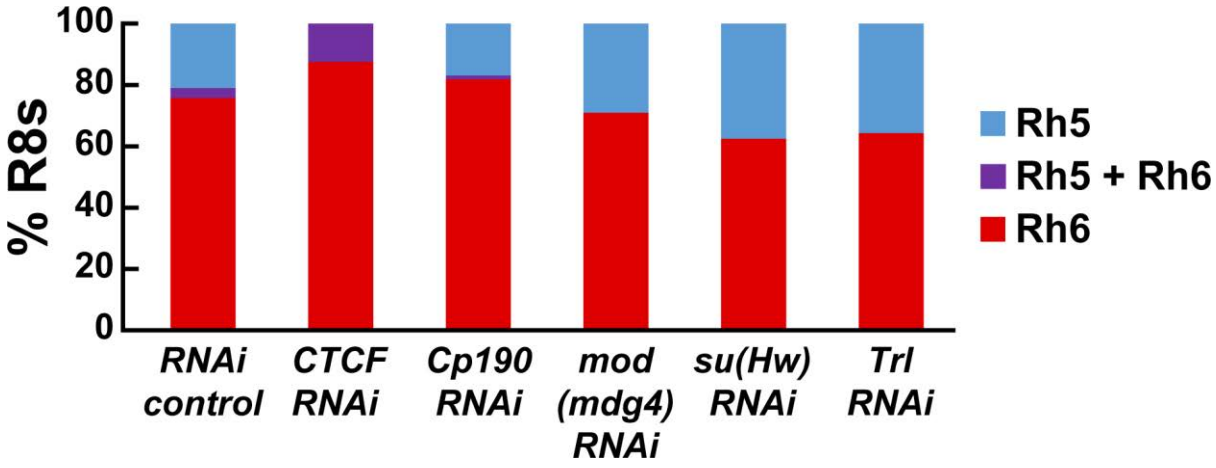


Fig. S1. Rh5 is not upregulated upon RNAi knockdown of other insulators
Rh5 is not dramatically upregulated upon RNAi knockdown of insulators *CTCF, Cp190, mod(mdg4), su(Hw),* or *Trl* compared to controls.

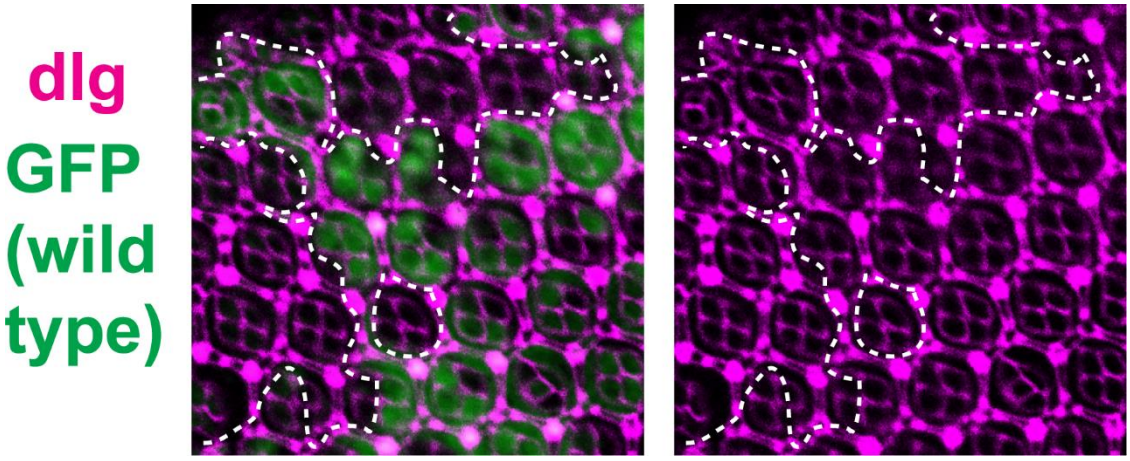


Fig. S2. Pupal interommatidial cell number is unaffected in BEAF mutants
Pupal interommatidial cell number, marked by Dlg, is similar in BEAF null mutant (GFP-) and wild type (GFP+) clones.

Bibliography

1. T. Cremer, C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**, 292-301 (2001).
2. M. Cremer *et al.*, Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res* **9**, 541-567 (2001).
3. H. B. Sun, J. Shen, H. Yokota, Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys J* **79**, 184-190 (2000).
4. J. A. Croft *et al.*, Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* **145**, 1119-1131 (1999).
5. M. Simonis *et al.*, Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-1354 (2006).
6. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).
7. T. J. Stevens *et al.*, 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59-64 (2017).
8. M. J. Rowley *et al.*, Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* **67**, 837-852 e837 (2017).
9. E. P. Nora *et al.*, Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012).
10. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).

11. T. Sexton *et al.*, Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472 (2012).
12. O. Symmons *et al.*, The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Dev Cell* **39**, 529-543 (2016).
13. Y. Guo *et al.*, CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-910 (2015).
14. S. Ahmed *et al.*, DNA zip codes control an ancient mechanism for gene targeting to the nuclear periphery. *Nat Cell Biol* **12**, 111-118 (2010).
15. W. H. Light, D. G. Brickner, V. R. Brand, J. H. Brickner, Interaction of a DNA zip code with the nuclear pore complex promotes H2A.Z incorporation and INO1 transcriptional memory. *Mol Cell* **40**, 112-125 (2010).
16. M. Capelson *et al.*, Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. *Cell* **140**, 372-383 (2010).
17. M. Kohwi, J. R. Lupton, S. L. Lai, M. R. Miller, C. Q. Doe, Developmentally regulated subnuclear genome reorganization restricts neural progenitor competence in Drosophila. *Cell* **152**, 97-108 (2013).
18. S. T. Kosak *et al.*, Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**, 158-162 (2002).
19. K. L. Reddy, J. M. Zullo, E. Bertolino, H. Singh, Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243-247 (2008).

