8-2013

Chromatin Insulators: Master Regulators of the Eukaryotic Genome

Todd Andrew Schoborg
tschobor@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Todd Andrew Schoborg entitled "Chromatin Insulators: Master Regulators of the Eukaryotic Genome." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biochemistry and Cellular and Molecular Biology.

Mariano Labrador, Major Professor

We have read this dissertation and recommend its acceptance:

Bruce McKee, Albrecht von Arnim, Ranjan Ganguly, Juan Jurat-Fuentes

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Chromatin Insulators: Master Regulators Of The Eukaryotic Genome

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Todd Andrew Schoborg

August 2013

ACKNOWLEDGEMENTS

This work would not have been possible without the guidance of a number of individuals. First and foremost, I would like to thank my advisor, Dr. Mariano Labrador, for his advice and direction throughout my tenure. Starting from the first day of my rotation, he provided steady leadership and encouragement while treating me as a colleague rather than a student, for which I will always be grateful. Dr. Howard Whiteman, my former undergraduate mentor, probably deserves much of the credit for where I am today, as if it weren't for him giving me my first opportunity to do research, along with his constant enthusiasm for science and encouragement that I could succeed as an academic, I would have found something else to do entirely. Dr. Hyuck-Joon Kang, a former lab member, was a model citizen for how graduate students should conduct themselves and I cannot thank him enough for the positive influence he had on me early in my graduate career. Dr. Tom Dockendorff also deserves mention for being an excellent resource of Drosophila Biology knowledge. Additionally, I have been fortunate to have had a number of talented undergraduate researchers assist me in many aspects of my work. Ryan Rickels and Josh Barrios, who have both moved on with their own graduate studies at highly respected institutions deserve particularly mention, as their hard work and enthusiasm for science have set a high standard for what undergraduate researchers are capable of achieving. I would also like to thank my committee members, Drs. Bruce McKee, Albrecht von Arnim, Ranjan Ganguly and Juan Jurat-Fuentes along with past and current members of the Labrador Lab, including Dr. Heather Wallace, Dr. Shaofei Zhang, Dr. Srilalitha Kuruganti, Cherie Hsu and Ran An for collaboration and discussion.

Also, I would like to thank the National Science Foundation (NSF) for providing me with a Graduate Research Fellowship for the last three years of my tenure here, and I am required to provide the following disclosure statement: “any opinions, findings, and conclusions or recommendations expressed in this dissertation are mine and do not necessarily reflect the views of the National Science Foundation”.

Finally, there are also a number of other people who did not have a direct role in the scientific aspect of this work, yet have been equally important in its success. My Mom and Dad have provided unwavering support and encouragement during my time here and are solely responsible for instilling values of hard work, grit and determination, all of which have served me well as a scientist. Additionally, I will always be grateful to Andrew Batman West and Michael O'Brien for constantly reminding me not to take life too seriously, Nate Jones for inspiration and last, but certainly not least, my wife, Dr. Rebecca Kendall Peebles. Her constant support, patience and understanding during this time has been remarkable and I will forever owe her a debt of gratitude for putting up with me the last 5 years.

ABSTRACT

Proper organization of the chromatin fiber within the three dimensional space of the eukaryotic nucleus relies on a number of DNA elements and their interacting proteins whose structural and functional consequences exert significant influence on genome behavior. Chromatin insulators are one such example, where it is thought that these elements assist in the formation of higher order chromatin loop structures by mediating long-range contacts between distant sites scattered throughout the genome. Such looping serves a dual role, helping to satisfy both the physical constraints needed to package the linear DNA polymer within the small volume of the nucleus while simultaneously orchestrating or excluding contacts between regulatory elements, such as enhancers and promoters, in order to direct the proper gene regulatory outputs needed to maintain cellular homeostasis. As a result of its central role in chromatin structure, insulators have been linked to a number of nuclear processes, although many aspects of their biology remain unanswered. The collection of work presented here addresses three of these concerns. Chapter I outlines the phylogenetic distribution of these elements, highlighting the lineage specificity of the *Drosophila melanogaster* insulator protein BEAF-32 and suggesting that insulator function poses a more significant agent for selection than conservation of the proteins themselves. Chapter II addresses a central debate in the insulator field regarding the function of insulator bodies, exposing an unexpected link between their formation, osmotic stress and cell death, while disproving the prevailing hypothesis set forth over a decade ago that essentially formed the foundation for how these elements function *in vivo*. Finally, in Chapter III, their contribution to inter-allelic complementation, or transvection, is addressed, where context- and dose-dependent effects on enhancer-promoter communication in *trans* were observed, suggesting

that chromatin structure is the ultimate determinant of whether enhancer-promoter communication in *trans* leads to a sustained transcriptional output. Such findings provide a new perspective for a classic genetic phenomenon while highlighting a conserved feature of genome function. Taken collectively, this body of work reflects the broad nuclear functions attributed to these elements and suggests that chromatin insulators function as master regulators of the eukaryotic genome.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ATTACHMENTS

INTRODUCTION

Chromatin is biochemically defined as an assembly of DNA and associated proteins that exert both structural and functional influences on the genome. This complex, along with a network of other nuclear factors, help package linear DNA polymers within the three dimensional space of the nucleus, ensure chromosome fidelity throughout the cell cycle and coordinate gene regulatory programs in a precise spatio-temporal manner. Chromatin, or 'stainable material', was first coined by the German anatomist Walther Flemming in the late 1800's to describe the filamentous, matrix-like scaffold structures observable in eukaryotic nuclei following treatment with basophilic dyes (Flemming, 1882). While the dynamic morphological changes observed for these structures were important for elucidating the fundamental principles underlying mitosis and cell division, Flemming could not have imagined that chromatin would ultimately hold the key to understanding heredity and the functional behavior of eukaryotic genomes. Although the wealth of chemical information generating by whole genome sequencing projects has provided an unparalleled view of the blueprint of life, understanding how a single genome can generate the multitude of differentiated cell types in a multicellular organism ultimately requires knowledge of how DNA-interacting proteins interpret this chemical information to modulate genome dynamics accordingly. Indeed, chromatin has been shown to play a key role in tissue homeostasis, stem cell biology and cancer while serving as the de facto member in epigenetic inheritance (Frye and Benitah, 2012; Suva et al., 2013; Tollervey and Lunyak, 2012; Watanabe et al., 2013), and as a result has become a central focus for research aimed at understanding how genomes function *in vivo* within the three dimensional space of the nucleus.

Recently, the Encyclopedia of DNA Elements Consortium (ENCODE) was established to address this very goal by mapping all the functional elements of various eukaryotic genomes, providing a comprehensive view of chromatin structure and function (Celniker et al., 2009; Dunham et al., 2012). To date, a variety of transcription factor and regulatory protein binding sites, histones and other chromatin proteins have been mapped in humans (*Homo sapiens*) (Gerstein et al., 2012; Neph et al., 2012; Sanyal et al., 2012; Thurman et al., 2012; Whitfield et al., 2012), fruit flies (*Drosophila melanogaster*) (Negre et al., 2011; Negre et al., 2010; Kharchenko et al., 2011) and nematodes (*Caenorhabditis elegans*) (Henikoff et al., 2011; Liu et al., 2011; Niu et al., 2011), resulting in an enormous amount of descriptive data that has been a critical first step towards understanding conserved mechanisms of eukaryotic genome function. As a result, nearly 80% of the underlying DNA sequence from humans now has at least one biochemical function associated with it (Dunham et al., 2012), while from an evolutionary perspective, comparisons of datasets have provided a glimpse into the properties that potentially differentiate one species from the next (Brooks et al., 2011; Deng et al., 2011; Spivakov et al., 2012). Additionally, subsequent data mining and bioinformatic analysis has been crucial for identifying cell-type specific differences in chromatin, such as transcription factor occupancy, nucleosome position and density, histone modifications and long-range looping contacts that have provided a framework for understanding how a single genome can direct multiple gene regulatory programs to establish cell specificity, when coupled with genome-wide mRNA transcription profiles (Bushey et al., 2009; Cherbas et al., 2011; Dong et al., 2012; Frietze et al., 2012; Heintzman et al., 2009; Natarajan et al., 2012; Ni et al., 2012; Nordman et al., 2011; Riddle et al., 2011; Spencer et al., 2010; Steiner et al., 2012; Arvey et al., 2012; Hou et al.,

2010b; Junier et al., 2012; Kagey et al., 2010; Kundaje et al., 2012; Ren et al., 2012; Vernot et al., 2012; Wang et al., 2012; Phillips-Cremin et al., 2013). However, it should be noted that most of these findings are the result of correlation rather than causation, a drawback further compounded by the fact that a number of the chromatin/regulatory elements implemented in cell-type differences are poorly understood in terms of their *in vivo* function, limiting our understanding of their influences on the genome, and ultimately, how the genome functions *in vivo.*

Chromatin insulators are one group of functional elements whose nuclear role remains poorly understood. These short DNA motifs display two properties when bound by insulator proteins: enhancer blocking and barrier/boundary activity (Wallace and Felsenfeld, 2007; West et al., 2002; Yang and Corces, 2012). Placement of a protein-bound insulator sequence between a regulatory enhancer and gene promoter disrupts communication between the two elements, in addition to preventing the spread of repressive heterochromatin along the chromatin fiber. However, both of these properties were described in Drosophila based on transgenic reporter systems with insulators taken out of their genomic context (Kellum and Schedl, 1991; Kellum and Schedl, 1992), limiting our ability to understand the *in vivo* role of these elements. High-throughput ChIP-Seq studies have since mapped thousands of endogenous insulators located throughout eukaryotic genomes (Bushey et al., 2009; Cuddapah et al., 2009; Jiang et al., 2009; Kim et al., 2007; Negre et al., 2010), where protein-protein contacts between distant insulator sites lead to the formation of higher order chromatin loop structures (Blanton et al., 2003; Hou et al., 2012; Hou et al., 2008; Kurukuti et al., 2006; Sexton et al., 2012). Such loops are thought to

be critical for partitioning the genome into domains that serve both structural and functional purposes, acting to optimize the physical constraints required for packaging DNA within the small volume of the nucleus while allowing for proper execution of gene regulatory programs, thus making chromatin insulators potential master regulators of genome organization and nuclear dynamics.

Early observations of chromatin structure in the interphase nucleus lead to the suggestion that higher order chromatin might form discrete structures that represent independent domains of active or repressed genes (Benyajati and Worcel, 1976; Weisbrod, 1982). Active genes were thought to be located in euchromatic regions of the genome, which consist of loosely packaged chromatin, while silent genes were thought to occupy densely packaged heterochromatic regions. Initially, chromatin insulators were thought to be responsible for establishing and maintaining these domains, functioning as boundary elements that could buffer against position effects resulting from the stochastic spread of heterochromatin along the chromatin fiber (Kellum and Schedl, 1991; Udvardy et al., 1985). Subsequent immunostaining of Drosophila polytene chromosomes supported this hypothesis, in which insulator proteins appeared to demarcate the boundaries between euchromatic interbands and heterochromatic bands (Labrador and Corces, 2002; Wallace et al., 2010; Zhao et al., 1995). Additionally, the promiscuous behavior of enhancers, known to be able to communicate with both distal and proximal promoters, sometimes even over hundreds of kilobases (Bellen et al., 1989; O'Kane and Gehring, 1987), requires that other regulatory mechanisms be in place in order to prevent misexpression. Indeed, insulators were also shown to restrict enhancer-promoter communication in a directional manner

when located between the two elements (Geyer and Corces, 1992; Kellum and Schedl, 1992), with most efforts since then focused on understanding how the ability to mediate chromatin looping might account for both of these properties (Blanton et al., 2003; Cai and Shen, 2001; Hou et al., 2012; Melnikova et al., 2004; Muravyova et al., 2001).

Although DNA sequences that confer insulator activity have been identified in nearly all eukaryotes, including yeast, plants, sea urchins and vertebrates (Chung et al., 1993; Chung et al., 1997; Farrell et al., 2002; Guo et al., 2008; Hily et al., 2009; Ishii et al., 2002; Ishii and Laemmli, 2003; Palla et al., 1997; Yang et al., 2011), most insulators identified thus far have come from Drosophila given its amenable genetic tools and polytene chromosomes. The *gypsy, scs/scs'* and *Fab7* insulators were the first insulators to be identified, with *gypsy* being the most well-characterized of the three (Galloni et al., 1993; Gdula et al., 1996; Geyer and Corces, 1992; Gyurkovics et al., 1990; Karch et al., 1994; Kellum and Schedl, 1991; Kellum and Schedl, 1992; Modolell et al., 1983; Parkhurst and Corces, 1986; Parkhurst et al., 1988; Udvardy et al., 1985; Zhao et al., 1995). Although high-throughput studies have since revealed the existence of thousands of insulator sites throughout the Drosophila genome (Bushey et al., 2009; Negre et al., 2010), almost everything known about insulator function has derived from extensive characterization of these three elements.

The *gypsy* insulator, located in the 5' LTR of the *gypsy* retrotransposon, was originally cloned and identified based on mutant analysis of the *yellow*^{2} *(y*^{2}*)* allele. Flies carrying *y*^{2} lack pigment in the body and wing as a result of a *gypsy* insertion between the wing and body enhancers and

the *yellow* promoter, a phenotype which was suppressed in a *suppressor of Hairy wing* s*u(Hw)* mutant background (Geyer and Corces, 1992; Modolell et al., 1983; Parkhurst and Corces, 1986). This suggested that the *gypsy* insulator could block enhancer-promoter communication through a mechanism dependent on the presence of a functional Su(Hw) protein. s*u(Hw)* was later shown to encode a DNA binding protein containing 12 zinc finger domains, which recognize a 367 bp segment containing 12 copies of a 12-bp motif near the 5' LTR of *gypsy* (Parkhurst et al., 1988; Spana et al., 1988). However, further mutant analysis in y^2 and cut^6 (ct^6), also the result of a *gypsy* insertion between the *ct* enhancer and promoter, uncovered two other genes that are necessary for the enhancer blocking ability of the *gypsy* insulator, *Mod(mdg4)* and *CP190.* The 67.2 isoform of Modifier of mdg4 (Mod(mdg4)67.2) (Buchner et al., 2000) physically interacts with Su(Hw) (Gdula and Corces, 1997; Ghosh et al., 2001) and acts as an enhancer of position effect variegation, imparting directionality on the enhancer blocking ability of Su(Hw) (Georgiev and Gerasimova, 1989; Gerasimova et al., 1995). Centrosomal Protein 190 (CP190), originally identified as a constituent of the pericentriolar material (PCM) of centrosomes during M-phase but later shown to have a nuclear function as well, is also required for *gypsy* function and physically interacts with both Su(Hw) and Mod(mdg4)67.2 (Butcher et al., 2004; Pai et al., 2004; Whitfield et al., 1988; Whitfield et al., 1995) . Two other genes, *dTOPORS* and *Ey(2)/Sus1*, have also been shown to have a more nuanced influence on *gypsy* activity with Su(Hw). Mutations in *Ey(2)/Sus1*, a highly conserved eukaryotic transcription factor, have been shown to affect the barrier activity of Su(Hw) without interfering with its ability to block enhancer-promoter communication (Kurshakova et al., 2007). *dTOPORS*, a ubiquitin ligase that is required for the enhancer blocking ability of the *gypsy* insulator, is

thought to mediate the interaction of Su(Hw) with the nuclear lamina (Capelson and Corces, 2005; Ramos et al., 2011).

However, genome-wide analysis for *gypsy* proteins suggests a much more complex regulatory landscape. Su(Hw) offers perhaps the most perplexing question as to its *in vivo* function as an insulator protein outside of *gypsy* (Bushey et al., 2009; Negre et al., 2010). Su(Hw) localizes mainly to intergenic regions and is often the only known insulator protein present, where a handful of these 'endogenous' Su(Hw) binding sites possess weak or non-functional enhancer blocking ability in transgenic assays (Schwartz et al., 2012). Additionally, the few genes that are associated with Su(Hw) show no commonality with regard to molecular function, unlike other insulator proteins. This has lead to the suggestion that Su(Hw) functions as a general facilitator of higher order chromatin structure, acting to organize the chromatin fiber into broad domains that may be further subdivided by the action of the other insulators (Bushey et al., 2009; Negre et al., 2010; Sexton et al., 2012). As for CP190, although some sites overlap with Su(Hw) and Mod(mdg4)67.2, the majority colocalize with two other insulator proteins, BEAF-32 and CTCF within or near active promoters (Bartkuhn et al., 2009; Bushey et al., 2009). Furthermore, CP190 recruitment to chromatin has been shown to be regulatable during heat shock and the ecdysone response, leading to an alteration in chromatin looping contacts that dictate expression at target loci depending on the chromatin-bound state of the protein (Wood et al., 2011). These findings, along with the fact that CP190 overlaps with all of the other insulator proteins at at least one genomic location, suggests that it may function as the master mediator of interactions

between different insulator-bound proteins in order to form specific higher-order chromatin structures.

In addition to Su(Hw), CP190 and Mod(mdg4)67.2, four other major insulator proteins have been identified in Drosophila that bind to sequences other than *gypsy.* GAGA Factor, originally identified as a Trithorax-like protein that regulates homeotic gene expression by antagonizing the repressive behavior of the Polycomb Group proteins (PcGs), binds to the *Fab-7* and *Mcp* insulator elements and has been shown to modulate the behavior of the *gypsy* insulator and mediate insulator-bypass through interactions with Mod(mdg4)67.2 (Busturia et al., 2001; Cavalli and Paro, 1998; Farkas et al., 1994; Gerasimova and Corces, 1998; Melnikova et al., 2004; Strutt et al., 1997). BEAF-32 and Zw5 were identified following characterization of the boundary elements at the 87A7 chromomere (*scs'* and *scs*), while *in vivo* interactions between the two proteins bound to these elements provided the first physical evidence for endogenous insulator pairing to generate chromatin loop structures (Blanton et al., 2003). BEAF-32, bound to *scs'* and Zw5 (*scs)*, are required to confer both enhancer blocking and boundary activity of these elements in transgenic assays (Gaszner et al., 1999; Kellum and Schedl, 1991; Kellum and Schedl, 1992; Udvardy et al., 1985; Zhao et al., 1995). Two isoforms of BEAF-32 (A & B) have been identified that differ only in their zinc finger DNA-binding domain at the N-terminus (Hart et al., 1997), and both localize to hundreds of endogenous insulator sites on polytene chromosomes at the interface between bands and interbands, in agreement with their initial characterization as boundary elements (Cuvier et al., 1998; Zhao et al., 1995). BEAF-32 has also been shown to interact with components of the nuclear lamina (Pathak et al., 2007). However,

high-throughput ChIP-Seq has revealed that BEAF-32 localizes primarily to gene promoters and 5' UTRs particularly for genes located in a head-to-head fashion along the chromosome, acting to maintain specific histone marks that are thought to be conducive for transcription. Many of these genes are tightly regulated and are involved in cell cycle progression, cell polarity, proliferation and differentiation, whose misregulation in a *BEAF-32* null background leads to unrestricted cell growth and formation of neoplastic tumors (Bushey et al., 2009; Emberly et al., 2008; Gurudatta et al., 2013; Jiang et al., 2009; Negre et al., 2010; Roy et al., 2007; Yang et al., 2012). Whether these phenotypes result from impairment of insulator function or a more direct role in gene regulation remains to be elucidated, although tissue-specific expression of a dominant-negative form of BEAF-32 leads to abrupt changes in overall chromatin structure in interphase nuclei and likely reflects a combination of some or all of these factors (Gilbert et al., 2006; Roy et al., 2007) .

Lastly, the CCCTC-Binding Factor (CTCF) is the only insulator protein found in both Drosophila and higher vertebrates (Schoborg and Labrador, 2010). The 5' hypersensitive site (5'HS4) of the chicken β-globin locus was the first vertebrate insulator element to be characterized, where it was shown in transgenic assays to block enhancer-promoter communication in a directional manner and protect the *Drosophila* mini-white gene from position effects (Chung et al., 1993). Further analysis of the 5'HS4 revealed the presence of a binding site for CTCF (Bell et al., 1999; Chung et al., 1997). To date, all vertebrate insulators that have been characterized bind CTCF. In *D. melanogaster,* the CTCF ortholog (dCTCF) was characterized and identified as the protein component of the *Fab-8* insulator from the Bx-C Hox

complex (Moon et al., 2005). dCTCF was later identified as the major protein component of other *Drosophila* insulators located within the Hox complex, such as *Fab-7* and *Fab-2* (Holohan et al., 2007), and immunostaining of polytene chromosomes and genome-wide ChIP analysis has also revealed thousands of other binding sites located throughout the Drosophila genome (Bushey et al., 2009; Gerasimova et al., 2007; Negre et al., 2010). Also, similar to its role with Su(Hw) at *gypsy* insulators, CP190 is also found at a large number of CTCF sites, where it appears to be important for either recruitment and/or stable binding of CTCF (Mohan et al., 2007).

Much like Su(Hw), the *in vivo* role of vertebrate and Drosophila CTCF remains enigmatic. The *in vivo* binding sites for CTCF has been mapped in humans (Cuddapah et al., 2009; Kim et al., 2007) in addition to *Drosophila* (Bushey et al., 2009; Negre et al., 2010; Smith et al., 2009). CTCF in humans had been linked to both transcriptional activation and repression (Baniahmad et al., 1990; Vostrov and Quitschke, 1997) before its role as an enhancer blocker was elucidated (Bell et al., 1999; Chung et al., 1993). In mice, CTCF has been shown to be responsible for loop formation at the *H19/Igf2* imprinting and β-globin loci that direct cell-specific contacts between appropriate regulatory elements (Hou et al., 2008; Kurukuti et al., 2006), while analysis of active and repressed regions of the human genome in relation to CTCF binding sites revealed distinct regions of histone modifications sharply demarcated by the protein and it has also been shown to establish CG-methylation domains (Bell and Felsenfeld, 2000; Cuddapah et al., 2009; Holohan et al., 2007; Kim et al., 2007). In Drosophila, a similar broad role in chromatin structure and gene function has been established, with dCTCF in many cases found within 200 bp of gene

promoters, particularly between divergently transcribed genes that play a role in developmental processes (Bushey et al., 2009; Smith et al., 2009). However, many other CTCF binding sites appear far from gene promoters, suggesting that these sites might function as traditional chromatin boundaries, such as in the *Drosophila* Hox complex (Holohan et al., 2007). Furthermore, recent high throughput analysis of dCTCF binding sites in conjunction with Histone H3 Lys27 trimethylation (H3K27me3) modifications and RNAi-depletion of insulator proteins revealed a positive role for dCTCF in maintaining these silenced regions of chromatin, analogous to its behavior in vertebrate cells (Cuddapah et al., 2009; Kim et al., 2007; Van Bortle et al., 2012). This suggests that the function of CTCF is highly conserved in both flies and vertebrates and has lead others to suggest that all of these functions (transcriptional activation/repression, insulation/boundary activity) derive from a common mechanism involving higher-order chromatin loop formation mediated by this protein (Phillips and Corces, 2009).

Taken collectively, these findings suggests that chromatin insulators play a central role in both genome and nuclear function. However, although the massive influx of high-throughput data has been critical in potentially uncovering the true *in vivo* role of these elements, many questions remain unanswered. This is partly due to the lack of a hypothesis-driven approach in most of these studies, providing instead a collection of descriptive data whose interpretation relies primarily on correlation. The collection of work provided here addresses a number of issues regarding insulator function from a hypothesis-first perspective, consisting of: (1) computational approaches aimed at elucidating the phylogenetic distribution of these elements; (2) resolving a long-standing debate within the insulator field by identifying the physiological basis for insulator

body formation; and (3) testing the relative contribution of these elements in transvection, a phenomenon resulting from enhancer-promoter communication between homologous chromosomes. Though seemingly disparate, each chapter is instead a reflection of the breadth of molecular functions these elements possess, an approach that highlights chromatin insulators as master regulators of the eukaryotic genome.

CHAPTER 1

The Phylogenetic Distribution Of Non-CTCF Insulator Proteins Is Limited To Insects And Reveals That BEAF-32 Is *Drosophila* Lineage Specific

This chapter has been published:

Schoborg T.A., and M Labrador. 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.

My contributions included: (1) devising experiments, (2) performing experiments, (3) data collection and analysis, (4) writing the manuscript and making figures. Mariano Labrador assisted with (1) and to a lesser extent (4). Only small revisions to the original figures have been made for the purposes of this dissertation.

Abstract

Chromatin insulators are DNA sequences found in eukaryotes that may organize genomes into chromatin domains by blocking enhancer-promoter interactions and preventing heterochromatin spreading. Considering that insulators play important roles organizing higher order chromatin structure and modulating gene expression, very little is known about their phylogenetic distribution. To date, six insulators and their associated proteins have been characterized, including Su(Hw), Zw5, CTCF, GAF, Mod(mdg4) and BEAF-32. However all insulator

proteins, with the exception of CTCF, which has also been identified in vertebrates and worms, have been exclusively described in *Drosophila melanogaster*. In this work, we have performed database searches utilizing each *D. melanogaster* insulator protein as a query to find orthologs in other organisms, revealing that except for CTCF all known insulator proteins are restricted to insects. In particular, the boundary element-associated factor of 32 kDa (BEAF-32), which binds to thousands of sites throughout the genome, was only found in the Drosophila lineage. Accordingly, we also found a significant bias of BEAF-32 binding sites in relation to transcription start sites (TSSs) in *D. melanogaster* but not in *Anopheles gambiae, Apis mellifera,* or *Tribolium castaneum*. These data suggest that DNA binding proteins such as BEAF-32 may have a dramatic impact in the genome of single evolutionary lineages. A more thorough evaluation of the phylogenetic distribution of insulator proteins will allow for a better understanding of whether the mechanism by which these proteins exert their function is conserved across phyla and their impact on genome evolution.

Introduction

Boundary elements, or insulators, are specific DNA sequences that when bound by insulator proteins play important roles in gene regulation, chromatin packing and nuclear organization. Insulators possess two important properties: position-dependent enhancer blocking and barrier activity. When placed between an enhancer sequence and a promoter sequence, protein-bound insulators are able to repress transcription by blocking promoter-enhancer communication in a directional manner, whereas when flanking transgenes, they are able to prevent heterochromatic spreading into gene loci and therefore offer protection from position effects (West et al., 2002). Recent evidence suggests that insulators are able to accomplish both roles by organizing the

chromatin fiber into higher-order structures (or domains) by initiating long-range contacts between insulator elements located throughout the genome, functionally forming a series of chromatin loops that might establish specific gene expression domains, preventing promiscuous enhancers located outside these domains from activating the promoters within (Wallace and Felsenfeld, 2007).

In Drosophila, five main proteins (and their associated protein factors) have been identified that bind to and give insulators their functional properties: Suppressor of Hairy Wing [Su(Hw)], dCTCF, GAGA Factor (GAF), boundary element associated factor of 32 kDa (BEAF-32), and Zeste-white 5 (Zw5), nearly all of which have been shown to be able to organize intervening chromatin into loop structures possibly via both homotypic and heterotypic interactions (Bushey et al., 2008). A variety of proteins have been found to be important for proper Su(Hw) insulator function, including CP190, Mod(mdg4)67.2, dTOPORS, and E(y)2/Sus1 (Capelson and Corces, 2005; Gerasimova et al., 1995; Kurshakova et al., 2007; Pai et al., 2004). Furthermore, CP190 has since been shown to be critical for the insulating function of dCTCF (Gerasimova et al., 2007; Mohan et al., 2007). CP190 was also found to be associated with BEAF-32 at a subset of BEAF-32 binding sites, leading to a model in which CP190 acts as a master facilitator of higher-order chromatin structure by organizing BEAF-32, Su(Hw) and dCTCF-bound insulators into intricate gene expression domains (Bushey et al., 2009). Interestingly, the only insulator protein that has a functional ortholog in vertebrates is CTCF, which like its Drosophila counterpart (Moon et al., 2005) has been shown to form long-range interactions in vertebrate cells (Ling et

al., 2006). However, the actual mechanism by which these insulators are able to organize chromatin into higher order domains is still poorly understood.

The boundary element-associated factor of 32 kDa (BEAF-32) plays an important role in regulating gene expression by modulating higher order chromatin structure, in accordance with its role as an insulator protein (Jiang et al., 2009). BEAF-32 was originally characterized as a novel protein with a high affinity for the *scs'* boundary element located distally to *hsp70* at the 87A7 locus in Drosophila, where it localizes to one end of the transcriptionally active puff incurred during heat shock (Udvardy et al., 1985; Zhao et al., 1995). Further characterization of the protein revealed the presence of two isoforms, BEAF-32A and BEAF-32B that differ only in the first 80 amino acids located at the N-terminus. Both N-termini possess an unusual zinc finger that is used for DNA binding (Hart et al., 1997), termed the BED-finger, after the two Drosophila proteins from which it was identified (BEAF-32 and DREF). This domain is characterized by an atypical C_2H_2 zinc-coordinating motif ($Cx_2Cx_nHx_{3\text{-}5}$[H/C]) that is predicted to form a zinc finger flanked by a conserved pair of aliphatic/aromatic residues located near the N-terminal portion of the BED domain (Aravind, 2000). Both isoforms also contain an identical C-terminus, which harbors another unusual domain, called the BESS domain that is predicted to form two or three alpha helices and is responsible for facilitating protein-protein interactions (Bhaskar and Courey, 2002; Hart et al., 1997) as well as a nuclear matrix binding domain (Pathak et al., 2007). Furthermore, the BESS domain has only been identified and characterized in five other Drosophila proteins (Bhaskar and Courey, 2002; Clark and McKearin, 1996; Cutler et al., 1998; Delattre et al., 2002; England et al., 1992; Reuter et al., 1990). BEAF-32's critical

role in regulating gene expression via chromatin organization in Drosophila is underscored by the fact that a dominant-negative form of *BEAF-32* which lacks the BED domain is embryonic lethal, and tissue-specific expression of the mutant protein leads to abrupt changes in chromatin structure (Gilbert et al., 2006).

Recent genome-wide analysis of BEAF-32 isoforms utilizing both high-throughput ChIP-chip techniques in parallel with computational analysis has revealed an interesting pattern of BEAF-32 distribution throughout the Drosophila genome, providing crucial insight into the role of BEAF-32 as well as how the protein may have shaped genome organization during the diversification of the Drosophila lineage. Bushey et al. (Bushey et al., 2009) and Jiang et al. (Jiang et al., 2009) found that the majority of BEAF-32 binding sites were highly enriched near the transcription start site (TSS) of actively transcribed target genes. Expression of BEAF-32-associated genes also decreased in BEAF-32 mutants, revealing BEAF-32 as a transcriptional activator at a subset of genes presumably by providing a chromatin environment conducive for transcription.

Herein we provide evidence that both isoforms of BEAF-32 are highly conserved within all 12 Drosophila species that have had their whole genomes sequenced (Clark et al., 2007), but absent in all other taxa, ranging from other insects to vertebrates. Both BEAF-32 isoforms appear to be functional in all Drosophilids, given their high sequence similarity, particularly within their DNA-binding BED, nuclear matrix-binding and protein-interacting BESS domains. Furthermore, we find a significant correlation between clusters of BEAF-32 binding motifs

relative to transcription start sites in Drosophila compared to other insect species. We also provide data regarding the distribution of other insulator proteins found in Drosophila across a wide range of eukaryotic taxa that suggests nearly all of the insulator proteins with the exception of CTCF are limited to insects. Although the enhancer-blocking and heterochromatin boundary mechanisms of insulators are still unknown, our results suggest that even if a common insulator mechanism exists, different insulator proteins perform this function across phyla, to facilitate chromatin structure and regulate gene expression in perhaps a species-specific manner.

Results

The Taxonomic Distribution Of D. Melanogaster Insulator Protein Orthologs Suggests That Nearly All Of Them Are Insect Specific

We were interested in the possibility that the proteins identified in *D. melanogaster* that confer sequence specific insulator activity (CTCF, Su(Hw), Mod(mdg4)67.2, CP190, BEAF-32, GAF, and Zw5) might also function in other eukaryotic taxa. We therefore searched for orthologs in a wide variety of eukaryotic organisms in which whole-genome sequence information was available (Figure A1 in Appendix). Ortholog searches revealed that the distribution of insulator proteins appears to be restricted to insects, with the exception of CTCF (Table 1). It is noticeable, however, that this distribution in both insects and in other eukaryotes is not widespread. Particularly, using our search methods we could not identify CTCF in the honey bee, *Apis mellifera*, or in *Arabidopsis thaliana*, fungi, and *C. elegans.*

As for the other insulators restricted solely to insects, Su(Hw) and CP190 were the only proteins that were found to have orthologs in every insect species. This is consistent with experimental evidence that CP190 is necessary for Su(Hw) enhancer-blocking activity (Pai et al., 2004),

Table 1. Distribution of Drosophila Insulator Proteins in Eukaryote Genomes.

SPECIES	Su(Hw)	CTCF	BEAF-32	CP190	Mod(Mdg4) 67.2	GAF	ZW5
D. melanogaster	*	*	*	*	*	*	*
D. sechellia	*	*	*	*	*	*	*
D. simulans	*	*	*	*	*	*	*
D. erecta	*	*	*	*	*	*	*
D. yakuba	*	*	*	*	*	*	*
D. ananassae	*	*	*	*	*	*	*
D. persimilis	*	*	*	*	*	*	*
D. pseudoobscura	*	*	*	*	*	*	*
D. mojavensis	*	*	*	*	*	*	*
D. grimshawi	*	*	*	*	*	*	*
D. willistoni	*	*	*	*	*	*	*
D. virilis	*	*	*	*	*	*	*
C. quinquefasciatus	*	*	-	*	*	*	-
A. aedypti	*	*	-	*	*	*	-
A. gambiae	*	*	-	*	*	*	-
T. castaneum	*	*	-	*	-	-	-
A. pisum	*	*	-	*	-	-	-
A. mellifera	*	-	-	*	-	*	-
N. vitripennis	*	*	-	*	-	*	-
P. hum. corporis	*	*	-	*	-	-	-
B. taurus	-	*	-	-	-	-	-
C. lupus familiaris	-	*	-	-	-	-	-
R. norvegicus	-	*	-	-	-	-	-
M. musculus	-	*	-	-	-	-	-
H. sapiens	-	*	-	-	-	-	-
P. troglodytes	-	*	-	-	-	-	-
T. guttata	-	*	-	-	-	-	-
G. gallus	-	*	-	-	-	-	-
X. tropicalis	-	*	-	-	-	-	-
T. rubripes	-	*	-	-	-	-	-
D. rerio	-	*	-	-	-	-	-
C. intestinalis	-	*	-	-	-	-	-
S. purpuratus	-	*	-	-	-	-	-
C.elegans	-	-	-	-	-	-	-
S. cerevisiae	-	-	-	-	-	-	-
S. pombe	-	-	-	-	-	-	-
O. sativa	-	-	-	-	-	-	-
A. thaliana	-	-	-	-	-	-	-

InParanoid and reciprocal best hits (RBH) were utilized to identify potential orthologs as described in the text. Positive orthologs are marked with an asterisk (*)

suggesting that this protein complex has a similar function in other species of insects. However, the other critical component of the gypsy insulator complex, Mod(mdg4)67.2, could only be identified among mosquitoes. We suspect that this could be an annotation issue, due to the complexity of the *mod(mdg4)* locus which in *D. melanogaster* can encode >25 isoforms (Buchner et al., 2000). Indeed, we obtained sufficient hits to other Mod(mdg4) isoforms in other insect species, but not to the 67.2 isoform. A similar situation is true for the poorly annotated GAF protein (Soeller et al. 1993), which appears absent in a few insect species but present in others (Table 1).

Finally, both BEAF-32 and Zw5 appear to be restricted to Drosophila, particularly BEAF-32 in which no significant ortholog was identified with InParanoid outside of Drosophila. Both tBLASTn and tBLASTx were also unsuccessful in identifying an ortholog. Interestingly, Zw5 returned no hits to other insects outside of Drosophila, but it did return hits to zinc finger proteins in higher vertebrates, such as *B. taurus* and *H. sapiens*. However, further dissection of these results using the reciprocal best hits (RBH) method did not support the conclusion that these proteins were true orthologs. Such lineage specificity for the Zw5 family of transcription factors has been described previously (Lespinet et al., 2002), but to our knowledge the specificity of BEAF-32 to the Drosophila lineage has not been described.

***BEAF-32A And BEAF-32B Isoforms Are Unique To* Drosophila**

Given that BEAF-32 plays a critical role in chromatin organization and gene regulation, and the apparent absence of nearly all other insulator protein orthologs in vertebrates except CTCF (Moon et al., 2005), we suspected that BEAF-32 might be absent from higher vertebrates like the

other insulator proteins; however, it was surprising that no significant hits outside of the Drosophila genus were obtained when using either isoform as the query. Successive iterations using PSI-BLAST returned hits to putative transposases, consistent with the BED domain's relationship with proteins derived from transposable elements (Aravind, 2000). A few hits were also obtained to the BESS domain, although these were limited to insects. However, hits to other proteins containing both a BED zinc finger domain and the BESS domain were not observed outside of Drosophila, suggesting that both isoforms of BEAF-32 are unique to the Drosophila lineage.

BEAF-32A & BEAF-32B Isoforms Are Highly Conserved

Examination of the BEAF-32 gene structure for each species revealed that all are capable of coding for both isoforms (Figure 1.1A). Both isoforms have two exons, with the last exon (coding for the nuclear matrix binding domain and the BESS domain) being shared by both. The first exon of isoform B, which codes for the BED II domain, is found entirely within the intron of isoform A. A comparison of the 5' splice junctions for each isoform revealed that all were functional, as they all retained the canonical 5' GU at the exon/intron boundary in addition to several conserved nucleotides at the 5' end of the intron. A similar situation is true for the 3' splice junction that is shared by both isoforms; in this case the 3' AG dinucleotide at the intron/exon boundary is retained in addition to other intronic nucleotides. Since the splice junctions are conserved and appear functional, both isoforms of BEAF-32 are most likely present in all Drosophila species analyzed. To further validate this conclusion, we generated a multiple alignment of BEAF-32A an BEAF-32B isoforms from twelve Drosophila species (Figure 1.2). The BED domains (BED I in isoform A and BED II in isoform B), the lamin-association domain

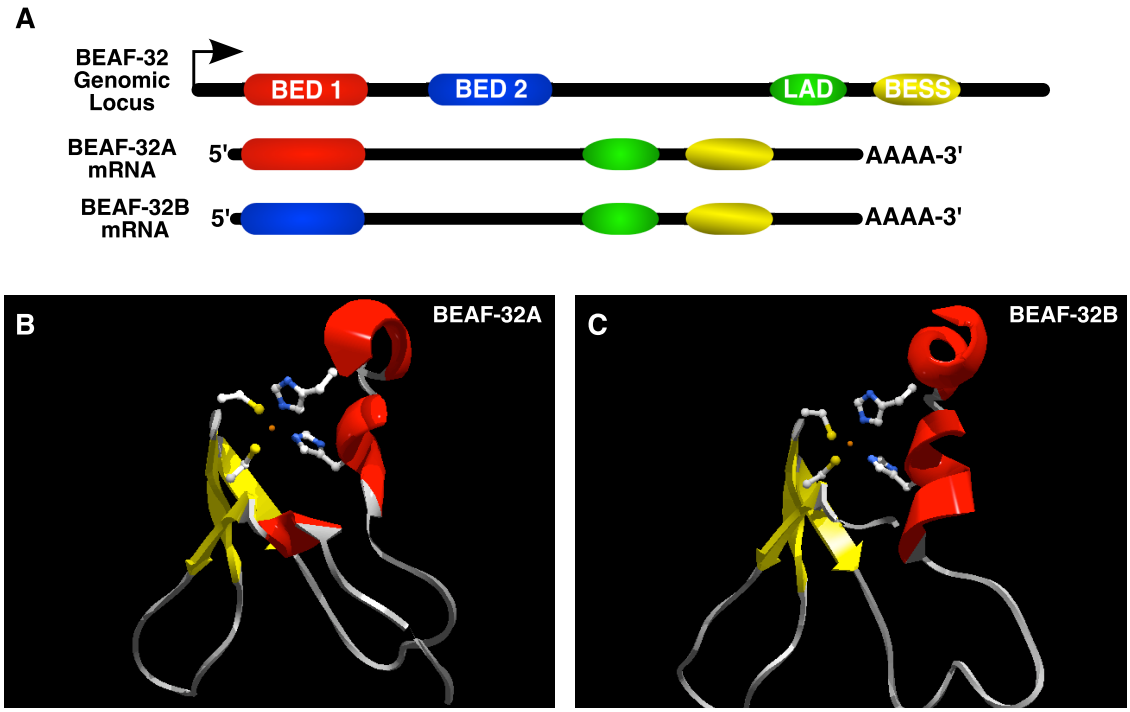

Figure 1.1. Gene Structure of BEAF-32 & Homology Modeling of the BED Domains. Schematic of the BEAF-32 locus from *Drosophila melanogaster* labeled with the position of the BED I, BED II, lamin association (LAD) and BESS domains and the corresponding mRNA of BEAF-32A and BEAF-32B (A). Homology model of the BED I domain from BEAF-32A (B) and BEAF-32B (C). Alpha helices are shown as red ribbons and beta sheets as yellow ribbons. Cys51, Cys54, His72 and His76 are shown in ball and stick coordinating the Zn^{2+} atom.