20. D. G. Lupianez *et al.*, Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).

21. K. J. Meaburn, P. R. Gudla, S. Khan, S. J. Lockett, T. Misteli, Disease-specific gene repositioning in breast cancer. *J Cell Biol* **187**, 801-812 (2009).

22. T. Wiech *et al.*, Human archival tissues provide a valuable source for the analysis of spatial genome organization. *Histochem Cell Biol* **123**, 229-238 (2005).

23. J. M. Koeman *et al.*, Somatic pairing of chromosome 19 in renal oncocytoma is associated with deregulated EGLN2-mediated [corrected] oxygen-sensing response. *PLoS Genet* **4**, e1000176 (2008).

24. A. R. Barutcu *et al.*, Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol* **16**, 214 (2015).

25. S. J. Dalrymple, J. F. Herath, T. J. Borell, C. A. Moertel, R. B. Jenkins, Correlation of cytogenetic and fluorescence in situ hybridization (FISH) studies in normal and gliotic brain. *J Neuropathol Exp Neurol* **53**, 448-456 (1994).

26. K. N. Thatcher, S. Peddada, D. H. Yasui, J. M. Lasalle, Homologous pairing of 15q11-13 imprinted domains in brain is developmentally regulated but deficient in Rett and autism samples. *Hum Mol Genet* **14**, 785-797 (2005).

27. W. Schwarzer *et al.*, Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).

28. S. S. P. Rao *et al.*, Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324 (2017).

29. E. P. Nora *et al.*, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922 (2017).

30. E. M. Smith, B. R. Lajoie, G. Jain, J. Dekker, Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *Am J Hum Genet* **98**, 185-201 (2016).

31. C. G. Spilianakis, M. D. Lalioti, T. Town, G. R. Lee, R. A. Flavell, Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637-645 (2005).

32. C. G. Spilianakis, R. A. Flavell, Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* **5**, 1017-1027 (2004).

33. K. Monahan, A. Horta, S. Lomvardas, LHX2- and LDB1-mediated *trans* interactions regulate olfactory receptor choice. *Nature* **doi:** https://doi.org/10.1038/s41586-018-0845-0, (2019).

34. E. J. Clowney *et al.*, Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell* **151**, 724-737 (2012).

35. A. Pombo, N. Dillon, Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* **16**, 245-257 (2015).

36. S. Schoenfelder *et al.*, Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**, 53-61 (2010).

37. F. J. Iborra, A. Pombo, D. A. Jackson, P. R. Cook, Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J Cell Sci* **109 (Pt 6)**, 1427-1436 (1996).

38. S. Melnik *et al.*, The proteomes of transcription factories containing RNA polymerases I, II or III. *Nat Methods* **8**, 963-968 (2011).

39. K. Van Bortle, V. G. Corces, Nuclear organization and genome function. *Annu Rev Cell Dev Biol* **28**, 163-187 (2012).

40. V. K. Tiwari, L. Cope, K. M. McGarvey, J. E. Ohm, S. B. Baylin, A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res* **18**, 1171-1179 (2008).

41. E. F. Joyce, J. Erceg, C. T. Wu, Pairing and anti-pairing: a balancing act in the diploid genome. *Curr Opin Genet Dev* **37**, 119-128 (2016).

42. P. Avner, E. Heard, X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet* **2**, 59-67 (2001).

43. S. Rastan, Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre. *J Embryol Exp Morphol* **78**, 1-22 (1983).

44. S. Rastan, E. J. Robertson, X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *J Embryol Exp Morphol* **90**, 379-388 (1985).

45. C. P. Bacher *et al.*, Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol* **8**, 293-299 (2006).

46. N. Xu, C. L. Tsai, J. T. Lee, Transient homologous chromosome pairing marks the onset of X inactivation. *Science* **311**, 1149-1152 (2006).

47. N. Xu, M. E. Donohoe, S. S. Silva, J. T. Lee, Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet* **39**, 1390-1396 (2007).

48. S. Augui *et al.*, Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science* **318**, 1632-1636 (2007).

49. M. E. Donohoe, S. S. Silva, S. F. Pinter, N. Xu, J. T. Lee, The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128-132 (2009).

50. H. V. Crouse, The Controlling Element in Sex Chromosome Behavior in Sciara. *Genetics* **45**, 1429-1443 (1960).

51. M. S. Bartolomei, A. C. Ferguson-Smith, Mammalian genomic imprinting. *Cold Spring Harb Perspect Biol* **3**, (2011).

52. J. M. LaSalle, M. Lalande, Homologous association of oppositely imprinted chromosomal domains. *Science* **272**, 725-728 (1996).

53. L. Riesselmann, T. Haaf, Preferential S-phase pairing of the imprinted region on distal mouse chromosome 7. *Cytogenet Cell Genet* **86**, 39-42 (1999).

54. N. M. Stevens, A Study of the Germ Cells of Certain Diptera, With Reference to the Heterochromosomes and the Phenomena of Synapsis. *J Exper Zool* **5**, 359-374 (1906).

55. E. B. Lewis, The Theory and Application of a New Method of Detecting Chromosomal Rearrangements in *Drosophila melanogaster*. *Am Nat* **88**, 225-239 (1954).