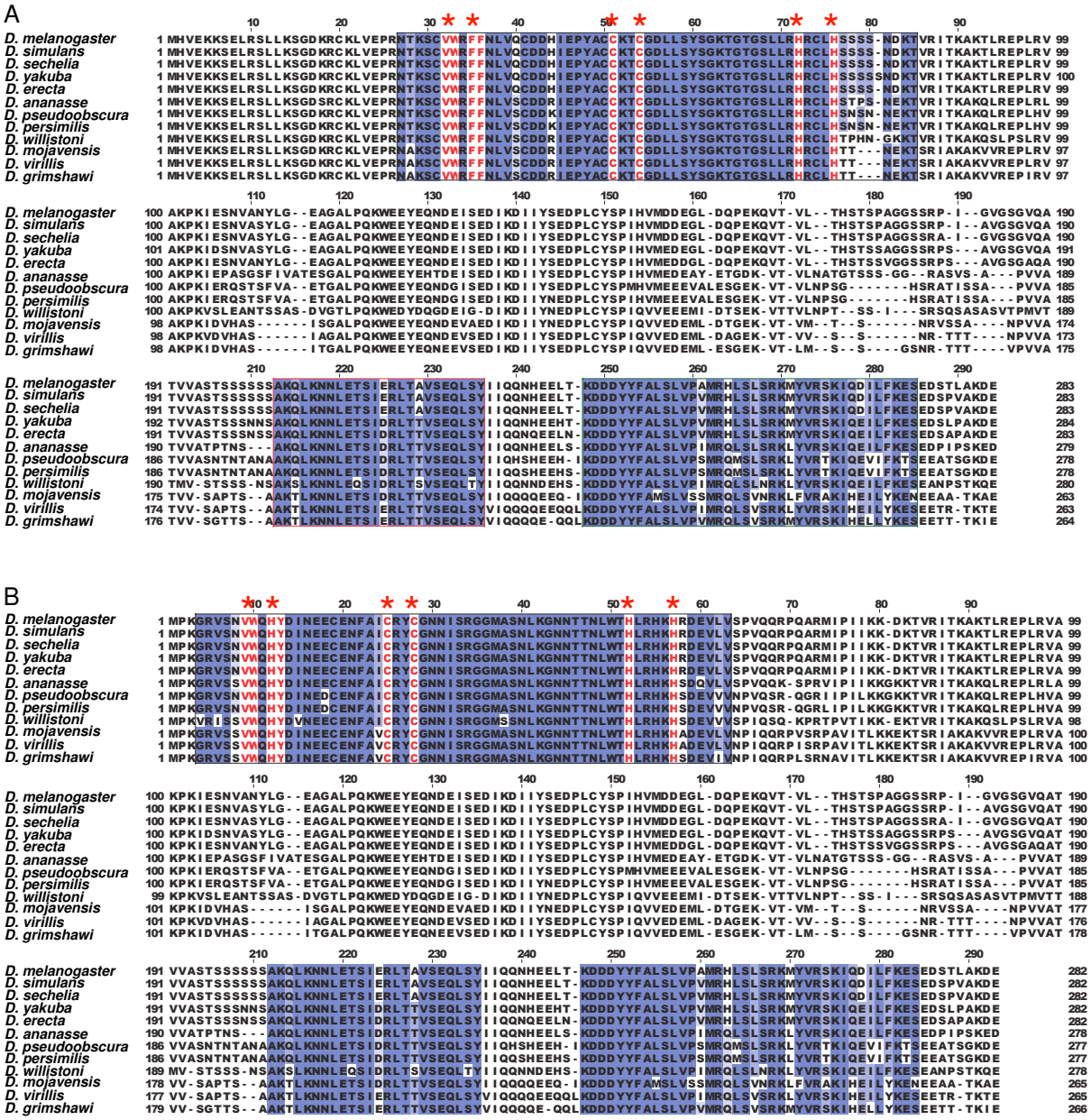

Figure 1.2. Alignment of BEAF-32A and BEAF-32B from 12 Drosophila Species.
The BED I domain of BEAF-32A (residues 27-85) (A) and BEAF-32B (residues 4-63) (B) is bounded by a black box, with coloring corresponding to percent identity. The Cys and His residues, as well as the conserved N-terminal aromatic/aliphatic residues which give the BED domain its signature, are marked by red asterisks. The lamin association domain (residues 203-223) and the C-terminal BESS domain (residues 240-275), respectively, are also denoted by percent identity.

(LAD) and BESS domain show remarkable sequence conservation; particularly, both BED domains retain the characteristic aromatic/aliphatic amino acid motif N-terminal to the $Cx_2Cx_nHx_{3-5}$[H/C] zinc finger DNA binding motif (Aravind, 2000), suggesting that both isoforms in all species are capable of DNA binding. A small indel is present within the BED I domain of isoform A in a few species along with a few amino acid substitutions, although these changes do not appear to affect the secondary or tertiary structure of the zinc finger. The BED II domain of isoform B is even more conserved, with no indels present and only a few amino acid substitutions. Homology modeling of both BED domains from *D. melanogaster* reveals that each can form a characteristic zinc finger, and are therefore capable of DNA binding (Figure 1.1B & 1.1C).

The C-terminal BESS domain is also highly conserved. Secondary structure prediction using the Jpred3 webserver (Cole et al., 2008) revealed that the motif in each species can fold into its characteristic alpha helical motif, suggesting that its ability to mediate protein-protein interactions between the two isoforms and possibly other proteins remains intact in all Drosophila species. Furthermore, amino acids 203-223 that have been shown to form a coiled-coiled domain responsible for BEAF-32's interaction with the nuclear lamina (Pathak et al., 2007) is also conserved across all species. Outside of these domains, the sequence conservation is also significant despite a few indels and more amino acid substitutions.

It is clear from this alignment that neither isoform appears to be rapidly evolving suggesting that purifying selection is acting to maintain both isoforms as functional proteins in all Drosophila

species. Indeed, a comparison of the global d_N/d_S ratio for each isoform using the SLAC maximum likelihood method (Kosakovsky Pond and Frost, 2005a; Kosakovsky Pond and Frost, 2005b) suggested that both isoforms are under purifying selection (BEAF32A: d_N/d_S= 0.11; BEAF32B: d_N/d_S= 0.13). SLAC, FEL, and REL estimates for individual codons revealed that all are either under neutral or negative selection, with no evidence for adaptive selection (Figure 1.3). With the apparent absence of any other protein containing both a BESS and BED domains in other insects and higher vertebrates, as well as the remarkable sequence conservation of both BEAF-32 isoforms in all Drosophila species analyzed, the data suggests that BEAF-32 is a novel protein playing a novel role in chromatin organization and gene expression only in Drosophila.

Genome Wide Analysis Reveals A Bias In The Association Of BEAF-32 Binding Motifs To Gene Promoters In **Drosophila** *But Not In Other Insect Species*

In addition to BEAF-32's localization to the scs' element found at one end of the *hsp70* gene at the 87A7 locus, immunostaining of polytene chromosomes revealed hundreds of BEAF-32 binding sites, primarily localized to interbands and flanking puff boundaries (Zhao et al., 1995). However, considerable debate remains over how BEAF-32 recognizes its binding site, as attempts to identify a definitive binding signature have remained elusive (Cuvier et al., 1998; Jiang et al., 2009). Recently, Bushey et al. suggested that the presence of 4-7 CGATA motifs within a 1 kb window was sufficient to predict *in vivo* BEAF-32 binding sites (Bushey et al., 2009), despite the fact that a large number of these clusters do not bind BEAF-32. However, nearly 75% of all CGATA clusters physically bound by BEAF-32 were within 200 bp of a TSS. As promoter structure is typically conserved to mediate a common function, evidence of BEAF -32 binding at or near promoters in Drosophila but not other insect species would lend support to

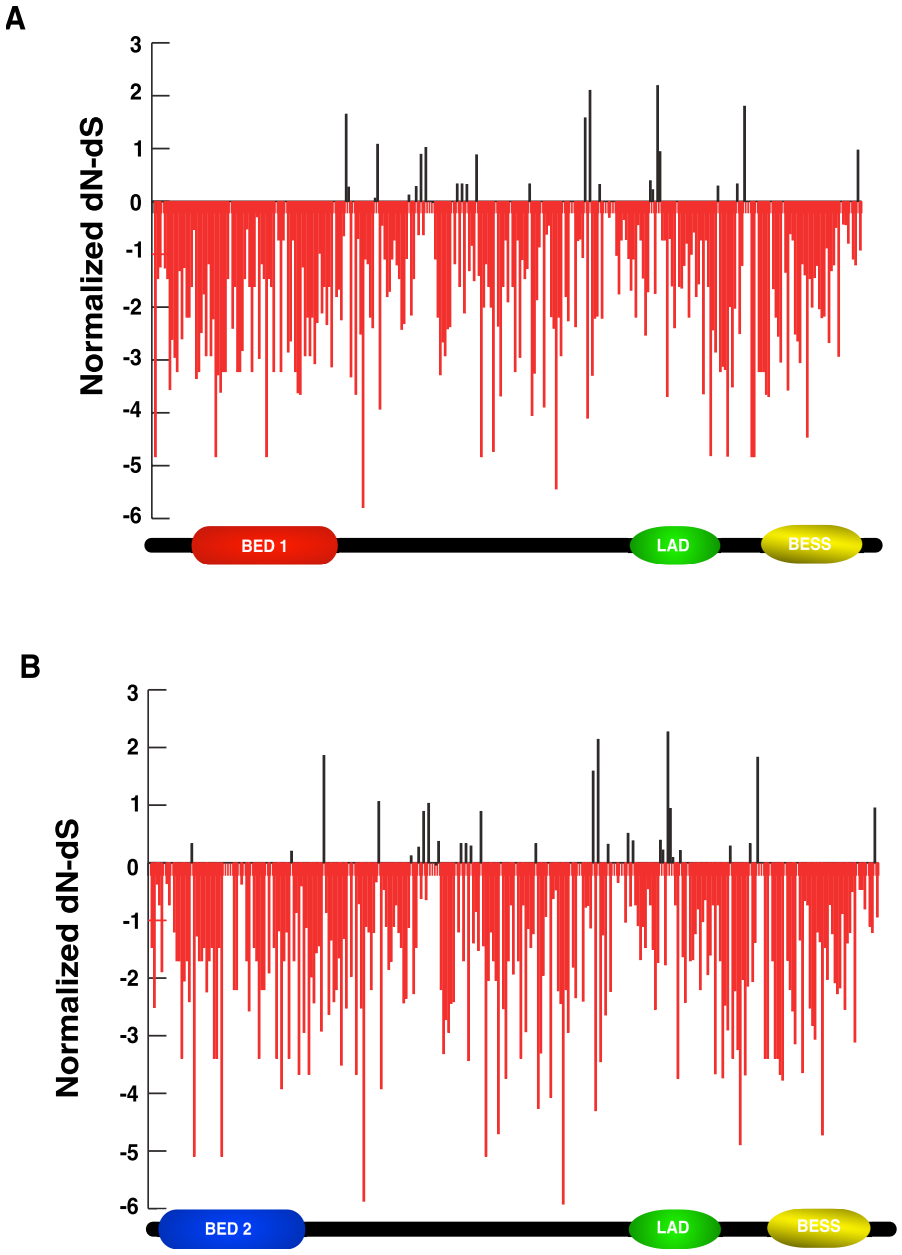

Figure 1.3. Selection Pressure Estimates for BEAF-32A/B.
dN-dS calculations for each codon in BEAF-32A (A) and BEAF-32B (B) normalized for branch lengths.

the idea that BEAF-32 is unique to Drosophila. We therefore used the Fly Enhancer program (Markstein et al., 2002) to search the genome sequences of *D. melanogaster, Anopheles gambiae* (African malaria mosquito), *Apis mellifera* (honey bee), and *Tribolium castaneum* (red flour beetle) for clusters of CGATA motifs. A clear bias of CGATA clusters in relation to TSSs could be seen in *D. melanogaster* compared to the other insect species (Figure 1.4A). Over 25% of all CGATA clusters (n=1,115) were found within 1 kb of a TSS in *D. melanogaster*, compared to <5% in all other insects. In order to validate that some of the clusters identified in our analysis represented true BEAF-32 binding sites, we used publicly available BEAF-32 ChIP data (GEO Accession: GSE15661) with a stringent 1% FDR to verify BEAF-32 occupancy at each of the genes identified by Fly Enhancer that had the required number of motifs <1 kb from the TSS. Using this approach, we found that nearly 85% of the genes identified by Fly Enhancer analysis with CGATA clusters <1 kb from a TSS are physically bound by BEAF-32 *in vivo* (n=302 genes). Furthermore, we searched the Mosquito Enhancer output for the corresponding orthologs of each *D. melanogaster* gene. Only 1% of the *A. gambiae* orthologs (n=193 genes) had a cluster <1 kb from a TSS (Yates's Corrected χ^2=338.52, P<<0.001).

Also, we ruled out the possibility that the DRE element, which contains a CGATA motif, may have biased our results. When the DRE element is removed from the analysis, the number of total clusters <1 kb is proportionally reduced by a quarter. However, the percentage of clusters found <1 kb from a TSS was just under 25%, nearly identical to the percentage of CGATA clusters <1 kb from a TSS when the DRE motif was included (Figure 1.4B). Taken together, these results suggest that the bias observed in CGATA distribution between insect species is not

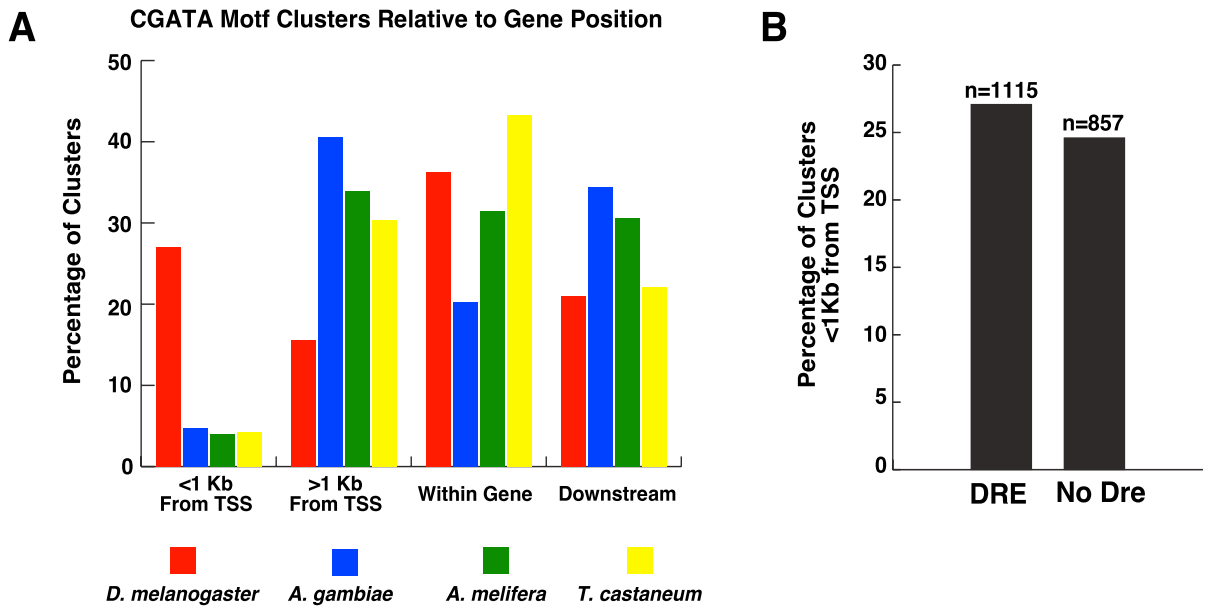

Figure 1.4. Distribution of BEAF-32 Binding Sites Relative To Promoters.
Percentage of binding sites for each species (A). Same analysis outlined in (A) with or without the DREF (DRE) motif included (B).

due to the presence of DREF in Drosophila. The clear bias of clustered motifs relative to gene position in *D. melanogaster* compared to the apparent random distribution in other insect species supports the conclusion that BEAF-32 is unique to Drosophila and may have played a role in reshaping genome structure and organization during evolution.

Discussion

Like many of the other insulator proteins found in Drosophila, BEAF-32's role in chromatin organization and gene regulation is still poorly understood. Although the mechanism by which these proteins are able to create higher-order chromatin domains via possible loop formation is still unclear, knowledge of the taxonomic distribution of these proteins might provide insight into whether these mechanisms are conserved across different phyla, and hence, whether a unifying mechanism of insulator action exists.

Computational data presented here suggests that the majority of Drosophila insulator proteins are restricted mainly to insects, with the exception of CTCF which has been shown previously to have a vertebrate ortholog (Moon et al., 2005). However, we note that our method of ortholog detection is conservative, and a more exhaustive computational approach might identify other putative orthologs, particularly for Mod(mdg4)67.2 and GAF in other insects. It appears unlikely, however, that a true *D. melanogaster* insulator ortholog outside of CTCF exists in vertebrates. We acknowledge the possibility that other unrelated proteins in these species might possess a function similar to that of insulator proteins and perform the same or related biochemical function as Drosophila insulator proteins. For example, although CTCF is present in both insects and vertebrates, it is still not certain how these similar proteins accomplish their insulating

function in their respective species (Moon et al., 2005). Although they both possess similar functions *in vivo*, they might interact with completely different protein partners. This is supported by the lack of a CP190 ortholog in other eukaryotes outside of the insects, which has been shown to be important for the insulating function of CTCF in Drosophila (Gerasimova et al., 2007; Mohan et al., 2007). This suggests that different proteins or different combinations of known insulator proteins can be utilized to achieve the same insulator function in different species.

Our ortholog data also suggests that BEAF-32 is Drosophila lineage specific. The most convincing data stems from the fact that no proteins outside of Drosophila were identified in BLAST searches that contained both a BED domain and a BESS domain with significant identity. Also, the remarkable sequence homology at the amino acid level of both isoforms in all 12 Drosophila species analyzed, particularly in the BED domains, nuclear matrix binding domain, and the BESS domain, suggests that all isoforms are functional, capable of organizing chromatin into higher-order structures and/or facilitating a chromatin environment to modulate gene expression accordingly. Furthermore, the significant bias of putative CGATA BEAF-32 binding motifs to less than 1 kb of transcription start sites in *D. melanogaster,* but not in *A. gambiae, T. castaneum, or A. mellifera* reinforces the notion that BEAF-32 is found only in Drosophila.. We acknowledge, however, that other factors outside of the CGATA motif might also be responsible for BEAF-32 recruitment. This is supported by our data and others (Bushey et al., 2009; Jiang et al., 2009) that suggest many CGATA clusters found far from TSSs are not bound by BEAF-32. Therefore, chromatin structure and transcription factor binding at

promoters may facilitate BEAF-32 binding in conjunction with the CGATA motif. Nonetheless, over 85% of the genes identified by Fly Enhancer with 7 CGATA motifs within 1 kb of a TSS are physically bound to BEAF-32 in vivo, suggesting that the motif data near promoters is reflective of *in vivo* BEAF-32 binding sites. This is in agreement with Bushey et al. who found that nearly 75% of physically bound clusters are either 200 bp upstream or 200 bp downstream of the TSS (Bushey et al., 2009). Furthermore, the fact that less than 1% of the corresponding *A. gambiae* orthologs were identified that contain the same CGATA motif suggests that the promoter bias observed is significant and provides further support suggesting that BEAF-32 is present only in Drosophila.

As for BEAF-32's evolutionary origin, the data presented here raises the question of how this protein might have originated and came to acquire such an important role in one specific lineage. Although it is likely to have arisen from an ancient transposase given its DNA-binding BED domains (Aravind, 2000), whether it acquired this directly from some transposon or indirectly from a cellular protein remains to be elucidated. Furthermore, the question of how it acquired each BED domain is also unclear. It is certain that these domains did not result from an exon duplication, given the fact that BED I and BED II retain no sequence identity outside of the residues responsible for zinc coordination and at the aromatic/aliphatic position. As for the BESS domain, which is annotated as being unique to Drosophila proteins, we identified four other insect proteins outside of Drosophila that have domains that can form an identical secondary structure, suggesting that the BESS domain is not unique to Drosophila, although it may be limited to insects. Thus, additional work is necessary to determine how the gene structure of

BEAF-32 evolved to its present day form within Drosophila and became unique to this genus.

BEAF-32 might play a significant role in shaping genome organization. Such a hypothesis has been put forth for CTCF in *D. melanogaster* (Smith et al., 2009), and given that both BEAF-32 and CTCF show a preference for binding just upstream of promoters and between divergently transcribed genes, it is possible that both proteins may function in a similar manner. This is underscored by the fact that both BEAF-32 and CTCF localize with CP190, and these three proteins overlap at nearly 500 sites throughout the genome (Bushey et al., 2009). However, a more interesting question is how possession of these proteins may have shaped genome architecture and organization during evolution. Although evidence remains limited, a recent study that addressed the distribution of CTCF across different nematode phyla suggested that possession of CTCF might have had drastic implications for genome organization and architecture of present day nematode species (Heger et al., 2009). As CTCF and BEAF-32 have a similar function in the cell, it will be interesting to determine how BEAF-32 may have influenced the evolution of the Drosophila genome. Such an influence might be reflected in our motif searching analysis, in which a clear bias of BEAF-32 binding clusters is represented in Drosophila only.

Here we have provided evidence that both isoforms of BEAF-32 are unique to all Drosophilids but absent in all other known species. Although it is not unusual for transcription factors to undergo lineage specific expansion (Lespinet et al., 2002), this finding is particularly interesting since BEAF-32 plays a global role in regulating expression of a large number of genes

controlling crucial cellular functions (Emberly et al., 2008; Jiang et al., 2009). The phylogenetic distribution of BEAF-32 suggests that the gene underwent a period of rapid evolution within the Drosophila lineage, but how such a protein may have arisen and acquired its cellular function is still uncertain. A more thorough characterization of the distribution of these proteins across all relevant taxa might provide insight into whether the mechanism of insulator action is conserved across phyla. It might also shed light on the role these proteins might have had in shaping genome organization and could consequently provide clues into how differential gene expression is achieved from one species to the next.

Materials and Methods

Sequence Acquisition

Nucleotide and protein sequences of *D. melanogaster* Su(Hw), CP190, Mod(mdg4)67.2, dCTCF, GAF, Zw5 and both BEAF-32 isoforms were obtained from FlyBase, corresponding to the R5.18 release. BEAF-32A and BEAF-32B nucleotide and protein sequences from 11 other Drosophila species were also obtained from FlyBase corresponding to the following releases: *D. ananassae,* R1.3; *D. erecta,* R1.3; *D. grimshawi,* R1.3; *D. mojavensis,* R1.3; *D. persimilis,* R1.3; *D. pseudoobscura pseudoobscura,* R2.4; *D. sechellia,* R1.3; *D. simulans,* R1.3; *D. virilis,* R1.2; *D. willistoni,* R1.3; *D. yakuba,* R1.3.

Ortholog Identification

Putative orthologs were identified using the amino acid sequence of each insulator protein as a query at the InParanoid7 database (Berglund et al., 2007). In cases in which there were multiple isoforms of the protein, only the sequence of the protein implicated in insulator function was

used. Significant hits were verified manually using the reciprocal BLAST hit method (Altschul et al., 1997), where two proteins from two genomes were considered orthologs if each protein used as a query returned the other protein as the highest scoring BLASTp match. Both tBLASTn and tBLASTx were used to verify that BEAF-32 contained no orthologs outside of Drosophila.

Alignment

Protein sequences were aligned using ClustalX 2.0.10 (Larkin et al., 2007). In some cases, the annotated BEAF-32 protein given by FlyBase either did not include both isoforms, or had incorrectly included both N-terminal BED domains in some of the Drosophila species. BEAF-32 nucleotide sequences were conceptually translated where necessary (using the gene structure of *BEAF-32* from *D. melanogaster* as a reference) to obtain both isoforms in each species. The Jalview program (Waterhouse et al., 2009) was used to perform manual adjustments to the resulting alignment where applicable.

Homology Modeling

Modeling was performed using Swiss-PDB Viewer (Deep View) software (v. 4.0.1) (Guex and Peitsch, 1997), using the NMR structure of a human BED-containing protein (PDB Code: 2DJR) as the template.

Motif Analysis

The motif-searching tool Fly Enhancer (Markstein et al., 2002) was used to search the *D. melanogaster* genome (FlyBase Release 4. 1) for the CGATA BEAF-32B binding motif. Mosquito Enhancer, Beetle Enhancer, and Bee Enhancer (beta 2), which use the same motif-

searching algorithm as Fly Enhancer, were used to search the genomes of *Anopheles gambiae, Tribolium castaneum, and Apis mellifera*, respectively, for the same BEAF-32B binding motif (Markstein et al., 2002). We searched each genome with the following parameter: seven CGATA motifs within a 1000 bp window. If such a motif was identified, then we considered this a single BEAF-32 binding cluster. Other parameters utilizing more or less motifs within a smaller bp window were also used for searching: 10 motifs in a 1000 bp window; 7 motifs in a 600 bp window; 5 motifs in a 600 bp window; 7 motifs in a 300 bp window; 5 motifs in a 300 bp window. A single factor ANOVA was used to ensure that the number of motifs and the window size did not bias the analysis (not shown). In other words, the distribution of CGATA binding motifs in relation to transcription start sites was consistent for all search parameters utilized. The position of each cluster in relation to genes were then counted, based on the following parameters: >1000 bp from a transcription start site (TSS); <1000 bp from a TSS; within a gene (both in coding and noncoding regions, such as UTRs and introns); or downstream. In this case, if the cluster was upstream of the TSS, then the first CGATA motif within the binding cluster was used to measure the distance to the nearest TSS.

To rule out the possibility that our motif search was biased, we performed an additional search with Fly Enhancer. The Drosophila transcription factor DREF binds to the DRE element, which contains the 5-mer CGATA motif (TATCGATA) and is highly enriched near target promoters (Hirose et al., 1993; Ohler et al., 2002). We used a Boolean operator to exclude any CGATA clusters that contained a TATCGATA motif that were identified in our original analysis. The

percentage of clusters <1kb from the TSS in *D. melanogaster* was then determined and compared to the original motif search.

We verified that the majority of genes with CGATA clusters identified by Fly Enhancer that were <1 kb from a TSS were physically bound to BEAF-32 *in vivo* using publicly available ChIP-chip data (GEO Accession: GSE15661; (Bushey et al., 2009)). Using a stringent false discovery rate (FDR) for BEAF-32 peak detection (<1%), we individually searched for each gene and ensured that bound BEAF-32 was within 1 kb of the promoter. We also searched the Mosquito Enhancer output for the corresponding *A. gambiae* orthologs, to verify whether the clusters were conserved at the same genes between the two species.

Tests Of Selection

The Datamonkey webserver (http://www.datamonkey.org/) was used to assess global d_N/d_s ratios (where d_N= rate of nonsynonymous codon substitutions and d_s= rate of synonymous substitutions) as well as individual d_N/d_s for each codon in each isoform (Kosakovsky Pond and Frost, 2005a). Three maximum likelihood-based counting methods were used to infer such rates: Single-Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL) and Random Effects Likelihood (REL) (Kosakovsky Pond and Frost, 2005b). In all cases, the HKY85 substitution model was used for data fitting.

CHAPTER 2

Chromatin Insulator Bodies Are Nuclear Structures That Form In Response To Osmotic Stress And Cell Death

This chapter has been published:

Schoborg, T.A., R. Rickels, J. Barrios, and M Labrador. 2013. Chromatin Insulator Bodies are Nuclear Structures that Form in Response to Osmotic Stress and Cell Death. *J. Cell Biol.* 202(2):261-276.

My contributions included: (1) devising experiments, (2) performing experiments, (3) data collection and analysis, (4) writing the manuscript and making figures. Ryan Rickels and Josh Barrios assisted with (1)-(3) while Mariano Labrador contributed to (1) and to a lesser extent (4). No revisions to the published manuscript have been made for the purposes of this dissertation.

Abstract

Chromatin insulators assist in the formation of higher order chromatin structures by mediating long-range contacts between distant genomic sites. It has been suggested that insulators accomplish this task by forming dense nuclear foci termed insulator bodies that result from the coalescence of multiple protein-bound insulators. However, these structures remain poorly understood, particularly the mechanisms triggering body formation and their role in nuclear function. Here we show that insulator proteins undergo a dramatic and dynamic spatial reorganization into insulator bodies during osmostress and cell death in a HOG/p38 MAPK-independent manner, leading to a large reduction in DNA-bound insulator proteins that rapidly repopulate chromatin as the bodies disassemble upon return to isotonicity. These bodies occupy distinct nuclear territories and contain a defined structural arrangement of insulator proteins. Our findings suggest insulator bodies are novel nuclear stress foci that can be used as a proxy to

monitor the chromatin-bound state of insulator proteins and provide new insights into the effects of osmostress on nuclear and genome organization.

Introduction

Packaging DNA in the nucleus requires the formation of higher order chromatin structures that function as both structural and functional regulators of the genome. Central to this process is the formation of long range contacts between distant genomic sites, resulting in the formation of loop structures that establish physical, topological and gene regulatory domains in addition to facilitating contacts between promoters and distant regulatory elements. Although a number of chromatin binding proteins have been implicated in this process, chromatin insulators are of particular interest given their broad role in chromatin structure and nuclear function. Despite their initial characterization from transgenic assays in Drosophila as enhancer and heterochromatin blockers, the in vivo function of these DNA elements more generally involves mediating long-range contacts. Seven insulator binding proteins have been identified in Drosophila*,* including Su(Hw), CP190, BEAF-32, Mod(mdg4)67.2, dCTCF, GAF and Zw5, with mammals containing only the CTCF ortholog (Schoborg and Labrador, 2010). In both taxa, these proteins bind to thousands of insulator sites scattered throughout the genome (Bushey et al., 2009; Cuddapah et al., 2009; Negre et al., 2010) where they participate in a plethora of long-range contacts with enhancers, promoters and other insulators, acting to both facilitate and repress transcription, maintain regions of histone modifications, and establish physical domains (reviewed in (Krivega and Dean, 2012; Van Bortle and Corces, 2012; Yang and Corces, 2012)).

It has been suggested that insulators spatially accomplish these tasks through the formation of multiple chromatin loop structures, mediated by contacts between multiple insulator-bound proteins, that physically manifest themselves as insulator bodies (Labrador and Corces, 2002). Drosophila insulator bodies consist of 10-30 punctate nuclear signals corresponding to Su(Hw), CP190, Mod(mdg4)67.2 and dCTCF (Gerasimova et al., 2000; Gerasimova and Corces, 1998; Gerasimova et al., 2007; Pai et al., 2004). Though early indirect evidence supported a functional role in *gypsy* insulator activity (Byrd and Corces, 2003; Gerasimova et al., 2000), recent work has suggested that these structures do not contribute to *gypsy* enhancer-blocking directly, and instead function as storage sites for insulator proteins poised for insulator activity (Golovnin et al., 2008; Golovnin et al., 2012). However, many fundamental aspects about these structures remain poorly understood, particularly how and why they form, whether they might contribute to other aspects of insulator function independently of enhancer blocking and the consequences of such behavior on nuclear organization and genome dynamics.

Here we show that insulator bodies are nuclear stress bodies that form in response to osmostress and cell death. Insulator proteins coalesce from diffusely distributed speckles into punctate insulator bodies rapidly in response to osmotic stress, exhibit dynamic behavior throughout the duration of stress and rapidly recover to their pre-stressed state upon return to isotonicity. This correlates with a reduction in chromatin-bound insulator proteins during the duration of stress that is restored within minutes during recovery. Insulator bodies localize primarily to the nuclear periphery where they show transient associations with lamin, in addition to chromatin lacunas within the condensed chromatin mass. Interestingly, this behavior is independent of the

HOG/p38 MAPK osmostress sensing pathway. In larval tissue, CP190 and Mod(mdg4)67.2 can form bodies independently of one another, while Mod(mdg4)67.2 is required for Su(Hw) entry into these structures. Our findings reveal novel insights into the role of stress on nuclear dynamics, provide a framework for elucidating the consequences of such behavior on genome function and organization and establish a model system in which to study various aspects of nuclear body biogenesis, maintenance and behavior.

Results

Insulator Bodies Form In Response To Hyperosmolarity

Previous work has primarily focused on insulator body behavior in 3rd instar larval tissues and S2 cells (Capelson and Corces, 2005; Capelson and Corces, 2006; Gerasimova et al., 2000; Gerasimova and Corces, 1998; Gerasimova et al., 2007; Ghosh et al., 2001; Golovnin et al., 2008; Golovnin et al., 2012; Lei and Corces, 2006; Pai et al., 2004; Ramos et al., 2011; Wood et al., 2011; Xu et al., 2004). Using antibodies directed against CP190 and Mod(mdg4)67.2, we were unable to identify structures that resembled insulator bodies in these same cells and tissues. Rather than exhibiting 10-30 nuclear periphery-associated punctate dots as observed in previous reports, our diploid cells displayed a diffuse distribution that appears speckled after image deconvolution. This pattern consists of numerous small foci, reminiscent of tiny speckles distributed throughout the entire volume of nucleus, with the exception of the nucleolus (Figure 2.1A , 2.1C & 2.1E)). Both proteins formed distinct bands on polytene chromosomes as expected (Figure A2B). Occasionally, one or two small punctate dots resembling insulator bodies were observed for CP190 in larval tissue and S2 cells; however, the majority of the signal remained distributed throughout the nucleus.

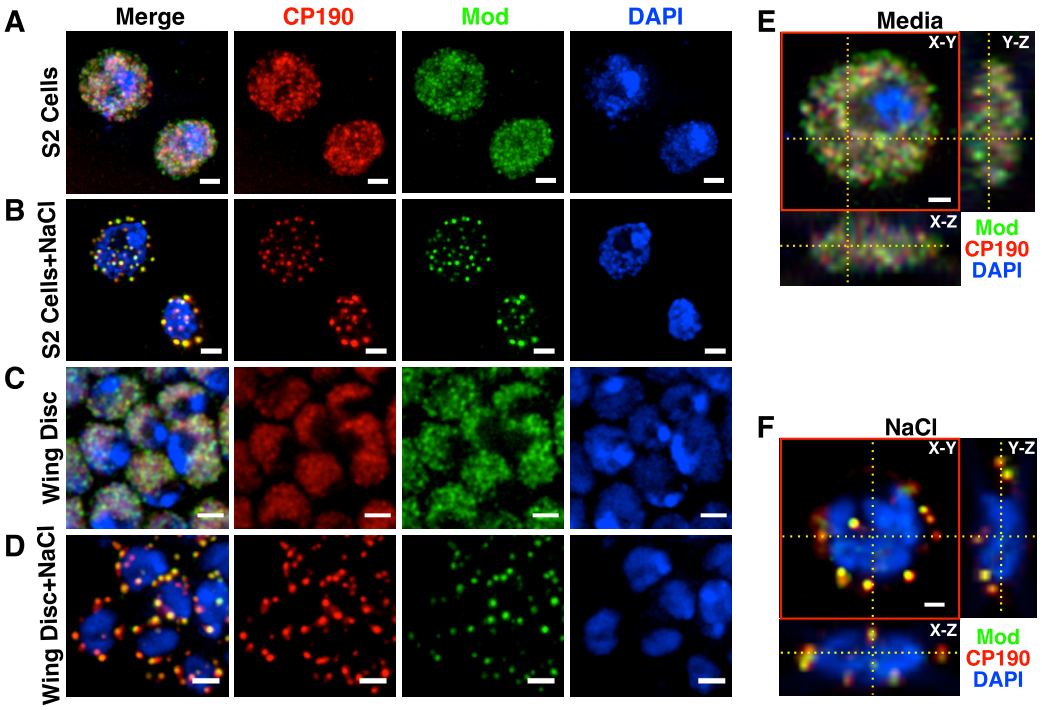

Figure 2.1. Insulator Bodies Form In Response To Osmostress.
S2 cells stained with CP190 and Mod(mdg4) under normal cellular conditions (A) or following treatment with 250mM NaCl (B). Wing discs from 3rd instar larvae stained for CP190 and Mod(mdg4) under normal cellular conditions (C) or following treatment with 250mM NaCl (D). Orthogonal projections along the indicated axes (yellow dashed lines) in an unstressed (E) and stressed (F) S2 cell. Note: (A)-(D) are maximum projections of 1 μm z-slices, while (E) & (F) are a single z-slice (X-Y plane only). All scale bars are 2 μm.

Our inability to observe insulator bodies in cells and tissues under normal cellular conditions led us to next determine the effects of various stressors on insulator body formation. Previous work has implicated certain stress-induced cues as regulators of body behavior, particularly heat shock (Gerasimova et al., 2000; Golovnin et al., 2012; Wood et al., 2011). Treatment of S2 cells with a 37°C HS for 20 or 60 min did not change the distribution of CP190 and Mod(mdg4)67.2 compared to non-HS controls, despite an obvious rearrangement of chromatin in the heat shocked cells (Figure A2A). However, subjecting cells and tissue to NaCl-induced osmotic stress resulted in the disruption of the diffusely speckled pattern and the formation of large CP190 and Mod(mdg4)67.2 foci in >99% of nuclei that was distinct from the unstressed control pattern, irrespective of cell/tissue type (Figure 2.1B, 2.1D &2.1F). These structures matched the description of insulator bodies given in previous reports, both in terms of number of bodies per nucleus and their localization to the nuclear periphery. Additionally, two chemically distinct osmolytes, sorbitol and sucrose, also induced body formation (Figure A2C). This appears to be a graded response, as CP190 gradually transitions from diffusely speckled to more punctate and numerous bodies as the salt concentration is increased up to 500 mM (Figure A2D). Cells permeabilized with detergent prior to addition of 250 mM NaCl failed to form bodies and instead maintained the diffusely speckled pattern observed in the absence of osmostress, verifying that insulator body formation occurs in response to increased osmotic loads (Figure A2E). Taken collectively, these data suggest that insulator bodies are novel nuclear stress bodies that form in response to osmostress.

Interestingly, this response appears to be relegated specifically to insulator proteins and their interacting partners. Other chromatin proteins, such as Polycomb group proteins (PcG) found in both Drosophila and mammals, have been shown to form speckle-like foci termed PcG bodies that may function as hubs involved in silencing developmental genes (Alkema et al., 1997; Bantignies et al., 2011; Messmer et al., 1992). PcG bodies in S2 cells marked with Polycomb (Pc) are not significantly altered during osmostress, remaining identical in size and nuclear distribution as compared to untreated media controls, while CP190 undergoes a substantial reorganization into bodies (Figure 2.2A). Furthermore, Heterochromatin Protein 1 (HP1), which binds to H3K9 methylated histone tails primarily in heterochromatin (Vermaak and Malik, 2009) is not disrupted during osmostress (Figure 2.2B). Given the lack of a similar response by other nuclear proteins, this data suggest that insulator body formation is not the result of a general biophysical effect on globular protein structure under conditions of hyperosmolarity and instead may be the result of a targeted response directed to insulator proteins.

Insulator Bodies Are Highly Ordered Structures With A Distinct Nuclear Distribution

The location of known insulator proteins within these bodies suggests that they have a defined structural organization. We observed extensive colocalization between Su(Hw), Mod(mdg4)67.2, CP190 and dCTCF proteins in stressed nuclei, which manifest themselves as irregular spherical structures (Figures 2.3A & 2.3B) in agreement with previous reports. Such results are not surprising, given that CP190 is a common component of both *gypsy* and dCTCF insulators and has been shown to colocalize to these structures previously (Gerasimova et al., 2007; Pai et al., 2004). However, BEAF-32 forms donut-shaped halos around the spherical bodies in stressed nuclei (Figure 2.3C) rather than colocalizing with the rest of the insulator proteins, a surprising

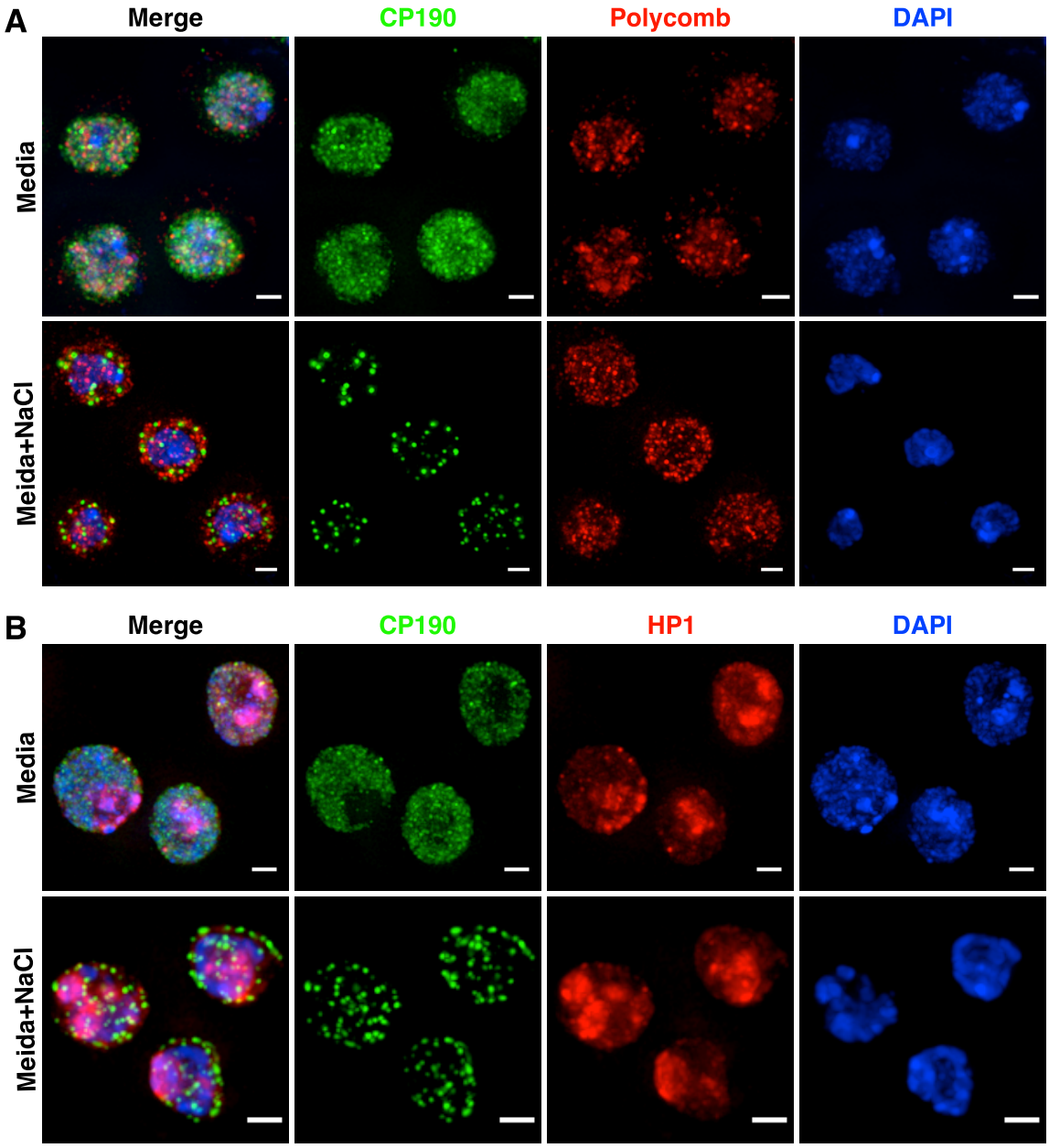

Figure 2.2. Effect Of Osmostress On Non-Insulator Chromatin Proteins.
S2 cells treated with or without 250mM NaCl and stained for CP190 and Polycomb (A) or HP1 (B). Scale bars are 2 μm.