56. W. M. Gelbart, Synapsis-dependent allelic complementation at the decapentaplegic gene complex in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **79**, 2636-2640 (1982).

57. Y. Hiraoka *et al.*, The onset of homologous chromosome pairing during Drosophila melanogaster embryogenesis. *J Cell Biol* **120**, 591-600 (1993).

58. J. C. Fung, W. F. Marshall, A. Dernburg, D. A. Agard, J. W. Sedat, Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations. *J Cell Biol* **141**, 5-20 (1998).

59. M. J. Gemkow, P. J. Verveer, D. J. Arndt-Jovin, Homologous association of the Bithorax-Complex during embryogenesis: consequences for transvection in Drosophila melanogaster. *Development (Cambridge, England)* **125**, 4541-4552 (1998).

60. E. F. Joyce, B. R. Williams, T. Xie, C. T. Wu, Identification of genes that promote or antagonize somatic homolog pairing using a high-throughput FISH-based screen. *PLoS Genet* **8**, e1002667 (2012).

61. B. R. Williams, J. R. Bateman, N. D. Novikov, C. T. Wu, Disruption of topoisomerase II perturbs pairing in drosophila cell culture. *Genetics* **177**, 31-46 (2007).

62. J. AlHaj Abed, J. Erceg, A. Goloborodko, S. C. Nguyen, R. B. McCole, W. Saylor, G. Fudenberg, B. R. Lajoie, J. Dekker, L. A. Mirny, C.-ting Wu, Highly Structured Homolog Pairing Reflects Functional Organization of the *Drosophila* Genome. *bioRxiv* **doi:** https://doi.org/10.1101/443887, (2018).

63. J. Erceg, J. AlHaj Abed, A. Goloborodko, B.R. Lajoie, G. Fudenberg, N. Abdennur, M. Imakaev, R.B. McCole, S.C. Nguyen, W. Saylor, E.F. Joyce, T.N. Senaratne, M.A. Hannan, G. Nir, J. Dekker, L.A. Mirny, C.-ting Wu, The genome-wide, multi-layered architecture of chromosome pairing in early *Drosophila* embryos. *bioRxiv* **doi:** https://doi.org/10.1101/443028, (2018).

64. D. W. Buster *et al.*, SCFSlimb ubiquitin ligase suppresses condensin II-mediated nuclear reorganization by degrading Cap-H2. *J Cell Biol* **201**, 49-63 (2013).

65. H. B. Li, K. Ohno, H. Gui, V. Pirrotta, Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. *PLoS Genet* **9**, e1003436 (2013).

66. T. A. Hartl, H. F. Smith, G. Bosco, Chromosome alignment and transvection are antagonized by condensin II. *Science* **322**, 1384-1387 (2008).

67. F. Bantignies, C. Grimaud, S. Lavrov, M. Gabut, G. Cavalli, Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. *Genes Dev* **17**, 2406-2420 (2003).

68. T. I. Gerasimova, K. Byrd, V. G. Corces, A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* **6**, 1025-1035 (2000).

69. H. B. Li *et al.*, Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in Drosophila melanogaster. *Mol Cell Biol* **31**, 616-625 (2011).

70. M. Ronshaugen, M. Levine, Visualization of trans-homolog enhancer-promoter interactions at the Abd-B Hox locus in the Drosophila embryo. *Dev Cell* **7**, 925-932 (2004).

71. J. Vazquez, M. Muller, V. Pirrotta, J. W. Sedat, The Mcp element mediates stable long-range chromosome-chromosome interactions in Drosophila. *Mol Biol Cell* **17**, 2158-2165 (2006).

72. K. Viets, M. Sauria, C. Chernoff, C. Anderson, S. Tran, A. Dove, R. Goyal, L. Voortman, A. Gordus, J. Taylor, and R.J. Johnston, Jr. , TADs pair homologous chromosomes to promote interchromosomal gene regulation. *bioRxiv* **doi:** https://doi.org/10.1101/445627, (2018).

73. Q. Szabo *et al.*, TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Sci Adv* **4**, eaar8082 (2018).

74. A. F. Dernburg *et al.*, Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell* **85**, 745-759 (1996).

75. H. Herman *et al.*, Trans allele methylation and paramutation-like effects in mice. *Nat Genet* **34**, 199-202 (2003).

76. J. B. Hollick, Paramutation and related phenomena in diverse species. *Nat Rev Genet* **18**, 5-23 (2017).

77. V. L. Chandler, Paramutation: from maize to mice. *Cell* **128**, 641-645 (2007).

78. E. H. Coe, The properties, origin, and mechanism of conversion-type inheritance at the B locus in maize. *Genetics* **53**, 1035-1063 (1966).

79. M. Stam *et al.*, The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. *Genetics* **162**, 917-930 (2002).

80. M. Louwers *et al.*, Tissue- and expression level-specific chromatin looping at maize b1 epialleles. *Plant Cell* **21**, 832-842 (2009).

81. G. I. Patterson, C. J. Thorpe, V. L. Chandler, Paramutation, an allelic interaction, is associated with a stable and heritable reduction of transcription of the maize b regulatory gene. *Genetics* **135**, 881-894 (1993).

82. J. E. Dorweiler *et al.*, mediator of paramutation1 is required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell* **12**, 2101-2118 (2000).

83. M. Alleman *et al.*, An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442**, 295-298 (2006).

84. M. Rassoulzadegan *et al.*, RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469-474 (2006).

85. S. T. Bennett *et al.*, Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. The IMDIAB Group. *Nat Genet* **17**, 350-352 (1997).

86. R. Aramayo, R. L. Metzenberg, Meiotic transvection in fungi. *Cell* **86**, 103-113 (1996).

87. H. Liu *et al.*, Transvection mediated by the translocated cyclin D1 locus in mantle cell lymphoma. *J Exp Med* **205**, 1843-1858 (2008).

88. J. Bollmann, R. Carpenter, E. S. Coen, Allelic interactions at the nivea locus of Antirrhinum. *Plant Cell* **3**, 1327-1336 (1991).

89. E. S. Coen, R. Carpenter, A semi-dominant allele, niv-525, acts in trans to inhibit expression of its wild-type homologue in Antirrhinum majus. *EMBO J* **7**, 877-883 (1988).

90. I. W. Duncan, Transvection effects in Drosophila. *Annu Rev Genet* **36**, 521-556 (2002).

91. L. Sipos *et al.*, Transvection in the Drosophila Abd-B domain: extensive upstream sequences are involved in anchoring distant cis-regulatory regions to the promoter. *Genetics* **149**, 1031-1050 (1998).

92. J. R. Morris, P. K. Geyer, C. T. Wu, Core promoter elements can regulate transcription on a separate chromosome in trans. *Genes Dev* **13**, 253-258 (1999).

93. D. Gohl, M. Muller, V. Pirrotta, M. Affolter, P. Schedl, Enhancer blocking and transvection at the Drosophila apterous locus. *Genetics* **178**, 127-143 (2008).

94. A. Martinez-Laborda, A. Gonzalez-Reyes, G. Morata, Trans regulation in the Ultrabithorax gene of Drosophila: alterations in the promoter enhance transvection. *EMBO J* **11**, 3645-3652 (1992).

95. F. Casares, W. Bender, J. Merriam, E. Sanchez-Herrero, Interactions of Drosophila Ultrabithorax regulatory regions with native and foreign promoters. *Genetics* **145**, 123-137 (1997).

96. J. E. Hendrickson, S. Sakonju, Cis and trans interactions between the iab regulatory regions and abdominal-A and abdominal-B in Drosophila melanogaster. *Genetics* **139**, 835-848 (1995).

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

98. K. Viets, C. Chernoff, E. Urban, J. Han, A. Chen, C. Anderson, S. Tran, D. Konzman, R.J. Johnston Jr., Activating and repressing stochastic gene expression between chromosomes. *Manuscript in preparation*.

99. M. F. Wernet *et al.*, Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174-180 (2006).

100. T. F. Mackay *et al.*, The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173-178 (2012).

101. C. Anderson *et al.*, Natural variation in stochastic photoreceptor specification and color preference in Drosophila. *Elife* **6**, (2017).

102. M. Fujioka, H. Mistry, P. Schedl, J. B. Jaynes, Determinants of Chromosome Architecture: Insulator Pairing in cis and in trans. *PLoS Genet* **12**, e1005889 (2016).

103. M. Fujioka, X. Wu, J. B. Jaynes, A chromatin insulator mediates transgene homing and very long-range enhancer-promoter communication. *Development (Cambridge, England)* **136**, 3077-3087 (2009).

104. B. Lim, T. Heist, M. Levine, T. Fukaya, Visualization of Transvection in Living Drosophila Embryos. *Mol Cell* **70**, 287-296 e286 (2018).

105. S. Henikoff, T. D. Dreesen, Trans-inactivation of the Drosophila brown gene: evidence for transcriptional repression and somatic pairing dependence. *Proc Natl Acad Sci U S A* **86**, 6704-6708 (1989).

106. J. W. Jack, B. H. Judd, Allelic pairing and gene regulation: A model for the zeste-white interaction in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **76**, 1368-1372 (1979).

107. S. Bickel, V. Pirrotta, Self-association of the Drosophila zeste protein is responsible for transvection effects. *EMBO J* **9**, 2959-2967 (1990).

108. J. A. Kassis, Pairing-sensitive silencing, polycomb group response elements, and transposon homing in Drosophila. *Adv Genet* **46**, 421-438 (2002).

109. J. A. Kassis, E. P. VanSickle, S. M. Sensabaugh, A fragment of engrailed regulatory DNA can mediate transvection of the white gene in Drosophila. *Genetics* **128**, 751-761 (1991).

110. J. A. Kassis, Unusual properties of regulatory DNA from the Drosophila engrailed gene: three "pairing-sensitive" sites within a 1.6-kb region. *Genetics* **136**, 1025-1038 (1994).

111. M. O. Fauvarque, J. M. Dura, polyhomeotic regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in Drosophila. *Genes Dev* **7**, 1508-1520 (1993).

112. M. Fujioka, Y. Emi-Sarker, G. L. Yusibova, T. Goto, J. B. Jaynes, Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development (Cambridge, England)* **126**, 2527-2538 (1999).

113. J. G. Gindhart, Jr., T. C. Kaufman, Identification of Polycomb and trithorax group responsive elements in the regulatory region of the Drosophila homeotic gene Sex combs reduced. *Genetics* **139**, 797-814 (1995).

114. A. M. Kapoun, T. C. Kaufman, Regulatory regions of the homeotic gene proboscipedia are sensitive to chromosomal pairing. *Genetics* **140**, 643-658 (1995).

115. M. J. Shimell, A. J. Peterson, J. Burr, J. A. Simon, M. B. O'Connor, Functional analysis of repressor binding sites in the iab-2 regulatory region of the abdominal-A homeotic gene. *Dev Biol* **218**, 38-52 (2000).