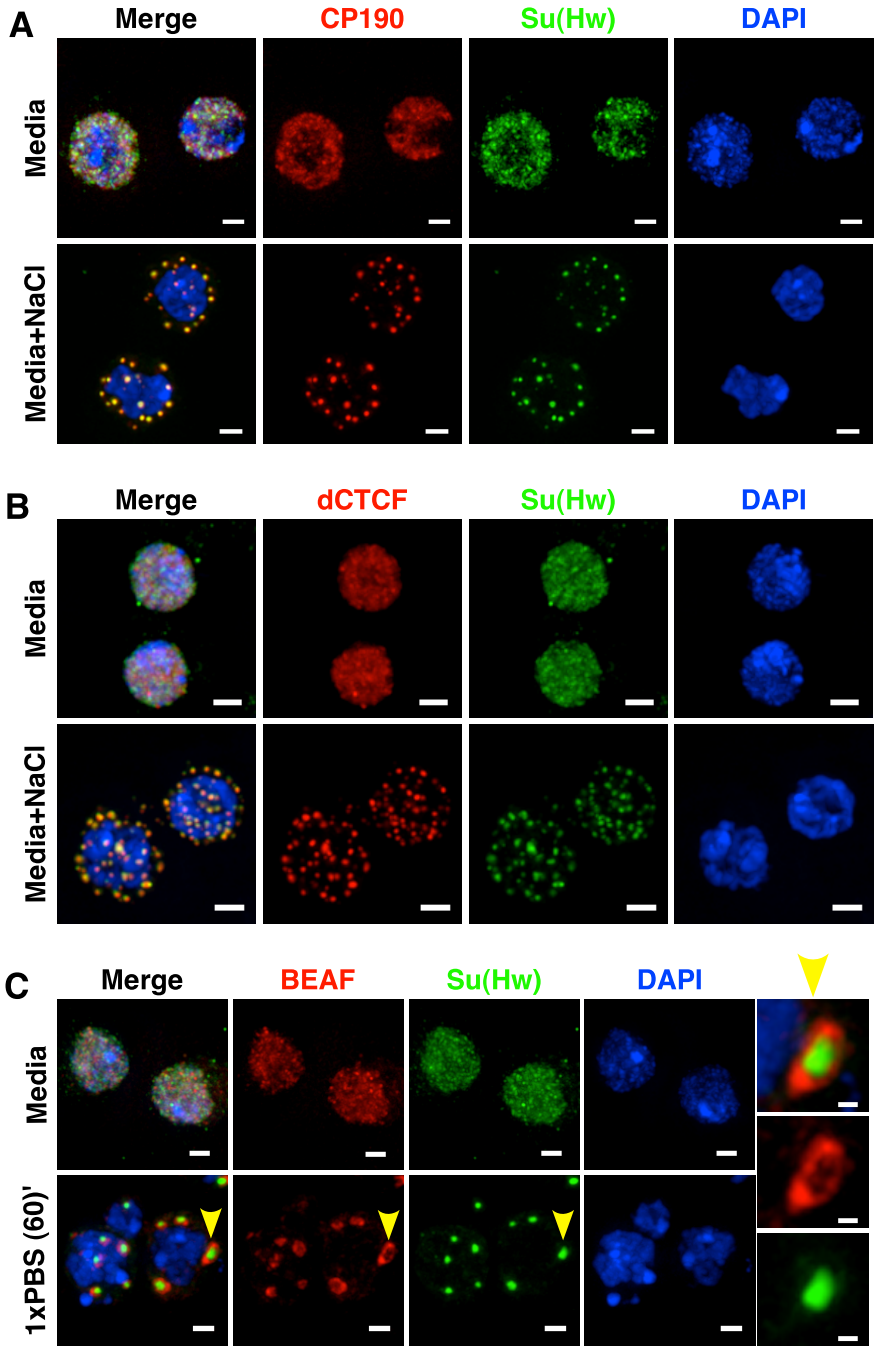

Figure 2.3. Insulator Bodies Have A Defined Structural Organization.
S2 cells treated with or without 250mM NaCl and stained for CP190 and Su(Hw) (A), dCTCF::mCherry and Su(Hw)::eGFP (B) and BEAF-32 and Su(Hw) after 60' incubation in 1xPBS. BEAF-32 forms large donut structures around the spherical structures (C, arrowhead+inset). Scale bars are 2 μm (inset scale bars in (C) are 0.5 μm).

finding given the substantial overlap between BEAF-32, CP190, dCTCF at multiple genomic sites (Bushey et al., 2009). This arrangement of insulator proteins is maintained despite the overall size of the bodies, with the diameter of the spherical portion ranging from ~200 nm to nearly 1 μm and the diameter of the surrounding BEAF-32 donuts being roughly proportionally double in size, meaning that these structures can approach sizes of over 2 μm in extreme cases. Identical structures are also observed in S2 cells over-expressing BEAF-32::mCherry and Su(Hw)::eGFP, ruling out potential antibody artifacts (Figure A3A). Such findings suggest that although insulator bodies can vary widely in number and size, even within the same cell, they are highly ordered structures.

The position of these structures within the diploid nucleus is also peculiar. Most of the bodies appear to be in defined territories in the nuclear periphery (near the edges of the condensed chromatin mass) and in DAPI-less lacunas within the mass, suggesting these structures form in regions devoid of chromatin (Figure 2.4A) and might be anchored to other nuclear structures, such as the nuclear matrix or the nuclear pore complex. Intensity Correlation Analysis revealed potential overlap between CP190 and lamin for a subset of insulator bodies in diploid cells; however, not all bodies are lamina associated, and small, punctate CP190 signals in unstressed cells also overlap with lamin (Figure 2.4A). Furthermore, no significant colocalization between insulator bodies and nuclear pore components were observed (Figure 2.4B), suggesting that associations with lamin or NUPs are not a requisite for insulator body formation. Furthermore, stressed S2 cells extracted with 2M NaCl to isolate insoluble nuclear components (Byrd and Corces, 2003) revealed a loss of lamin-associated nuclear bodies, particularly in nuclei

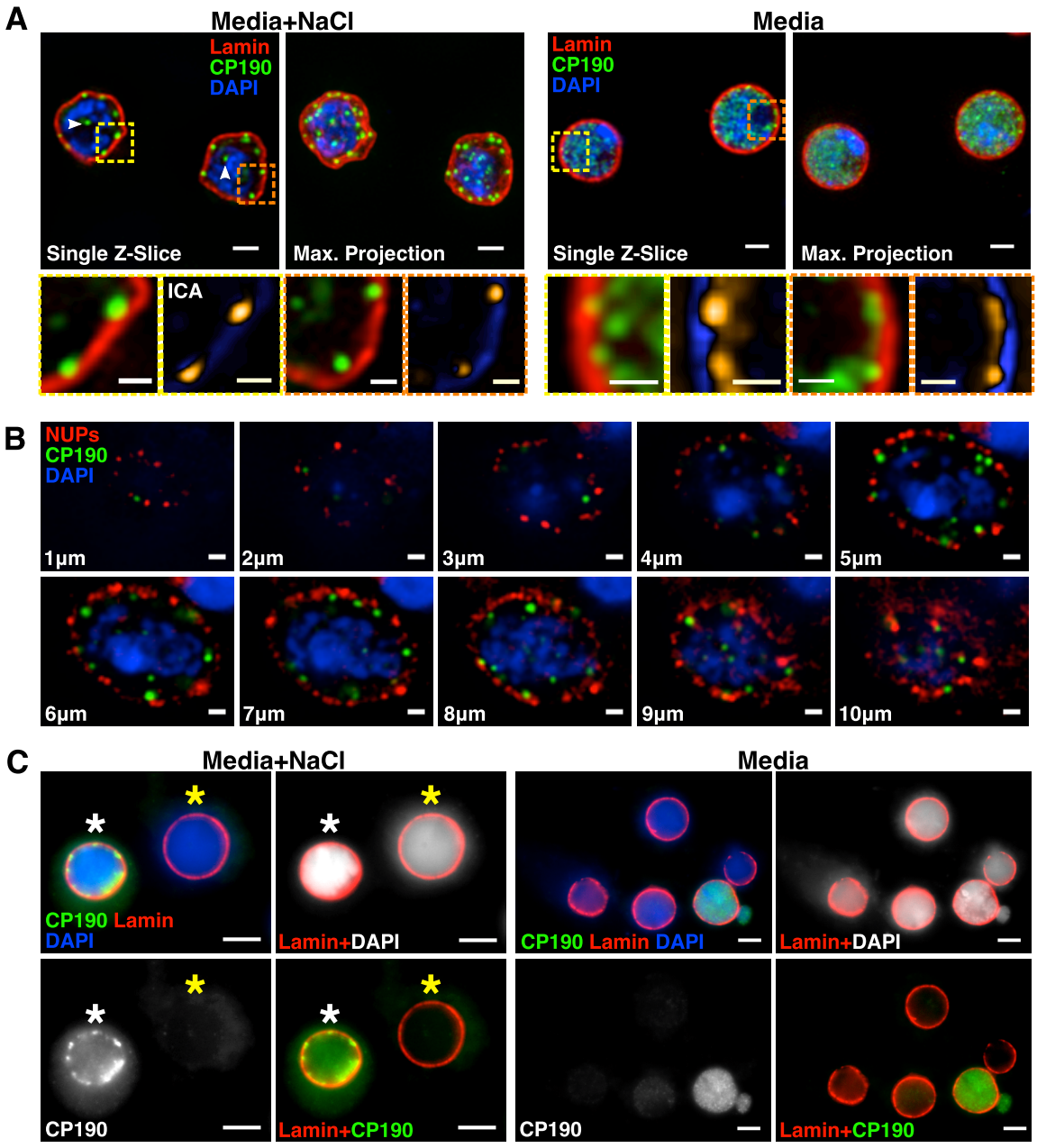

Figure 2.4. Nuclear Distribution of Insulator Bodies.
S2 cells treated with or without 250mM NaCl and stained for CP190 and lamin. Bodies localized to the interior form in DAPI-less lacunas (white arrowheads). Intensity Correlation Analysis (Boxed Regions, Inset) reveals regions of high overlap (Gold) between CP190 bodies in the nuclear periphery and lamin (A). Serial 1 µm z-slices through a S2 nucleus stressed with 250mM NaCl stained for CP190 and nuclear pore complex components (NUPs) (B). Nuclear Halos generated from 250mM NaCl-stressed S2 Cells showing a highly extracted nucleus with no CP190 signal (yellow asterisk) and a less efficiently extracted nucleus (white asterisk) showing remnants of CP190 bodies colocalized with lamin (C). Scale Bar is 2 µm and 0.5µm (inset) (A), 1 µm in (B) and 4 µm in (C).

displaying a high extraction efficiency (>95% of soluble protein removed, large DAPI halo) (Figure 2.4C). These data confirm that insulator bodies located in the nuclear periphery remain soluble and associate only transiently with the nuclear lamina.

Osmostress-Induced Insulator Body Formation Can Account For Previously Published Reports Of These Structures

A comparison of our data with descriptions of these structures given in previous work, such as number, size and nuclear distribution strongly suggests that the initially-described insulator bodies are identical to the osmostress-induced insulator bodies described here (Capelson and Corces, 2005; Capelson and Corces, 2006; Gerasimova et al., 2000; Gerasimova and Corces, 1998; Gerasimova et al., 2007; Ghosh et al., 2001; Golovnin et al., 2008; Golovnin et al., 2012; Lei and Corces, 2006; Pai et al., 2004; Ramos et al., 2011; Wood et al., 2011; Xu et al., 2004). If this is true, an obvious question arises: how might these structures have arisen in previous reports? We found that both the choice of buffer and time of dissection until fixation dictated whether tissue displayed insulator body formation or not (See Methods). Larval tissue dissected rapidly in 100 µl SFX Media or 1X PBS (<5 min) retained the diffusely speckled pattern, whereas dissection in Drosophila Ringer's Solution led to the rearrangement of CP190 and Mod to the nuclear periphery and formation of insulator bodies (Figure 2.5A). Interestingly, incubating tissues in either SFX or 1X PBS for >30 minutes under non-humidifying conditions lead to the formation of insulator bodies (Figure 2.5B). SFX media-treated tissues formed a single large body, contiguous with diffusely speckled protein located primarily in the nuclear periphery. 1X PBS incubation led to the formation of smaller, but more numerous bodies in the periphery contiguous with a smaller proportion of diffusely speckled protein, a morphology identical to previously published reports and to tissue treated with 250mM NaCl (Figure 2.1D).

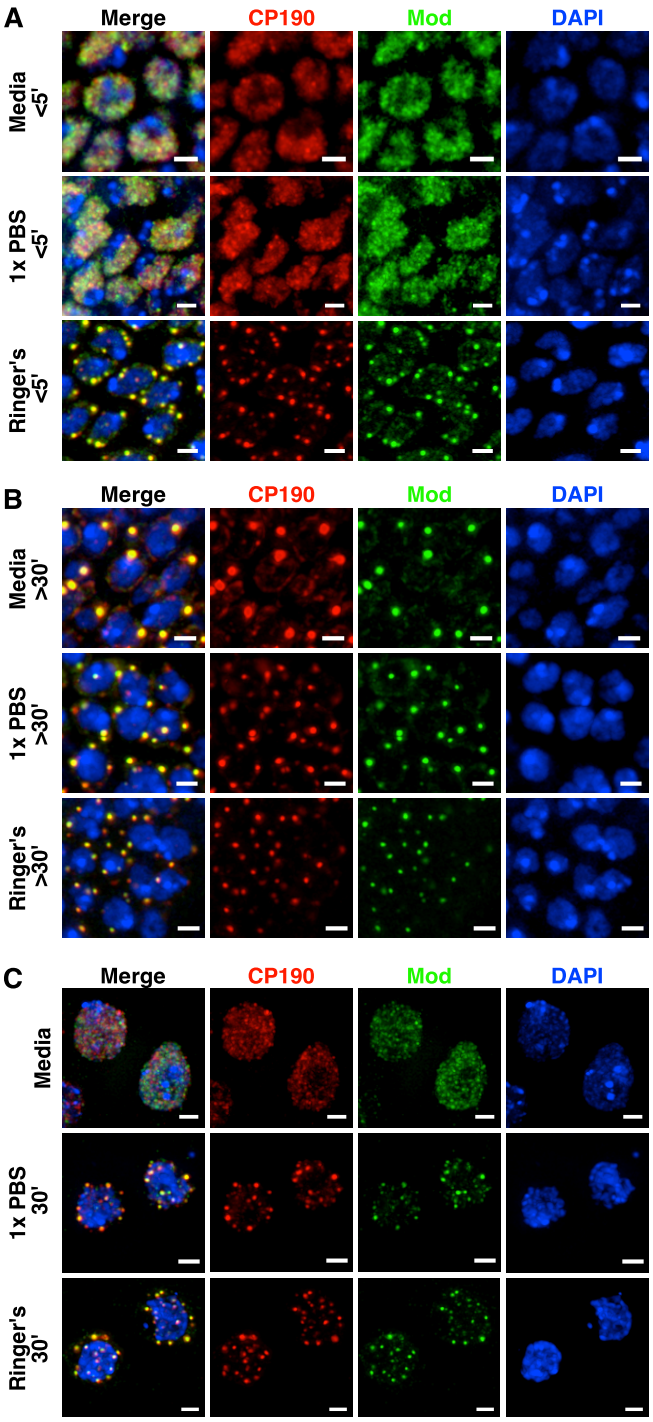

Figure 2.5. Dissection Buffer Effects On Body Formation.
Wing discs from 3rd instar larvae dissected in SFX Media, 1xPBS or Drosophila Ringer's Solution and fixed in <5' (A) or following a 30' incubation in non-humidified conditions (B) stained for CP190 and Mod(mdg4). S2 cells show a similar response following a 30' incubation in 1xPBS or Ringer's Solution (media controls were kept humidified to prevent evaporation) (C). Scale bars are 2 µm.

Similar results were obtained in S2 cells incubated with 1XPBS or Ringer's Solution, with cells near the edge of the liquid showing a more robust response (Figure 2.5C). This observation, coupled with the considerable H_2O evaporation and subsequent increase in solute concentration noted following >30 min. tissue incubations under non-humidifying conditions irrespective of buffer, likely explains how these structures were previously generated in the absence of purposeful induction.

Insulator Bodies Form Rapidly Following Stress, Display Highly Dynamic Behavior During The Duration Of Stress, And Are Readily Reversible

The drastic change in the nuclear distribution of insulator proteins in osmotically stressed versus unstressed nuclei suggested a highly dynamic transition between the two states. Using fluorescently tagged versions of BEAF-32 and Su(Hw) we were able to track the progression of insulator body formation in S2 cells during osmostress. Both BEAF-32 and Su(Hw) appear to nucleate from smaller speckles, creating larger structures—this correlates with the gradual disappearance of diffusely speckled signal throughout the nucleus as the bodies become larger and their fluorescent intensity increases, over an order of minutes as the salt concentrations gradually increases to 250 mM (Figure 2.6A and File 1). As the duration of time exposed to salt increases, the bodies remain roughly the same size and exhibit highly variable dynamics. Some bodies remain localized close to their sites of nucleation, with minimal movement, while others move readily and undergo rounds of fusion to create larger bodies, whose movement throughout the nuclear periphery appears constrained by the nuclear matrix and the chromatin mass (Figure 2.6B & 2.6C). FRAP analysis of stationary Su(Hw)::eGFP bodies suggest that these structures undergo rapid protein turnover, with recovery half times on the order of seconds (4-15 sec)

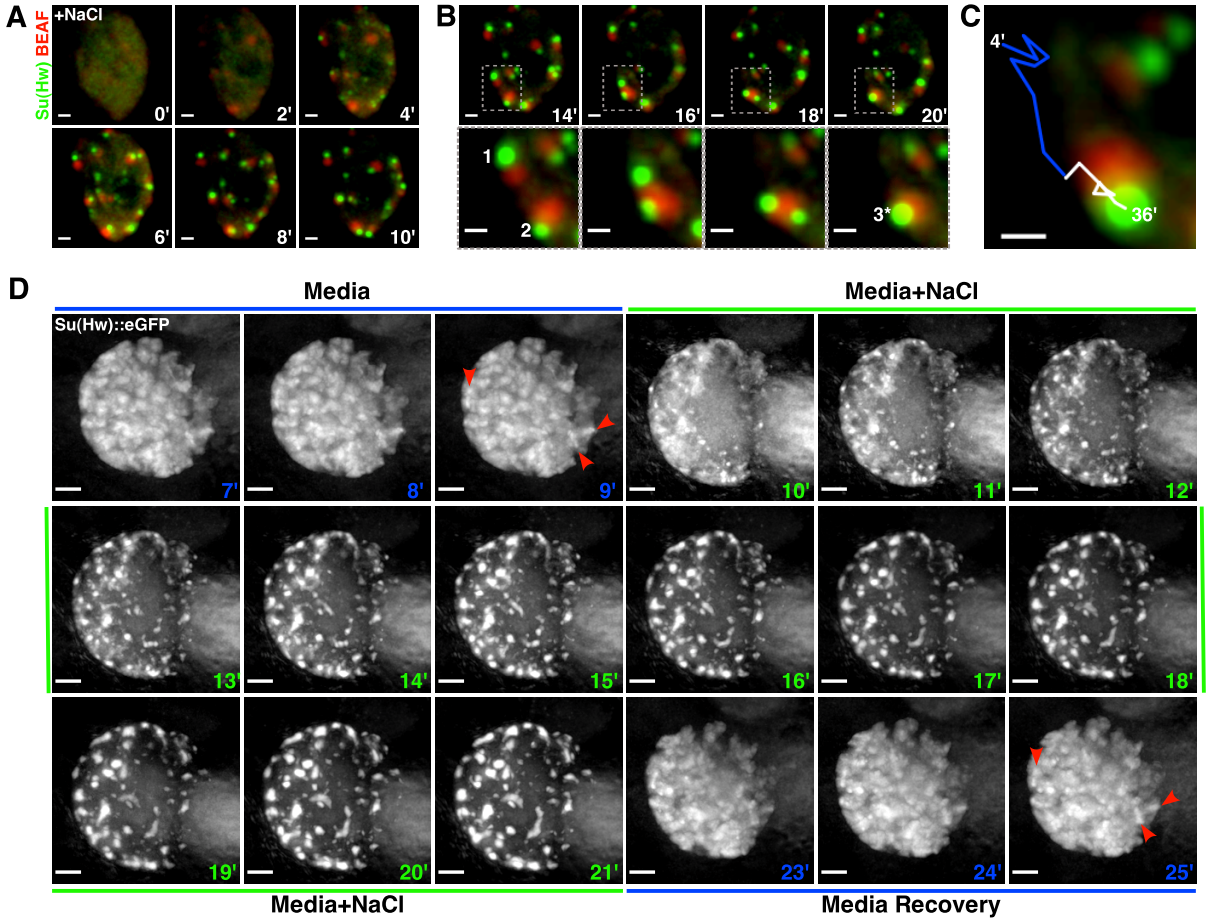

Figure 2.6. Kinetics of Insulator Body Formation and Disassembly.
Frames taken at 2 minute intervals following gradual 250 mM NaCl media addition at time 0' in S2 cells expressing Su(Hw)::eGFP and BEAF-32::mCherry. Bodies form in a matter of minutes from diffuse speckles (A) and can undergo rounds of fusion (Bodies 1 & 2) to produce larger structures (Body 3*) (B). The dynamic movement of Body 1 starting with its formation at 4' until its fusion with Body 2 at 20' (blue line) and the movement of the fused body (white line) until the final frame was acquired (36') (C). A polytene nucleus from a 3rd instar salivary gland expressing Su(Hw)::eGFP subjected to 250mM NaCl osmostress (10') followed by recovery in isotonic media (23'). Arrowheads (9', 25') mark bands of Su(Hw) (D). Scale bars are 3 µm except the lower panel in (B) and (C) which are 2 µm.

despite a relatively small fraction of free protein (M_F 30-45%) (Figure A3B). In addition to their rapid formation, insulator bodies disappear equally quickly once cells are returned to isotonic media. Using the same C-terminally tagged Su(Hw)::eGFP used in S2 cells, we generated transgenic flies containing this construct under *UAS*/Gal4 control. Polytene chromosome squashes and chromatin immunoprecipitation with α-GFP verified its DNA binding ability, while expression in the wing margin of cut^{6}; $su(Hw)^{e04061}$ restores *gypsy* insulator function, confirming that the tagged construct accurately reproduces the enhancer-blocking behavior of endogenous Su(Hw) (Figure A3C-A3E). Using explanted salivary glands dissected from 3rd instar larvae expressing this construct, we tested whether insulator body formation is reversible once osmostress is alleviated. Prior to salt addition, DNA-bound Su(Hw)::eGFP is distributed exclusively along polytene chromosomes from salivary glands (Figure 2.6D). Within ~60 seconds of salt addition, this pattern is disrupted, and throughout the duration of stress, Su(Hw) continues to relocate into bodies, with some individual foci drawing together to produce larger fusions. Remarkably, by the time the first recovery frame is acquired (~2 min.), these bodies have disappeared and the Su(Hw) signal is once again distributed on the chromosomes, which persists as the chromosomes continue to expand to their pre-stressed state. Interestingly, bands of Su(Hw) visible prior to stress are restored with a nearly identical spatial distribution in the nucleus following recovery (Figure 2.6D and File 2). Furthermore, diploid tissue subjected to two rounds of salt treatment and recovery show identical behavior, with body formation and disassembly kinetics nearly identical between both rounds of treatment (File 3 and File 4).

Insulator Body Formation Correlates With A Reduction Of Chromatin-bound Insulator Proteins

Given the distinct localization of these structures to DAPI-less regions of the stressed diploid nucleus, we hypothesized that insulator bodies may not be attached to chromatin as previously thought. We first compared the distribution of CP190 on polytene chromosomes from osmostressed and control salivary glands from third instar larvae. Whole mount staining of intact nuclei from media controls revealed multiple bands of CP190 that overlapped extensively with the chromosome arms (Figure 2.7A), reflecting the chromatin-bound state of this protein. However, these bands were absent from osmostressed nuclei and virtually all of the CP190 was instead confined to insulator bodies located in the nuclear periphery and interior spaces between the chromosome arms (Figure 2.7B), strongly suggesting that insulator proteins are removed from chromatin in order to form bodies.

To test this hypothesis, we used chromatin immunoprecipitation (ChIP) to biochemically measure chromatin removal during osmostress. Using S2 cells, we tested chromatin enrichment during stress at 3 types of Su(Hw) insulators: the *gypsy* insulator (Su(Hw), CP190 and Mod(mdg4)67.2); the *homie* super insulator (all known insulator proteins) (Fujioka et al., 2009) and an endogenous intragenic insulator (3L:12247800) that binds only to Su(Hw). All stressed samples show a ~50-80% decrease in the amount of chromatin bound Su(Hw) compared to media-only controls, depending on the insulator. Both *gypsy* and the Su(Hw)-only insulator show the largest decrease (~80%), while *homie* shows less of a reduction (~50%) (Figure 2.7C-2.7E). We have also observed similar reductions in Su(Hw) enrichment at these insulators in cells treated with sucrose, in addition to seven other Su(Hw) binding sites, suggesting that this

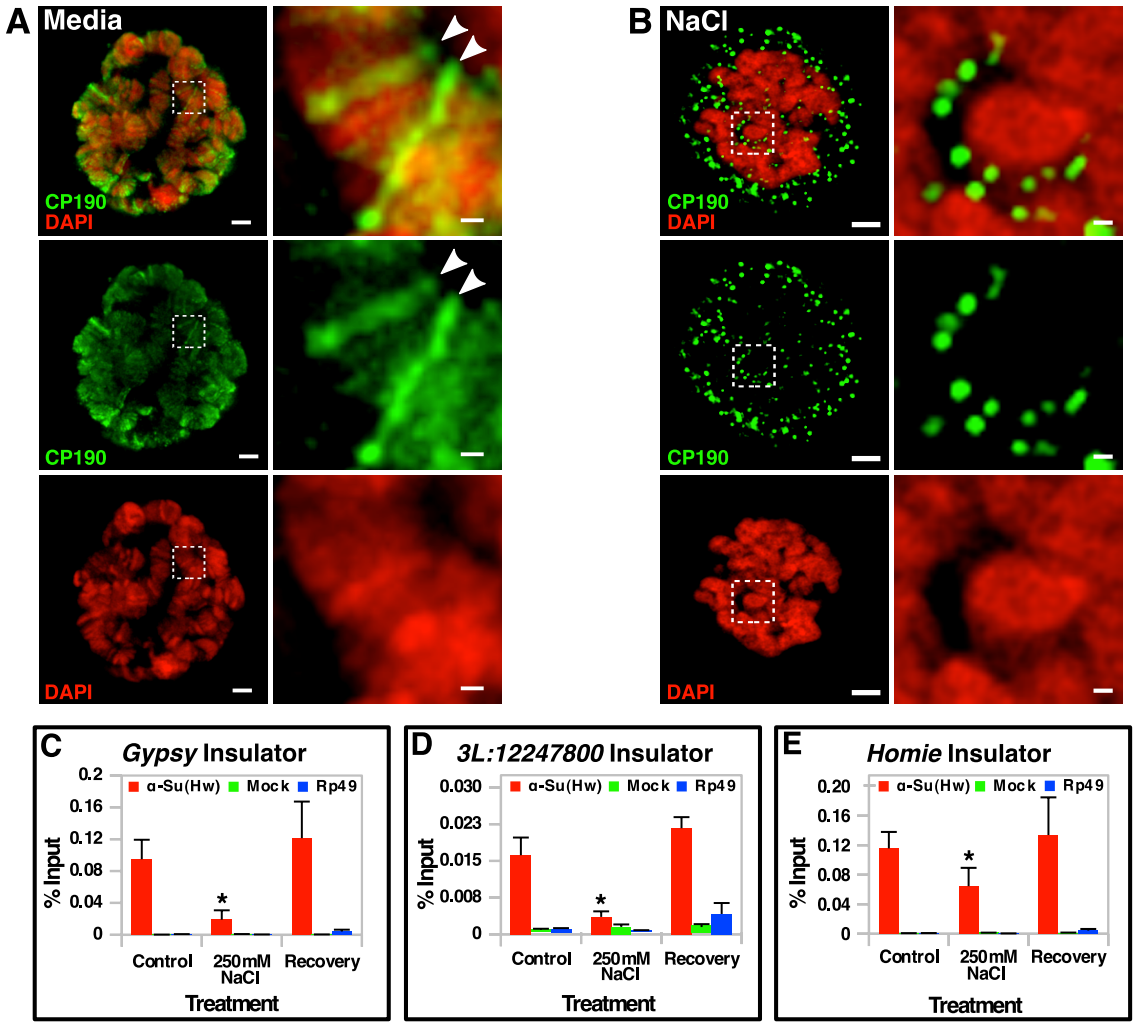

Figure 2.7. Insulator Bodies Are Not Bound To Chromatin.
A media-treated salivary gland polytene nucleus labeled with CP190 showing the expected band pattern (inset, arrowheads) (A). A polytene nucleus stressed with 250mM NaCl labeled with CP190 shows bodies in the nuclear periphery and interchromosomal spaces lacking DAPI (inset) (B). Chromatin Immunoprecipitation (ChIP) of Su(Hw) at gypsy (C), 3L:12247800 (D) and homie insulators (E) in media, stressed, and recovery S2 cells. Asterisks mark reductions significantly different from media controls (Student's paired t-test, p=0.05. Error bars represent S.E.M.). Scale bars in (A) and (B) are 3 μm and 0.5 μm (insets).

behavior is not restricted to a specific subset of insulators or is caused by the biophysical effects of NaCl. Additionally, chromosome conformation capture (3C) experiments further support the idea of chromatin removal during stress, as looping contacts at the *muscleblind* (*mbl)* locus that are disrupted in Su(Hw)-knockdown cells are similarly disrupted following osmostress (Figure A3F).

This data, combined with our Su(Hw)::eGFP live imaging stress+recovery results suggested a model in which insulator body formation correlates with a reduction in chromatin-bound insulator proteins that is restored upon recovery as the bodies disassemble and the normal chromatin architecture is restored. To test this hypothesis, we measured Su(Hw) enrichment at each insulator following 2.5 minute recovery in isotonic media after 20 min osmostress. Not surprisingly, Su(Hw) enrichment after recovery is restored to levels greater than or equal to those observed for media controls (Figures 2.7C-2.7E), verifying that insulator body formation can be used as a proxy to monitor the chromatin-bound state of insulator proteins.

Differential Requirement For Insulator Protein Recruitment To Insulator Bodies

Given that insulator bodies are highly ordered structures containing a reproducible arrangement of insulator proteins, we wondered whether removal of any one protein would disrupt their formation. Previous work has suggested that full length CP190 is required for formation of insulator bodies marked with Su(Hw) and Mod(mdg4)67.2 in both S2 cells and larval tissue (Golovnin et al., 2012; Pai et al., 2004). However, shRNA-mediated knockdown of CP190 in the posterior compartment of wing discs from 3rd instar larvae using a *UAS-DCR2; engrailed-Gal4* driver did not disrupt the ability of Mod(mdg4)67.2 to form bodies under conditions of

osmostress, which were morphologically identical to those formed in the anterior compartment containing wildtype levels of CP190 (Figure 2.8A). Conversely, DsRNA-mediated knockdown of CP190 in S2 cells did impair the ability of Mod(mdg4)67.2 to form bodies during osmostress, remaining diffusely speckled despite extensive chromatin condensation. Interestingly, even small amounts of CP190 present following incomplete knockdown leads to nucleation of small bodies marked by both CP190 and Mod(mdg4)67.2 in these cells (Figure A4A).

To confirm that CP190 does not significantly influence Mod(mdg4)67.2 insulator body behavior in tissue, we took advantage of two trans-heterozygous CP190 allele combinations ($CP190^{H31-2}/CP190^{P11}$ and $CP190^{4-1}/CP190^{P11}$) given that flies carrying CP190 homozygous null mutations are embryonic lethal (Pai et al., 2004). $CP190^{P11}$ is a large deletion removing the entire CP190 locus, $CP190^{H31-2}$ produces a truncated CP190 protein possessing only the N-terminal BTB domain due to a point mutation that truncates a splice junction and the $CP190^{4-1}$ allele produces a larger truncation missing only part of the C-terminal Glu-rich domain as a result of a nonsense mutation. In either CP190 allele combination, Mod(mdg4)67.2 body formation was readily observable in all larval tissue (brains, imaginal discs and salivary glands) and indistinguishable from balanced controls (Figure 2.8B). Such findings are in agreement with shRNA-depletion of CP190 (Figure 2.8A) and suggests that Mod(mdg4)67.2 can form bodies independently of CP190 in larval tissue.

Interestingly, null mutations in *mod(mdg4)67.2* ($mod(mdg4)^{u1}$) disrupted the ability of Su(Hw), but not CP190, to enter insulator bodies in wing discs during osmostress (Figure 2.8C). In the

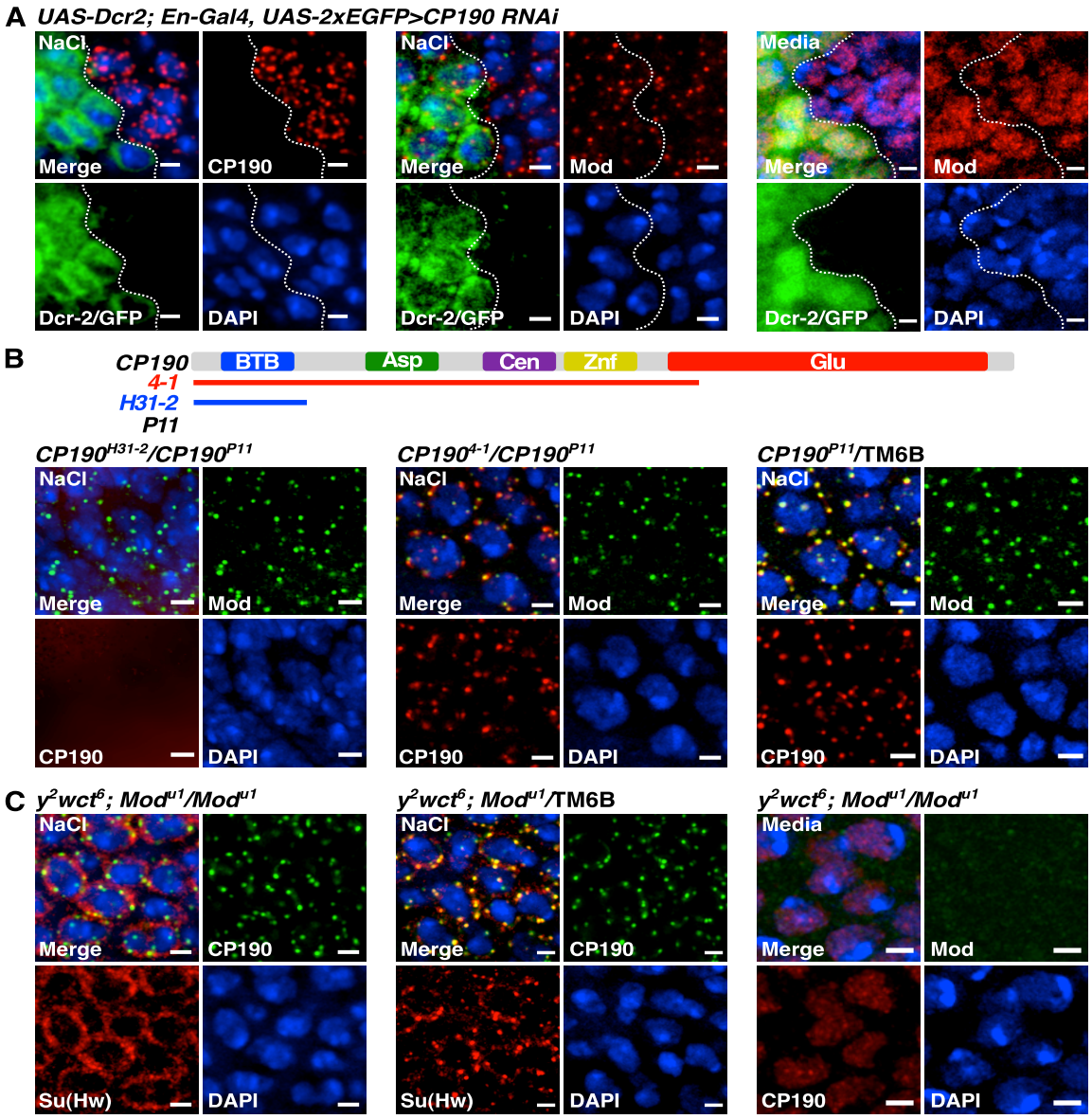

Figure 2.8. Differential Protein Requirements For Body Formation.
Wing discs from CP190-RNAi larvae stressed with 250mM NaCl and labeled with CP190 (left panel) or Mod(mdg4) (middle panel). Unstressed controls labeled with Mod(mdg4) are shown in the right panel. GFP marks Dcr-2+ knockdown cells and the dashed line demarcates the anterior-posterior axis of the wing disc (A). Wing discs from two transheterozygous CP190 mutant larvae, $CP190^{H31\text{-}2}/CP190^{P11}$ (left panel) and $CP190^{4\text{-}1}/CP190^{P11}$ (middle panel) and a balanced control containing full length CP190 (right panel) stressed with 250mM NaCl and stained with CP190 and Mod(mdg4). Note that our CP190 antibody recognizes the $CP190^{4\text{-}1}$ isoform but not the $CP190^{H31\text{-}2}$ isoform. Domains of CP190 present in each truncated allele is indicated by the colored line (B). Wing discs from null mod^{u1} homozygotes (left panel) and balanced heterozygotes (middle panel) stressed with 250 mM NaCl and stained with Su(Hw) and CP190. Mod staining (right panel) verifies absence of protein in the mod^{u1} mutant (C). Scale bars are 2 μm.

absence of Mod(mdg4)67.2, Su(Hw) remained diffusely distributed exclusively in the nuclear periphery, surrounding the condensed chromatin mass while CP190 formed insulator bodies. Only when Mod(mdg4)67.2 was present did Su(Hw) enter CP190-marked bodies, suggesting that interactions between Mod(mdg4)67.2 and Su(Hw), but not CP190 and Su(Hw), are required for Su(Hw) to enter insulator bodies. Finally, mutations in s*u(Hw)* did not alter the ability of CP190 or Mod(mdg4)67.2 to form insulator bodies in larval tissue (Figure A4B), while BEAF-32 recruitment to CP190 and Mod(mdg4)67.2 marked bodies was not impaired by reductions in any of the three *gypsy* components. Taken collectively, these data suggest that protein recruitment to insulator bodies relies on a complex network of protein-protein interactions that may be cell/tissue specific.

Insulator Body Formation Is Independent Of The DMEKK1/p38 Osmostress-Sensing Pathway

Next, we attempted to elucidate the mechanism responsible for controlling insulator body formation. We focused on the highly conserved HOG/MAPK pathway, given its central role in mediating the osmostress response in virtually all eukaryotes (Saito and Posas, 2012). Activation leads to cell cycle arrest, increased synthesis of intracellular osmolytes, and fine tuning of transcription and translation in order to allow cells to tolerate hyperosmotic conditions that would otherwise trigger cell death. At the core of this pathway is a MAP Kinase cascade that in flies includes the upstream MAPKKK, *dMekk1,* and the downstream effector MAPK, *p38.* Drosophila contains two *p38* genes, *p38a* and *p38b* that mediate the response to a variety of environmental stressors in a partially redundant manner. *p38b* and *dMekk1* are required for osmostress tolerance, while p*38a* appears to be dispensable (Craig et al., 2004; Han et al., 1998;

Inoue et al., 2001). However, mutations in either *dMekk1* (*dMekk1*^{UR36}) (Inoue et al., 2001) or p38a and p38b (*p38a*^{del}, *p38b*$^{\Delta 25}$, *p38b*$^{\Delta 45}$) (Vrailas-Mortimer et al., 2011) failed to suppress CP190 insulator body formation (Figure 2.9A-2.9D, A5A-A5C). Neither did RNAi-mediated knockdown of *JNK* (*basket*), another MAPK that is activated by Mekk1 under conditions of hyperosmolarity in mammalian cells (Yujiri et al., 1999) (Figure 2.9E, A5D). Taken collectively, these findings suggest that insulator body formation is independent of the canonical HOG/MAPK osmostress sensing pathway.

Insulator Bodies Are Also Evident In Apoptotic Nuclei

Given that insulator proteins form bodies readily in response to osmostress independently of the HOG/MAPK pathway, we wondered whether other cellular pathways might also trigger formation. We focused on cell death, particularly apoptosis, given the morphological similarities between cells in the initial stages of apoptosis and those under osmotic shock (Burg et al., 2007). To test this hypothesis, we examined eye/antennal discs from Drop^{Mio} 3rd instar larvae, in which retinal precursor cells undergo cell death due to arrested furrow progression (Mozer, 2001). In death regions, CP190 forms bodies in a subset of apoptotic nuclei (marked with cleaved caspase-3) reminiscent of those induced during osmostress, while those cells not marked as apoptotic contain diffusely speckled CP190 signal distributed throughout the nucleus (Figure A6A and A6B). Similar results were obtained with *Bar*^{S} eye discs, with BEAF-32 foci readily observable in these apoptotic tissues as well. Thus, both osmostress and cell death trigger formation of insulator bodies.