116. J. L. Chen *et al.*, Enhancer action in trans is permitted throughout the Drosophila genome. *Proc Natl Acad Sci U S A* **99**, 3723-3728 (2002).

117. P. K. Geyer, M. M. Green, V. G. Corces, Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. *EMBO J* **9**, 2247-2256 (1990).

118. J. R. Morris *et al.*, An analysis of transvection at the yellow locus of Drosophila melanogaster. *Genetics* **151**, 633-651 (1999).

119. M. Savitsky, T. Kahn, E. Pomerantseva, P. Georgiev, Transvection at the end of the truncated chromosome in Drosophila melanogaster. *Genetics* **163**, 1375-1387 (2003).

120. J. R. Morris, D. A. Petrov, A. M. Lee, C. T. Wu, Enhancer choice in cis and in trans in Drosophila melanogaster: role of the promoter. *Genetics* **167**, 1739-1747 (2004).

121. S. A. Ou *et al.*, Effects of chromosomal rearrangements on transvection at the yellow gene of Drosophila melanogaster. *Genetics* **183**, 483-496 (2009).

122. P. M. Bingham, The Regulation of White Locus Expression: A Dominant Mutant Allele at the White Locus of DROSOPHILA MELANOGASTER. *Genetics* **95**, 341-353 (1980).

123. W. M. Gelbart, C. T. Wu, Interactions of zeste mutations with loci exhibiting transvection effects in Drosophila melanogaster. *Genetics* **102**, 179-189 (1982).

124. Z. Zachar, C. H. Chapman, P. M. Bingham, On the molecular basis of transvection effects and the regulation of transcription. *Cold Spring Harb Symp Quant Biol* **50**, 337-346 (1985).

125. S. M. Smolik-Utlaut, W. M. Gelbart, The effects of chromosomal rearrangements on the zeste-white interaction in Drosophila melanogaster. *Genetics* **116**, 285-298 (1987).

126. R. Hopmann, D. Duncan, I. Duncan, Transvection in the iab-5,6,7 region of the bithorax complex of Drosophila: homology independent interactions in trans. *Genetics* **139**, 815-833 (1995).

127. J. Zhou, H. Ashe, C. Burks, M. Levine, Characterization of the transvection mediating region of the abdominal-B locus in Drosophila. *Development (Cambridge, England)* **126**, 3057-3065 (1999).

128. W. M. Leiserson, N. M. Bonini, S. Benzer, Transvection at the eyes absent gene of Drosophila. *Genetics* **138**, 1171-1179 (1994).

129. A. B. Coulthard, N. Nolan, J. B. Bell, A. J. Hilliker, Transvection at the vestigial locus of Drosophila melanogaster. *Genetics* **170**, 1711-1721 (2005).

130. M. Ashburner, Gene activity dependent on chromosome synapsis in the polytene chromosomes of Drosophila melanogaster. *Nature* **214**, 1159-1160 (1967).

131. G. Korge, Direct correlation between a chromosome puff and the synthesis of a larval saliva protein in Drosophila melanogaster. *Chromosoma* **62**, 155-174 (1977).

132. J. S. Kornher, D. Brutlag, Proximity-dependent enhancement of Sgs-4 gene expression in D. melanogaster. *Cell* **44**, 879-883 (1986).

133. T. D. Dreesen, S. Henikoff, K. Loughney, A pairing-sensitive element that mediates trans-inactivation is associated with the Drosophila brown gene. *Genes Dev* **5**, 331-340 (1991).

134. A. M. Pattatucci, T. C. Kaufman, The homeotic gene Sex combs reduced of Drosophila melanogaster is differentially regulated in the embryonic and imaginal stages of development. *Genetics* **129**, 443-461 (1991).

135. X. Bing, T. Z. Rzezniczak, J. R. Bateman, T. J. Merritt, Transvection-based gene regulation in Drosophila is a complex and plastic trait. *G3 (Bethesda)* **4**, 2175-2187 (2014).

136. M. C. Marin, J. R. Rodriguez, A. Ferrus, Transcription of Drosophila troponin I gene is regulated by two conserved, functionally identical, synergistic elements. *Mol Biol Cell* **15**, 1185-1196 (2004).

137. J. B. Gibson, D. S. Reed, S. Bartoszewski, A. V. Wilks, Structural changes in the promoter region mediate transvection at the sn-glycerol-3-phosphate dehydrogenase gene of Drosophila melanogaster. *Biochem Genet* **37**, 301-315 (1999).

138. N. Juni, D. Yamamoto, Genetic analysis of chaste, a new mutation of Drosophila melanogaster characterized by extremely low female sexual receptivity. *J Neurogenet* **23**, 329-340 (2009).

139. J. R. Morris, J. L. Chen, P. K. Geyer, C. T. Wu, Two modes of transvection: enhancer action in trans and bypass of a chromatin insulator in cis. *Proc Natl Acad Sci U S A* **95**, 10740-10745 (1998).

140. D. J. Mellert, J. W. Truman, Transvection is common throughout the Drosophila genome. *Genetics* **191**, 1129-1141 (2012).

141. A. J. Blick *et al.*, The Capacity to Act in Trans Varies Among Drosophila Enhancers. *Genetics* **203**, 203-218 (2016).