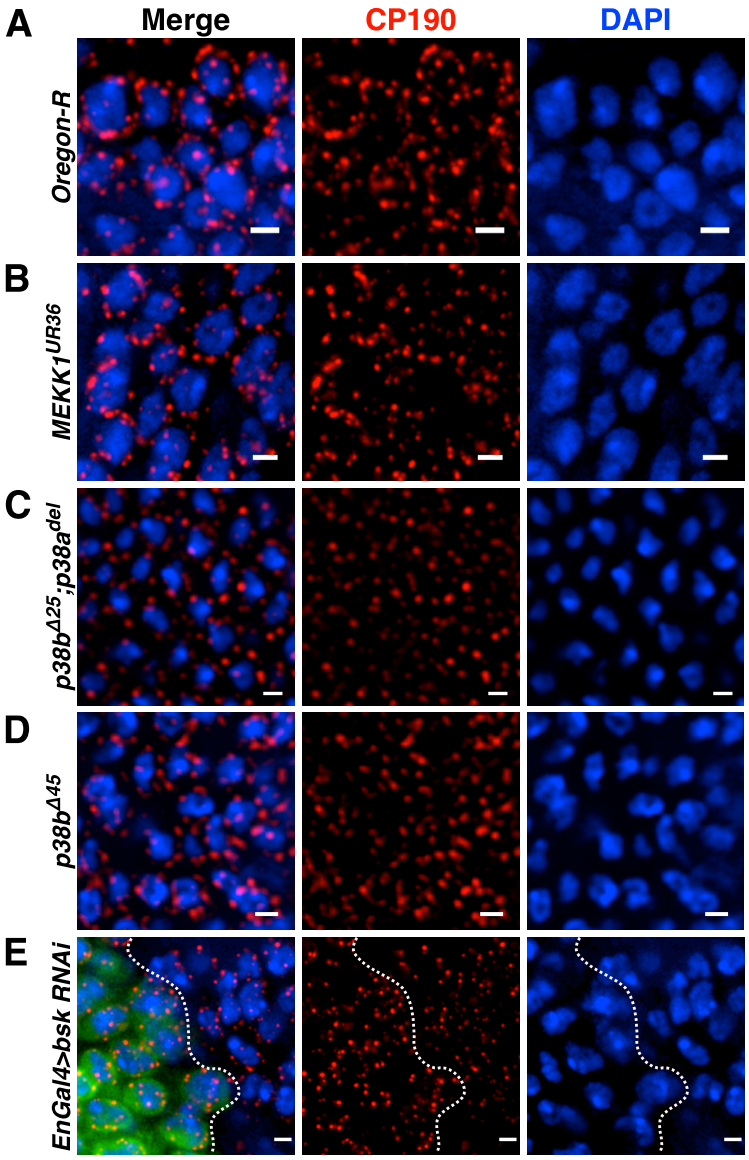

Figure 2.9. Effects Of HOG/MAPK Mutations On Body Formation.
Wing discs from Oregon-R (A), *dMEKK1*^{UR36}/*dMEKK*^{UR36} (B), *p38b*$^{\Delta 25}$;*p38a*^{del}/ *p38b*$^{\Delta 25}$;*p38a*^{del} (C), *p38b*$^{\Delta 45}$/*p38b*$^{\Delta 45}$ (D) and *bsk (JNK)-RNAi* (GFP marks Dcr-2^{+} knockdown cells and the dashed line demarcates the anterior-posterior axis of the wing disc) (E) stressed with 250mM NaCl and stained with CP190. Scale bars are 2 μm.

Discussion

Our results demonstrate that insulator bodies are a novel class of nuclear stress bodies, which to our knowledge has yet to be described in any eukaryote in response to osmostress. Our data suggest a model in which insulator proteins are removed from chromatin and form bodies in distinct nuclear territories, which are maintained throughout the duration of osmostress by constant turnover of proteins. Once the stress response is alleviated, the bodies rapidly disassemble as the individual proteins migrate back to their cognate binding sites, restoring the normal chromatin architecture observed prior to stress (Figure 2.10). Other nuclear stress bodies have been described in both Drosophila and mammals in response to heat shock, consisting of hnRNPs and transcription factors required for rapid induction and processing of heat shock responsive genes which allow the cell to adapt to elevated temperatures (Biamonti and Vourc'h, 2010). Whether insulator bodies play a functional role in the cellular response to osmostress remains to be elucidated; however, given the potential epigenetic consequences of both heat shock and osmostress (Seong et al., 2011), it is likely that a better understanding of the relationship between stress and nuclear dynamics will reveal additional mechanisms underlying environmentally-induced changes in genome function.

Our findings in light of previous work raises the question of whether osmostress-induced insulator bodies are the “same” as those identified in past reports (Capelson and Corces, 2005; Capelson and Corces, 2006; Gerasimova et al., 2000; Gerasimova and Corces, 1998; Gerasimova et al., 2007; Ghosh et al., 2001; Golovnin et al., 2008; Golovnin et al., 2012; Lei and Corces, 2006; Pai et al., 2004; Ramos et al., 2011; Wood et al., 2011; Xu et al., 2004). Previous

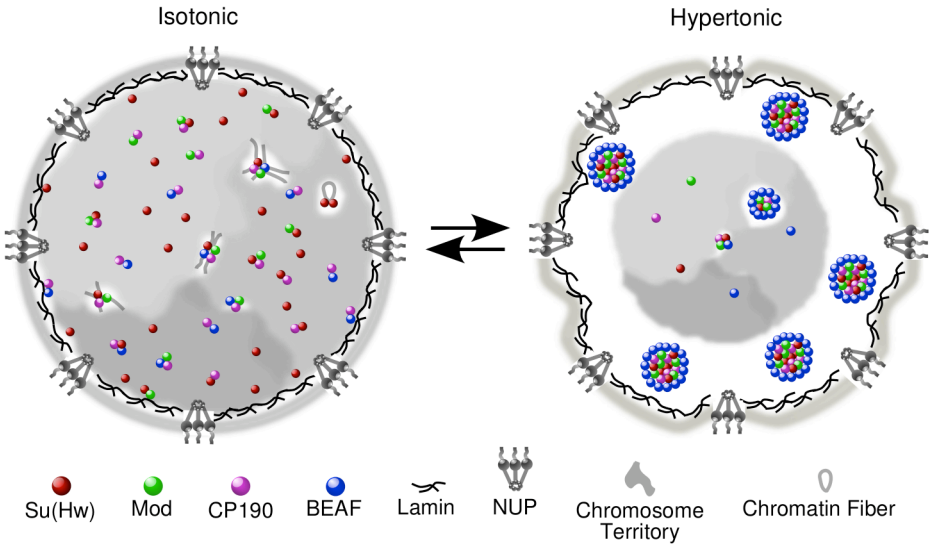

Figure 2.10. Model For Insulator Body Formation During Osmostress And Cell Death.

characterization of these structures has relied on three main criteria: the number of bodies per diploid nucleus (10-30), their nuclear distribution (nuclear periphery) and extensive colocalization between insulator proteins. Our data satisfies all of these requirements, and we argue that osmostress-induced insulator bodies are identical to those published previously—if not for the simple reason that we were unable to observe these structures in any other cellular context. Furthermore, we have provided a likely explanation of how these structures may have arisen in the absence of purposeful induction. Insulator body formation does not occur in small volumes of 1x PBS if tissues are dissected and fixed rapidly (<5 minutes); however, extended tissue dissections prior to fixation in small volumes of 1x PBS under non-humidifying conditions (i.e., on the benchtop/under the stereoscope) leads to the formation of insulator bodies that are identical to those purposefully treated with elevated NaCl, sorbitol or sucrose making it simple to envisage how these structures formed in previous reports. Perhaps most importantly, however, is that this also creates the potential for misinterpretation of data. Our NaCl-gradient results suggests that the robustness of the insulator body response correlates with the severity of the osmostress, which could lead to a range of insulator body phenotypes if the initial dissection and incubation conditions prior to fixation are not properly controlled between samples being compared. We therefore urge caution in using these structures as a metric of insulator function/activity, unless such controls are implemented.

Along these lines, we have provided direct biochemical evidence that these structures are not the physical manifestations of multiple DNA-bound insulator proteins as chromatin-bound Su(Hw) is dramatically reduced during body formation. Given that insulator properties, such as

enhancer/heterochromatin blocking and chromatin looping are conferred by insulator proteins physically bound to DNA, it is extremely unlikely that these structures contribute to these processes during osmostress. Although we cannot rule out that there may be a handful of insulator sites that do not lose their proteins and instead act to nucleate and tether these structures to chromatin, we argue that this is unlikely for the following reasons: first, immunostaining in stressed polytene chromosomes reveals complete loss of the characteristic CP190 band-like signals rather than an increase in signal at a handful of binding sites, which would be expected if they acted as nucleation hubs. Additionally, unlike diploid cells in which DAPI would likely lack the sensitivity to detect the presence of a few peripherally-associated chromatin fibers that might still remain attached to bodies following global chromatin condensation, the organization of polytene chromosomes ensures that individual chromatids remain in close association with one another and there is a clear demarcation between peripherally localized bodies and chromatin when polytene chromosomes condense. Nonetheless, ChIP-Seq coupled with immuno-FISH of potential nucleation sites will be needed to accurately address this possibility.

It is interesting to point out that the distribution of insulator proteins in unstressed diploid cells, consisting of hundreds of tiny speckles distributed throughout the volume of the nucleus, are morphologically similar to Polycomb (PcG) bodies which have been shown to be functionally relevant given that they physically colocalize with DNA to contribute to Hox gene silencing (Bantignies et al., 2011). We favor the idea that insulator speckles under unstressed conditions also reflect the functional state of DNA-bound insulator proteins, and are likely the physical manifestations of localized chromatin looping between insulator sites to establish chromatin

domains. This is supported by the fact that there are distinct speckles for each insulator protein that overlap with other insulator proteins at some, but not all speckles, likely to be a reflection of the combinatorial binding of insulator proteins to different sites as measured by ChIP (Bushey et al., 2009; Negre et al., 2010). Now that high-resolution looping maps of the Drosophila genome are available (Hou et al., 2012; Sexton et al., 2012), it would be possible to test this using immuno-FISH, particularly between physical domain borders that have been shown to undergo long-range looping contacts and are demarcated by specific combinations of insulator proteins (Hou et al., 2012).

As for the structures themselves, it is important to reiterate that they do not appear to be insoluble aggregates of randomly associated proteins. Insulator bodies do not colocalize with mCherry-tagged version of Hsp70, Hsp40 or Pros54 (a 26S proteasome subunit) suggesting these structures are not sites of unfolded proteins or those targeted for degradation. They contain a reproducible arrangement of insulator proteins within these structures, exemplified by the presence of BEAF-32 as a donut-shaped pattern around a spherical core of CP190, dCTCF, Su(Hw) and Mod(mdg4)67.2. Other donut-shaped nuclear bodies have been described using electron microscopy, such as PML bodies and clastosomes (Lafarga et al., 2002; Zhong et al., 2000), and it is possible that even the spherical proteins also manifest themselves as ring or donut structures that are not readily observable given the resolution limits of light microscopy. Future super resolution imaging and electron microscopy will be critical for understanding the organization of these structures. Furthermore, biochemical isolation of these structures followed by mass spectrometry will be required to identify the large number of proteins involved, which

will be crucial for identifying novel insulator proteins and other interacting partners, potential posttranslational modifications required for body formation and perhaps providing insight into what the functional role of these structures might be.

Given this relatively ordered structural arrangement dictated by protein-protein contacts, it might be expected that a loss of certain core "scaffolding" proteins would suppress insulator body formation. Though this is difficult to assess globally, since we still do not know the full complement of proteins that are in these structures, it was recently shown in S2 cells that DsRNA-mediated knockdown of CP190 disrupts the ability of Mod(mdg4) to enter bodies, while similar reductions in Mod(mdg4) impairs the ability of Su(Hw), but not CP190 to enter bodies (Golovnin et al., 2012). Interestingly, we found a similar CP190-dependence for Mod(mdg4) in S2 cells, but not larval tissue, where Mod(mdg4) was still able to enter bodies independently of CP190. This was confirmed not only in CP190 shRNA-depleted tissue, but in two other CP190 truncated mutants as well that had previously been shown to disrupt Mod(mdg4) localization in tissue (Lei and Corces, 2006), though this was likely due to misinterpretation as outlined above. However, Mod(mdg4) was required for Su(Hw) to localize to CP190 bodies in tissues, in agreement with previous work in S2 cells (Golovnin et al., 2012). One interpretation of our results would be that there are tissue/cell type-specific requirements for proteins to be recruited to insulator bodies, such as the availability of other proteins, posttranslational modifications, or even RNA. For example, Mod(mdg4) could be recruited to bodies in a redundant manner, by either CP190 or some tissue-specific protein/RNA that is present in larval tissue but not S2 cells. A more likely possibility, however, involves differences in posttranslational modifications to the

proteins themselves. We find that CP190 is SUMOylated in response to osmostress in S2 cells that is removed upon recovery (Figure A5E-A5G); however, SUMOylated Mod(mdg4) or Su(Hw) was not detected under the same conditions with our antibodies, which might be the result of epitope masking. Interestingly, CP190, Mod(mdg4) and Su(Hw) all contain computationally predicted SUMO interacting motifs (SIM), which for Mod(mdg4) happens to be within the Q-rich domain that has been shown to be necessary for its localization to bodies in S2 cells and also contains a SUMOylation motif required for body formation (Golovnin et al., 2008; Golovnin et al., 2012). Perhaps this SIM mediates Mod(mdg4)'s interaction with SUMOylated CP190 in S2 cells, which would explain its dependence on CP190 in this cell type. Nonetheless, it will be important for future work addressing the role of SUMO in body formation to take into account potential discrepancies between cell/tissue types.

Additionally, we have yet to identify a signal transduction pathway that might coordinate these potential posttranslational modifications. We have ruled out the canonical HOG/p38 MAPK osmostress sensing pathway; however, it is interesting that these structures are also present in a subset of apoptotic nuclei. The phenotypic similarities between the two processes, such as cell volume reduction, chromatin condensation and disrupted lamin suggests that the two might not be mutually exclusive and may share similar signaling pathways (Burg et al., 2007), which might involve both biological and mechanical/biophysical cues. Future characterization of the link with apoptosis may be critical to resolving these issues, in addition to future studies addressing the role of molecular crowding in this phenomenon (Richter et al., 2008).

Finally, the physiological significance of chromatin removal and insulator body formation remains unknown. It appears not to be required for chromatin compaction or to directly induce changes in gene expression (Figure A6C-A6E). Heat shock has been shown to reduce chromatin bound CP190, but presumably not other insulator proteins and does not lead to body formation (Oliver et al., 2010; Wood et al., 2011), suggesting that this phenomenon may be restricted to osmostress. It is intriguing that other types of chromatin binding proteins do not show a dramatic reorganization during osmostress, and given the central role insulators play organizing the chromatin fiber into higher order structures, we favor the idea that insulators do play a functional role in the osmostress response and are specifically targeted to form bodies. Perhaps removal of insulators from DNA is needed to disrupt or reorganize chromatin domains that are needed for the genome to execute otherwise quiescent gene regulatory programs to adapt to osmostress. In such a state, the nucleus would be primed for rapid recovery once the stress is alleviated, as the insulator proteins stored in the bodies would be readily available to re-bind chromatin, reestablishing the domains present in the unstressed state and restoring the default chromatin architecture for that cell type. Future high throughput studies, including RNA/ChIP-Seq and Hi-C to examine global changes in transcript levels, chromatin bound insulators and looping contacts, will be necessary for testing such a hypothesis.

Materials And Methods

Fly Stocks & Husbandry

All stocks and crosses were maintained on standard cornmeal-agar media at 25°C.

Microinjection to generate transgenic lines *yw; P{SuHw::eGFP, w^+}* was performed by Genetivision (Houston TX). Bloomington Stock Center: *$y^1 sc^* v^1$; P{TRiP.HMS00845}attP2*

(CP190 RNAi, Stock #33903). *y^1v^1; P{TRiP.JF01275}attP2* (JNK RNAi, Stock #31323). *P{UAS-Dcr-2.D}^{1}, w^{1118}; P{en2.4-GAL4}^{e16E}, P{UAS-2xEGFP}AH2* (Stock #25752). *w*; P{GAL4-vg.M}2; TM2/TM6B, Tb^1* (Stock # 6819). *w^{1118}; PBac{RB]Su(Hw)e04061/TM6B, Tb^1* (Stock #18224). CP190 mutants (Victor Corces): *y^2wct^6; CP190^{H31-2}/TM6B, Tb^1, y^2wct^6; CP190$^{4-1}$/TM6B, Tb^1* and *CP190^{P11}/TM6B Tb^1*. Gal4 Drivers: *yw; Hsp70-Gal4/Cyo* (Bruce McKee) and *w*; GMR-Gal4* (Tom Dockendorff). *Mekk1* mutant (Hyung Don Ryoo): *FRT82B, MeKK1^{UR36}/TM6B, Tb^1*. *p38* mutants (Alysia Vrailas-Mortimer): *p38b$^{\Delta 45}$* and *p38b$^{\Delta 25}$/Cyo, GFP; p38a^{del}*. Our lab: *y^2wct^6; mod(mdg4)^{u1}/TM6B, Tb^1*. *yw; Dr^{Mio}/TM6B, Tb^1*. *y^2wct^6; P{Su(Hw)::eGFP}/Cyo; Su(Hw)^{e04061}/TM6B, Tb^1* and *y^2wct^6; P{GAL4-vg.M}2; Su(Hw)^{e04061}/TM6B, Tb^1*. RNAi crosses were maintained at 29°C to induce high levels of knockdown.

Expression Vector Construction

The pMK33-CTAP vector backbone (Veraksa et al., 2005) was used to generate dual-expression constructs containing both SuHw-eGFP and mCherry coding sequences under the control of the copper-responsive Metallothionein (Mt) promoter. Su(Hw) (amplified from ovary cDNA) and eGFP were fused in frame and inserted into the XhoI/SpeI sites of pMK33-CTAP using the InFusion HD Cloning System (Clontech) creating a C-terminally tagged construct. From this vector, the Metallothionein promoter was amplified with primers designed with a 3' NheI site and stitched back into the pMK33 NotI site using the InFusion system. Next, the mCherry coding sequence was amplified from pAN583 (gift from Andreas Nebenführ) with primers containing 5' AvrII/BswiI/AgeI sites and 3' EcoRV/KpnI/MluI sites and fused into the 3' NheI site downstream of the newly inserted Mt promoter. Final construction of the dual expression vector was

achieved by simply amplifying a coding sequence of interest (CP190, BEAF-32, CTCF, H2Av, etc.) and inserting it into either the 5' or 3' cut sites flanking mCherry to create C or N-terminally tagged fusions. Fly expression constructs were generated using the pUAST-Y vector backbone containing a 5X UAS, minimal Hsp70 promoter and w^+. *SuHw::eGFP* was amplified from pMK33 and inserted into XhoI/XbaI sites using the InFusion HD Cloning system.

Antibodies

Rat and rabbit polyclonal IgG antibodies generated against full length Su(Hw) and CP190 and Mod(mdg4)67.2 lacking only the BTB domains were previously generated in our laboratory (Wallace et al., 2010) and used at the following dilutions for immunostaining: Su(Hw) 1:50-1:300; CP190: 1:500-1:1000; Mod(mdg4)67.2: 1:250. α-Lamin Dm0 (1:100), α-BEAF-32 (1:20) and α-HP1 (C1A9) (1:100) (Developmental Studies Hybridoma Bank); α -Polycomb (1:200, Santa Cruz Biotechnology); α-Cleaved Caspase-3 (1:200, Cell Signaling Technologies). Secondary antibodies labeled with FITC or Texas Red were obtained from Jackson ImmunoResearch and used at 1:500-1:1000.

S2 Cell Culture, Transfection And DsRNA Treatment

Cells were maintained in HyClone SFX insect media (Thermo Scientific) supplemented with 1X Penicillin/Streptomycin at 25°C. Transfection of S2 cells was achieved using Lipofectin (Invitrogen). Briefly, 1-3 µg of vector was combined with 15 µl Lipofectin in 1 ml of SFX media and overlayed on $2x10^6$ cells for 24 hours. After 3-4 days, SFX media containing 300 µg/ml Hygromycin (Invitrogen) was added to select stable lines. Expression was induced with 500 µM $CuSO_4^{2-} \cdot 5H_2O$ added to each flask 14-16 hrs prior to imaging. For DsRNA treatment, ~10^6 S2

cells were seeded in a 6 well plate and treated with 15 µg CP190 DsRNA (Butcher et al., 2004) and prepped for imaging as described (Rogers and Rogers, 2008). Knockdown levels were monitored by lysing ~10^7 S2 cells on ice in 100 µl RIPA buffer supplemented with 1x Protease Inhibitor (Roche). 12 µg of lysate was resolved in a 7.5% acrylamide gel, wet transferred at 4°C overnight (10V) and probed with α-CP190 (1:1500) and α-Su(Hw) (1:1000).

Stress Treatment And Immunostaining

S2 cells 3-5 days post subculture were allowed to adhere to a Poly-L lysine coverslip for 30 minutes in a covered 35 mm cell culture dish. To induce osmostress, media was removed and quickly replaced with fresh SFX media supplemented with the indicated concentration of osmolyte (from a 5M stock). Controls were kept in conditioned media. Cells were stressed for 20 minutes and then immunostained as described (Rogers and Rogers, 2008). For larval and ovary tissue, Oregon-R 3rd instar larvae and adult females were quickly dissected in SFX media and transferred to 0.5 ml tubes containing 500 µl of SFX Media supplemented with 0.5% BSA and the required concentration of osmolyte. Controls were treated similarly, with the exception that the osmolyte was excluded. Tubes were rotated at RT for 20 minutes to induce osmostress and immunostained as described (Saint Phalle, 2003) with the following adjustments: tissues were fixed for 25 minutes in 0.5% Triton-X/4% PFA and block-permeabilized in 0.5% Triton-X/1%BSA for at least 2 hrs with rotation to speed fixation and increase antigen accessibility in salivary glands.

For heat shock (HS), 1 ml of ~50% confluent S2 cells were seeded onto a coverslip containing Poly-L lysine in a 35 mm cell culture dish and allowed to adhere overnight at 25°C. Dishes were

then placed in a 37°C H_2O bath such that the level of water outside the dish and media inside the dish were equal and incubated for either 20 or 60 minutes, then fixed immediately for immunostaining. To induce chromatin condensation independently of osmotic stress, S2 cells adhered to coverslips were treated with 50 mM Na Azide in media for 20 minutes at room temperature.

Buffer/Dissection Condition Stress Test

To test the effects of dissection buffer/conditions on insulator body formation, 3rd instar larvae were dissected in shallow depression slides containing 100 µl of either HyClone SFX insect media (Thermo Scientific), 1x PBS (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, pH 7.4) or Drosophila Ringer's Solution (3mM $CaCl_2 \cdot H_2O$, 182mM KCl, 46mM NaCl, 10mM Tris-HCl, pH 7.2). Dissections were performed in ~5 minutes and either fixed immediately with 4% PFA/0.3% Triton-X/1xPBS or left to incubate on the benchtop under non-humidifying conditions for 30 minutes prior to fixation. Subsequent immunostaining was then carried out as described. For S2 cells, 100 µl of S2 cells ~50% confluent were pelleted at 500 rpm and resuspended in 100 µl of either 1xPBS or Ringer's Solution and distributed evenly over a 22x22mm Poly-L lysine coverslip and allowed to adhere for 30 minutes in a 35mm cell culture dish with the lid removed prior to fixation. Controls were kept covered in conditioned media and allowed to adhere for the same amount of time. Cells were then stained as described.

Detergent Permeabilization Prior To Osmostress

S2 cells were allowed to adhere to Concavalin-A treated coverslips in conditioned media for 30 minutes at room temperature. The media was then aspirated and Triton-X diluted to 0.2% in

conditioned media was then quickly overlaid on the cells and incubated for 5 minutes at room temperature. Controls were treated with conditioned media without detergent. Following aspiration 250mM NaCl+Media was overlaid on the cells, treated for 15 minutes at room temperature and then fixed and immunostained as described.

Nuclear Halos

Nuclear halos from S2 were prepared as described (Byrd and Corces, 2003; Pathak et al., 2007) with the following exceptions. First, S2 cells were allowed to attach to Poly-L lysine coverslips for 45 minutes at room temperature. Cells were then either treated with media containing 250 mM NaCl or left untreated for 20 minutes, followed by extraction with 2M NaCl (2M NaCl, 5mM HEPES pH 7.5, 2mM KCl, 2mM EDTA, 0.05% Triton-X and 1X Protease Inhibitor (Roche)) for 5 minutes at RT. Slips were briefly rinsed 3X in PBS and then fixed with 4% PFA for 10 minutes. Subsequent immunostaining was performed as described above.

Microscopy And Live Imaging

All immunostaining and live imaging experiments were performed on a Leica DM6000B widefield epifluorescent microscope equipped with Leica HCX PLAN APO 63x/1.4NA and 100x/1.35NA oil immersion objectives and a Hamamatsu ORCA-ER CCD camera. Simple PCI (v6.6) was used for image acquisition. Image processing of raw Z-stacks was performed using AutoQuant's 3D Deconvolution Algorithm utilizing an adaptive (blind) PSF implemented into Leica Deblur (v2.3.2) software. Final brightness/contrast adjustments following deconvolution were performed using ImageJ (v1.47b). For live imaging experiments, S2 cells were placed into an Ibidi μ-Slide upright$^{0.8}$ imaging chamber and allowed to adhere to the top of the chamber for

20 minutes. A perfusion system utilizing gravity flow allowed for the gradual addition of SFX media containing osmolyte to induce osmostress. All experiments were performed at room temperature (~23°C). Lamp output (100W) for each channel was reduced to 10% and experiments were kept under 2 hrs to minimize photobleaching, toxicity and focus drift. For salivary glands & imaginal discs, tissues were dissected in SFX media and anchored to a coverslip containing Poly-L lysine. This coverslip was then oriented tissue-side down over top of a depression slide filled with SFX media, leaving one edge open to allow for gas exchange and access to the media pool. A thin layer of nail polish applied to one corner prevented movement of the coverslip during imaging. Salt treatment and recovery were performed by carefully removing the existing media in the depression slide with a pipette and slowly adding back the media of interest. Z-stacks were taken at the indicated time intervals and each raw stack was then processed using AutoQuant software as described for fixed samples and assembled using ImageJ. Final brightness/contrast adjustments and further image analysis was also performed using ImageJ and the plugins MTrackJ (E. Meijering, 2012) and Intensity Correlation Analysis (Li et al., 2004). DAPI, FITC and Texas Red fluorochromes were used for fixed samples whereas eGFP and mCherry were used for live imaging.

FRAP Analysis

FRAP was performed using the Marianas spinning disc confocal platform (Intelligent Imaging Innovations) consisting of an inverted Zeiss Axio Observer microscope outfitted with a CSU-X1 spinning disc, Photometrics Evolve 512 EMCCD camera, Vector high speed point scanner and a Plan-Apochromat 100x/1.4NA oil objective. S2 cells expressing Su(Hw)-eGFP were stressed with 250 mM NaCl+SFX media and cells expressing low amounts of transfected protein were

imaged at room temperature (~23°C). Roughly 3 bodies per cell were bleached simultaneously using a 488 nm laser set to 100% (50 mW), frames acquired every 250 milliseconds and recoveries were recorded and monitored in real time using Slidebook 5.0 software and terminated once the curve plateaued. ImageJ was used to extract intensity measurements from each ROI. Raw intensities were corrected for photobleaching and subtracted from background as described (Zheng et al., 2011) and normalized, with the final prebleach frame intensity taken to be 1. Recovery curves were plotted and fitted to a 1-phase association exponential function using GraphPad Prism 6 software. The mobile fraction (M_F) and half-life of recovery ($t_{1/2}$) were calculated from this curve as described (Reits and Neefjes, 2001).

Chromatin Immunoprecipitation

10^7 S2 cells were used for ChIP. Osmolyte stress was performed in 1.5 ml tubes containing 250 mM NaCl in SFX media and rotated for 20 min at RT. Controls were kept in the same conditioned media. For recovery treatments, cells were stressed with osmolyte for 20 min, pelleted at 2500 g for 2.5 min and then gently resuspended in 1 ml of fresh SFX media and rotated for 2.5 minutes at room temperature. ChIP was performed essentially as described (Wu et al., 2003) as follows: crosslinking was performed by adding 16% paraformaldehyde to a final concentration of 1% and tubes rotated for 10 minutes at room temperature. Cells were then pelleted and resuspended in 1% SDS Lysis Buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl pH 8.1 and 1X Protease Inhibitor (Roche)) and placed on ice for 10 min. Chromatin was sheared to an average size of 500 bp using a Diagenode Bioruptor (Denville, NJ) coupled to a continual flow 4°C H_2O bath using the following parameters: high power, 30x cycles of 30 sec. on, 30 sec. off. Insoluble debris were pelleted and the supernatant diluted 10 fold in IP buffer (0.01% SDS,

1.1% Triton-X, 1.2mM EDTA, 16.7mM Tris-HCl pH 8.1, 16.7mM NaCl, 1X Protease Inhibitor (Roche)). Diluted chromatin extracts were precleared using 100 µl of Protein A-agarose beads (Invitrogen) for 30 min at 4°C. 300 µl of this solution was used for each pulldown, using 5 µl of α-SuHw (previously ChIP-validated, (Wallace et al., 2010)) overnight at 4°C; mock controls were also included. Antibody-antigen complexes were recovered using 35 µl Protein A-agarose beads for 2 hrs at 4°C, the beads harvested by centrifugation and serially washed for 20 minutes each at 4°C with 1 ml of the following wash buffers: Low Salt Wash (0.1% SDS, 1% Triton, 2mM EDTA, 20 mM Tris-HCl pH 8.1, 150mM NaCl); High Salt Wash (0.1% SDS, 1% Triton, 2mM EDTA, 20 mM Tris-HCl pH 8.1, 500mM NaCl); Lithium Wash (0.25M LiCl, 1%NP-40, 1% Na Deoxycholate, 1mM EDTA, 10mM Tris-HCl pH 8.1); 1X TE. Beads were then resuspended in 1 ml of TE and transferred to a new tube. Antibody-Antigen complexes were eluted from the beads using 500 µl of Elution Buffer (1% SDS, 0.1M $NaHCO_3$) and incubated at RT for 30 min. 10% Input Controls were also diluted in Elution Buffer to final volume of 300 µl, and 20 µl of 5M NaCl added to all samples and placed at 65°C overnight to reverse formaldehyde crosslinks. 2 µl of 10mg/ml proteinase K was then added and incubated 1 hr at 65°C. Solutions were extracted once with an equal volume of phenol:chloroform:isoamyl alcohol, EtOH precipitated, washed and resuspended in 25 µl nuclease free H_2O.

Real Time PCR Quantification Of ChIP Samples

Runs were performed on a BioRad iQ5 cycler using iQ SYBR Green Supermix (BioRad). Three biological replicates for each treatment (control, osmolyte stress and recovery) were included in addition to 3 technical replicates for each. Primer sets for each insulator were designed based on ChIP-chip data (Negre et al., 2010) and all gave 98-101% amplification efficiencies. Rp49 was

used as a negative control region. Data were normalized using the Percent Input Method and a paired Student's t-test was used to assess statistical significance.

Quantitative Chromosome Conformation Capture

3C was carried out in ~10^8 S2 cells treated with or without 250 mM NaCl for 5 minutes at RT as described (Comet et al., 2011) with the following adjustments: cells were crosslinked for 10 min at RT on a rotating platform using 10 ml of a 1%PFA/SFX media solution, dounce homogenized (20 strokes) in NP buffer on ice, and digested at 37°C, 400rpm overnight with EcoRI (NEB, 1500 units). 100 units of T4 ligase (NEB) were used for ligation, which was carried out for 4 hours at RT with gentle shaking. Crosslinks were reversed at 65°C, 400 rpm overnight, incubated with 25 µl 10mg/ml Proteinase K at 56°C for 4 hours at 400 rpm and extracted with a single round of phenol:chloroform:isoamyl alcohol. DNA was EtOH precipitated and resuspended in 50-75 µl H_2O. Concentrations were determined using a fluorometer and all samples were diluted to ~50 ng/µl. Sample purity was assessed via qPCR SYBR Green quantification using a 10-fold serial dilution of each template and amplifying with RP49 primers; samples showing >110% amplification efficiencies were re-purified with phenol:chloroform. Digestion efficiency calculations and data analysis/normalization was performed as described (Hagege et al., 2007). Two minimally overlapping BAC clones used for normalization were obtained from Children's Hospital Oakland Research Institute (CHORI, BACR48A11 & BACR28012). A Student's paired t-test was used to assess statistical significance based on 3 biological replicates per treatment.

RNA Extraction, CDNA Synthesis And QPCR

Oregon-R, *yw, dMEKK1^{UR36}* and *JNK-RNAi +/- UAS-Dcr2; eng-Gal4* wing discs (~6 pairs) were dissected in triplicate in SFX media and RNA extracted using 300 µl Trizol (Invitrogen). Entire *p38b$^{\Delta 45}$* and *p38b$^{\Delta 25}$; p38a^{del}* 3rd instar larvae (4-6) were homogenized and similarly extracted with 300 µl Trizol. Samples were treated with Turbo-free DNase (Ambion) and 500 ng of RNA was used for cDNA synthesis using the iScript cDNA synthesis kit with oligo dT primers (BioRad). qPCR runs were performed on a BioRad iQ5 cycler using iQ SYBR Green Supermix (BioRad) using 1 µl of cDNA. 10 µl of each representative genotype was resolved on a 1.5% agarose gel and imaged using an UVP EpiChemi^3 Gel Documentation System (Upland, CA). To measure JNK-RNAi knockdown, fold enrichment was calculated by comparing gene specific C_t values to Rp49 C_t values following the $\Delta\Delta C_t$ method. A paired Student's t-test was used to assess statistical significance based on 3 biological replicates per treatment.

Western Blotting

10^7 S2 cells were lysed on ice in 100 µl RIPA buffer supplemented with 1x Protease Inhibitor (Roche) and the SUMO isopeptidase inhibitors N-ethylmaleimide (80mM, Sigma) and iodoacetamide (.2mM, Acros Organics). For stress/recovery experiments, S2 cells in media were rotated in 1.5 ml tubes at room temperature with or without 250 mM NaCl in SFX media for 20 minutes. For recovery, stressed cells were pelleted at 2500 g for 2 minutes and gently resuspended in fresh SFX media for the indicated amount of time before being lysed. 12 µg of lysate was resolved in a 7.5% acrylamide gel, wet transferred at 4°C overnight (10V) and probed with α-CP190 (1:1500), α-Su(Hw) (1:1000) or α-Mod(mdg4)67.2 (1:1000).

CHAPTER 3

The Drosophila *Gypsy* Insulator Can Mediate Transvection In The Presence Of The *Vestigial* Enhancer

This chapter has been submitted for publication:

Schoborg, T.A., S. Kuruganti, R. Rickels, and M. Labrador. 2013. The Drosophila *gypsy* Insulator Can Mediate Transvection in the Presence of the *vestigial* Enhancer. **Submitted.**

S. Kuruganti and myself will be considered joint first-authors. My contributions included: (1) performing experiments, (2) data collection and analysis, (3) writing the manuscript and making figures and to a lesser extent (4) devising experiments. Srilalitha Kuruganti's contribution includes (1), (2) and (4) while Ryan Rickels assisted with (1) & (2). Mariano Labrador contributed primarily to (4) and to a lesser extent (3).

Abstract

Though operationally defined as *cis*-regulatory elements, enhancers can also communicate with promoters on a separate homolog in *trans*, a mechanism that has been suggested to account for the ability of certain alleles of the same gene to complement one another in a process otherwise known as transvection. This homolog-pairing dependent process is facilitated in Drosophila by chromatin-associated pairing proteins, many of which remain unknown and their mechanism of action uncharacterized. Here we have tested the role of the *gypsy* chromatin insulator in facilitating pairing and communication between enhancers and promoters in *trans* using a transgenic *eGFP* reporter system engineered to allow for targeted deletions in the *vestigial* Boundary Enhancer (*vg*BE) and the *hsp70* minimal promoter, along with one or two flanking *gypsy* elements. We found a modest 2.5-3x increase in *eGFP* reporter levels from homozygotes carrying an intact copy of the reporter on each homolog compared to unpaired hemizygotes,

although this behavior was independent of *gypsy.* However, detectable levels of GFP protein along the DV wing boundary in trans-heterozygotes lacking a single enhancer and promoter was only observed in the presence of two flanking *gypsy* elements. Our results demonstrate that *gypsy* can stimulate enhancer-promoter communication in *trans* throughout the genome in a context-dependent manner, likely through modulation of local chromatin dynamics once pairing has been established by other elements.

Introduction

Unlocking the mechanism by which gene regulatory elements (enhancers, promoters and other transcription factor binding sites) coordinate gene expression in a precise spatio-temporal manner is critical to understanding how eukaryotic genomes function in vivo. Chromatin structure plays a central role in this process, where nucleosome position and density, chromatin insulators, histone modifications and their associated proteins function to modulate the properties of these regulatory elements. Most of the effort in understanding this interplay has focused on their behavior in *cis*—that is, interactions occurring along the same chromosome. Indeed, enhancers are defined as *cis*-regulatory elements that function to stimulate gene expression when located either distal or proximal to their cognate promoters, and can function over large distances (Bulger and Groudine, 2010; Krivega and Dean, 2012; Smallwood and Ren, 2013; Szutorisz et al., 2005). This is achieved through physical association between the enhancer and the promoter, mediated by a number of regulatory proteins, general transcription factors, RNA Pol II and chromatin binding proteins that result in the formation of chromatin loop structures that are critical for transcriptional activation (Nolis et al., 2009; Ronshaugen and Levine, 2004; Su et al., 1990; Kagey et al., 2010; Sanyal et al., 2012).

However, the ability of enhancers to also act in *trans* (i.e., on a separate DNA molecule) on promoters has been observed both *in vitro* and *in vivo* (Bateman et al., 2012; Chen et al., 2002; Dunaway and Droge, 1989; Mueller-Storm et al., 1989; Ronshaugen and Levine, 2004; Wedel et al., 1990) where such behavior has been suggested to account for a number of homolog pairing-dependent phenotypes, such as in the phenomenon of transvection (see (Duncan, 2002; Kennison and Southworth, 2002; Muller and Schaffner, 1990; Tartof and Henikoff, 1991; Wu and Morris, 1999) for review).

Transvection was first coined by E.B. Lewis in 1954 to describe the ability of certain Drosophila alleles of *Ultrabithorax (Ubx)* to complement one another, leading to partial rescue of the mutant phenotype (Lewis, 1954). Importantly, this rescue was nullified if the locus on either homolog was relocated to a new position on the chromosome, suggesting that somatic pairing between homologous chromosomes is essential for transvection. This intragenic complementation has been reported almost exclusively in Drosophila, with transvection effects observed at a number of loci: *yellow (y)* (Geyer et al., 1990; Morris et al., 1999)*, decapentaplegic (dpp)* (Gelbart, 1982)*, eyes absent* (eya) (Zimmerman et al., 2000), *white (w)* (Babu and Bhat, 1980), *Gpdh* (Gibson et al., 1999)*, hedgehog (hh)* (Lee et al., 1992)*, wingless (wg)* (Buratovich et al., 1997), *engrailed (en)* (Condie and Brower, 1989)*, pointed (pnt)* (Scholz et al., 1993)*, cubitus interrruptus (ci)* (Locke and Tartof, 1994), *sex combs reduced (scr)* (Pattatucci and Kaufman, 1991)*, brown (bw)* (Dreesen et al., 1991; Henikoff and Dreesen, 1989) and *vestigial (vg)* (Williams et al., 1991). Recent studies have concluded that transvection is pervasive throughout

the entire Drosophila genome (Chen et al., 2002; Mellert and Truman, 2012), a finding not surprising given that homologs remain paired in somatic nuclei during interphase in flies (McKee, 2004). In other eukaryotes, homologs in somatic tissue do not remain in close synapse throughout interphase, yet quite a few cases of transvection and other pairing dependent phenomena have been reported in model systems other than Drosophila, including yeast (Aramayo and Metzenberg, 1996), plants (Matzke et al., 2001) and mammals (Liu et al., 2008; Rassoulzadegan et al., 2002; Sandhu et al., 2009), suggesting that eukaryotes possess evolutionarily conserved mechanisms that allow homologous chromosomes to communicate in *trans*.

Given the need for physical associations between enhancers and promoters to generate a sustained transcriptional output, homolog pairing in *trans* can facilitate contacts between a functional enhancer located on one homolog and a functional promoter located on the other, increasing the frequency of collisions between the two elements and thus the probability that a stable ternary complex is established. A number of proteins have been shown to be required for homolog pairing in Drosophila meiosis, including the multi-subunit cohesin complex (SMC1, SMC3, SCC1/RAD21, SCC3); however, it is not required for somatic paring (Joyce *et al.* 2012). In mammals, this complex has been shown to be required for stable *cis* looping contacts between enhancers and promoters, suggesting it may play a more direct role in gene regulation (Hadjur et al. 2009; Nativio et al. 2009; Dorsett 2011). Only a handful of Drosophila genes, mainly involved in mitotic functions, cell cycle control and chromatin organization, including *Topoisomerase II (Top2)* have been shown to promote somatic pairing (Joyce et al., 2012;

Williams et al., 2007), although other chromatin binding proteins, such as Zeste and members of the Polycomb Group Complex, have been shown to be required for transvection in a number of cases, including those involving communication between enhancers and promoters in *trans* (Babu and Bhat, 1980; Jack and Judd, 1979; Leiserson et al., 1994; Pattatucci and Kaufman, 1991; Pirrotta et al., 1985; Sipos et al., 1998; Wu et al., 1989).

In theory, however, any DNA element and its associated proteins that can mediate stable long-range contacts between distant genomic sites could potentially function to stabilize homolog pairing to facilitate enhancer-promoter interactions in *trans*. Chromatin insulators are well-suited for this task, given their ability to mediate long-distance contacts along the chromatin fiber in vivo. These DNA elements were first identified in Drosophila based on their ability to block enhancer-promoter communication and heterochromatin spreading along the chromatin fiber in transgenic assays. Such properties are conferred by insulator-binding proteins, seven of which have been characterized in Drosophila, including Su(Hw), CP190, BEAF-32, Mod(mdg4)67.2, dCTCF, GAF and Zw5, with mammals containing only a CTCF ortholog highly divergent in amino acid similarity to its Drosophila counterpart (Schoborg and Labrador, 2010). The *gypsy* insulator located within the 5' LTR of the *gypsy* retrotransposon is perhaps the most well characterized insulator, whose enhancer and heterochromatin blocking properties are conferred by Su(Hw), CP190 and Mod(mdg4)67.2. However, in addition to *gypsy* there are thousands of endogenous insulator sites located throughout the genome, where combinatorial binding of insulator proteins to each of these sites suggests a complex landscape whose functional consequences remain poorly understood (Bushey et al., 2009; Cuddapah et al., 2009; Negre et

al., 2010). Recent work has shown that these elements help mediate long-range contacts between enhancers, promoters and other insulator sites in order to direct transcriptional outputs, maintain regions of histone modifications, and establish gene regulatory and physical domains (Handoko et al., 2011; Hou et al., 2012; Krivega and Dean, 2012; Sexton et al., 2012; Yang and Corces, 2012).

Though most of these interactions are thought to occur in *cis,* potential interactions in *trans* are not out of the question, even for long range contacts between non-homologous sites*,* analogous to the behavior observed for olfactory receptor choice by the *H* enhancer and various *olfactory* gene promoters located on different chromosomes in mice (Lomvardas et al., 2006). Interestingly, *gypsy* insulators have been previously implicated in transvection, both directly and indirectly, at the *yellow* locus (Georgiev and Corces, 1995; Geyer et al., 1990). Furthermore, reduction of Su(Hw) has been shown to reduce somatic pairing by ~30% in embryos (Fritsch et al., 2006), suggesting that insulators might contribute to pairing dependent enhancer-promoter communication in *trans.*

Here we have used a reporter construct designed to elucidate the role of the *gypsy* insulator in transvection. We engineered our system with the *vestigial (vg)* boundary enhancer (*vg*BE) and a minimal *hsp*70 promoter to drive *eGFP* expression, flanked by a *gypsy* insulator upstream and downstream (2-insulator), upstream only (1-insulator) or no *gypsy* insulator (0-insulator). Using the Cre/*loxP* and Flp/*FRT* system to delete the promoter or enhancer, respectively, "promoterless" and "enhancerless" flies were created and crossed to measure transvection

effects. Quantitative fluorescent microscopy and qPCR of wing discs from 3rd instar larvae reveal a pairing dependence in the non-deletion constructs that is independent of *gypsy*, whereas transvection was only visually observed along the DV wing boundary in lines containing two flanking insulators. Interestingly, the *vg*BE alone can drive expression of a large pool of e*GFP* transcripts in *cis* in the absence of a functional promoter, the majority of which are not translated into protein. Taken collectively, our results demonstrate that the *gypsy* insulator can contribute to transvection in a dose-dependent manner, likely through modulation of local chromatin dynamics once other chromatin elements have established homolog pairing.

Results

A P-element EGFP Reporter System Engineered To Induce Artificial Mutations In Enhancers & Promoters Shows Insertion Bias To Endogenous Insulators

The somatic pairing dependence required for transvection ultimately derives from the ability of enhancers to act on promoters in close proximity in *trans* when preferential interactions with promoters in *cis* are lost. This feature is supported by the fact that a majority of classical complementing mutant alleles disrupt either the enhancer or promoter (Duncan, 2002). Our P-element reporter constructs were designed with site-specific recombination sites flanking both these elements in order to selectively delete either the enhancer or promoter. We chose the *vestigial* boundary enhancer (*vg*BE), which drives expression in a small stripe of cells at the dorsal/ventral boundary in developing wing and haltere imaginal discs (Williams et al., 1994) and flanked it with *FRT* sites. The minimal hsp70 promoter was flanked by *loxP* sites, followed by the *eGFP* reporter. Two other derivatives of this construct were created by adding either a single *gypsy* insulator upstream of the *vg*BE or two *gypsy* insulators flanking the entire construct (Figure 3.1A). To test the *gypsy* insulator's contribution to transvection, flies lacking either the

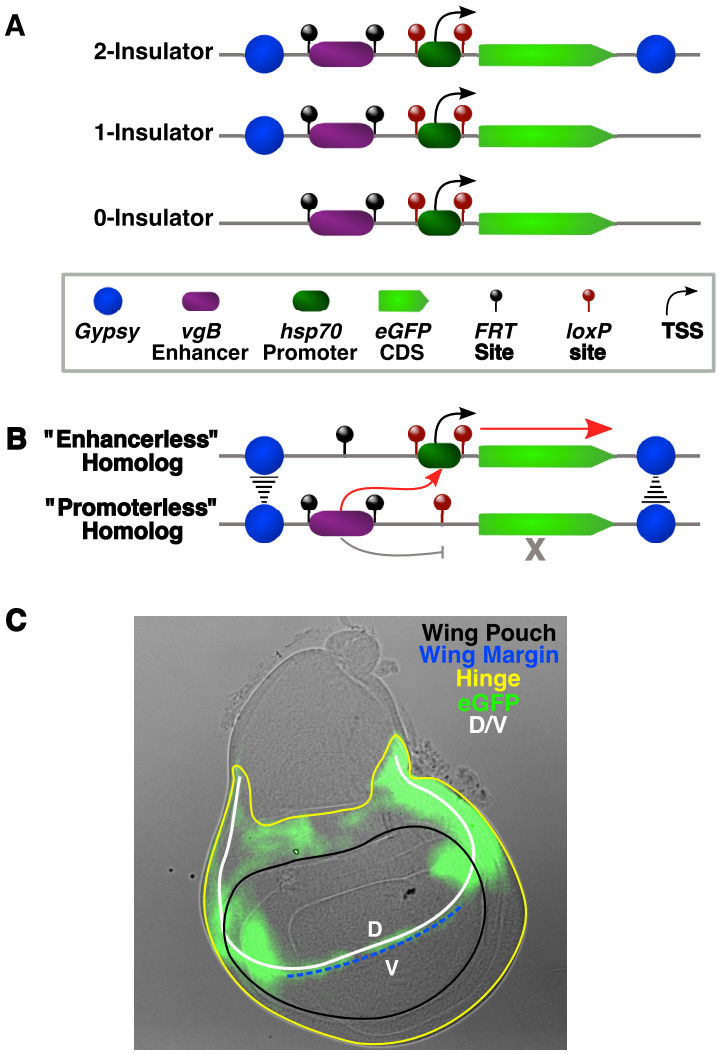

Figure 3.1. Schematic Of The Reporter System.
All constructs contained the vgBE enhancer flanked by FRT sites, a minimal hsp70 promoter flanked by loxP sites, an *eGFP* coding sequence, and either zero, one or two gypsy insulators (A). Communication in trans between a functional enhancer ("promoterless") on one homolog and a functional promoter ("enhancerless") on the other homolog might be facilitated by trans interactions between flanking gypsy insulators, leading to expression only from the "enhancerless" homolog (B). Differential Interference Contrast (DIC)-GFP overlay of a 3rd instar wing disc showing the expression pattern of the vgBE along the dorsal-ventral (D/V) boundary, particularly within the hinge and wing margin (C).

*vg*BE or the *hsp70* promoter were crossed to generate trans-heterozygous progeny for each construct. We hypothesized that loss of communication between the *vg*BE and the *hsp70* promoter in *cis* due to promoter removal could be restored in *trans* by stable interactions between *gypsy* components, inducing eGFP expression from the "enhancerless" homolog along the D/V boundary of the wing disc (Figure 3.1B).

Following microinjection, we recovered 11 independent insertions for the 2-insulator construct, 18 insertions for the 1-insulator construct and 42 insertions for the 0-insulator construct. All eleven 2-insulator lines displayed strong eGFP expression in the D/V compartment boundary of the wing discs, particularly along the wing margin within the wing pouch and the hinge region (Figure 3.1C), whereas only nine 1-insulator lines and seven 0-insulator lines displayed detectable eGFP levels, likely the result of position effects (Markstein et al., 2008). Mapping of these lines revealed a noticeable bias towards the 5' end of genes, transposable elements and other transposon "hotspots", similar to observations in previous reports (Bellen et al., 2004; Oh et al., 2003; Spradling et al., 1999). Although no correlation with chromatin states/domains were observed (Filion et al., 2010), nearly all insertions outside of transposons were within 1 Kb of an endogenous protein-bound insulator, particularly those bound by GAF and/or CP190 regardless of whether a *gypsy* insulator was present in the construct (Table A1-A3 In Appendix).

***Homolog Pairing Increases Reporter Levels Independently Of* Gypsy**

We first examined reporter levels in homozygous and hemizygous flies carrying an intact (i.e., non-deleted enhancer/promoter) construct. Immunostaining of wing discs revealed a significant decrease in the amount of GFP protein in hemizygous larvae carrying the construct on only one

homolog compared to homozygous larvae carrying a copy on each homolog, a pattern independent of gypsy presence (Figure 3.2A-3.2C). In terms of dosage, the expression level of GFP in homozygotes would be expected to be twice the amount observed in hemizygotes, particularly if levels were influenced solely by enhancer-promoter communication in *cis*. Using qPCR, we measured the levels of *GFP* expression and found that transcript levels were reduced 2.5-3X in hemizygotes compared to homozygotes, suggesting that pairing in *trans* can stimulate increased transcription (Figure 3.2A-3.2C). However, this behavior was not significantly influenced by the *gypsy* insulator, as the 2.5x, 3x, and 2.8x reduction observed for the 0-, 1- and 2-insulator hemizygotes, respectively, rules out any synergistic effect that would be expected if this element contributed to pairing.

The VgBE Can Drive Reporter Expression In The Absence Of A Functional Promoter

Next, we tested whether *gypsy* might influence enhancer-promoter communication in *trans* by removing either the *vg*BE or the *hsp70* minimal promoter to generate "enhancerless" and "promoterless" lines (Figure A7A) and combining them to create trans-heterozygous individuals. However, qPCR revealed a large amount of *GFP* transcript present in promoterless lines, with levels in the promoterless homozygote (P^-/P^-) equal or greater to the levels observed in the non-deleted hemizygote ($P^+E^+/+$). This pattern of expression was consistent in every single line examined, independent of *gypsy* insulator presence (Figure 3.3A-3.3C), suggesting that the *vg*BE can drive expression in the absence of a functional promoter. We suspected that perhaps *vg*BE's proximity to the coding region of *GFP* (~183 bp after *hsp70* promoter deletion) might explain this result, as any RNA Pol II recruited to the *vg*BE and able to find a suitable transcription start site (TSS) might be able to generate a transcript. We found two TSSs in potential promoters

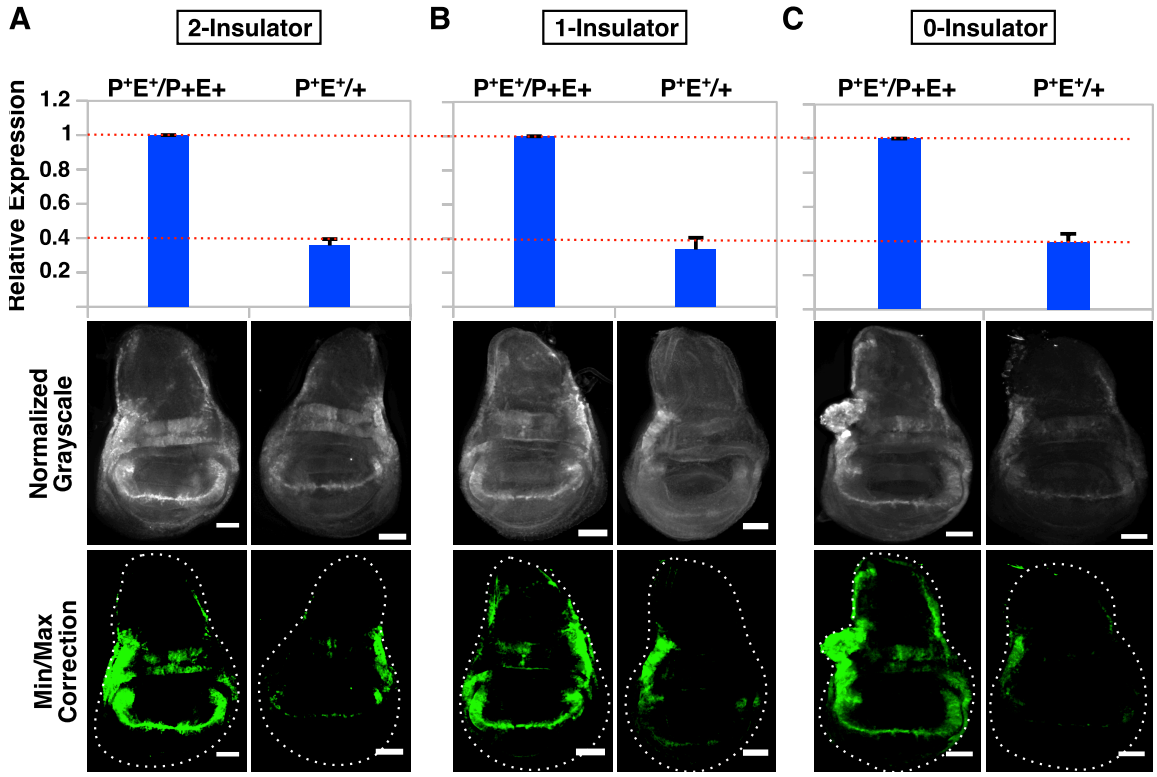

Figure 3.2. Pairing Dependent Influences On Reporter Expression
QPCR analysis (top graph) and immunostaining (bottom panels) of wing discs from intact homozygotes (P^+E^+/P^+E^+) and intact hemizygotes ($P^+E^+/+$) for 2-insulator (A), 1-insulator (B) and 0-insulator (C) lines. For microscopy, images were normalized to P^+E^+/P^+E^+ for each respective line and minimum/maximum level corrections were applied equally to both genotypes using ImageJ and false-colored green. Error bars represent standard error of the mean (S.E.M) and scalebars are 50 μm.

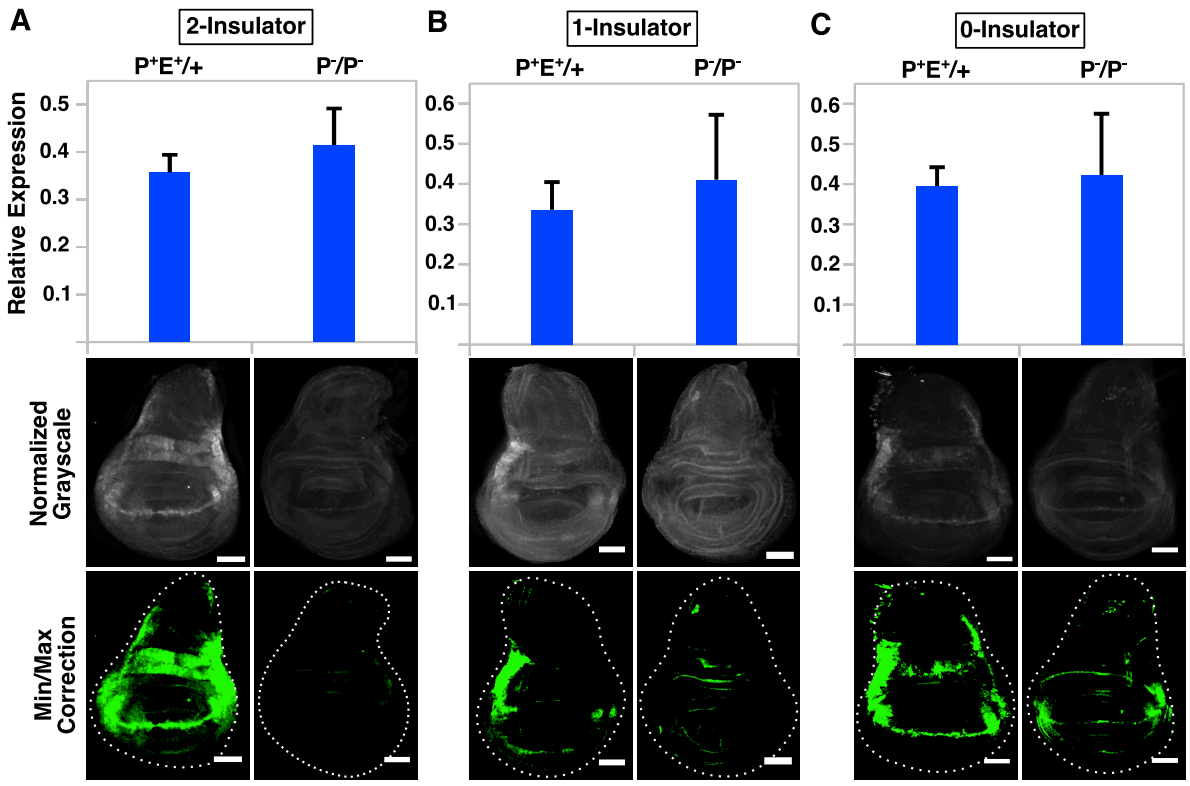

Figure 3.3. The *vg*BE Can Function As a Promoter.
QPCR analysis (top graph) and immunostaining (bottom panels) of wing discs from intact hemizygotes ($P^+E^+/+$) and "promoterless" homozygotes (P^-/P^-) for 2-insulator (A), 1-insulator (B) and 0-insulator (C) lines. For microscopy, images were normalized to $P^+E^+/+$ for each respective line and minimum/maximum level corrections were applied equally to both genotypes using ImageJ and false-colored green. Error bars represent standard error of the mean (S.E.M) and scalebars are 50 μm.

within the *vg*BE using a neural network prediction algorithm (minimum promoter score=0.8) (Reese, 2001); however, using qPCR primers designed to measure mRNAs arising from these TSSs only, we could not detect a sufficient amount of transcripts to fully account for the total pool of *GFP* (Figure A7B). It is worth noting that our *GFP* qPCR primers are located at the 3' end of the transcript (within 130 bp of the stop codon), which would fail to distinguish whether other cryptic TSSs located downstream of our *vg*BE qPCR primers might be utilized and therefore contribute to the pool of transcripts as well.

Interestingly, *GFP* transcript levels as measured by qPCR did not correlate with the amount of GFP protein. Analysis of wing discs stained with α-GFP revealed only a small amount of signal, barely above background levels, or no signal at all along the DV boundary within the wing pouch and hinge in promoterless homozygotes (P^{-}/P^{-}). Non-deleted hemizygotes ($P^{+}E^{+}/+$), on the other hand, displayed strong signal along the boundary (Figure 3.3A-3.3C). This pattern was consistent in all lines examined and independent of *gypsy* insulator presence. Background levels of expression in other parts of the wing disc were similar between the two genotypes, ruling out the possibility that misexpression by other regulatory elements might contribute to the large amount of *GFP* transcript in promoterless individuals. Although we have not measured mRNA or protein stability directly in these lines, it should be noted that cDNA generated using two different priming methods (a mix of random hexamers and oligo dT primers or oligo dT primers only) gave identical results (Figure A7C), suggesting that our results are not due to primer bias during cDNA synthesis and that the majority of transcripts being measured were polyadenylated. Taken collectively, our results suggests that although the *vg*BE can drive reporter expression in

the absence of a functional promoter, many of these transcripts do not give rise to functional protein.

*Two G*ypsy *Insulators Can Facilitate Enhancer-Promoter Communication In* Trans

We next determined whether *gypsy* could promote enhancer-promoter communication in *trans* by staining wing discs with α-GFP from larvae containing a single functional enhancer on one homolog and a single functional promoter on the other. GFP signal was barely above background or undetectable along the wing margin and hinge region in 0-insulator and 1-insulator trans-heterozygotes (P^-/E^-) (Figure 3.4 and Figure 3.5). However, moderate levels of GFP were observed in 2-insulator trans-heterozygotes, with most of the expression concentrated in the hinge region as staining along the wing margin was considerably weaker and variegated (Figure A8). No signal was observed for either the promoterless ($P^-/+$) or enhancerless ($E^-/+$) hemizygote, and although a small amount of protein could be detected in the promoterless homozygote (P^-/P^-), the GFP signal was much stronger in the trans-heterozygote (P^-/E^-) (Figure 3.6). Taken collectively, these data suggest that the presence of two flanking *gypsy* insulators is sufficient to facilitate enhancer-promoter communication in *trans*.

Strangely, however, our qPCR results did not agree with our image analysis—we were unable to detect the additional transcripts that should have been present in the 2-insulator trans-heterozygote. Instead, the 2-insulator, 1-insulator and 0-insulator lines all displayed the same behavior: the trans-heterozygote expression level was always the sum of the promoterless ($P^-/+$) and enhancerless ($E^-/+$) hemizygote expression levels, a finding that would be expected if each homolog were acting independently of one another (i.e., in cis) and transvection was absent.

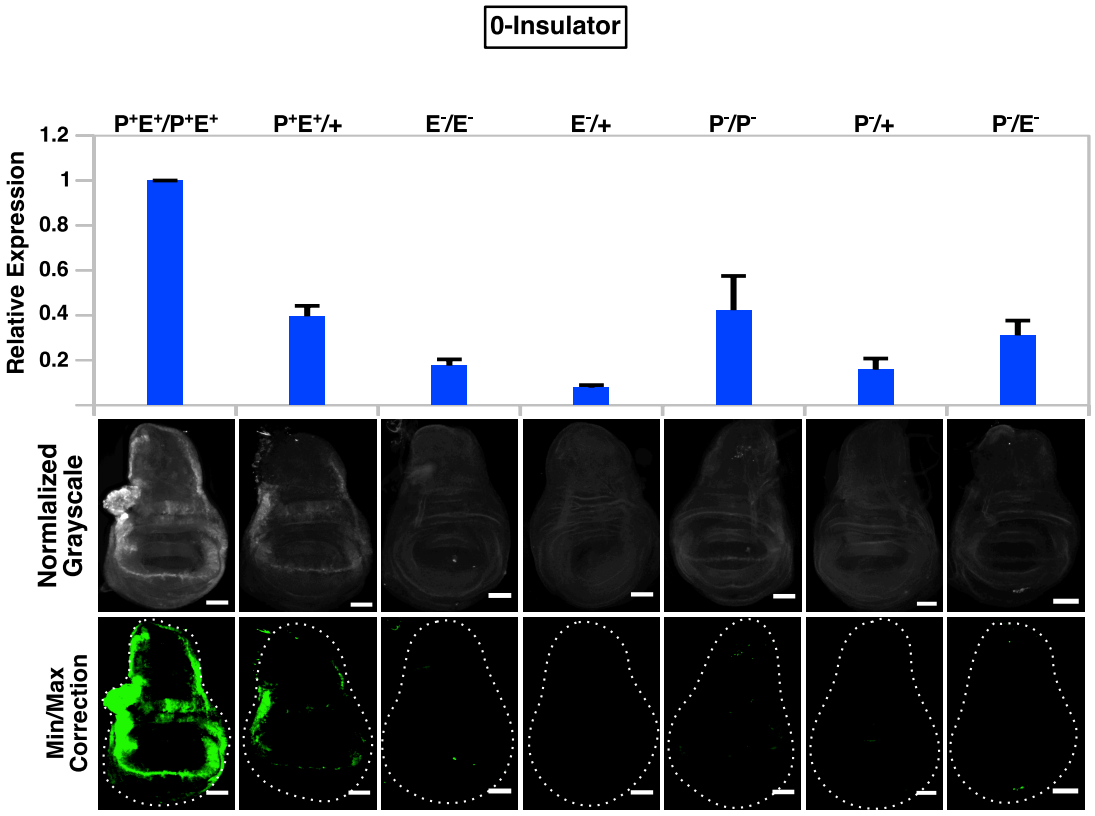

Figure 3.4. Absence Of *gypsy* Does Not Promote Transvection.
QPCR analysis (top graph) and immunostaining (bottom panels) of wing discs from all seven 0-insulator genotypes. For microscopy, all images were normalized to P^+E^+/P^+E^+ and minimum/maximum level corrections were applied equally to all genotypes using ImageJ and false-colored green. Genotype Nomenclature: P^+E^+/P^+E^+, the intact construct containing enhancer and promoter homozygote; $P^+E^+/+$, the intact construct containing enhancer and promoter hemizygote; E^-/E^-, deleted *vg*BE "enhancerless" homozygote; $E^-/+$, deleted *vg*BE "enhancerless" hemizygote; P^-/P^-, deleted "promoterless" homozygote; $P^-/+$, deleted "promoterless" hemizygote; P^-/E^-, Trans-heterozygote (Transvection). Error bars represent standard error of the mean (S.E.M) and scalebars are 50 µm.

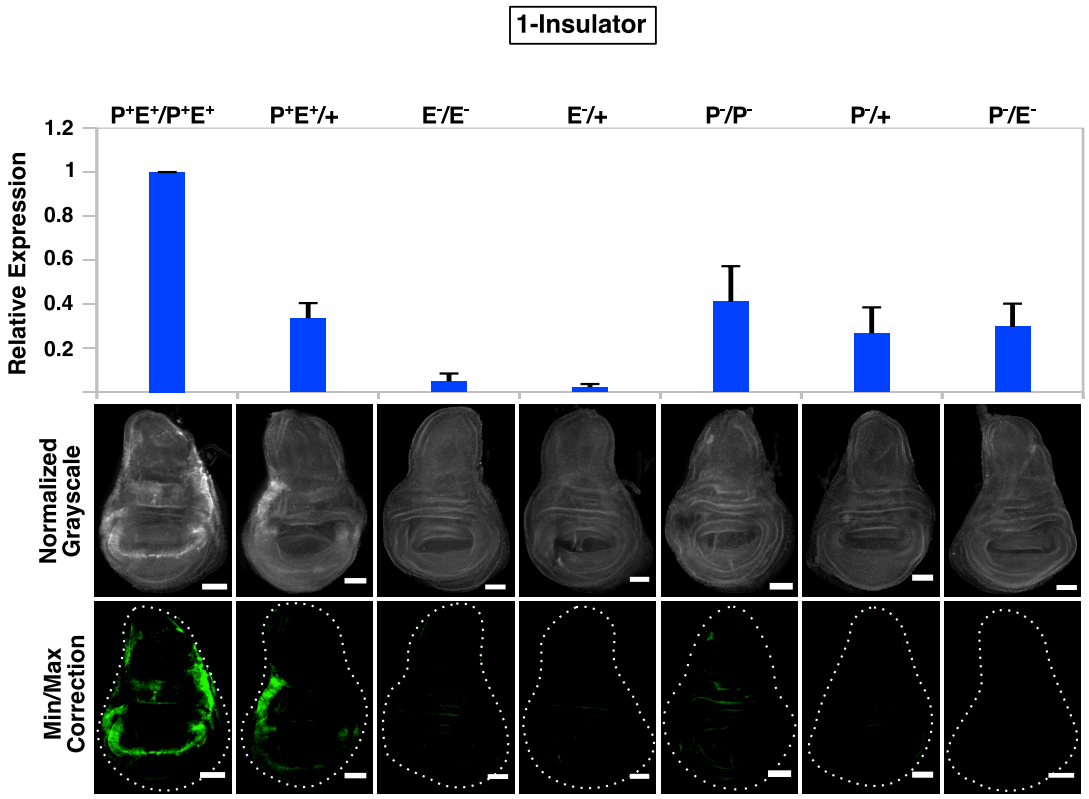

Figure 3.5. A Single *gypsy* Does Not Promote Transvection.
QPCR analysis (top graph) and immunostaining (bottom panels) of wing discs from all seven 1-insulator genotypes. For microscopy, all images were normalized to P^+E^+/P^+E^+ and minimum/maximum level corrections were applied equally to all genotypes using ImageJ and false-colored green. Genotype Nomenclature: P^+E^+/P^+E^+, the intact construct containing enhancer and promoter homozygote; $P^+E^+/+$, the intact construct containing enhancer and promoter hemizygote; E^-/E^-, deleted *vg*BE "enhancerless" homozygote; $E^-/+$, deleted *vg*BE "enhancerless" hemizygote; P^-/P^-, deleted "promoterless" homozygote; $P^-/+$, deleted "promoterless" hemizygote; P^-/E^-, Trans-heterozygote (Transvection). Error bars represent standard error of the mean (S.E.M) and scalebars are 50 µm.

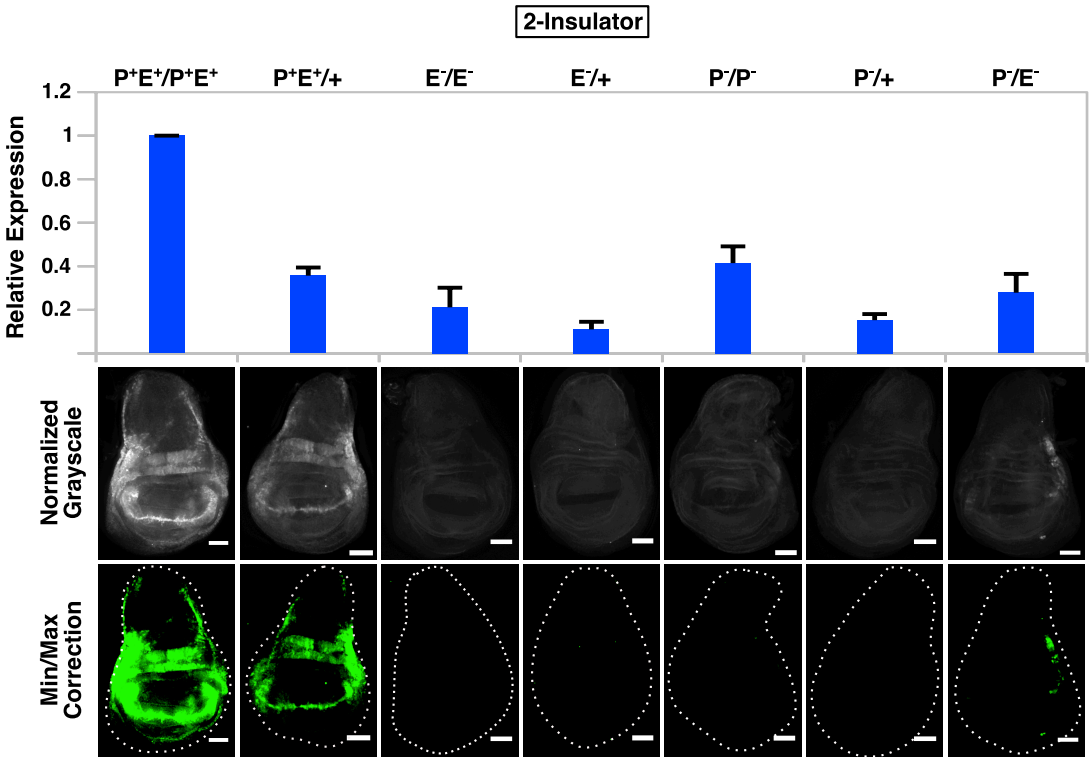

Figure 3.6. Flanking *gypsy* Insulators Promote Transvection.
QPCR analysis (top graph) and immunostaining (bottom panels) of wing discs from all seven 2-insulator genotypes. For microscopy, all images were normalized to $P^{+}E^{+}/P^{+}E^{+}$ and minimum/maximum level corrections were applied equally to all genotypes using ImageJ and false-colored green. Genotype Nomenclature: $P^{+}E^{+}/P^{+}E^{+}$, the intact construct containing enhancer and promoter homozygote; $P^{+}E^{+}/+$, the intact construct containing enhancer and promoter hemizygote; E^{-}/E^{-}, deleted *vg*BE "enhancerless" homozygote; $E^{-}/+$, deleted *vg*BE "enhancerless" hemizygote; P^{-}/P^{-}, deleted "promoterless" homozygote; $P^{-}/+$, deleted "promoterless" hemizygote; P^{-}/E^{-}, Trans-heterozygote (Transvection). Error bars represent standard error of the mean (S.E.M) and scalebars are 50 μm.

However, given the caveats of qPCR analysis of our system described earlier, we believe that our immunostaining and subsequent image analysis is the most accurate reflection of whether trans-interactions are occurring or not and therefore conclude that two *gypsy* insulators are sufficient to mediate transvection.

Mutations In Su(Hw) Reduce Reporter Levels In A Pairing-Independent Manner

Finally, to confirm that the *gypsy* insulator plays a direct role in mediating transvection, we combined a subset of our 2-insulator lines into a genetically null *su(Hw)^{e04061}* mutant background and measured GFP reporter levels by both immunostaining and qPCR. Su(Hw) was the first protein identified shown to be critical for the insulator properties of *gypsy* and has been shown to be required for somatic pairing in embryos (Fritsch et al., 2006; Modolell et al., 1983; Parkhurst and Corces, 1986; Parkhurst et al., 1988), suggesting that Su(Hw) and other insulator proteins distributed throughout the genome might contribute to transvection in a global manner. If the reduction observed in mutants is attributable to pairing influences, then we would expect that only homozygous individuals carrying the reporter on each homolog would be affected, as hemizygous individuals lacking a suitable pairing region on the opposite homolog should not be affected by loss of such pairing. Immunostaining and qPCR analysis of mutant wing discs revealed a significant decrease in reporter expression for both homozygotes and hemizygotes (Figure A9A). Since transcription does not appear to be globally perturbed in a *su(Hw)^{e04061}* background (Figure A9B), it is likely that the reduced reporter expression we observe is due to position effects resulting from the failure of the flanking *gypsy* insulators to prevent repressive chromatin from spreading into the reporter locus in *cis* due to loss of Su(Hw) (Markstein et al., 2008), rather than a reduction or loss of homolog pairing.

Discussion

In this work we have utilized a transgenic reporter engineered to induce artificial mutations in the *vg*BE enhancer and *hsp70* promoter in order to test the role of the *gypsy* insulator in the phenomenon of transvection. We find that although pairing does appear to modestly increase the amount of *eGFP* transcript arising from intact constructs containing functional regulatory elements in *cis* on each homolog, this behavior is not dependent on *gypsy.* However, flanking *gypsy* insulators are required for transvection when enhancerless and promoterless homologs were combined in *trans*. Interestingly, we also found that the *vg*BE can drive transcription of the reporter in the absence of a functional promoter. From these findings, we conclude that *gypsy* can mediate interactions in *trans* in a dose-dependent manner*,* although we cannot rule out that other insulator binding sites may contribute to pairing in cooperation with other factors at specific genomic locations, depending on the local chromatin landscape.

Recent reports have utilized a reporter scheme analogous to ours, whose main advantage lies in the fact that transvection can be tested in a tightly controlled manner using fluorophores that allow for single-cell analysis of enhancer-promoter communication in *trans,* as opposed to phenotypic analysis of whole animals (Bateman et al., 2012; Mellert and Truman, 2012). However, unlike those reports, which took advantage of the φC31 integration system to test only a handful of characterized integration sites in the genome (Groth et al., 2004), we utilized a P-element transformation method to integrate our reporters into different regions of the genome, allowing for a more global analysis of transvection effects. Despite a definite homing bias, particularly to other repetitive sequences, transposon hotspots and the 5' end of genes in

agreement with other reports (Bellen et al., 2004; Oh et al., 2003; Spradling et al., 1999), a common theme among all insertion sites was the presence of insulator binding sites, particularly CP190 and GAF. As insulators are primarily nucleosome-free (Negre et al., 2010), it is possible that these elements maintain a chromatin architecture conducive for insertion and hence may be responsible for the homing bias observed for P-elements.

The most important point to address involves the apparent discrepancy between our qPCR results and immunostaining. Our image analysis strongly supports the conclusion that the 2-insulator construct can support transvection, as moderate amounts of GFP along the DV boundary, primarily in the hinge region, was readily observable at levels much higher than either promoterless ($P^-/+$) or enhancerless ($E^-/+$) hemizygote, which displayed no signal whatsoever. Perhaps most tellingly, the amount of signal in the promoterless homozygotes (P^-/P^-) was much lower than the trans-heterozygote, suggesting that pairing a functional enhancer with a functional promoter in *trans* can positively stimulate transcription above background levels if aided by two *gypsy* insulators. However, using qPCR we were unable to detect the expected increase in transcripts in the 2-insulator trans-heterozygous lines as compared to either the 1- or 0-insulator lines. We have shown that transcript levels, as measured by qPCR, do not accurately reflect the amount of protein present in the promoterless genotypes—in all lines, regardless of insulator presence, transcript levels were consistently and reproducibly equal to or higher in all promoterless homozygotes (P^-/P^-) as compared to non-deleted hemizygotes ($P^+E^+/+$), despite weak levels of GFP protein staining for P^-/P^- and high levels for $P^+E^+/+$. As protein levels are the ultimate determinant of an organism's phenotype, we argue that our immunostaining and image

analysis are the most appropriate metric by which to evaluate *gypsy* insulator function in transvection. With this in mind, nearly all examples of transvection have relied on phenotypic descriptions rather than quantitative analysis of transcript or protein levels as a result of the non-linear relationship between gene expression and phenotype (Raser and O'Shea, 2005). Therefore, an alternative way to address *gypsy's* role in transvection would be to re-engineer our construct to contain the native *vestigial* (*vg*) promoter and gene region and assess phenotypically whether full or partial rescue of the *vg* mutant phenotype is observed.

Nonetheless, the question still remains: where are all the extra transcripts in the promoterless lines coming from and why are they not translated into functional protein? It is important to reiterate that we were measuring polyadenylated transcripts that would be considered mature and stable and that background levels in other cells of the wing discs and even other tissues appear to be equal in all genotypes. The likeliest explanation is that the *vg*BE itself functions as a promoter to facilitate transcription of 5'-truncated messages that are not translated, or translated into non-functional protein. However, the mechanism of how it might do so remains speculative. Although RNA Pol II has been shown to be recruited directly to enhancers (De Santa et al., 2010; Kim et al., 2010; Wang et al., 2011) and intergenic transcripts between enhancers and promoters have been detected both *in vivo* and *in vitro* (Tchurikov et al., 2009; Zhu et al., 2007), we could not detect enough enhancer-originating transcripts to account for the total pool of *GFP* transcripts. We note that another TSS, whose utilization would not have been detected using our enhancer-specific qPCR primers, is predicted near the 5' end of *GFP.* This would still be detectable with our *GFP*-specific qPCR primers and there is a downstream ATG triplet that could

function as a start codon for this truncated transcript; however, this would not be translated in the correct reading frame, hence giving rise to a non-functional protein. Nonetheless, 5' RACE will be critical for determining all of the potential transcripts with different 5' ends that contribute to the total *GFP* transcript pool in promoterless individuals.

With this possibility in mind, having multiple types of *GFP* transcripts that differ in their 5' ends that constitute the total pool might explain why we do not observe an increase in transcript levels in the 2-insulator trans-heterozygote as compared to the 1- and 0-insulator trans-heterozygotes. Rather than observing an increase in the relative number of total transcripts, perhaps it is simply a shift in the relative ratio of the different types of messages. For simplicity, one could imagine two types of *GFP* transcripts that differ in their 5' ends: one that gives rise to a functional GFP protein and the other that gives rise to a non-functional protein. In this case, the functional transcript would only arise from the homolog containing the intact promoter (which would only be possible in *trans*) whereas the non-functional transcript would arise from the promoterless homolog in *cis*. For 1- and 0-insulator trans-heterozygotes, only 10% of the total pool might consist of functional transcript, with the overwhelming majority (90%) consisting of non-functional transcript driven in *cis* due to the failure of the *vg*BE to stably communicate with the functional promoter in *trans*. However, in 2-insulator trans-heterozygous individuals, this proportion would be reversed, with 90% of the transcripts being functional, facilitated by communication in *trans* between the *vg*BE and functional promoter as a result of the flanking *gypsy* insulators. Note that this argument is only valid if we assume that the total output of the *vg*BE is equal regardless of whether *cis* or *trans* interactions predominate and only if it is acting

in either conformation at a given point in time, not both simultaneously. This idea is supported by the fact that although regulatory elements prefer to act in *cis* (Casares et al., 1997; Geyer et al., 1990; Gohl et al., 2008; Martinez-Laborda et al., 1992; Mellert and Truman, 2012; Morris et al., 1999), competition between promoters for a single enhancer ultimately dictates whether the enhancer functions primarily in *cis* or in *trans* within the same cell (Bateman et al., 2012). Nonetheless, more work will be needed to address this intriguing possibility.

Finally, why do two flanking *gypsy* insulators, but not a single upstream *gypsy* insulator, support transvection? One might assume that if *gypsy* contributes to homolog pairing, then a single insulator located just upstream of the enhancer and promoter would still be more than capable of ensuring that those two elements remain in close proximity in *trans.* However, perhaps the most critical determinant is simply chromatin structure itself—it is widely accepted as a key regulator of transcription in *cis,* so the same principles would also apply in *trans.* Even if pairing were to bring enhancers and promoters in close proximity, the underlying chromatin must still be permissible in order for transcription to occur. Insulators were originally identified based on their ability to buffer the effects of surrounding chromatin influences (i.e., position effects) on transgene expression (Kellum and Schedl, 1991; Udvardy et al., 1985), and regardless of the mechanism by which insulators accomplish this task (chromatin looping, nucleosome position/density modulation, etc.) it is likely that the transvection we observe is simply due to the flanking insulators establishing a permissive chromatin environment favorable for transcription. The single *gypsy,* on the other hand would not be able to establish the same environment and therefore even if pairing were established, transcription would still be unlikely to occur given the

lack of a suitable chromatin landscape. Our *su(Hw)* mutant data supports this hypothesis, as significant reductions in GFP expression was observed in both homozygotes and hemizygotes, highlighting the importance of chromatin structure on transgene expression regardless of pairing influences. Our findings, along with a number of other studies linking chromatin proteins to transvection (Babu and Bhat, 1980; Jack and Judd, 1979; Leiserson et al., 1994; Pattatucci and Kaufman, 1991; Pirrotta et al., 1985; Sipos et al., 1998; Wu et al., 1989) and the failure of other studies to observe transvection except when their reporters were located in defined PhiC31 genomic sites that are highly permissible to transcription (Markstein et al., 2008; Mellert and Truman, 2012), suggests that chromatin itself is the master regulator of this phenomenon.

Materials And Methods

Fly Stocks & Husbandry

Flies were cultured on standard cornmeal-agar media and maintained at 25ºC. Flippase (FLP) (y^1w^{1118} *P{ry[+t7.2]=hsFLP}1;* Dr^{Mio}/TM3*,* ry^*Sb^1), Cre recombinase (y^1w^{67c23}; noc^{Sco}/CyO*,* *P{*w^{+mC}*=Crew}DH1*) and *su(Hw)* mutant flies (w^{1118}*; PBac{RB}su(Hw)*^{e04061}/TM6B*,* Tb^1) were obtained from the Bloomington Stock Center. Microinjection to generate 2-insulator (*yw; P{Ins-vgBE-eGFP-Ins,* w^{+mC}*}*), 1-insulator (*yw; P{Ins-vgBE-eGFP,* w^{+mC}*}*) and 0-insulator (*yw; P{vgBE-eGFP,* w^{+mC}*}*) transgenics was performed by Genetivision (Houston, TX). The *w;* noc^{Sco}*/*CyO*;* MKRS, Sb^1/TM6B*,* Tb^1 double balancer line was a gift from Bruce McKee.