142. T. Cremer, M. Cremer, Chromosome territories. *Cold Spring Harb Perspect Biol* **2**, a003889 (2010).

143. K. P. Eagen, Principles of Chromosome Architecture Revealed by Hi-C. *Trends Biochem Sci*, (2018).

144. J. Blanton, M. Gaszner, P. Schedl, Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev* **17**, 664-675 (2003).

145. C. J. Sigrist, V. Pirrotta, Chromatin insulator elements block the silencing of a target gene by the Drosophila polycomb response element (PRE) but allow trans interactions between PREs on different chromosomes. *Genetics* **147**, 209-221 (1997).

146. M. Muller, K. Hagstrom, H. Gyurkovics, V. Pirrotta, P. Schedl, The mcp element from the Drosophila melanogaster bithorax complex mediates long-distance regulatory interactions. *Genetics* **153**, 1333-1356 (1999).

147. E. Kravchenko *et al.*, Pairing between gypsy insulators facilitates the enhancer action in trans throughout the Drosophila genome. *Mol Cell Biol* **25**, 9283-9291 (2005).

148. J. R. Bateman, J. E. Johnson, M. N. Locke, Comparing enhancer action in cis and in trans. *Genetics* **191**, 1143-1155 (2012).

149. B. J. Beliveau *et al.*, Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc Natl Acad Sci U S A* **109**, 21301-21306 (2012).

150. C. Fritsch, G. Ploeger, D. J. Arndt-Jovin, Drosophila under the lens: imaging from chromosomes to whole embryos. *Chromosome Res* **14**, 451-464 (2006).

151. N. Negre *et al.*, A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet* **6**, e1000814 (2010).

152. A. M. Wood *et al.*, Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. *Mol Cell* **44**, 29-38 (2011).

153. C. T. Ong, K. Van Bortle, E. Ramos, V. G. Corces, Poly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in Drosophila. *Cell* **155**, 148-159 (2013).

154. K. Van Bortle *et al.*, Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res* **22**, 2176-2187 (2012).

155. S. Cuartero, U. Fresan, O. Reina, E. Planet, M. L. Espinas, Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *EMBO J* **33**, 637-647 (2014).

156. G. Morata, P. A. Lawrence, Development of the eye-antenna imaginal disc of Drosophila. *Dev Biol* **70**, 355-371 (1979).

157. C. B. Hug, A. G. Grimaldi, K. Kruse, J. M. Vaquerizas, Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216-228 e219 (2017).

158. C. H. Lee, T. Herman, T. R. Clandinin, R. Lee, S. L. Zipursky, N-cadherin regulates target specificity in the Drosophila visual system. *Neuron* **30**, 437-450 (2001).

159. K. J. Venken *et al.*, Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. *Nat Methods* **6**, 431-434 (2009).

160. K. J. Venken *et al.*, A molecularly defined duplication set for the X chromosome of Drosophila melanogaster. *Genetics* **186**, 1111-1125 (2010).

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

162. F. Payne, An experiment to test the nature of the variations on which selection acts. *Indiana Univ Studies* **5**, 1-45 (1918).

163. A. L. Parks *et al.*, Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. *Nat Genet* **36**, 288-292 (2004).

164. D. M. Duncan, E. A. Burgess, I. Duncan, Control of distal antennal identity and tarsal development in Drosophila by spineless-aristapedia, a homolog of the mammalian dioxin receptor. *Genes Dev* **12**, 1290-1303 (1998).

165. R. J. Johnston, Jr. *et al.*, Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the Drosophila eye. *Cell* **145**, 956-968 (2011).

166. C. H. Waddington, A note on some alleles of aristapedia. *J. Genet.* **51**, 123-129 (1952).

167. S. U. Thanawala *et al.*, Regional modulation of a stochastically expressed factor determines photoreceptor subtypes in the Drosophila retina. *Dev Cell* **25**, 93-105 (2013).

168. D. Jukam *et al.*, The insulator protein BEAF-32 is required for Hippo pathway activity in the terminal differentiation of neuronal subtypes. *Development (Cambridge, England)* **143**, 2389-2397 (2016).

169. H. Y. Hsiao *et al.*, Dissection and immunohistochemistry of larval, pupal and adult Drosophila retinas. *J Vis Exp*, 4347 (2012).

170. B. J. Beliveau *et al.*, Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat Commun* **6**, 7147 (2015).

171. M. R. S. Qikai Xu, Gregory J. Hannon, Stephen J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes. *PNAS* **106**, 2289-2294 (2008).

172. J. Yan *et al.*, Regulatory logic driving stable levels of defective proventriculus expression during terminal photoreceptor specification in flies. *Development (Cambridge, England)* **144**, 844-855 (2017).

173. S. J. Gratz, J. Wildonger, M. M. Harrison, K. M. O'Connor-Giles, CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. *Fly (Austin)* **7**, 249-255 (2013).

174. F. Port, H. M. Chen, T. Lee, S. L. Bullock, Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc Natl Acad Sci U S A* **111**, E2967-2976 (2014).

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

176. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).

177. C. Hou, L. Li, Z. S. Qin, V. G. Corces, Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* **48**, 471-484 (2012).

178. L. Li *et al.*, Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol Cell* **58**, 216-231 (2015).

179. B. Schuettengruber *et al.*, Cooperativity, specificity, and evolutionary stability of Polycomb targeting in Drosophila. *Cell Rep* **9**, 219-233 (2014).

180. M. R. Stadler, J. E. Haines, M. B. Eisen, Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. *Elife* **6**, (2017).

181. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).

182. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

183. L. A. Lettice *et al.*, A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-1735 (2003).

184. Y. Ahn, H. E. Mullan, R. Krumlauf, Long-range regulation by shared retinoic acid response elements modulates dynamic expression of posterior Hoxb genes in CNS development. *Dev Biol* **388**, 134-144 (2014).