Construct Design And Plasmid Generation

The reporter construct consisted of 5 core elements: the *vestigial* boundary enhancer (*vg*BE), *loxP* and *FRT* sites, *5xUAS/hsp-70* minimal promoter, and an *eGFP* coding sequence. Additional

gypsy insulator sequences were present as required to generate either the 1- or 2-insulator construct. The pGREEN pelican plasmid, consisting of an eGFP reporter with a 5' MCS flanked by two *gypsy* insulators, served as the vector backbone for these constructs (Barolo et al., 2000). First, the *5xUAS/hsp-70* minimal promoter was amplified from the pUAST vector using specific primers designed to insert *NheI* and *XhoI* cut sites and a single *loxP* site oriented in the same direction on both sides of the promoter. The PCR product was digested with *Nhe*I and *Xho*I and cloned into the pGREEN pelican vector digested with *Nhe*I and *Xho*I to obtain the pGP-hsp70 plasmid. The wing and haltere disc *vestigial* Boundary Enhancer (*vg*BE) present in the second intron of the *vg* gene (Williams et al., 1994) was amplified from *yw* gDNA with primers engineered with *BbvCI* and *BamHI* sites, digested and cloned into pCR 2.1-λ- FRT that was available in the lab. This plasmid consists of λ DNA (~800bp) and *FRT* sequences in the same orientation flanking the multiple cloning site. The cloned *vg*BE enhancer and its flanking *FRT* sites was digested as a *Kpn*I-*Sac*II fragment and cloned into pGP-hsp70 to obtain the 2-insulator construct *Ins-VgBE-eGFP-Ins*. To obtain the 1-insulator construct (*Ins-VgBE-eGFP*), the 3' insulator downstream of eGFP was deleted from this vector by restriction digestion with *Spe*I and *Eco*47III. The sticky ends generated by *Spe*I digestion were end-filled using *Pfu* DNA polymerase and blunt ends were ligated. To generate the 0-insulator construct (*VgBE-eGFP*), both the insulators were deleted from the pGP-hsp70 plasmid, generating a *Kpn*I-λ-*vg*BE-eGFP-*Sac*II cassette that was reinserted into the insulator-less pGP-hsp70 plasmid. All constructs were microinjected into y^1w^{67c23} flies and individual lines established by w^+ selection.

Insertion Mapping & Inverse PCR

Individual lines were mapped using both classical and molecular genetics. A single male homozygous for the transgene (w^+) from each line was crossed to *w; noc*^{Sco}/CyO; MKRS, *Sb*^{1}/TM6B, *Tb*^{1} virgins. Progeny males carrying w^+, *CyO* and *TM6B* were then crossed to *yw* virgins and the resulting offspring scored to determine how w^+ segregated with respect to the dominantly-marked *CyO* and *TM6B* balancers. Inverse PCR to identify the precise insertion position in the genome was carried out as described (Berkley Drosophila Genome Project). Genomic DNA was extracted from adult flies with DNAzol (Invitrogen), ~1µg digested with *HinP*I, *Sau3A*I or *Msp*I in a 30 µl reaction volume for 3 hrs at 37ºC and ligated (T4 DNA ligase, NEB) overnight at 4ºC in a 400 µl reaction volume. DNA was EtOH precipitated and washed followed by resuspension in 10 µl H_2O and amplified with both the *Pry1/Pry4* and *PwhtI/Plac1* primer sets. Samples producing a single strong band with minimum background were PCR purified and sequenced with either the *Sp1* or *Spep1* primer. Sequences were mapped to the latest version of the Drosophila genome using the BLAST algorithm at FlyBase.

Inducing Artificial Mutations In The Enhancer And Promoter

To delete the promoter, transgenic males carrying an intact construct were crossed with virgin females carrying Cre recombinase. Individual progeny males (w^+) were then crossed to *yw* virgins and gDNA extracted from adult progeny using DNAzol (Invitrogen), followed by PCR to screen for promoter removal. A similar crossing scheme was used to remove the *vg*BE enhancer, with transgenic males carrying an intact construct crossed with virgins carrying *flippase* under control of the *hsp-70* promoter. Larvae were subjected to daily heat shock by submerging vials in

a 37ºC H_2O bath for 1 hr. Individual males (w^+) were then crossed to *yw* virgins and gDNA extracted from adult progeny to screen for enhancer removal by PCR.

Genotype Nomenclature

For each line, a total of seven genotypes were analyzed: (1) the intact construct containing enhancer and promoter homozygote (P^+E^+/P^+E^+); (2) the intact construct containing enhancer and promoter hemizygote ($P^+E^+/+$); (3) deleted "promoterless" homozygote (P^-/P^-); (4) deleted "promoterless" hemizygote ($P^-/+$); (5) deleted *vg*BE "enhancerless" homozygote (E^-/E^-); (6) deleted *vg*BE "enhancerless" hemizygote ($E^-/+$) and (7) Trans-heterozygote (transvection) (P^-/E^-), derived from crossing "promoterless" and "enhancerless" homozygotes. All hemizygotes were obtained by crossing the homozygote to *yw* or $y^2wct^6; su(Hw)^{e04061}$/TM6B, Tb^1.

Immunostaining And Microscopy

Immunostaining was performed as described (Saint Phalle, 2003). Wing imaginal discs were dissected from the late third instar larvae in SFX media and fixed with 500µl Fixation Buffer (4% PFA/0.5%Triton/1xPBS) for 30 min at RT with rotation. The discs were washed 3X with Block-Permeabilization solution (1% BSA/0.5% Triton/1X PBS) for 10 min each, and then incubated in the same solution for 1 hr at RT with rotation. Discs were then incubated with α-GFP (Invitrogen) diluted 1:350 in wash solution (1% BSA/0.1% Triton/1X PBS) for 1 hr at RT. After washing 3X for 10 min each, discs were incubated with α-rabbit IgG-Texas Red secondary antibody (Jackson ImmunoResearch) at 1:500 for 1 hr at RT, washed 3X with PBST for 10 min each, counterstained with DAPI and rinsed with H_2O. Discs were then oriented on coverslips containing Poly-L lysine and mounted in Vectashield.

Images were obtained on a Leica DM6000B widefield epifluorescent microscope equipped with a Hamamatsu ORCA-ER CCD camera and a HC PL FLUOTAR 20x/.50NA objective. Simple PCI (v6.6) was used for acquisition of raw images, which were processed using AutoQuant's 3D Deconvolution Algorithm utilizing an adaptive (blind) PSF implemented into Leica Deblur (v2.3.2) software. All seven genotypes for each line were processed and imaged at the same time using identical immunostaining, microscope, camera and software settings. Image level normalization, minimum/maximum correction and false coloring were performed using ImageJ (v1.47n).

RNA Isolation & cDNA Synthesis

Total RNA was isolated from wing imaginal discs dissected from late third instar larva. 10-15 discs were dissected in SFX media and homogenized in 300 µl TRIzol (Invitrogen) by vortexing for 30 sec. 60 µl chloroform was added and vortexed for 15 s, centrifuged at 12,000g for 10 min at 4ºC and the upper aqueous layer precipitated with 150µl isopropanol. Samples were incubated at RT for 10 min and centrifuged at 12,000g for 10 min at 4ºC. RNA pellets were washed with 80% EtOH and resuspended in 8.5 µl nuclease-free H_2O. Genomic DNA was removed by DNAse treatment (*TURBO DNA-Free*, Ambion/Life Technologies) by incubating at 37ºC for 20 min. Concentrations and purity were determined using a NanoDrop spectrophotometer, and 500ng of RNA was used for cDNA synthesis using either the iScript™ cDNA synthesis kit containing a blend of oligo dT and random hexamers or the iScript™ Select cDNA synthesis kit with oligo dT primers only (s*u(Hw)*^{e04061} mutant analysis) (BioRad) for 1 hr at 42ºC.

Real Time PCR & Data Normalization

qPCR runs were performed on a BioRad iQ5 cycler using iQ SYBR Green Supermix (BioRad) using 1 µl of cDNA and primers specific for *eGFP* and *Rp49*. Both primers sets displayed 99-101% efficiencies. Three biological replicates for each genotype and 3 technical replicates were used and the relative expression was calculated by comparing *eGFP* C_t values to *Rp49* C_t values following the ΔC_t method. For each line, the relative expression for the intact homozygote genotype (P^+E^+/P^+E^+) was taken to be 1 and the other genotypes normalized accordingly. To derive the final data, the normalized values from all available lines were averaged. Error bars represent standard error of the mean (S.E.M.).

CONCLUSIONS

The collection of work outlined here attempts to address three important questions regarding insulator biology from a hypothesis-driven perspective. First, what does the lineage restrictions for BEAF-32 tell us about the evolutionary consequences of insulator function? Secondly, what is the physiological basis for insulator body formation, and what functional purpose (if any) do they serve? Lastly, how important are these elements in facilitating interactions between enhancers and promoters when located on homologous chromosomes? In terms of the chromatin insulator field as a whole, however, perhaps the most important consideration is whether these studies provide significant and novel insight into how these elements regulate genome dynamics *in vivo*.

Although a number of molecular processes are required for proper nuclear function and genome behavior, long-range chromatin looping plays a central role in virtually all of them (spatial organization of the chromatin fiber, chromosome architecture, transcriptional regulation at multiple levels, etc.), functioning as a global regulator of these events. The need for chromatin looping is easiest to conceptualize when considering the following: if each of the 4 chromosome pairs of the *D. melanogaster* genome were stretched and arranged end to end, nearly 10 cm of linear DNA would be present, a number more enormous when considering that it all has to fit within a nucleus with a diameter of 4-6 μm and spherical volume of ~50-110 μm^3. Chromatin looping helps maximize the amount of DNA that can occupy a given volume and it was originally thought that packaging chromatin into higher order loop structures was based solely on the need to satisfy these physical or structural requirements. However, it is becoming apparent

that this organization is non-random and results in the partitioning of the genome into a hierarchy of domains, some of which provide the master physical or structural organization required for packaging, whereas others might serve in a more direct functional capacity, establishing domains that directly influence gene expression (i.e., histone modification domains, regions of co-expressed/co-regulated genes, etc.). Similar findings have been drawn from analysis of yeast, Drosophila and human genomes, suggesting that domain hierarchy established by chromatin looping is a fundamental principle underlying eukaryotic genome function, lending support to the idea that the proteins that mediate this process function as master regulators of the genome (Dixon et al. 2012; Lieberman-Aiden et al. 2009; Duan et al. 2010; Hou et al. 2012; Sexton et al. 2012). Although the full complement of proteins and other factors that are responsible for orchestrating and maintaining these domains remain unknown, recent studies from Drosophila have suggested that chromatin insulators function as the architects of this domain hierarchy, explaining why 'long-range chromatin loop mediators' has become the favored hypothesis for the *in vivo* role of these elements (Hou et al. 2012; Sexton et al. 2012).

Probably the most unexpected finding from early genome-wide profiling studies was the enormous amount of combinatorial binding among different insulator proteins depending on the insulator—for example, one insulator might be enriched in BEAF-32, CP190 and CTCF, another in Su(Hw), CP190 and Mod(mdg4)67.2, another in Su(Hw) only, etc. Although the functional significance of this distribution remains poorly understood, such a complex landscape might suggest the existence of an “insulator code”, analogous to the histone code (Jenuwein and Allis, 2001), that is responsible for directing the correct looping contacts among the thousands of

possibilities and thereby establishing and maintaining the aforementioned hierarchy of chromatin domains. Recent studies lend support to this hypothesis, as cell-type specific binding of insulator proteins have been observed in Drosophila and vertebrate CTCF has been shown to mediate different chromatin loops in a cell-type specific manner to direct distinct gene regulatory programs (Bushey et al., 2009; Hou et al., 2010b; Junier et al., 2012; Ren et al., 2012; Phillips-Cremins et al., 2013), whereas domains of coexpressed genes in Drosophila lack both CP190 and BEAF-32 (Wallace et al., 2010). Furthermore, only the insulators that bind all known chromatin proteins (super insulators) appear to demarcate physical domains (those that serve a structural role in helping to properly fold the chromatin fiber) and are also enriched at borders of H3K27me3 domains, where CTCF has been shown to play an active role in their maintenance (Hou et al., 2012; Van Bortle et al., 2012). Others have also noticed the correlation between physical domain borders and insulators bound by specific combinations of insulator proteins, and have also suggested that other types of chromatin domains might be established by different classes of insulators (Sexton et al., 2012). This data also suggests that insulator placement within the genome is not random, a finding augmented by the fact that insulator binding site swapping in the Hox Complex leads to partial homeotic transformations (Iampietro et al., 2008). Taken collectively, this data supports the conclusion that the function bestowed upon different classes of insulators is likely to be critical for establishing a hierarchy of chromatin domains that exert master control over the genome.

The computational analysis outlined in Chapter I lends support to this idea; however, the lack of suitable orthologs for many Drosophila insulator proteins in other eukaryotes suggests that the

ability to mediate loops, rather than the proteins directly involved, has conferred the selective advantage throughout the course of the evolution. This is supported by the fact that a stalled RNA Pol II can function as an effective insulator that can block enhancer-promoter communication in Drosophila (Chopra et al., 2009). Additionally, a number of insulator sequences in yeast, primarily coding for tRNAs, have been identified that lack similar behavior in animals (Ishii et al., 2002; Ishii and Laemmli, 2003). Furthermore, although it has been assumed for years that plants lack insulators all together, recent studies have suggested that the plant-specific DNA binding proteins Asymmetric Leaves1 and 2 (AS1/AS2) repress *KNOX* expression in a looping-dependent manner, while the *TBS* sequence from *Petunia* can function as a bona-fide enhancer blocker in transgenic assays in *A. thaliana* (Guo et al., 2008; Hily et al., 2009; Yang et al., 2011). Also, tobacco specific matrix-attachment regions (MARS) have also been shown to increase reporter gene expression when flanking plant-specific loci, even when heterologous, analogous to the boundary property of animal insulators (Allen et al., 1996; Allen et al., 1993; Breyne et al., 1992; Mlynarova et al., 2003). Although the complement of plant proteins that confer these properties are unknown, the phylogenetic distribution of Drosophila insulator proteins suggests that they are likely to be plant specific.

Interestingly, however, the *gypsy* insulator can function as an insulator in *A. thaliana* to boost transgene expression in the absence of *gypsy* proteins. Coexpression of *su(Hw),* but presumably none of the other *gypsy* components (*CP190* and *Mod(mdg4)67.2*) increased reporter expression (She et al., 2010), a somewhat bizzare finding given that both CP190 and Mod(mdg4)67.2 are required for proper *gypsy* function in Drosophila, with mutations in *Mod(mdg4)67.2* in particular

converting *gypsy* into a Su(Hw)-dependent bi-directional silencer (Gdula et al., 1996; Gerasimova et al., 1995; Pai et al., 2004). Nonetheless, the data outlined in Chapter I, combined with the presence of plant and yeast-specific sequences that possess insulator properties yet have no homology to animals support the conclusion that insulator function (i.e., looping), rather than the proteins that confer them, has been the main agent for selection throughout the course of evolution.

Based on this conclusion, it is interesting to speculate that perhaps "insulation" may in fact be a ubiquitous phenomenon conferred solely by topological arrangement of the chromatin fiber, and that a number of proteins can display insulator properties depending on the context. The aforementioned promoter-proximal stalled RNA Pol II, which also depends on a number of general transcription factors and negative elongation factors (Chopra et al., 2009), is one example that supports this idea. Furthermore, although it has been assumed that the loss of other Drosophila insulator proteins in vertebrates was compensated by CTCF acquiring their functions, it is more likely that other proteins that help mediate chromatin looping, but have yet to be identified as possessing insulator behavior, are involved in vertebrates. The mammalian cohesin complex is one likely candidate, given that the SA2 (SCC3 in Drosophila and yeast) cohesin subunit physically interacts with CTCF, colocalize at thousands of genomic sites and is required for stabilizing CTCF-specific loop formation between regulatory elements (Hadjur et al., 2009; Hou et al., 2010a; Kagey et al., 2010; Mishiro et al., 2009; Nativio et al., 2009; Parelho et al., 2008; Rubio et al., 2008; Xiao et al., 2011). Therefore, perhaps the vertebrate cohesin complex functions analogously to the other Drosophila insulator proteins to assist CTCF in establishing

the correct long distance contacts. In support of this conclusion, direct interactions between CTCF and SCC3 are not observable in Drosophila, and only a small fraction of CTCF sites colocalize with cohesins, which appears to be more dependent on CP190 than CTCF (Bartkuhn et al., 2009).

Finally, the apparent lineage specificity of BEAF-32 and Zw5 to Drosophila lends further support to the idea that insulation is simply a consequence of chromatin loop formation, one that can be mediated by a number of different types of proteins. The biased distribution of BEAF-32 to promoters along with the finding that the genes that are misregulated in a *BEAF-32* mutant background show a very defined Gene Ontology (GO) classification, such as cell cycle progression, cell polarity, proliferation and differentiation (Bushey et al., 2009; Emberly et al., 2008; Gurudatta et al., 2013; Jiang et al., 2009; Negre et al., 2010; Roy et al., 2007; Yang et al., 2012), suggests that BEAF-32 functions primarily as a transcription factor for a subset of genes rather than as a true insulator, which might be expected to have a more global effect on transcription as a whole. This is supported by the fact that a hallmark of transcription factors is their ability to undergo rapid evolution within single evolutionary lineages (Lespinet et al., 2002), something that appears to be true for BEAF-32. It might be of interest for future work to computationally identify other lineage-specific transcription factors from various eukaryotes, including higher vertebrates, and test their insulating abilities in transgenic assays, which would lend further evidence to the idea of insulation being a property shared by many different types of DNA- and chromatin-binding proteins.

Perhaps of all the work presented here, the conclusions derived from the characterization of insulator body formation might represent the most significant contribution towards understanding the *in vivo* role of these elements, as these structures were a primary reason why the chromatin looping model was suggested in the first place, prior to any type of high-throughput/genome-wide analysis (Gerasimova and Corces, 1998; Gerasimova and Corces, 2001). The observation that chromatin insulator proteins appeared to concentrate into discrete foci within the diploid nucleus in small numbers (10-30) that did not match with the total number of binding sites identified by polytene chromosome spreads (>500) lead to the suggestion that these foci were the physical manifestations of multiple protein-bound insulators interacting with one another, and as a physical consequence, creating loops of chromatin analogous to rosette structures. Although in reality the model was very simplistic and only had the support of a handful of experiments that provided indirect evidence for loop formation (Byrd and Corces, 2003; Gerasimova et al., 2000), the hypothesis achieved dogma-like status in the field and only a few publications on these structures followed, though none ever addressed directly whether the basic premise of the model (chromatin loop formation) was actually valid. Most troublesome was that these structures continued to be used as a metric for insulator function even though no one had ever shown if this were true, either. Coupled with the fact that prior to the work described here, even the most fundamental biology of these structures, such as how and why they form, whether they might play a functional role in other aspects of insulator behavior and their contributions to genome organization were poorly understood, suggesting that any contribution to any of these questions could potentially fill large gaps in our understanding of these structures.

Admittedly, the work described here did not begin with the goal of testing the basic tenets of the model either. Instead, it gradually evolved to that stage, only after it had been identified what was actually causing insulator body formation in the first place. It was realized shortly after beginning to work on these structures that their inconsistent presence in fixed cells hinted that they were not likely to be the rosette structures responsible for establishing and organizing interphase chromatin within the nucleus. Nonetheless, is was difficult to comprehend exactly why they might be forming, and the lack of consistency displayed by cells within the population as a whole prevented any type of biochemical analysis that would be necessary for providing direct evidence for this. However, a chance observation that a higher proportion of cells displayed body formation after extended incubations in phosphate buffered saline soon lead to osmostress being identified as the mediator of this phenomenon. Though unexpected, its robust effect allowed for direct biochemical analysis of DNA binding via chromatin immunoprecipitation that confirmed insulator bodies were not the rosette structures originally suggested, ultimately disproving a hypothesis that had formed the foundation for how these elements behave *in vivo.*

It is interesting to consider, however, that in terms of the big picture, the conclusions drawn from these experiments haven't changed the way in which we think about insulator function *in vivo.* At the very least, however, it should prevent future misguided analysis of insulator function based on these structures, something that the field as a whole can benefit from. Although it is clear that insulator bodies are not bound to chromatin, genome-wide analysis of these elements in Drosophila and vertebrates over the last few years has independently verified that individual

elements do participate in mediating long range contacts between distant genomic sites that generate chromatin loops under non-stress conditions. To put it in a much simpler perspective, even though the initial hypothesis put forth by Gerasimova and Corces was based on a dissection artifact, ultimately they got it right, even if it was for all the wrong reasons.

Nonetheless, two major questions, perhaps not mutually exclusive, regarding insulator body formation remain unanswered: what is their physiological relevance and what are the consequences for chromatin structure and function as a whole? The lack of targeted degradation of insulator proteins and the rapid repopulation of chromatin following recovery suggests that they might function in an osmoprotective manner, directly involved in allowing the cell to adapt to changes in osmolarity. How they might do so, however, remains speculative and could involve a number of different nuclear processes, including those in which a link with insulators have yet to be identified (cell cycle arrest, DNA replication and repair, etc.). Addressing this question globally from the perspective of chromatin structure and function seems most applicable, since chromatin has the ability to exert a significant influence on virtually every aspect of nuclear biology. From a structural standpoint, it is likely that a number of biochemical and topological changes ensue as insulator proteins are removed from chromatin. Insulator binding sites are primarily nucleosome free, a biochemical property that may be dependent on chromatin remodeling complexes, such as Brahma, NURF and NuRD. Interestingly, both NURF and NuRD have context-dependent effects on the enhancer blocking ability of different classes of Drosophila insulators, although it remains unknown whether insulator proteins play a direct role in modulating the behavior of NURF and NuRD. However, loss of Su(Hw) leads to a reduction

in Brahma, followed by a subsequent increase in nucleosome density at Su(Hw) binding sites (Li et al. 2010; Negre et al. 2010; Vorobyeva et al. 2013). It is possible that removal of insulator proteins from chromatin is needed to modulate the behavior of specific remodeling complexes, inducing high nucleosome density at these sites which might be required for full condensation and compaction of chromatin observed during osmostress. Additionally, changes in histone modifications to these and surrounding nucleosomes is also likely to accompany removal of insulator proteins from chromatin, given that CTCF appears to demarcate distinct regions of histone modifications in both Drosophila and vertebrates. This is further corroborated by the finding that dCTCF in Drosophila appears to play an active, positive role in maintaining domains of H3K27me3 (Cuddapah et al., 2009; Kim et al., 2007; Van Bortle et al., 2012). These changes in histone modifications might also be needed to neutralize any charges on the histone tails themselves in order to prevent electrostatic repulsion between closely packaged nucleosomes, allowing for more efficient compaction of the chromatin mass.

However, the most dramatic structural change likely accompanies rearrangement of the topological organization of the chromatin fiber due to loss of the long range looping contacts between distant insulator sites as the proteins are removed from chromatin. Given the hierarchy of loop domains that insulators have been hypothesized to establish, this would likely disrupt the entire spatial architecture of the genome, from the physical domains responsible for establishing the overall topological arrangement in the nucleus to the smaller functional domains required for proper gene expression. However, it is interesting to note that the ChIP analysis outlined in Chapter II revealed that the class of insulators responsible for establishing these physical

domains (Hou et al. 2012; Sexton et al. 2012) showed less of a reduction (though still significant) in response to stress compared to the other insulator sites expected to be responsible for controlling the functional domains. This might suggest that a 'memory' of the default topological or physical conformation of the chromatin fiber is maintained throughout the stress response by these types of insulators whose proteins are not readily removed, while the functional domains are disrupted or reorganized due to significant loss of insulator proteins at these sites. This type of mechanism might allow the cell to rapidly recover from osmostress, requiring only the functional domains to be re-established since the overall topological organization of the chromatin fiber is likely to be maintained.

But what about the functional consequences of these structural changes? The most obvious is changes in gene expression, as nucleosome position and density, histone modifications and looping contacts can all significantly influence gene regulation, particularly if those looping contacts are responsible for establishing clusters of co-expressed genes (i.e., the functional domains) or directly modulating contacts between enhancers and promoters. This might also be a mechanism by which insulators contribute to cell cycle progression, given that both BEAF-32 and CTCF bind to promoters of genes involved in cell cycle progression and development and appear to play a significant role in controlling their expression to prevent uncontrolled cell division (Bushey et al. 2009; Emberly et al. 2008; Jinag et al. 2009; Yang et al. 2012). However, this does not appear to be true for all promoter-bound insulator proteins, as removal of Su(Hw) altered the expression of some genes (*CanB*), but not others (*Rph*), during osmostress. Other studies have suggested that removal of Su(Hw) and other insulator proteins does not lead to a

dramatic, global change in gene expression (Schwartz et al. 2012). However, it is important to point out that these types of studies have used RNAi-based methods to reduce protein levels on an individual basis—thus, removal of one insulator protein might be compensated by the presence of other insulator proteins at the same site, which would likely buffer any changes in gene expression that might be expected otherwise. Given that osmostress removes all insulator proteins from chromatin, it will be absolutely critical for future work to apply high-throughput ChIP-Sequencing, RNA-Sequencing, and Hi-C (chromosome conformation capture) methods to test this possibility. It is also likely that this approach will provide the most complete, comprehensive example of how changes in chromatin functionally alter genome behavior and possibly shed light on the physiological purpose of insulator body formation.

Although changes in gene expression would be the most likely functional consequence of removing insulator proteins from chromatin, two other nuclear processes, DNA replication and DNA repair, might also be linked to this response. Recently, high-throughput studies identified a handful of insulator binding sites that also colocalize with origin of replication (ORC) proteins. Removal of Su(Hw) from chromatin resulted in an increase in nucleosome density at insulator sites, which correlated with a reduction in binding by the ORC complex (Vorobyeva et al. 2013; Yang et al. 2013). This increase in nucleosome density as a result of insulator protein loss from chromatin might be a method the cell uses to prevent or delay S-phase entry, by ensuring that origins do not fire and initiate new rounds of DNA synthesis via sterical hindrance of ORC binding to DNA under conditions of osmostress. Origins of replication are known to be associated with nucleosome free regions (MacAlpine et al. 2010; Eaton et al. 2011), particularly

near active promoters that are likely to be enriched in BEAF-32, CP190 or CTCF, suggesting that insulators may play a yet-unidentified role in modulating DNA replication by directly altering the immediate chromatin environment and ultimately determining whether an origin is licensed for replication. Alternatively, insulator bodies could act as a sink to sequester DNA replication components to ensure that DNA replication does not ensue during osmostress, although no direct protein-protein contacts between insulator proteins and ORC proteins have yet been identified.

One reason why this would be a feasible method to prevent S-phase entry or delay further replication under conditions of osmostress is because it would afford a level of rapid and precise control to instantly shut replication down, rather than relying on some independent signal transduction pathway whose downstream effects might take minutes or longer to manifest. This becomes even more important when considering the large number of DNA breaks that occur following treatment with certain osmolytes, such as NaCl (Kültz and Chakravarty 2001). Replication at these sites could potentially be catastrophic for the cell, so it would make sense that multiple levels of rapid control might be employed by the cell in order to halt synthesis until the stress is alleviated and/or the damage has been repaired. Nonetheless, it is interesting to speculate that insulators might also play a role in the actual repair process, or, perhaps contribute to these osmostress-induced DNA breaks in the first place. In mammals, the majority of these NaCl-induced breaks occur in 'gene deserts', or regions of the genome that lack genes (Dmitrieva et al. 2011). Although it is not known why NaCl-induced breaks are located specifically in these regions of the genome, these 'gene deserts' might be analogous to the large intergenic regions in Drosophila that happen to be enriched in Su(Hw)-only sites. It would be interesting to determine

whether DNA breaks occur during insulator body formation and their location with respect to these and other insulator binding sites, which could be easily achieved via ChIP using a phospho-specific antibody against Histone H2Av that specifically marks regions of DNA damage.

If this is true, then the next logical step would be to determine why these insulator sites are prone to DNA damage and whether this occurs prior to or after insulator proteins are removed from chromatin. It might be expected that the reduced nucleosome occupancy of insulator binding sites while bound to chromatin could create a scenario in which removal of insulator proteins from chromatin during osmostress transiently exposes large regions of 'naked DNA' that would increase the probability that double-stranded breaks would occur at this site. In support of this idea, three of the genomic breakpoints corresponding to the major structural rearrangements of the Hox Complex in different Drosophila species, particularly those in the bithorax (Bx-C) Hox complex between *D. melanogaster* and *D. virilis,* align with known insulator binding sites (Negre et al. 2010). In addition, this pattern is not limited to the Hox complex, since breakpoints flanking syntenic chromosome segments between *D. melanogaster* and *Anopheles gambiae* occur within or near insulators as well, particularly those bound by BEAF-32. Taken collectively, this data might suggest that insulator sites are more prone to DNA breakage and it would be of interest for future work to address the role of insulator body formation in this process. Whether insulator proteins might also function in some aspect of the actual DNA repair process, either by allowing the repair machinery to have unhindered access to DNA damage in the unbound (insulator body) state or helping to recruit/stabilize the repair machinery at sites of damage once the stress is alleviated, is of course highly speculative, but still feasible nonetheless.

Finally, the role of these elements in contributing to the phenomenon of transvection, while seemingly disparate from the other chapters, lends further credence to idea of insulators being master regulators of the eukaryotic genome. Although limited by an unforeseen issue with the qPCR analysis, the most important conclusions could still be drawn from immunostaining, which verified that at the least, two flanking *gypsy* insulators are required for transvection. Although we could not rule out an additional contribution to homolog pairing, it appears that this is more analogous to the original boundary property of insulators, in which flanking a transgene with insulators can buffer the repressive effects of surrounding heterochromatin and hence boost its expression. Although it hasn't been tested directly, this most likely means that the flanking *gypsy* insulators are interacting with one another, shielding the reporter locus on both homologs within a chromatin loop mediated in *cis* rather than in *trans*. However, *trans* interactions between homologs, particularly between the enhancer and the promoter are readily observable, suggesting a much more complex chromatin landscape than can be explained by simply buffering a reporter locus from a repressive environment by creating a loop in *cis* in isolation.

Clearly, both pairing contributions and the spatial organization of chromatin fiber within the three dimensional space of the nucleus are playing significant roles as well. Homolog pairing, likely established by other unknown elements, is the critical first step, ensuring that the promoter on one homolog is in close proximity to the enhancer on the other. Where this occurs spatially in the nucleus is also critical—transcription is thought to occur in defined territories within the nucleus, called transcription factories, which are enriched in high local concentrations of RNA

Pol IIs, general transcription factors and other elongation, splicing and termination factors required for transcription and processing of mRNAs (Eskiw et al., 2008; Iborra et al., 1996a; Iborra et al., 1996b; Jackson et al., 1993; Jackson et al., 1998). Whether *gypsy* might play a role in targeting the paired loci to a transcription factory remains speculative; however its main role is to ensure a chromatin environment that is conducive for transcription within the transcription factory. This conducive chromatin state, whatever it may consist of, is established and facilitated through looping between *gypsy* elements, most likely in *cis* although transient interactions in *trans* cannot be excluded. Perhaps all four *gypsy* insulators (2 from each homolog) interact with one another to form an isolated double loop structure, facilitated by other pairing elements, that sequester the enhancer and promoter in close proximity within the transcription factory while buffering the negative effects of the surrounding chromatin environment, hence leading to a sustained transcriptional output. Given recent advances in imaging techniques, the whole process of transcription, from localization to transcription factories, binding of transcription factors and promoter firing can now be imaged in real time in single cells (Janicki et al., 2004; Revyakin et al., 2012), providing an powerful tool when combined with other biochemical techniques, such as chromosome conformation capture, in which to test this hypothesis.

Although the results suggests that a chromatin environment that is conducive for transcription is the deciding factor in gene expression even in *trans*, the importance of homolog pairing must not be overlooked. Transvection was not observed for 2-insulator lines when two separate lines carrying the reporter construct in a different location in the genome were combined in *trans*. Although this experiment was initially designed to test the possibility of extremely long-range

contacts between *gypsy* elements, it indirectly provided evidence for pairing dependence between homologous sequences as being required for transvection (as originally suggested by E.B. Lewis), in addition to having a chromatin environment suitable for transcription. However, does the failure of *gypsy* to interact over extreme distances (and even on separate chromosomes) provide evidence against these elements as long-range loop mediators? The answer is most certainly no, as extreme long-range (>100 Kb) contacts between the *homie* insulator and other regulatory elements have been demonstrated (Fujioka et al., 2009). It is also likely that the physical and topological domains that are responsible for genome organization within the three dimensional space of the nucleus also exert significant influence, as there are regions that despite being located far from one another in terms of the linear organization of a chromosome actually end up being within close spatial proximity when packaged inside the nucleus *in vivo.* These regions would have a higher propensity to interact with one another as a result, compared to two regions that are located on opposite sides of the nucleus. Nonetheless, this data suggests that both homolog pairing and chromatin structure play key roles in mediating transvection.

It is interesting to note that a number of studies probing the propensity of *trans* interactions in Drosophila have concluded that transvection is possible throughout the genome, a finding supported by the large number of loci that have demonstrated inter-allelic complementation and the ability of homologs to remain paired throughout interphase (Chen et al., 2002; McKee, 2004; Mellert and Truman, 2012). However, the failure of the 0- and 1-insulator lines to display evidence for transvection, even when located in genomic locations similar to the 2-insulator lines (i.e., gene promoters/5'UTRs) that did show transvection, suggests that the Drosophila genome is

not permissible to transvection by default and lends further evidence for chromatin structure as being the primary determinant of this phenomenon. Interestingly, a similar behavior has been demonstrated in plants, where a previous study utilizing a reporter system similar to our 0-insulator construct suggested that the tobacco genome is not permissible to transvection with the exception of defined locations that can form unique chromatin loop structures (i.e., inverted DNA repeats) (Matzke et al. 2001), suggesting that the functional behavior of the Drosophila and plant genomes might be more similar than previously thought. It would be of great interest to adapt our strategy to plants, using plant-specific insulator sequences (i.e., *TBS* from *Petunia* and tobacco Matrix Attachment Regions (MARS)) to test whether transvection is more readily observed in genomic locations previously shown to be negative for transvection. Not only would this add further support for the idea that chromatin structure is the ultimate determinant of enhancer-promoter communication in *trans,* but would also highlight yet another conserved mechanism of eukaryotic genome function dictated by chromatin insulators.

In conclusion, if chromatin insulators are master regulators of nuclear and genome behavior, what might future studies uncover? Chromatin plays a critical role in both DNA replication and repair, and it is possible that clever work in the near future will identify a link between insulators and these processes. Although the field of insulator biology is currently dominated by high-throughput 'fishing' analyses, it is evident that both traditional, hypothesis-driven wet lab and sophisticated computational approaches will be critical for making the transition from correlative to causative, which will be necessary for unlocking the secrets of these elements in the future. Excitingly, the field has also recognized this and it will only be a matter of time until the

mechanistic insight into how these enigmatic elements coordinate genome dynamics is ultimately revealed, likely with a number of unexpected findings along the way.

LIST OF REFERENCES

Alkema, M.J., M. Bronk, E. Verhoeven, A. Otte, L.J. van 't Veer, A. Berns, and M. van Lohuizen. 1997. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes Dev.* 11:226-240.

Allen, G.C., G. Hall, Jr., S. Michalowski, W. Newman, S. Spiker, A.K. Weissinger, and W.F. Thompson. 1996. High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. *Plant Cell*. 8:899-913.

Allen, G.C., G.E. Hall, Jr., L.C. Childs, A.K. Weissinger, S. Spiker, and W.F. Thompson. 1993. Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. *Plant Cell*. 5:603-613.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25:3389-3402.

Aramayo, R., and R.L. Metzenberg. 1996. Meiotic transvection in fungi. *Cell*. 86:103-113.

Aravind, L. 2000. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* 25:421-423.

Arvey, A., P. Agius, W.S. Noble, and C. Leslie. 2012. Sequence and chromatin determinants of cell-type–specific transcription factor binding. *Genome Res.* 22:1723-1734.

Babu, P., and S. Bhat. 1980. Effect of zeste on white complementation. *Basic Life Sci.* 16:35-40.

Baniahmad, A., C. Steiner, A.C. Köhne, and R. Renkawitz. 1990. Modular structure of a chicken lysozyme silencer: Involvement of an unusual thyroid hormone receptor binding site. *Cell*. 61:505-514.

Bantignies, F., V. Roure, I. Comet, B. Leblanc, B. Schuettengruber, J. Bonnet, V. Tixier, A. Mas, and G. Cavalli. 2011. Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell*. 144:214-226.

Barolo, S., L.A. Carver, and J.W. Posakony. 2000. GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. *Biotechniques*. 29:726, 728, 730, 732.

Bartkuhn, M., T. Straub, M. Herold, M. Herrmann, C. Rathke, H. Saumweber, G.D. Gilfillan, P.B. Becker, and R. Renkawitz. 2009. Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J.* 28:877-888.

Bateman, J.R., J.E. Johnson, and M.N. Locke. 2012. Comparing enhancer action in cis and in trans. *Genetics*. 191:1143-1155.

Bell, A., and G. Felsenfeld. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*. 405:482 - 485.

Bell, A.C., A.G. West, and G. Felsenfeld. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell.* 98:387-396.

Bellen, H.J., R.W. Levis, G. Liao, Y. He, J.W. Carlson, G. Tsang, et al. 2004. The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. *Genetics*. 167:761-781.

Bellen, H.J., C.J. O'Kane, C. Wilson, U. Grossniklaus, R.K. Pearson, and W.J. Gehring. 1989. P-element-mediated enhancer detection: a versatile method to study development in Drosophila. *Genes Dev.* 3:1288-1300.

Benyajati, C., and A. Worcel. 1976. Isolation, characterization, and structure of the folded interphase genome of Drosophila melanogaster. *Cell*. 9:393-407.

Berglund, A.-C., E. Sjolund, G. Ostund, and E.L.L. Sonnhammer. 2007. InParanoid 6: Eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res.* 36:D263-D266.

Bhaskar, V., and A.J. Courey. 2002. The MADF-BESS domain factor Dip3 potentiates synergistic activation by dorsal and twist. *Gene*. 299:173-184.

Biamonti, G., and C. Vourc'h. 2010. Nuclear stress bodies. *Cold Spring Harb. Perspect. Biol.* 2:a000695.

Blanton, J., M. Gaszner, and P. Schedl. 2003. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev.* 17:664-675.

Breyne, P., M. van Montagu, N. Depicker, and G. Gheysen. 1992. Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco. *Plant Cell*. 4:463-471.

Brooks, A.N., L. Yang, M.O. Duff, K.D. Hansen, J.W. Park, S. Dudoit, S.E. Brenner, and B.R. Graveley. 2011. Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Res.* 21:193-202.

Buchner, K., P. Roth, G. Schotta, V. Krauss, H. Saumweber, G. Reuter, and R. Dorn. 2000. Genetic and Molecular Complexity of the Position Effect Variegation Modifier mod(mdg4) in Drosophila. *Genetics*. 155:141-157.

Bulger, M., and M. Groudine. 2010. Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev. Biol.* 339:250-257.

Buratovich, M.A., R.G. Phillips, and J.R. Whittle. 1997. Genetic relationships between the mutations spade and Sternopleural and the wingless gene in Drosophila development. *Dev. Biol.* 185:244-260.

Burg, M.B., J.D. Ferraris, and N.I. Dmitrieva. 2007. Cellular response to hyperosmotic stresses. *Physiol. Rev.* 87:1441-1474.

Bushey, A.M., E.R. Dorman, and V.G. Corces. 2008. Chromatin insulators: Regulatory mechanisms and epigenetic inheritance. *Mol. Cell*. 32:1-9.

Bushey, A.M., E. Ramos, and V.G. Corces. 2009. Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. *Genes Dev.* 23:1338-1350.

Busturia, A., A. Lloyd, F. Bejarano, M. Zavortink, H. Xin, and S. Sakonju. 2001. The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development*. 128:2163-2173.

Butcher, R.D.J., S. Chodagam, R. Basto, J.G. Wakefield, D.S. Henderson, J.W. Raff, and W.G.F. Whitfield. 2004. The Drosophila centrosome-associated protein CP190 is essential for viability but not for cell division. *J. Cell Sci.* 117:1191-1199.

Byrd, K., and V.G. Corces. 2003. Visualization of chromatin domains created by the gypsy insulator of Drosophila. *J. Cell Biol.* 162:565-574.

Cai, H.N., and P. Shen. 2001. Effects of cis Arrangement of Chromatin Insulators on Enhancer-Blocking Activity. *Science*. 291:493-495.

Capelson, M., and V.G. Corces. 2005. The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol. Cell*. 20:105-116.

Capelson, M., and V.G. Corces. 2006. SUMO conjugation attenuates the activity of the gypsy chromatin insulator. *EMBO J.* 25:1906-1914.

Casares, F., W. Bender, J. Merriam, and E. Sanchez-Herrero. 1997. Interactions of Drosophila Ultrabithorax regulatory regions with native and foreign promoters. *Genetics*. 145:123-137.

Cavalli, G., and R. Paro. 1998. The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell*. 93:505-518.

Celniker, S.E., L.A. Dillon, M.B. Gerstein, K.C. Gunsalus, S. Henikoff, G.H. Karpen, et al. 2009. Unlocking the secrets of the genome. *Nature*. 459:927-930.

Chen, J.L., K.L. Huisinga, M.M. Viering, S.A. Ou, C.T. Wu, and P.K. Geyer. 2002. Enhancer action in trans is permitted throughout the Drosophila genome. *Proc. Natl. Acad. Sci. U. S. A.* 99:3723-3728.

Cherbas, L., A. Willingham, D. Zhang, L. Yang, Y. Zou, B.D. Eads, et al. 2011. The transcriptional diversity of 25 Drosophila cell lines. *Genome Res.* 21:301-314.

Chopra, V.S., J. Cande, J.-W. Hong, and M. Levine. 2009. Stalled Hox promoters as chromosomal boundaries. *Genes Dev.* 23:1505-1509.

Chung, J., M. Whiteley, and G. Felsenfeld. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell*. 74:505 - 514.

Chung, J.Ä., A.Ä. Bell, and G. Felsenfeld. 1997. Characterization of the chicken Beta-globin insulator. *Proc. Natl. Acad. Sci. U. S. A.* 94:575-580.

Clark, A.G., M.B. Eisen, D.R. Smith, C.M. Bergman, B. Oliver, T.A. Markow, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature*. 450:203-218.

Clark, K.A., and D.M. McKearin. 1996. The Drosophila stonewall gene encodes a putative transcription factor essential for germ cell development. *Development*. 122:937-950.

Cole, C., J.D. Barber, and G.J. Barton. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* 36:197-201.

Comet, I., B. Schuettengruber, T. Sexton, and G. Cavalli. 2011. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc. Natl. Acad. Sci. U. S. A.* 108:2294-2299.

Condie, J.M., and D.L. Brower. 1989. Allelic interactions at the engrailed locus of Drosophila: engrailed protein expression in imaginal discs. *Dev. Biol.* 135:31-42.

Craig, C.R., J.L. Fink, Y. Yagi, Y.T. Ip, and R.L. Cagan. 2004. A Drosophila p38 orthologue is required for environmental stress responses. *EMBO Rep.* 5:1058-1063.

Cuddapah, S., R. Jothi, D.E. Schones, T.Y. Roh, K. Cui, and K. Zhao. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* 19:24 - 32.

Cutler, G., K.M. Perry, and R. Tjian. 1998. Adf-1 Is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. *Mol. Cell. Biol.* 18:2252-2261.

Cuvier, O., C. Hart, and U. Laemmli. 1998. Identificaiton of a class of chromatin boundary elements. *Mol. Cell. Biol.* 18:7478-7486.

De Santa, F., I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B.K. Tusi, et al. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* 8:e1000384.

Delattre, M., A. Spierer, N. Hulo, and P. Spierer. 2002. A new gene in *Drosophila melanogaster*, ravus, the phantom of the modifier of position-effect variegation Su(var)3-7. *The International Journal of Developmental Biology*. 46:167-171.

Deng, X., J.B. Hiatt, D.K. Nguyen, S. Ercan, D. Sturgill, L.W. Hillier, et al. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nat. Genet.* 43:1179-1185.

Dixon, J.R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, et al. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485: 376–380.

Dmitrieva, N.I., K. Cui, D.A. Kitchaev, K. Zhao, and M.B. Burg. 2011. DNA double-stranded breaks induced by high NaCl occur predominately in gene deserts. *Proc. Natl. Acad. Sci. U.S.A.* 108:20796-20801.

Dong, X., M. Greven, A. Kundaje, S. Djebali, J. Brown, C. Cheng, et al. 2012. Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.* 13:R53.

Dorsett, D. 2011. Cohesin: genomic insights into controlling gene transcription and development. *Curr. Opin. Genet. Dev.* 21:199-206.

Dreesen, T.D., S. Henikoff, and K. Loughney. 1991. A pairing-sensitive element that mediates trans-inactivation is associated with the Drosophila brown gene. *Genes Dev.* 5:331-340.

Duan, Z., M. Andronescu, K. Schutz, S. Mcllwain, Y.J. Kim, C. Lee, et al. 2010. A three-dimensional model of the yeast genome. *Nature.* 465:363-367.

Dunaway, M., and P. Droge. 1989. Transactivation of the Xenopus rRNA gene promoter by its enhancer. *Nature*. 341:657-659.

Duncan, I.W. 2002. Transvection effects in Drosophila. *Annu. Rev. Genet.* 36:521-556.

Dunham, I., A. Kundaje, S.F. Aldred, P.J. Collins, C.A. Davis, F. Doyle, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489:57-74.

Eaton M.L., J.A. Prinz, H.K. MacAlpine, G. Tretyakov, P.V. Kharchenko, and D.M MacAlpine. 2011. Chromatin signatures of the Drosophila replication program. *Genome Res.* 21:164-174.

Emberly, E., R. Blattes, B. Schuettengruber, M. Hennion, N. Jiang, C. Hart, E. Kas, and O. Cuvier. 2008. BEAF regulates cell-cylce genes through the controlled deposition of H3K9 methylation marks into its conserved *dual-core* binding sites. *PLoS Biol.* 6:2896-2910.

England, B., A. Admon, and R. Tjian. 1992. Cloning of Drosophila transcription factor Adf-1 reveals homology to Myb oncoproteins. *Proc. Natl. Acad. Sci. U. S. A.* 89:683-687.

Eskiw, C.H., A. Rapp, D.R. Carter, and P.R. Cook. 2008. RNA polymerase II activity is located on the surface of protein-rich transcription factories. *J. Cell Sci.* 121:1999-2007.

Farkas, G., J. Gausz, M. Galloni, G. Reuter, H. Gyurkovics, and F. Karch. 1994. The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature*. 371:806-808.

Farrell, C.M., A.G. West, and G. Felsenfeld. 2002. Conserved CTCF insulator elements flank the mouse and human β-globin loci. *Mol. Cell. Biol.* 22:3820-3831.

Filion, G.J., J.G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L.D. Ward, et al. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell*. 143:212-224.

Flemming, W. 1882. Zellsubstanz, Kern und Zelltheilung. F. C. W. Vogel, Leipzig.

Frietze, S., R. Wang, L. Yao, Y. Tak, Z. Ye, M. Gaddis, H. Witt, P. Farnham, and V. Jin. 2012. Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.* 13:R52.

Fritsch, C., G. Ploeger, and D.J. Arndt-Jovin. 2006. Drosophila under the lens: imaging from chromosomes to whole embryos. *Chromosome Res.* 14:451-464.

Frye, M., and S.A. Benitah. 2012. Chromatin regulators in mammalian epidermis. *Semin. Cell Dev. Biol.* 23:897-905.

Fujioka, M., X. Wu, and J.B. Jaynes. 2009. A chromatin insulator mediates transgene homing and very long-range enhancer-promoter communication. *Development*. 136:3077-3087.

Galloni, M., H. Gyurkovics, P. Schedl, and F. Karch. 1993. The bluetail transposon: evidence for independent cis-regulatory domains and domain boundaries in the bithorax complex. *EMBO J.* 12:1087-1097.

Gaszner, M., J. Vazquez, and P. Schedl. 1999. The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes Dev.* 13:2098-2107.

Gdula, D.A., and V.G. Corces. 1997. Characterization of Functional Domains of the su(Hw) Protein That Mediate the Silencing Effect of mod(mdg4) Mutations. *Genetics*. 145:153-161.

Gdula, D.A., T.I. Gerasimova, and V.G. Corces. 1996. Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 93:9378-9383.

Gelbart, W.M. 1982. Synapsis-dependent allelic complementation at the decapentaplegic gene complex in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* 79:2636-2640.

Georgiev, P.G., and V.G. Corces. 1995. The Su(Hw) protein bound to gypsy sequences in one chromosome can repress enhancer-promoter interactions in the paired gene located in the other homolog. *Proc. Natl. Acad. Sci. U. S. A.* 92:5184-5188.

Georgiev, P.G., and T.I. Gerasimova. 1989. Novel genes influencing the expression of the yellow locus and mdg4 (gypsy) in Drosophila melanogaster. *Molecular and General Genetics MGG*. 220:121-126.

Gerasimova, T.I., K. Byrd, and V.G. Corces. 2000. A chromatin insulator determines the nuclear localization of DNA. *Mol. Cell*. 6:1025-1035.

Gerasimova, T.I., and V.G. Corces. 1998. Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell*. 92:511-521.

Gerasimova, T.I., and V.G. Corces. 2001. Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu. Rev. Genet.* 35:193-208.

Gerasimova, T.I., D.A. Gdula, D.V. Gerasimov, O. Simonova, and V.G. Corces. 1995. A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell*. 82:587-597.

Gerasimova, T.I., E.P. Lei, A.M. Bushey, and V.G. Corces. 2007. Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. *Mol. Cell*. 28:761-772.

Gerstein, M.B., A. Kundaje, M. Hariharan, S.G. Landt, K.-K. Yan, C. Cheng, et al. 2012. Architecture of the human regulatory network derived from ENCODE data. *Nature*. 489:91-100.

Geyer, P., and V. Corces. 1992. DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes Dev.* 6:1865 - 1873.

Geyer, P.K., M.M. Green, and V.G. Corces. 1990. 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.

Ghosh, D., T.I. Gerasimova, and V.G. Corces. 2001. Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. *EMBO J.* 20:2518-2527.

Gibson, J.B., D.S. Reed, S. Bartoszewski, and A.V. Wilks. 1999. Structural changes in the promoter region mediate transvection at the sn-glycerol-3-phosphate dehydrogenase gene of Drosophila melanogaster. *Biochem. Genet.* 37:301-315.

Gilbert, M.K., Y.Y. Tan, and C.M. Hart. 2006. The Drosophila boundary element-associated factors BEAF-32A and 32B affect chromatin structure. *Genetics*. 173:1365-1375.

Gohl, D., M. Muller, V. Pirrotta, M. Affolter, and P. Schedl. 2008. Enhancer blocking and transvection at the Drosophila apterous locus. *Genetics*. 178:127-143.

Golovnin, A., L. Melnikova, I. Volkov, M. Kostuchenko, A.V. Galkin, and P. Georgiev. 2008. 'Insulator bodies' are aggregates of proteins but not of insulators. *EMBO Rep.* 9:440-445.

Golovnin, A., I. Volkov, and P. Georgiev. 2012. SUMO conjugation is required for the assembly of Drosophila Su(Hw) and Mod(mdg4) into insulator bodies that facilitate insulator complex formation. *J. Cell Sci.* 125:2064-2074.

Groth, A.C., M. Fish, R. Nusse, and M.P. Calos. 2004. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics*. 166:1775-1782.

Guex, N., and M. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*. 18:2714-2723.

Guo, M., J. Thomas, G. Collins, and M.C.P. Timmermans. 2008. Direct Repression of KNOX Loci by the ASYMMETRIC LEAVES1 Complex of Arabidopsis. *The Plant Cell Online*. 20:48-58.

Gurudatta, B.V., J. Yang, K. Van Bortle, P.G. Donlin-Asp, and V.G. Corces. 2013. Dynamic changes in the genomic localization of DNA replication-related element binding factor during the cell cycle. *Cell Cycle*. 12:1605-1615.

Gyurkovics, H., J. Gausz, J. Kummer, and F. Karch. 1990. A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. *EMBO J.* 9:2579-2585.

Hadjur, S., L.M. Williams, N.K. Ryan, B.S. Cobb, T. Sexton, P. Fraser, A.G. Fisher, and M. Merkenschlager. 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*. 460:410-413.

Hagege, H., P. Klous, C. Braem, E. Splinter, J. Dekker, G. Cathala, W. de Laat, and T. Forne. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* 2:1722-1733.

Han, S.J., K.Y. Choi, P.T. Brey, and W.J. Lee. 1998. Molecular cloning and characterization of a Drosophila p38 mitogen-activated protein kinase. *J. Biol. Chem.* 273:369-374.

Handoko, L., H. Xu, G. Li, C.Y. Ngan, E. Chew, M. Schnapp, et al. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43:630-638.

Hart, C.M., K. Zhao, and U.K. Laemmli. 1997. The scs' boundary element: characterization of boundary element- associated factors. *Mol. Cell. Biol.* 17:999-1009.

Heger, P., B. Marin, and E. Schierenberg. 2009. Loss of the insulator protein CTCF during nematode evolution. *BMC Molecular Biology*. 10:84.

Heintzman, N.D., G.C. Hon, R.D. Hawkins, P. Kheradpour, A. Stark, L.F. Harp, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 459:108-112.

Henikoff, J.G., J.A. Belsky, K. Krassovsky, D.M. MacAlpine, and S. Henikoff. 2011. Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U. S. A*. 108:18318-18323.

Henikoff, S., and T.D. Dreesen. 1989. 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.

Hily, J.-M., S. Singer, Y. Yang, and Z. Liu. 2009. A transformation booster sequence (TBS) from Petunia hybrida functions as an enhancer-blocking insulator in Arabidopsis thaliana. *Plant Cell Reports*. 28:1095-1104.

Hirose, F., M. Yamaguchi, H. Handa, Y. Inomata, and A. Matsukage. 1993. Novel 8-base pair sequence (Drosophila DNA replication-related element) and specific binding factor involved in the expression of Drosophila genes for DNA polymerase and Proliferating Cell Nuclear Antigen. *J. Biol. Chem.* 268:2092-2099.

Holohan, E.E., C. Kwong, B. Adryan, M. Bartkuhn, M. Herold, R. Renkawitz, S. Russell, and R. White. 2007. CTCF Genomic Binding Sites in *Drosophila* and the Organisation of the Bithorax Complex. *PLoS Genet.* 3:1211-1222.

Hou, C., R. Dale, and A. Dean. 2010a. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U. S. A.* 107:3651-3656.

Hou, C., R. Dale, and A. Dean. 2010b. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U. S. A.*. 107:3651-3656.

Hou, C., L. Li, Z.S. Qin, and V.G. Corces. 2012. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Mol. Cell*. 48:471-484.

Hou, C., H. Zhao, K. Tanimoto, and A. Dean. 2008. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc. Natl. Acad. Sci. U. S. A.* 105:20398-20403.

Iampietro, C., F. Cleard, H. Gyurkovics, R.K. Maeda, and F. Karch. 2008. Boundary swapping in the Drosophila Bithorax complex. *Development*. 135:3983-3987.

Iborra, F.J., A. Pombo, D.A. Jackson, and P.R. Cook. 1996a. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J. Cell Sci.* 109 (Pt 6):1427-1436.

Iborra, F.J., A. Pombo, J. McManus, D.A. Jackson, and P.R. Cook. 1996b. The topology of transcription by immobilized polymerases. *Exp. Cell Res.* 229:167-173.

Inoue, H., M. Tateno, K. Fujimura-Kamada, G. Takaesu, T. Adachi-Yamada, J. Ninomiya-Tsuji, K. Irie, Y. Nishida, and K. Matsumoto. 2001. A Drosophila MAPKKK, D-MEKK1, mediates stress responses through activation of p38 MAPK. *EMBO J.* 20:5421-5430.

Ishii, K., G. Arib, C. Lin, G. Van Houwe, and U. Laemmli. 2002. Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell*. 109:551 - 562.

Ishii, K., and U.K. Laemmli. 2003. Structural and Dynamic Functions Establish Chromatin Domains. *Mol. Cell*. 11:237-248.

Jack, J.W., and B.H. Judd. 1979. 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.

Jackson, D.A., A.B. Hassan, R.J. Errington, and P.R. Cook. 1993. Visualization of focal sites of transcription within human nuclei. *EMBO J.* 12:1059-1065.

Jackson, D.A., F.J. Iborra, E.M. Manders, and P.R. Cook. 1998. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell*. 9:1523-1536.

Janicki, S.M., T. Tsukamoto, S.E. Salghetti, W.P. Tansey, R. Sachidanandam, K.V. Prasanth, et al. 2004. From silencing to gene expression: real-time analysis in single cells. *Cell*. 116:683-698.

Jenuwein, T., and C.D. Allis. 2001. Translating the Histone Code. *Science*. 293:1074-1080.

Jiang, N., E. Emberly, O. Cuvier, and C.M. Hart. 2009. Genome-wide mapping of BEAF binding sites in Drosophila links BEAF to transcription. *Mol. Cell. Biol.* 29:3556-3568.

Joyce, E.F., B.R. Williams, T. Xie, and C.T. Wu. 2012. Identification of genes that promote or antagonize somatic homolog pairing using a high-throughput FISH-based screen. *PLoS Genet.* 8:e1002667.

Junier, I., R.K. Dale, C. Hou, F. Képès, and A. Dean. 2012. CTCF-mediated transcriptional regulation through cell type-specific chromosome organization in the β-globin locus. *Nucleic Acids Res.*

Kagey, M.H., J.J. Newman, S. Bilodeau, Y. Zhan, D.A. Orlando, N.L. van Berkum, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 467:430-435.

Karch, F., M. Galloni, L. Sipos, J. Gausz, H. Gyurkovics, and P. Schedl. 1994. Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of Drosophila melanogaster. *Nucleic Acids Res.* 22:3138-3146.

Kellum, R., and P. Schedl. 1991. A position-effect assay for boundaries of higher order chromosomal domains. *Cell*. 64:941 - 950.

Kellum, R., and P. Schedl. 1992. A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell. Biol.* 12:2424-2431.

Kennison, J.A., and J.W. Southworth. 2002. Transvection in Drosophila. *Adv. Genet.* 46:399-420.

Kharchenko, P.V., A.A. Alekseyenko, Y.B. Schwartz, A. Minoda, N.C. Riddle, J. Ernst, et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature*. 471:480-485.

Kim, T., Z. Abdullaev, A. Smith, K. Ching, D. Loukinov, R. Green, M. Zhang, V. Lobanenkov, and B. Ren. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*. 128:1231 - 1245.

Kim, T.K., M. Hemberg, J.M. Gray, A.M. Costa, D.M. Bear, J. Wu, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 465:182-187.

Kosakovsky Pond, S.L., and S.D.W. Frost. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*. 21:2531-2533.

Kosakovsky Pond, S.L., and S.D.W. Frost. 2005b. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22:1208-1222.

Krivega, I., and A. Dean. 2012. Enhancer and promoter interactions-long distance calls. *Curr. Opin. Genet. Dev.* 22:79-85.

Kültz, D., and D. Chakravarty. 2001. Hyperosmolarity in the form of elevated NaCl but not urea causes DNA damage in murine kidney cells. *Proc. Natl. Acad. Sci. U.S.A.* 98:1999-2004.

Kundaje, A., S. Kyriazopoulou-Panagiotopoulou, M. Libbrecht, C.L. Smith, D. Raha, E.E. Winters, et al. 2012. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res.* 22:1735-1747.

Kurshakova, M., O. Maksimenko, A. Golovnin, M. Pulina, S. Georgieva, P. Georgiev, and A. Krasnov. 2007. Evolutionarily conserved E(y)2/Sus1 protein is essential for the barrier activity of Su(Hw)-dependent insulators in Drosophila. 27:332-338.

Kurukuti, S., V.K. Tiwari, G. Tavoosidana, E. Pugacheva, A. Murrell, Z. Zhao, V. Lobanenkov, W. Reik, and R. Ohlsson. 2006. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. U. S. A*. 103:10684-10689.

Labrador, M., and V.G. Corces. 2002. Setting the boundaries of chromatin domains and nuclear organization. *Cell*. 111:151-154.

Lafarga, M., M.T. Berciano, E. Pena, I. Mayo, J.G. Castano, D. Bohmann, J.P. Rodrigues, J.P. Tavanez, and M. Carmo-Fonseca. 2002. Clastosome: a subtype of nuclear body enriched

in 19S and 20S proteasomes, ubiquitin, and protein substrates of proteasome. *Mol. Biol. Cell*. 13:2771-2782.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23:2947-2948.

Lee, J.J., D.P. von Kessler, S. Parks, and P.A. Beachy. 1992. Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. *Cell*. 71:33-50.

Lei, E.P., and V.G. Corces. 2006. RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat. Genet.* 38:936-941.

Leiserson, W.M., N.M. Bonini, and S. Benzer. 1994. Transvection at the eyes absent gene of Drosophila. *Genetics*. 138:1171-1179.

Lespinet, O., Y.I. Wolf, E.V. Koonin, and L. Aravind. 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* 12:1048-1059.

Lewis, E.B. 1954. The Theory and Application of a New Method of Detecting Chromosomal Rearrangements in Drosophila melanogaster. *The American Naturalist*. 88:225-239.

Lieberman-Aiden, E., N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 326:289–293

Li, M., V.E. Belozerov, and H.N Cai. 2010. Modulation of chromatin boundary activities by nucleosome-remodeling activities in Drosophila melanogaster. *Mol. Cell Biol.* 30:1067-1076

Ling, J.Q., T. Li, J.F. Hu, T.H. Vu, H.L. Chen, X.W. Qiu, A.M. Cherry, and A.R. Hoffman. 2006. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*. 312:269-272.

Liu, H., J. Huang, J. Wang, S. Jiang, A.S. Bailey, D.C. Goldman, et al. 2008. Transvection mediated by the translocated cyclin D1 locus in mantle cell lymphoma. *J. Exp. Med.* 205:1843-1858.

Liu, T., A. Rechtsteiner, T.A. Egelhofer, A. Vielle, I. Latorre, M.-S. Cheung, et al. 2011. Broad chromosomal domains of histone modification patterns in C. elegans. *Genome Res.* 21:227-236.

Locke, J., and K.D. Tartof. 1994. Molecular analysis of cubitus interruptus (ci) mutations suggests an explanation for the unusual ci position effects. *Mol. Gen. Genet.* 243:234-243.

Lomvardas, S., G. Barnea, D.J. Pisapia, M. Mendelsohn, J. Kirkland, and R. Axel. 2006. Interchromosomal Interactions and Olfactory Receptor Choice. *Cell.* 126:403-413.

MacAlpine H.K., R. Gordan, S.K. Powell, A.J. Hartemink, and D.M. MacAlpine. 2010. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res.* 20:201-211.

Markstein, M., P. Markstein, V. Markstein, and M. Levine. 2002. Genome-wide analysis of clustered dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc. Natl. Acad. Sci. U. S. A.* 99:763-768.

Markstein, M., C. Pitsouli, C. Villalta, S.E. Celniker, and N. Perrimon. 2008. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat. Genet.* 40:476-483.

Martinez-Laborda, A., A. Gonzalez-Reyes, and G. Morata. 1992. Trans regulation in the Ultrabithorax gene of Drosophila: alterations in the promoter enhance transvection. *EMBO J.* 11:3645-3652.

Matzke, M., M.F. Mette, J. Jakowitsch, T. Kanno, E.A. Moscone, J. van der Winden, and A.J. Matzke. 2001. A test for transvection in plants: DNA pairing may lead to trans-activation or silencing of complex heteroalleles in tobacco. *Genetics*. 158:451-461.

McKee, B.D. 2004. Homologous pairing and chromosome dynamics in meiosis and mitosis. *Biochim. Biophys. Acta*. 1677:165-180.

Mellert, D.J., and J.W. Truman. 2012. Transvection is common throughout the Drosophila genome. *Genetics*. 191:1129-1141.

Melnikova, L., F.ß. Juge, N. Gruzdeva, A. Mazur, G. Cavalli, and P. Georgiev. 2004. Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 101:14806-14811.

Messmer, S., A. Franke, and R. Paro. 1992. Analysis of the functional role of the Polycomb chromo domain in Drosophila melanogaster. *Genes Dev.* 6:1241-1254.

Mishiro, T., K. Ishihara, S. Hino, S. Tsutsumi, H. Aburatani, K. Shirahige, Y. Kinoshita, and M. Nakao. 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* 28:1234-1245.

Mlynarova, L., A. Hricova, A. Loonen, and J.P. Nap. 2003. The presence of a chromatin boundary appears to shield a transgene in tobacco from RNA silencing. *Plant Cell*. 15:2203-2217.

Modolell, J., W. Bender, and M. Meselson. 1983. Drosophila melanogaster mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3-kilobase mobile element. *Proc. Natl. Acad. Sci. U. S. A.* 80:1678-1682.

Mohan, M., M. Bartkuhn, M. Herold, A. Philippen, N. Heinl, I. Bardenhagen, et al. 2007. The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *The EMBO Journal*. 26:4203-4214.

Moon, H., G. Filippova, D. Loukinov, E. Pugacheva, Q. Chen, S. Smith, et al. 2005. CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep.* 6:165 - 170.

Morris, J.R., J. Chen, S.T. Filandrinos, R.C. Dunn, R. Fisk, P.K. Geyer, and C. Wu. 1999. An analysis of transvection at the yellow locus of Drosophila melanogaster. *Genetics*. 151:633-651.

Mozer, B.A. 2001. Dominant drop mutants are gain-of-function alleles of the muscle segment homeobox gene (msh) whose overexpression leads to the arrest of eye development. *Dev. Biol.* 233:380-393.

Mueller-Storm, H.P., J.M. Sogo, and W. Schaffner. 1989. An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell*. 58:767-777.

Muller, H.P., and W. Schaffner. 1990. Transcriptional enhancers can act in trans. *Trends Genet.* 6:300-304.

Muravyova, E., A. Golovnin, E. Gracheva, A. Parshikov, T. Belenkaya, V. Pirrotta, and P. Georgiev. 2001. Loss of Insulator Activity by Paired Su(Hw) Chromatin Insulators. *Science*. 291:495-498.

Natarajan, A., G.G. Yardımcı, N.C. Sheffield, G.E. Crawford, and U. Ohler. 2012. Predicting cell-type–specific gene expression from regions of open chromatin. *Genome Res.* 22:1711-1722.

Nativio, R., K.S. Wendt, Y. Ito, J.E. Huddleston, S. Uribe-Lewis, K. Woodfine, et al. 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet.* 5:e1000739.

Negre, N., C.D. Brown, L. Ma, C.A. Bristow, S.W. Miller, U. Wagner, et al. 2011. A cis-regulatory map of the Drosophila genome. *Nature*. 471:527-531.

Negre, N., C.D. Brown, P.K. Shah, P. Kheradpour, C.A. Morrison, J.G. Henikoff, et al. 2010. A Comprehensive Map of Insulator Elements for the *Drosophila* Genome. *PLoS Genet.* 6:e1000814.

Neph, S., J. Vierstra, A.B. Stergachis, A.P. Reynolds, E. Haugen, B. Vernot, et al. 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*. 489:83-90.

Ni, Y., A. Weber Hall, A. Battenhouse, and V. Iyer. 2012. Simultaneous SNP identification and assessment of allele-specific bias from ChIP-seq data. *BMC Genetics*. 13:46.

Niu, W., Z.J. Lu, M. Zhong, M. Sarov, J.I. Murray, C.M. Brdlik, et al. 2011. Diverse transcription factor binding features revealed by genome-wide ChIP-seq in C. elegans. *Genome Res.* 21:245-254.

Nolis, I.K., D.J. McKay, E. Mantouvalou, S. Lomvardas, M. Merika, and D. Thanos. 2009. Transcription factors mediate long-range enhancer-promoter interactions. *Proc. Natl. Acad. Sci. U. S. A.* 106:20222-20227.

Nordman, J., S. Li, T. Eng, D. MacAlpine, and T.L. Orr-Weaver. 2011. Developmental control of the DNA replication and transcription programs. *Genome Res.* 21:175-181.

O'Kane, C.J., and W.J. Gehring. 1987. Detection in situ of genomic regulatory elements in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* 84:9123-9127.

Oh, S.W., T. Kingsley, H.H. Shin, Z. Zheng, H.W. Chen, X. Chen, et al. 2003. A P-element insertion screen identified mutations in 455 novel essential genes in Drosophila. *Genetics*. 163:195-201.

Ohler, U., L. Guo-chun, H. Niemann, and G. Rubin. 2002. Computational analysis of core promoters in the Drosophila genome. *Genome Biol.* 3:research0087.0081-0087.0012.

Oliver, D., B. Sheehan, H. South, O. Akbari, and C.Y. Pai. 2010. The chromosomal association/dissociation of the chromatin insulator protein Cp190 of Drosophila melanogaster is mediated by the BTB/POZ domain and two acidic regions. *BMC Cell Biol.* 11:101.

Pai, C.-Y., E.P. Lei, D. Ghosh, and V.G. Corces. 2004. The Centrosomal Protein CP190 Is a Component of the gypsy Chromatin Insulator. *Mol. Cell.* 16:737-748.

Palla, F., R. Melfi, L. Anello, M. Di Bernardo, and G. Spinelli. 1997. Enhancer blocking activity located near the 3' end of the sea urchin early H2A histone gene. *Proc. Natl. Acad. Sci. U. S. A.* 94:2272-2277.

Parelho, V., S. Hadjur, M. Spivakov, M. Leleu, S. Sauer, H.C. Gregson, et al. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*. 132:422-433.

Parkhurst, S.M., and V.G. Corces. 1986. Interactions among the gypsy transposable element and the yellow and the suppressor of hairy-wing loci in Drosophila melanogaster. *Mol. Cell. Biol.* 6:47-53.

Parkhurst, S.M., D.A. Harrison, M.P. Remington, C. Spana, R.L. Kelley, R.S. Coyne, and V.G. Corces. 1988. The Drosophila su(Hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. *Genes Dev.* 2:1205-1215.

Pathak, R.U., N. Rangaraj, S. Kallappagoudar, K. Mishra, and R.K. Mishra. 2007. Boundary element-associated factor 32B connects chromatin domains to the nuclear matrix. *Mol. Cell. Biol.* 27:4796-4806.

Pattatucci, A.M., and T.C. Kaufman. 1991. The homeotic gene Sex combs reduced of Drosophila melanogaster is differentially regulated in the embryonic and imaginal stages of development. *Genetics*. 129:443-461.

Phillips, J.E., and V.G. Corces. 2009. CTCF: Master Weaver of the Genome. *Cell*. 137:1194-1211.

Phillips-Cremins, J.E., M.E.G Sauria, A. Sanyal, T.I. Gerasimova, B.R. Lajoie, J.S.K. Bell, et al. 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell.* 153:1281-1295.

Pirrotta, V., H. Steller, and M.P. Bozzetti. 1985. Multiple upstream regulatory elements control the expression of the Drosophila white gene. *EMBO J.* 4:3501-3508.

Ramos, E., E.A. Torre, A.M. Bushey, B.V. Gurudatta, and V.G. Corces. 2011. DNA topoisomerase II modulates insulator function in Drosophila. *PLoS ONE*. 6:e16562.

Raser, J.M., and E.K. O'Shea. 2005. Noise in gene expression: origins, consequences, and control. *Science*. 309:2010-2013.

Rassoulzadegan, M., M. Magliano, and F. Cuzin. 2002. Transvection effects involving DNA methylation during meiosis in the mouse. *EMBO J.* 21:440-450.

Reese, M.G. 2001. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Comput. Chem.* 26:51-56.

Ren, L., Y. Wang, M. Shi, X. Wang, Z. Yang, and Z. Zhao. 2012. CTCF Mediates the Cell-Type Specific Spatial Organization of the *Kcnq5* Locus and the Local Gene Regulation. *PLoS ONE*. 7:e31416.

Reuter, G., M. Giarre, J. Farah, J. Gausz, A. Spierer, and P. Spierer. 1990. Dependence of position-effect variegation in Drosophila on dose of a gene encoding an unusual zinc-finger protein. *Nature*. 344:219-223.

Revyakin, A., Z. Zhang, R.A. Coleman, Y. Li, C. Inouye, J.K. Lucas, S.R. Park, S. Chu, and R. Tjian. 2012. Transcription initiation by human RNA polymerase II visualized at single-molecule resolution. *Genes Dev.* 26:1691-1702.

Richter, K., M. Nessling, and P. Lichter. 2008. Macromolecular crowding and its potential impact on nuclear function. *Biochim. Biophys. Acta*. 1783:2100-2107.

Riddle, N.C., A. Minoda, P.V. Kharchenko, A.A. Alekseyenko, Y.B. Schwartz, M.Y. Tolstorukov, et al. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res.* 21:147-163.

Rogers, S.L., and G.C. Rogers. 2008. Culture of Drosophila S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. *Nat. Protocols*. 3:606-611.

Ronshaugen, M., and M. Levine. 2004. Visualization of trans-homolog enhancer-promoter interactions at the Abd-B Hox locus in the Drosophila embryo. *Dev. Cell*. 7:925-932.

Roy, S., M.K. Gilbert, and C.M. Hart. 2007. Characterization of BEAF Mutations Isolated by Homologous Recombination in Drosophila. *Genetics*. 176:801-813.

Rubio, E.D., D.J. Reiss, P.L. Welcsh, C.M. Disteche, G.N. Filippova, N.S. Baliga, R. Aebersold, J.A. Ranish, and A. Krumm. 2008. CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 105:8309-8314.

Saint Phalle, B. 2003. Immunostaining of Whole-Mount Imaginal Discs. *In* Drosophila Cytogenetics Protocols. Vol. 247. 373-387.

Saito, H., and F. Posas. 2012. Response to hyperosmotic stress. *Genetics*. 192:289-318.

Sandhu, K.S., C. Shi, M. Sjolinder, Z. Zhao, A. Gondor, L. Liu, et al. 2009. Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev.* 23:2598-2603.

Sanyal, A., B.R. Lajoie, G. Jain, and J. Dekker. 2012. The long-range interaction landscape of gene promoters. *Nature*. 489:109-113.

Schoborg, T., and M. Labrador. 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.

Scholz, H., J. Deatrick, A. Klaes, and C. Klambt. 1993. Genetic dissection of pointed, a Drosophila gene encoding two ETS-related proteins. *Genetics*. 135:455-468.

Schwartz, Y.B., D. Linder-Basso, P.V. Kharchenko, M.Y. Tolstorukov, M. Kim, H.B. Li, et al. 2012. Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Res.* 22:2188-2198.

Seong, K.H., D. Li, H. Shimizu, R. Nakamura, and S. Ishii. 2011. Inheritance of stress-induced, ATF-2-dependent epigenetic change. *Cell*. 145:1049-1061.

Sexton, T., E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli. 2012. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*. 148:458-472.

She, W., W. Lin, Y. Zhu, Y. Chen, W. Jin, Y. Yang, et al. 2010. The gypsy insulator of Drosophila melanogaster, together with its binding protein suppressor of Hairy-wing, facilitate high and precise expression of transgenes in Arabidopsis thaliana. *Genetics*. 185:1141-1150.

Sipos, L., J. Mihaly, F. Karch, P. Schedl, J. Gausz, and H. Gyurkovics. 1998. 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.

Smallwood, A., and B. Ren. 2013. Genome organization and long-range regulation of gene expression by enhancers. *Curr. Opin. Cell Biol.* 25:387-394.

Smith, S.T., P. Wickramasinghe, A. Olson, D. Loukinov, L. Lin, J. Deng, et al. 2009. Genome wide ChIP-chip analyses reveal important roles for CTCF in Drosophila genome organization. *Dev. Biol.* 328:518-528.

Spana, C., D.A. Harrison, and V.G. Corces. 1988. The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. *Genes Dev.* 2:1414-1423.

Spencer, W.C., G. Zeller, J.D. Watson, S.R. Henz, K.L. Watkins, R.D. McWhirter, et al. 2010. A Spatial and Temporal Map of C. elegans Gene Expression. *Genome Res.* 2:325-341.

Spivakov, M., J. Akhtar, P. Kheradpour, K. Beal, C. Girardot, G. Koscielny, et al. 2012. Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome Biol.* 13:R49.

Spradling, A.C., D. Stern, A. Beaton, E.J. Rhem, T. Laverty, N. Mozden, S. Misra, and G.M. Rubin. 1999. The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics*. 153:135-177.

Steiner, F.A., P.B. Talbert, S. Kasinathan, R.B. Deal, and S. Henikoff. 2012. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res.* 22:766-777.

Strutt, H., G. Cavalli, and R. Paro. 1997. Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.* 16:3621-3632.

Su, W., S. Porter, S. Kustu, and H. Echols. 1990. DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial glnA promoter. *Proc. Natl. Acad. Sci. U. S. A.* 87:5504-5508.

Suva, M.L., N. Riggi, and B.E. Bernstein. 2013. Epigenetic reprogramming in cancer. *Science*. 339:1567-1570.

Szutorisz, H., N. Dillon, and L. Tora. 2005. The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem. Sci.* 30:593-599.

Tartof, K.D., and S. Henikoff. 1991. Trans-sensing effects from Drosophila to humans. *Cell*. 65:201-203.

Tchurikov, N.A., O.V. Kretova, E.D. Moiseeva, and D.V. Sosin. 2009. Evidence for RNA synthesis in the intergenic region between enhancer and promoter and its inhibition by insulators in Drosophila melanogaster. *Nucleic Acids Res.* 37:111-122.

Thurman, R.E., E. Rynes, R. Humbert, J. Vierstra, M.T. Maurano, E. Haugen, et al. 2012. The accessible chromatin landscape of the human genome. *Nature*. 489:75-82.

Tollervey, J.R., and V.V. Lunyak. 2012. Epigenetics: judge, jury and executioner of stem cell fate. *Epigenetics*. 7:823-840.

Udvardy, A., E. Maine, and P. Schedl. 1985. The 87A7 chromomere : Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J. Mol. Biol.* 185:341-358.

Van Bortle, K., and V.G. Corces. 2012. Nuclear Organization and Genome Function. *Annu. Rev. Cell Dev. Biol.* 28:163-187.

Van Bortle, K., E. Ramos, N. Takenaka, J. Yang, J.E. Wahi, and V.G. Corces. 2012. Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res.* 22:2176-2187.

Veraksa, A., A. Bauer, and S. Artavanis-Tsakonas. 2005. Analyzing protein complexes in Drosophila with tandem affinity purification–mass spectrometry. *Dev. Dyn.* 232:827-834.

Vermaak, D., and H.S. Malik. 2009. Multiple Roles for Heterochromatin Protein 1 Genes in Drosophila. *Annu. Rev. Genet.* 43:467-492.

Vernot, B., A.B. Stergachis, M.T. Maurano, J. Vierstra, S. Neph, R.E. Thurman, J.A. Stamatoyannopoulos, and J.M. Akey. 2012. Personal and population genomics of human regulatory variation. *Genome Res.* 22:1689-1697.

Vorobyeva, N.E., M.U. Mazina, A.K. Golovnin; D.V. Kopytova, D.Y. Gurskiy, E.N. Nabirochkina, et al. 2013. Insulator protein Su(Hw) recruits SAGA and Brahma complexes and constitutes part of Origin Recognition Complex-binding sites in the Drosophila genome. *Nucleic Acids Res.* 41:5717-5730

Vostrov, A.A., and W.W. Quitschke. 1997. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. *J. Biol. Chem.* 272:33353-33359.

Vrailas-Mortimer, A., T. del Rivero, S. Mukherjee, S. Nag, A. Gaitanidis, D. Kadas, C. Consoulas, A. Duttaroy, and S. Sanyal. 2011. A muscle-specific p38 MAPK/Mef2/MnSOD pathway regulates stress, motor function, and life span in Drosophila. *Dev. Cell*. 21:783-795.

Wallace, H.A., M.P. Plata, H.J. Kang, M. Ross, and M. Labrador. 2010. Chromatin insulators specifically associate with different levels of higher-order chromatin organization in Drosophila. *Chromosoma*. 119:177-194.

Wallace, J.A., and G. Felsenfeld. 2007. We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* 17:400-407.

Wang, D., I. Garcia-Bassets, C. Benner, W. Li, X. Su, Y. Zhou, et al. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*. 474:390-394.

Wang, H., M.T. Maurano, H. Qu, K.E. Varley, J. Gertz, F. Pauli, et al. 2012. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 22:1680-1688.

Watanabe, A., Y. Yamada, and S. Yamanaka. 2013. Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368:20120292.

Waterhouse, A.M., J.B. Procter, D.M.A. Martin, M. Clamp, and G.J. Barton. 2009. Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 25:1189-1191.

Wedel, A., D.S. Weiss, D. Popham, P. Droge, and S. Kustu. 1990. A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science*. 248:486-490.

Weisbrod, S. 1982. Active chromatin. *Nature*. 297:289-295.

West, A.G., M. Gaszner, and G. Felsenfeld. 2002. Insulators: many functions, many mechanisms. *Genes Dev.* 16:271-288.

Whitfield, T., J. Wang, P. Collins, E.C. Partridge, S. Aldred, N. Trinklein, R. Myers, and Z. Weng. 2012. Functional analysis of transcription factor binding sites in human promoters. *Genome Biol.* 13:R50.

Whitfield, W.G., M.A. Chaplin, K. Oegema, H. Parry, and D.M. Glover. 1995. The 190 kDa centrosome-associated protein of Drosophila melanogaster contains four zinc finger motifs and binds to specific sites on polytene chromosomes. *J. Cell Sci.* 108:3377-3387.

Whitfield, W.G., S.E. Millar, H. Saumweber, M. Frasch, and D.M. Glover. 1988. Cloning of a gene encoding an antigen associated with the centrosome in Drosophila. *J. Cell Sci.* 89:467-480.

Williams, B.R., J.R. Bateman, N.D. Novikov, and C.T. Wu. 2007. Disruption of topoisomerase II perturbs pairing in drosophila cell culture. *Genetics*. 177:31-46.

Williams, J.A., J.B. Bell, and S.B. Carroll. 1991. Control of Drosophila wing and haltere development by the nuclear vestigial gene product. *Genes Dev.* 5:2481-2495.

Williams, J.A., S.W. Paddock, K. Vorwerk, and S.B. Carroll. 1994. Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. *Nature*. 368:299-305.

Wood, A.M., K. Van Bortle, E. Ramos, N. Takenaka, M. Rohrbaugh, B.C. Jones, K.C. Jones, and V.G. Corces. 2011. Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. *Mol. Cell*. 44:29-38.

Wu, C.-H., Y. Yamaguchi, L.R. Benjamin, M. Horvat-Gordon, J. Washinsky, E. Enerly, et al. 2003. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev.* 17:1402-1414.

Wu, C.T., R.S. Jones, P.F. Lasko, and W.M. Gelbart. 1989. Homeosis and the interaction of zeste and white in Drosophila. *Mol. Gen. Genet.* 218:559-564.

Wu, C.T., and J.R. Morris. 1999. Transvection and other homology effects. *Curr. Opin. Genet. Dev.* 9:237-246.

Xiao, T., J. Wallace, and G. Felsenfeld. 2011. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol. Cell. Biol.* 31:2174-2183.

Xu, Q., M. Li, J. Adams, and H.N. Cai. 2004. Nuclear location of a chromatin insulator in Drosophila melanogaster. *J. Cell Sci.* 117:1025-1032.

Yang, J., and V.G. Corces. 2012. Insulators, long-range interactions, and genome function. *Curr. Opin. Genet. Dev.* 22:86-92.

Yang, J., E. Ramos, and V.G. Corces. 2012. The BEAF-32 insulator coordinates genome organization and function during the evolution of Drosophila species. *Genome Res.* 22:2199-2207.

Yang, J., E. Sung, P.G Donlin-Asp, and V.G. Corces. 2013. A subset of Drosophila Myc sites remain associated with mitotic chromosomes colocalized with insulator proteins. *Nat. Commun.* 4:1464.

Yang, Y., S. Singer, and Z. Liu. 2011. Evaluation and comparison of the insulation efficiency of three enhancer-blocking insulators in plants. *Plant Cell, Tissue and Organ Culture (PCTOC)*. 105:405-414.

Yujiri, T., G.R. Fanger, T.P. Garrington, T.K. Schlesinger, S. Gibson, and G.L. Johnson. 1999. MEK kinase 1 (MEKK1) transduces c-Jun NH2-terminal kinase activation in response to changes in the microtubule cytoskeleton. *J. Biol. Chem.* 274:12605-12610.

Zhao, K., C.M. Hart, and U.K. Laemmli. 1995. Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell*. 81:879-889.