185. G. Andrey *et al.*, A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).

186. K. S. Sandhu *et al.*, Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev* **23**, 2598-2603 (2009).

187. J. E. Phillips-Cremins, V. G. Corces, Chromatin insulators: linking genome organization to cellular function. *Mol Cell* **50**, 461-474 (2013).

188. M. Gans, Etude genetique et physiologique du mutant z de *Drosophila melanogaster*. *Bull. Biol. Fr. Belg. Suppl.* **38**, 1-90 (1953).

189. A. S. Goldsborough, T. B. Kornberg, Reduction of transcription by homologue asynapsis in Drosophila imaginal discs. *Nature* **381**, 807-810 (1996).

190. A. V. Oppenheim, R.W. Schafer, T.G. Stockham, Nonlinear filtering of multiplied and convolved signals. *Proceedings of the IEEE* **56**, 1264-1291 (1968).

191. J. Canny, A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell* **8**, 679-698 (1986).

192. C. B. Barber, D.P. Dobkin, H. Huhdanpaa. , The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software* **22**, 469-483 (1996).

193. T. F. Chan, L. A. Vese, Active contours without edges. *IEEE Trans Image Process* **10**, 266-277 (2001).

194. T. Sato, M. A. Russell, R. E. Denell, Homoeosis in Drosophila: a new enhancer of polycomb and related homoeotic mutations. *Genetics* **105**, 357-370 (1983).

195. J. Simon, A. Chiang, W. Bender, Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. *Development (Cambridge, England)* **114**, 493-505 (1992).

Kayla Viets

PROFESSIONAL PROFILE

Self-motivated cell and molecular biologist with 8+ years of experience in areas including DNA/RNA FISH, cellular imaging, RNA-seq, molecular cloning, and analysis of genomics data. Extensive innovation, communication, and leadership skills demonstrated through developing DNA Oligopaints at Johns Hopkins, authoring multiple publications, presenting at 18 conferences, and mentoring 16 students. Experienced in designing protocols to fit specific technical requirements and leading complex projects from conception to completion.

EDUCATION

Johns Hopkins University — Baltimore, MD
PhD in Cell, Molecular, Developmental Biology and Biophysics — March 2019

Northwestern University — Evanston, IL
B. A. with honors in Biological Sciences, minor in Spanish — June 2013

SKILLS AND AREAS OF EXPERIENCE

Laboratory Techniques

DNA and RNA FISH, confocal microscopy, tissue dissection and immunostaining, next generation RNA-seq, analysis of genomics datasets, tissue preparation and fixation for HiC, PCR, gel electrophoresis, DNA extraction, molecular cloning, scanning electron microscopy, RNAi, Western blotting, mammalian cell culture, RT-PCR

Software/Programming Languages

Adobe Illustrator and Photoshop, Microsoft Office, GraphPad Prism, FlyBase, ImageJ, Geneious, UCSC Genome Browser, Unix, R

Innovation, Communication, and Leadership Skills

- Developed DNA Oligopaints technique at Johns Hopkins, optimized it for use in larval fly tissue, and independently trained members of 15 labs in the technique
- 1 first author paper, 1 in revision, 2 in prep; 4 secondary author papers
- 3 funded grant proposals
- 18 previous conference presentations
- Established collaborations with multiple laboratories with expertise in bioinformatics, sequencing, HiC, and DNA/RNA FISH
- Designed projects, oversaw completion of experiments, and taught scientific techniques as primary lab mentor for 10 graduate rotation students, 5 undergraduates, and 1 high schooler

RESEARCH EXPERIENCE

Tempus Labs Chicago, IL

Variant Scientist Feb 2019-present

- Analyze tumor sequencing data from cancer patients
- Summarize somatic and inherited genetic test results and generate customized result reports for physicians
- Perform quality control functions for laboratory workflow and clinical reports
- Perform research on genes and sequence variants to determine clinical relevance
- Participate in research and development of new gene panels and tests, and updates to existing tests

Johns Hopkins University Baltimore, MD

Predoctoral Research Fellow, Johnston Lab March 2014-March 2019

- Discovered a novel mechanism by which specialized chromatin structures drive gene regulatory interactions between chromosomes using DNA FISH, RNA-seq, genetics, and analysis of HiC and ChIP datasets
- Designed and successfully generated two barcoded pools of 90,000 optimized FISH probes (total of 331 DNA and RNA targets) by improving an existing bioinformatics pipeline
- Worked with a team of four undergraduates and a fellow graduate student to determine additional DNA factors required for gene regulation between chromosomes

Graduate Rotation Student, Wendland Lab Jan-March 2014

- Examined proteins required for clathrin-independent endocytosis in yeast using molecular cloning and microscopy

Graduate Rotation Student, Bortvin Lab Oct 2013-Jan 2014

- Characterized the expression patterns of the LINE-1 transposon and its inhibitor protein, Maelstrom, using mammalian cell culture, Western blotting, and RT-PCR

Graduate Rotation Student, Chen Lab Aug-Oct 2013

- Tracked histone localization patterns in the fruit fly female germline using RNAi, immunohistochemistry, and genetics

Northwestern University Evanston, IL

Undergraduate Research Assistant, Brickner Lab Sept 2010 – June 2013

- Identified DNA elements responsible for targeting the *GAL1* gene to the nuclear periphery in *Saccharomyces cerevisiae* using molecular cloning and immunohistochemistry