Zhong, S., P. Salomoni, and P.P. Pandolfi. 2000. The transcriptional role of PML and the nuclear body. *Nat. Cell Biol.* 2:E85-90.

Zhu, X., J. Ling, L. Zhang, W. Pi, M. Wu, and D. Tuan. 2007. A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Res.* 35:5532-5544.

Zimmerman, J.E., Q.T. Bui, H. Liu, and N.M. Bonini. 2000. Molecular genetic analysis of Drosophila eyes absent mutants reveals an eye enhancer element. *Genetics*. 154:237-246.

APPENDIX

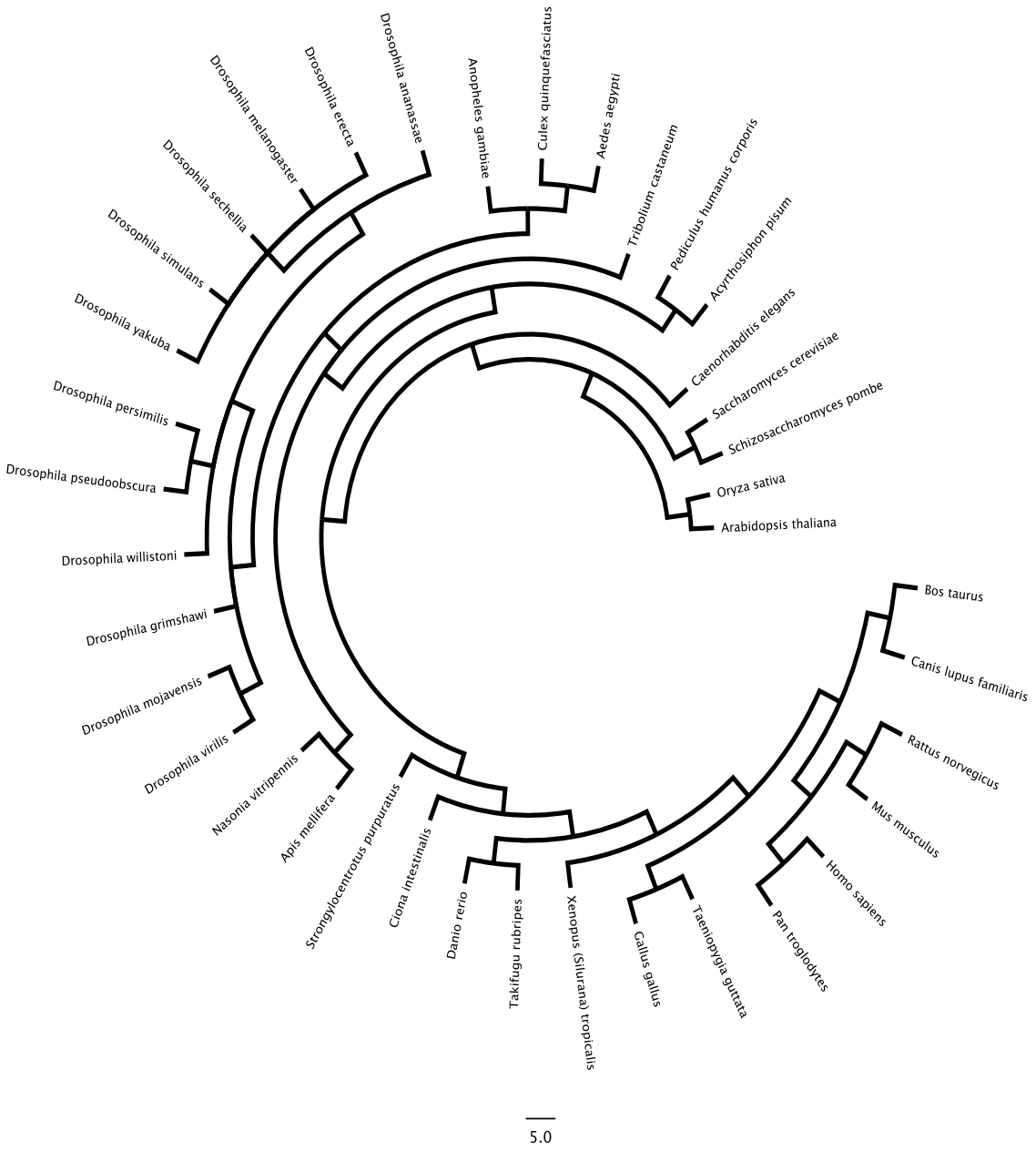

Figure A1. Taxanomic Distribution Of Species Used For Chapter I Analysis.
Sequences for tree construction were obtained from the Taxonomy Browser at NCBI and constructed using FigTree v1.2.2.

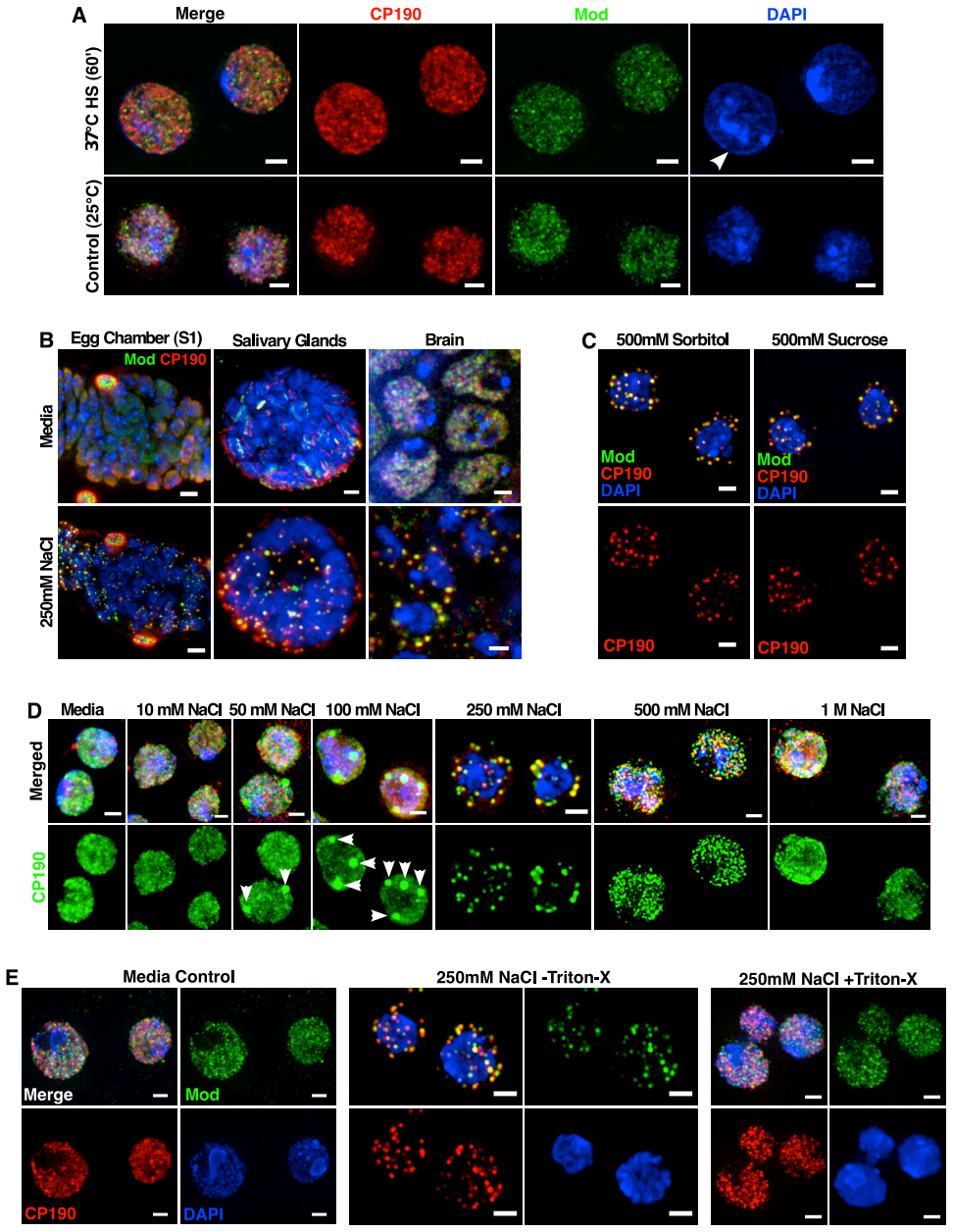

Figure A2. Confirmation of Osmostress As Inducer of Insulator Body Formation.
S2 cells heat shocked at 37°C for 60 minutes and untreated controls labeled with CP190 and Mod(mdg4). Arrowhead denotes chromatin rearrangement in HS nuclei (A). Egg chambers, salivary glands (polytene nuclei) and brains treated with or without 250 mM NaCl and labeled with CP190 and Mod(mdg4) (B). S2 cells treated with 500 mM Sorbitol or Sucrose and labeled with CP190 and Mod(mdg4) (C). S2 cells treated with an increasing gradient of NaCl and labeled with CP190. Arrowheads mark CP190 bodies. (D). S2 cells treated with or without 0.2% Triton-X followed by addition of 250 mM NaCl and labeled with CP190 and Mod(mdg4) (E). Scale bars are 2 μm in (A) and (C)-(E); in (B) egg chambers are 4 μm, salivary gland polytene nuclei are 3 μm and brains are 2 μm.

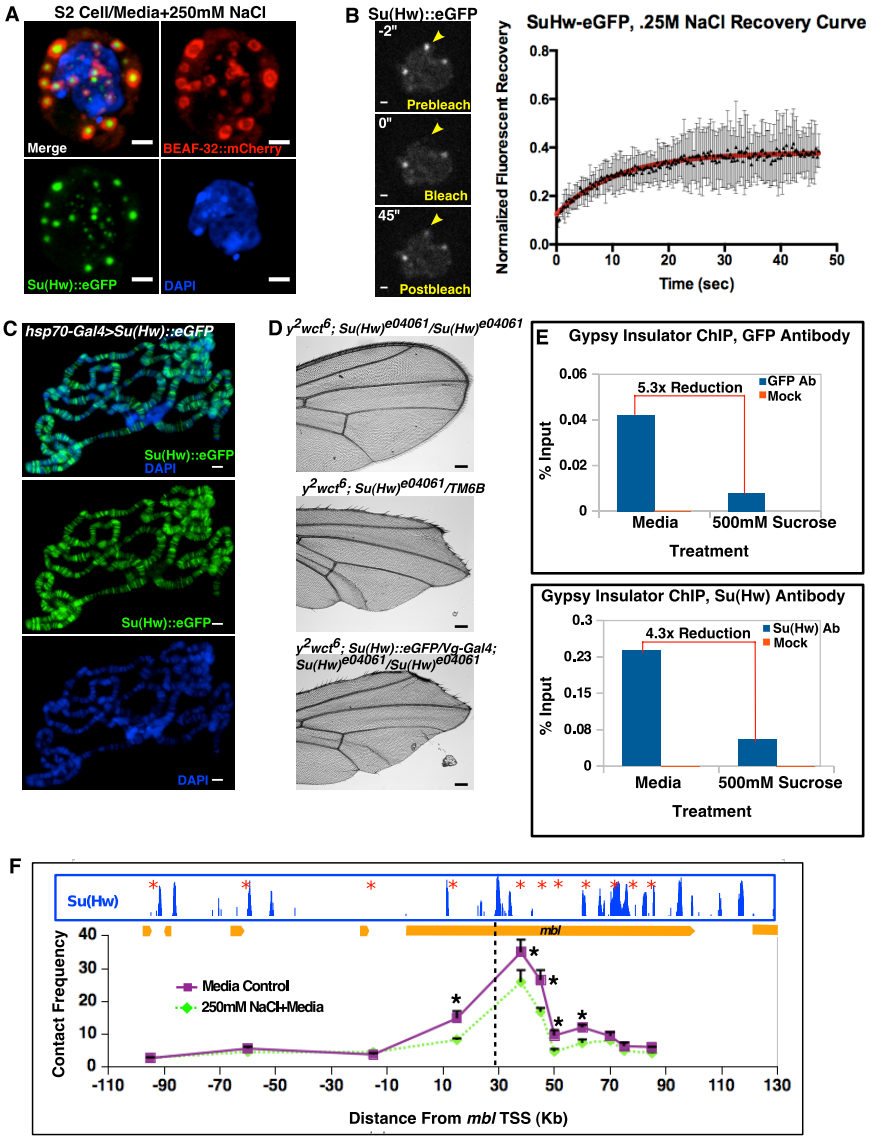

Figure A3. Antibody Check, Tagged Construct Confirmation, FRAP and 3C.
S2 cell transfected with BEAF-32::mCherry and Su(Hw)::eGFP and stressed with 250 mM NaCl (A). FRAP analysis of stationary bodies (error bars represent S.E.M of 3 bodies from the same cell) (B). Polytene chromosome spread from a Su(Hw)::eGFP-expressing 3rd instar salivary gland (C). Expression of this construct in a hypomorphic background of *su(Hw)* (*su(Hw)*^{e04061}) restores *gypsy* insulator function in *ct*^{6} (D). Chromatin immunoprecipitation using α-GFP from S2 cells expressing this construct treated with or without 500 mM sucrose shows a similar fold reduction at *gypsy* as the endogenous Su(Hw) protein (E). Chromosome conformation capture (3C) analysis of looping contacts throughout the *muscleblind (mbl)* locus in S2 cells following treatment with or without 250 mM NaCl for 5 minutes. The black dashed line marks the position of the 3C anchor primer with test primers noted by red asterisks. Blue peaks represent Su(Hw) binding sites. Statistically significant reductions are marked with black asterisks (Student's paired t-test, p=0.05. Error bars represent S.E.M) (F). Scale bars are 2 μm in (A) & (B), 3μm in (C) and 100 μm in (D).

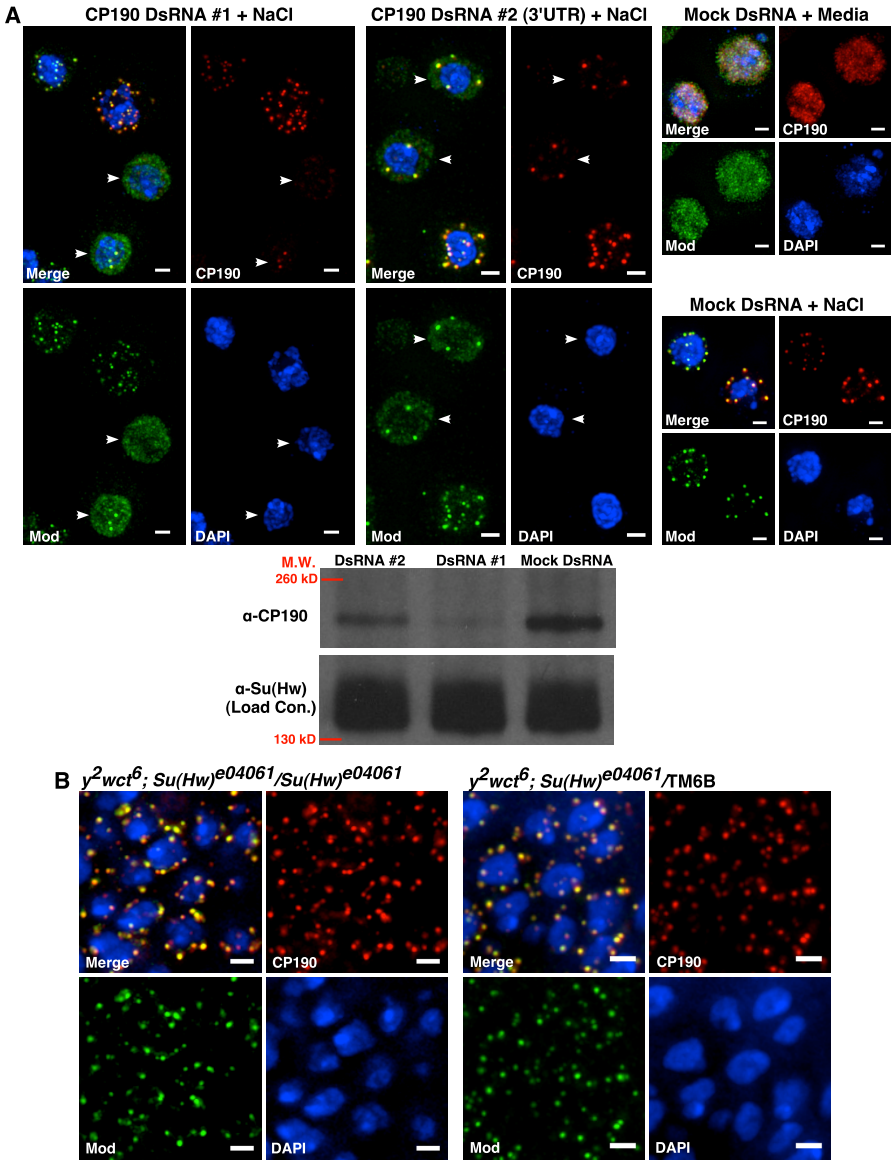

Figure A4. Effect of CP190 DsRNA-Knockdown in S2 Cells Following Osmostress.
S2 cells soaked with two different CP190 DsRNA constructs (#1 and #2) and treated with 250 mM NaCl and labeled with CP190 and Mod(mdg4). Arrowheads mark nuclei with significantly reduced levels of CP190. Mock DsRNA controls treated with or without 250 mM NaCl are also shown. Western blot of lysates from knockdown lines compared to mock controls verifies CP190 reduction for both DsRNA constructs (A). Wing Discs from Su(Hw)^{e04061} homozygotes (left panel) and balanced controls (right panel) treated with 250 mM NaCl and labeled with CP190 and Mod(mdg4) (B). Scale bars are 2 μm.

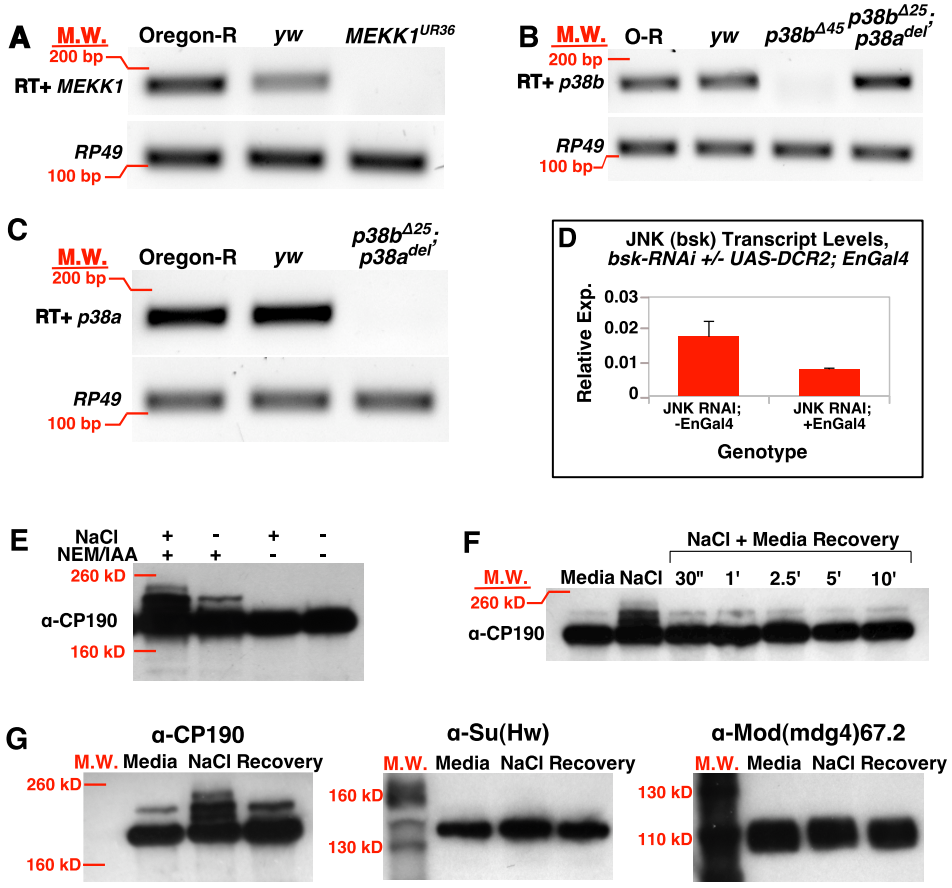

Figure A5. Confirmation Of Mutant Alleles For *dMEKK1* and *p38a*/*p38b*.
RT-PCR from the indicated genotype confirms the absence of transcript in *dMEKK1*^{UR36} (A), *p38b*$^{\Delta 45}$ (B) and *p38a*^{del} (C) mutants (Rp49 indicates loading control). qPCR analysis of *JNK* transcript levels in wing discs expressing *bsk-RNAi* under control of an *engrailed-Gal4; UAS-DCR2* driver which is only active in the posterior compartment of the wing disc (Error bars represent the S.E.M of 3 biological replicates) (D). Western Blot of CP190 from S2 cell lysates treated with or without 250 mM NaCl and in either the presence or absence of the SUMO isopeptidase inhibitors NEM and IAA (E). Western Blot of CP190 from S2 cell lysates treated with 250 mM NaCl and allowed to recover in isotonic media for the indicated period of time before lysing in the presence of NEM and IAA (F). Western Blot of CP190, Su(Hw) and Mod(mdg4) from S2 cells under conditions of media, osmostress or recovery in fresh media (NEM and IAA included in lysis buffer) (G). Molecular weight (M.W.) markers are indicated.

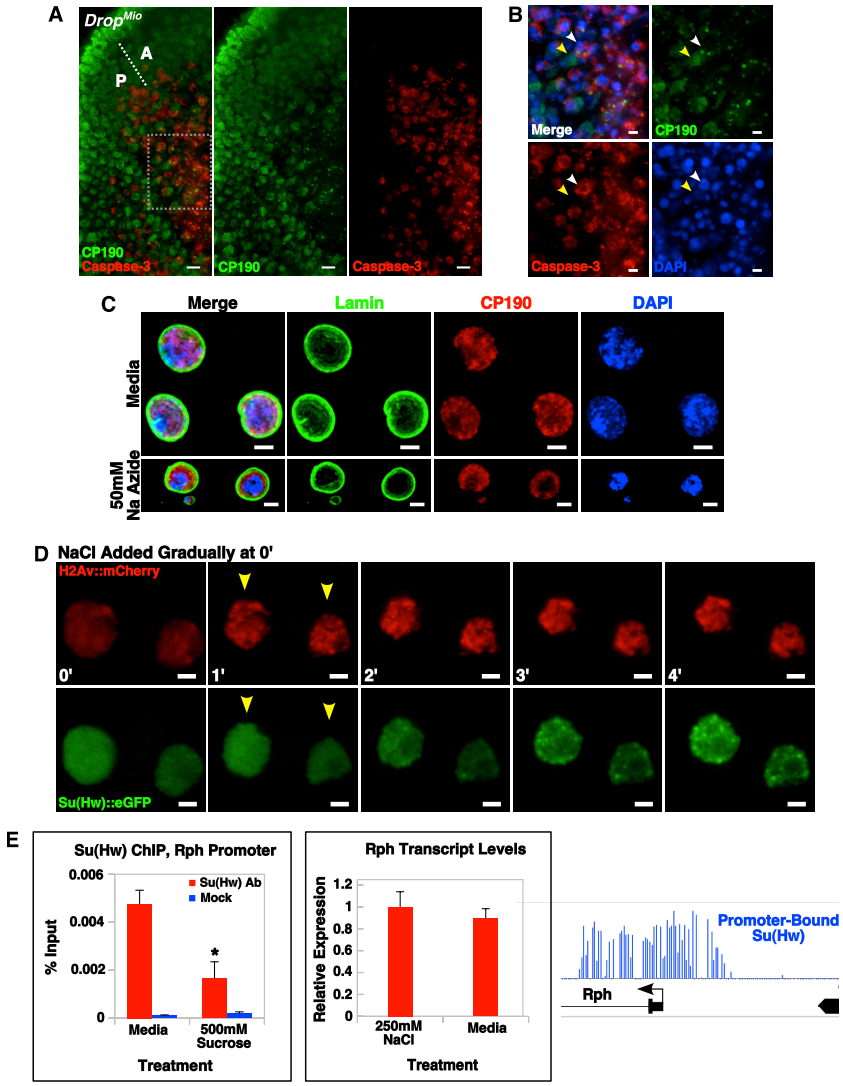

Figure A6. Insulator Bodies Are Also Evident In Tissues Undergoing Cell Death, marked by cleaved caspase-3 in eye discs derived from *Drop*^{Mio} larvae (A/P marks anterior-posterior orientation of the tissue) (A). A close up the gray boxed region from (A) shows diffuse staining for CP190, lack of cleaved caspase-3 and decondensed DAPI in healthy cells (yellow arrowhead) and formation of CP190 bodies in cells marked with caspase and condensed DAPI (white arrowhead) (B). S2 cells treated with or without low concentrations of Na Azide to induce chromatin compaction independently of osmostress and labeled with lamin and CP190 (C). Time-lapse imaging of a S2 cell transfected with H2Av::mCherry to mark chromatin and Su(Hw)::eGFP stressed with 250 mM NaCl at time 0 with frames taken at 1 minute intervals. First evidence of chromatin compaction is indicated by yellow arrowheads (D). ChIP of Su(Hw) from the promoter of *Rph* following osmostress and the expression levels of *Rabphilin (Rph).* Asterisk marks statistically significant reductions of Su(Hw) but expression levels are not significantly altered (Student's paired t-test, p=0.05. Error bars represent S.E.M of 3 biological replicates) (E). Scale bars are 5 μm in (A) and 2 μm in (B)-(D).

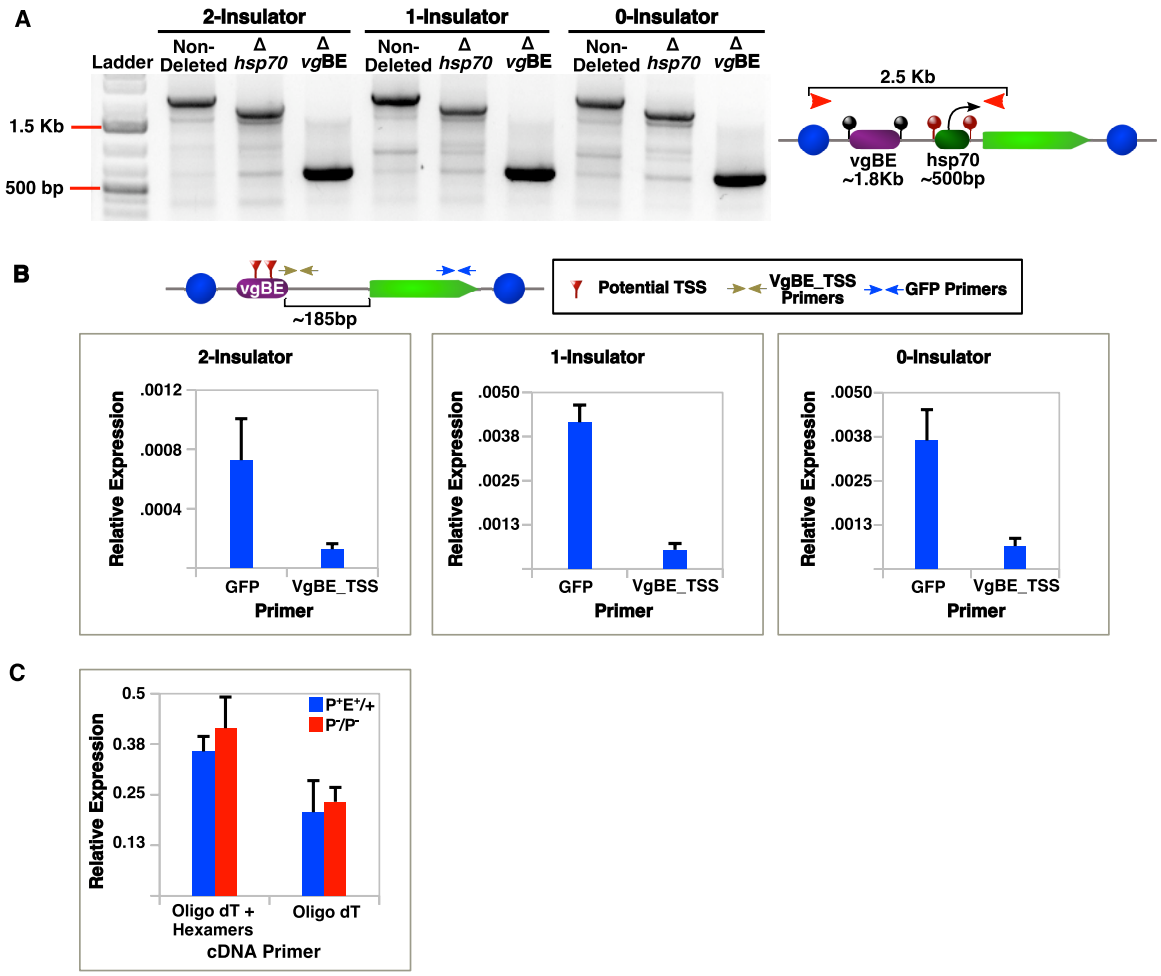

Figure A7. Confirmation Of *vg*be & *Hsp70* Promoter Deletion.
Schematic (right) shows position of PCR primers (red arrowheads) and the size of each respective element (A). Schematic showing potential position of cryptic transcription start sites (TSSs) in the *vg*BE (red triangles), the distance between the *vg*BE and *eGFP* start codon following promoter removal and the position of test primers (tan and blue arrows) used for qPCR. Graphs showing transcript levels based on these primers for a single "promoterless" (P^-/P^-) representative from 2-insulator, 1-insulator and 0-insulator lines are shown (B). qPCR analysis of *eGFP* transcript levels reverse-transcribed using either a mixture of random hexamers+oligo dT primers or oligo dT primers alone for cDNA synthesis for the indicated 2-insulator genotypes (C). All error bars represent the standard error of the mean (S.E.M.).

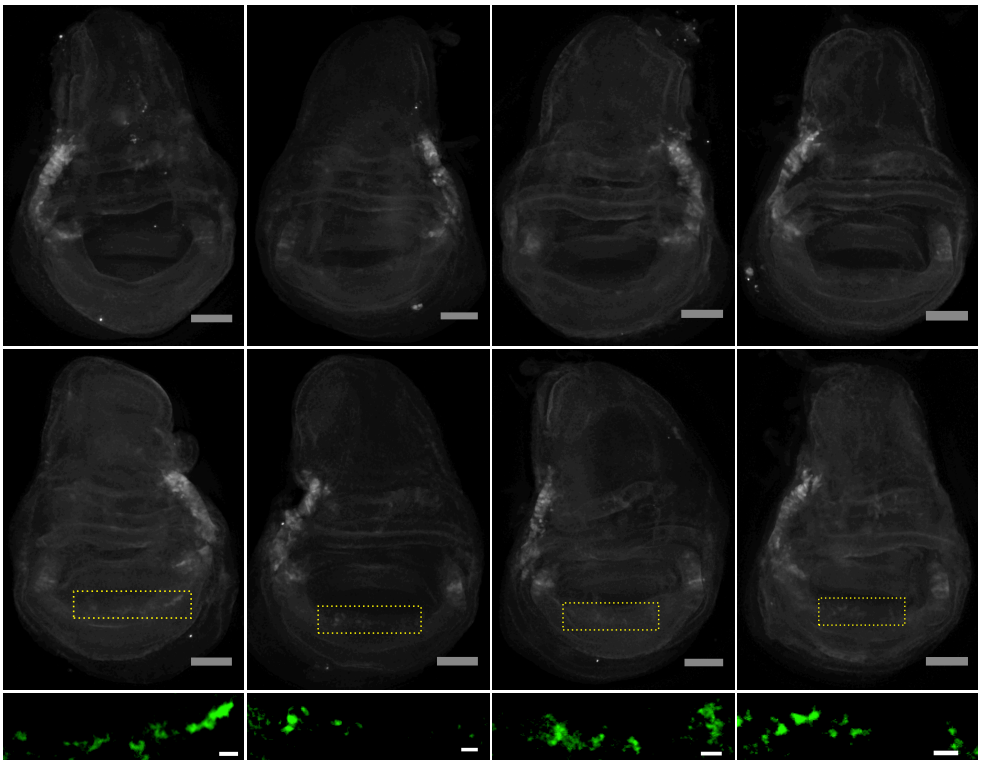

Figure A8. 2-insulator Trans-Heterozygote (P^-/E^-) GFP Protein Expression Pattern.
Shown is a collection of 8 different wing discs from a representative 2-insulator P^-/E^- line (grayscale images) with magnified panels below showing a closeup of the yellow boxed region false-colored in green. All grayscale levels were normalized equally, while min/max corrections to magnified panels were performed separately for each disc using ImageJ. Scalebars in grayscale panels are 50 μm and 10 μm in magnified panels.

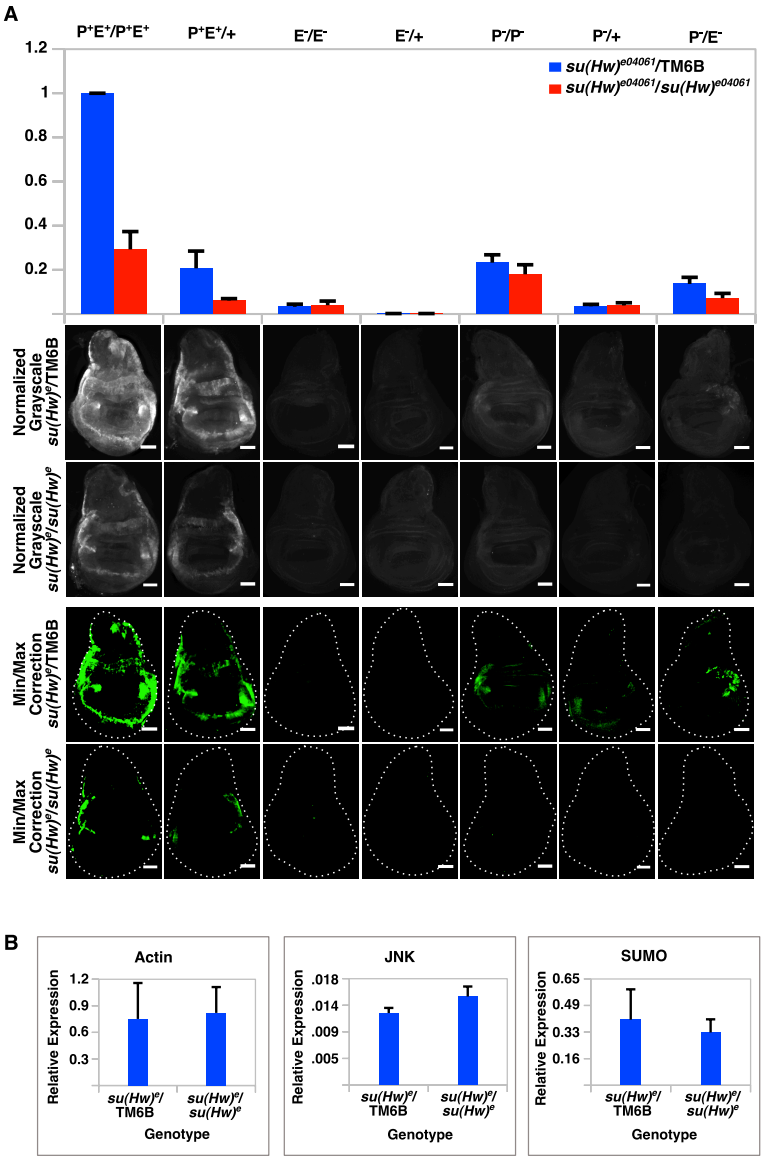

Figure A9. Effects of *su(Hw)* Mutations on 2-Insulator GFP Expression.
QPCR analysis (top graph) and immunostaining (bottom panels) of wing discs from all seven 2-insulator genotypes in a TM6B-balanced or *su(Hw)*^{e04061} mutant background. For microscopy, all grayscale images were normalized to the balanced P^+E^+/P^+E^+, while minimum/maximum level corrections were applied equally to both balanced and *su(Hw)*^{e04061} backgrounds based on each individual *reporter genotype.* Thus, each of the 14 genotypes irrespective of genetic background are directly comparable in the normalized grayscale panels, while only a single genotype (such as P^-/P^-) is directly comparable between backgrounds in the minimum/maximum corrected panels (A). qPCR analysis of *SUMO, Actin* and *JNK (bsk)* expression levels in a balanced or *su(Hw)*^{e04061} background (B). Error bars represent standard error of the mean (S.E.M) and scalebars are 50 µm.

Table A1. Chromosome And Genomic Coordinates For 2-Insulator Lines.

LINE	CHRM.	GENOMIC COORD.	INSERTION TYPE	HOTSPOT[b]	CHROMATIN TYPE[d]	INSULATOR PROTEINS
PGP-13B	3	3L: 7,127,316	Promoter (*melted*)	*	Blue	CP190, BEAF, Mod, CTCF, GAF, Su(Hw)
PGP-20	2	-[c]	Transposon (*Invader*)	-	-	-
PGP-22A	3	3L: 13,221,600	Promoter (*caps*)	*	Blue	CP190, BEAF, Mod, CTCF, GAF
PGP-23A	3	-[c]	Transposon (*F*)	-	-	-
PGP-25A	3	3R: 22,360,900	Promoter (*CG6490*)	*	Blue/Red	CP190, Mod, GAF
PGP-28[a]	2	2L: 12,018,983	Exon (*CG6734*)	-	Yellow	-
PGP-33A[a]	2	2L: 7,576,480	Promoter (*RapGAP1*)	*	Red	CP190, Mod, GAF, SuHw
PGP-39A[a]	2	2L: 8,989,200	Promoter (r*ost*)	-	Red	-
PGP-50A[a]	2	2L: 9,758,467	5' UTR (*zf30C*)	*	Black	CP190, BEAF, Mod, CTCF, GAF
PGP-104A1[a]	2	2L: 20,163,757	Transposon (*Invader*)	-	-	-
PGP-146C2	3	3R: 8,326,193	Intergenic	-	Black	-

[a]Used for Su(Hw) mutant analysis.
[b]Transposon hotspot if more than 5 insertions for other P-element or Pbacs were found at this position in FlyBase.
[c]Not enough flanking sequence recovered to accurately predict insertion position.
[d]From: Filion et al (2010). *Systematic protein location mapping reveals five principal chromatin types in* Drosophila *cells.* Cell *143,* 212-224.

Table A2. Chromosome And Genomic Coordinates For 1-Insulator Lines.

LINE	CHRM.	GENOMIC COORD.	INSERTION TYPE	HOTSPOT[a]	CHROMATIN TYPE[c]	INSULATOR PROTEINS
gi-ve 2	3	3L: 5,177,560	Intron (*shep)*	*	Red	GAF
gi-ve 2A	3	3R: 24,816,510	Intergenic	*	Black	GAF
gi-ve 9A	2	-[b]	Transposon (*gypsy*)	-		-
gi-ve 20	3	-[b]	Transposon (*Diver*)	-		-
gi-ve 29	3	3R: 7,392,923	Promoter (*sea/fabp*)	*	Red	CTCF, GAF, Su(Hw)
gi-ve 31	2	2R: 6,762,150	Promoter (*CG30015*)	*	Red/Yellow	CP190, BEAF, CTCF, GAF
gi-ve 40	3	3L: 13,470,406	Promoter (s*tv)*	*	Black	GAF
gi-ve 45	3	-[b]	Transposon (*Doc)*	-		-
gi-ve 7D	X	X: 13,072,760	Intron (*CG34411*)	-	Black	BEAF, CP190, CTCF

[a]Transposon hotspot if more than 5 insertions for other P-element or Pbacs were found at this position in FlyBase.
[b]Not enough flanking sequence recovered to accurately predict insertion position.
[c]From: Filion et al (2010). *Systematic protein location mapping reveals five principal chromatin types in* Drosophila *cells.* Cell *143,* 212-224.

Table A3. Chromosome And Genomic Coordinates For 0-Insulator Lines.

LINE	CHRM.	GENOMIC COORD.	INSERTION TYPE	HOTSPOT[a]	CHROMATIN TYPE[b]	INSULATOR PROTEINS
PNG-10	X	X: 19,671,651	Exon (*vfl*)	-	Blue	-
PNG-11	X	X: 17,793,173	Promoter (*CG32495*)	-	Yellow	CP190, BEAF, CTCF, Mod
PNG-20	2	2R: 17,557,761	5' UTR (*NC2alpha*)	-	Yellow	CP190, BEAF
PNG-44	3	3L: 6,957,768	Promoter (s*gl*)	*	Yellow	CP190, BEAF, GAF
PNG-46	3	3R: 24,953,620	Transposon (*Opus*)	-		-
PNG-50	3	3R: 16,886,068	Promoter (*Mvl*)	*	Yellow	CP190, BEAF, CTCF, Mod, GAF
PNG-1C	2	2L: 9,616,828	5' UTR (*GlcAT-S*)	*	Red	CP190, BEAF, CTCF, Mod, GAF

[a]Transposon hotspot if more than 5 insertions for other P-element or Pbacs were found at this position using FlyBase.
[b]From: Filion et al (2010). *Systematic protein location mapping reveals five principal chromatin types in* Drosophila *cells.* Cell *143,* 212-224.

VITA

Todd Andrew Schoborg grew up on a tobacco farm near the banks of the Licking River in Berry, Kentucky and graduated from Pendleton County High School in 2004. Though he didn't realize the significance at the time, Pendleton County had also been the childhood home of Phillip A. Sharp, a molecular biologist who was awarded a share of the 1993 Nobel Prize in Physiology or Medicine for his role in identifying introns in eukaryotic genes. He was drawn to Murray State University in Western Kentucky for his undergraduate studies, enrolling in the Fall of 2004 with every intention of becoming a Wildlife or Fisheries Biologist. Shortly after completing an introductory course in molecular cell biology, he switched his major to molecular biology and saw his first action as a researcher working in an evolutionary ecology lab under the auspices of Dr. Howard Whiteman as a National Science Foundation Biomathematics fellow. In 2008 he graduated from Murray State with honors and began his graduate work in the Department of Biochemistry and Cellular and Molecular Biology at the University of Tennessee, Knoxville and joined the lab of Dr. Mariano Labrador in 2009. In 2010 he was awarded a Graduate Research Fellowship from the National Science Foundation, which was later followed by the Science Alliance Award for Graduate Student Excellence in 2013.