HONORS AND AWARDS

- $87,000 National Institutes of Health Ruth L. Kirschstein National Research Service Award (NRSA) Predoctoral Fellowship

- Department of Biology Poster Award, Johns Hopkins University
- Poster Award, Sciences category, Northwestern University Undergraduate Research and Arts Exposition
- $3000 Northwestern University Weinberg College Grant
- $3000 Northwestern University Summer Undergraduate Research Grant

TEACHING EXPERIENCE

Johns Hopkins University Baltimore, MD
Teaching Assistant, Department of Biology Aug 2014-May 2015

- Led laboratory classes for undergraduate Genetics and Development courses

OUTREACH/VOLUNTEER EXPERIENCE

STEM Achievement in Baltimore Elementary Schools Baltimore, MD
STEM Mentor Aug 2015-June 2018

- Guided elementary schoolers in underserved communities in monthly activities to expand their knowledge of scientific problem-solving
- Helped students use STEM knowledge to develop engineering projects to address issues in their communities

Northwestern University Evanston, IL
Facilitator, Science Research Workshop Oct 2011-June 2013

- Led weekly workshops to train fellow students in joining a lab, writing grants, obtaining research funding, and presenting their research
- All students mentored received research fellowships

REVIEWER EXPERIENCE

Peer reviewer for: *PLOS One, Genetics*

PUBLICATIONS

Viets, K., Sauria, M., Chernoff, C., Anderson, C., Tran, S., Dove, A., Goyal, R., Voortman, L., Gordus, A., Taylor, J., and Johnston, R.J., Jr. "TADs pair homologous chromosomes to promote interchromosomal gene regulation." *In revision at Developmental Cell.* bioRxiv doi: https://doi.org/10.1101/445627

Viets, K.*, Urban, E.*, Chernoff, C.*, Han, J., Chen, A., Anderson, C., Tran, S., Konzman, D., and R.J. Johnston, Jr. "Activating and repressing stochastic gene expression between chromosomes." *Manuscript in preparation for submission to Developmental Cell.*
*indicates equal contribution

Viets, K., and R.J. Johnston, Jr. "Probing chromosome pairing and transvection to understand nuclear organization." *Manuscript in preparation*.

Duan, H., de Navas, L.F., Hu, F., Sun, K., Mavromatakis, Y.E., **Viets, K.**, Zhou, C., Kavaler, J., Johnston, R.J., Tomlinson, A., Lai, E.C. "The *mir-279/996* cluster represses receptor

tyrosine kinase signaling to determine cell fates in the *Drosophila* eye." 2018. *Development,* 145: doi: 10.1242.

Yan, J., Anderson, C., **Viets, K.**, Tran, S., Goldberg, G., Small, S., and Johnston, R.J., Jr. "Regulatory logic driving stable levels of *defective proventriculus* expression during terminal photoreceptor specification in flies." 2017. *Development,* 144:844-855.

Viets, K.*, Eldred, K.C.*, and Johnston, R.J., Jr. "Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates." 2016. *Trends Genet,* 32: 638-659. *indicates equal contribution

Brickner, D.G., Sood, V., Tutucci, E., Coukos, R., **Viets, K.**, Singer, R.H., and Brickner, J.H. "Subnuclear positioning and interchromosomal clustering of the *GAL1-10* locus are controlled by separable, interdependent mechanisms." 2016. *Mol Biol Cell*, 27: 2980-2993.

Jukam, D., **Viets, K.**, Anderson, C., Zhou, C., DeFord, P., Yan, J., Cao, J., Johnston, R.J., Jr. "The insulator protein BEAF-32 is required for Hippo pathway activity in the terminal differentiation of neuronal subtypes." 2016. *Development*, 143: 2389-2397.

CONFERENCE PRESENTATIONS

Research Talks

- Northeast Regional Chromosome Pairing Conference — Oct 2018
  Sudbury, Ontario, Canada
- Mid-Atlantic Regional Chromosome Pairing Conference — July 2018
  Philadelphia, PA
- CSHL Nuclear Organization and Function Meeting — May 2018
  Cold Spring Harbor, NY
- Genetics Society of America Drosophila Research Conference — April 2018
  Philadelphia, PA
- Northeast Regional Chromosome Pairing Conference — Oct 2017
  Brunswick, ME
- Mid-Atlantic Regional Chromosome Pairing Conference — June 2017
  Baltimore, MD
- Johns Hopkins Chromatin and Chromosomes Workshop — June 2017
  Baltimore, MD
- Northeast Regional Chromosome Pairing Conference — Oct 2016
  Boston, MA
- Mid-Atlantic Regional Chromosome Pairing Conference — June 2016
  Baltimore, MD
- Northeast Regional Chromosome Pairing Conference — Oct 2015
  Boston, MA
- Northeast Regional Chromosome Pairing Conference — Oct 2014
  Boston, MA

Poster Presentations

- 82nd Cold Spring Harbor Laboratory Symposium on Quantitative Biology: Chromosome Segregation and Structure — June 2017
  Cold Spring Harbor, NY
- Genetics Society of America Drosophila Research Conference — April 2017
  San Diego, CA
- Johns Hopkins University Department of Biology Retreat — Oct 2016
  Hershey, PA
- Genetics Society of America Allied Genetics Conference — July 2016
  Orlando, FL
- Johns Hopkins University Department of Biology Retreat — Oct 2015
  St. Michaels, MD
- Genetics Society of America Drosophila Research Conference — Mar 2015
  Chicago, IL
- Johns Hopkins University Department of Biology Retreat — Oct 2014
  Fairfield, PA