12-2010

The role of the Suppressor of Hairy-wing insulator protein in chromatin organization and expression of transposable elements in Drosophila melanogaster

Heather Anne Wallace
University of Tennessee - Knoxville, hwallac3@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Heather Anne Wallace entitled "The role of the Suppressor of Hairy-wing insulator protein in chromatin organization and expression of transposable elements in Drosophila melanogaster." 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, Ranjan Ganguly, Todd Reynolds, Albrecht von Arnim

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

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

To the Graduate Council:

I am submitting herewith a dissertation written by Heather Anne Wallace entitled “The role of the Suppressor of Hairy-wing insulator protein in chromatin organization and expression of transposable elements in *Drosophila melanogaster.*” 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

Ranjan Ganguly

Albrecht von Arnim

Todd Reynolds

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the
Graduate School

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

The role of the Suppressor of Hairy-wing insulator protein in chromatin organization and expression of transposable elements in *Drosophila melanogaster*

A Dissertation Presented for

the Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Heather Anne Wallace

December 2010

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Mariano Labrador, for his guidance, patience, and support, without which this work would not have been possible. His enthusiasm and passion for science are inspiring, and were a big factor in my decision to join his lab. I am also grateful to my committee members, Dr. Bruce McKee, Dr. Ranjan Ganguly, Dr. Albrecht von Arnim, and Dr. Todd Reynolds for their time and support, as well as their valuable suggestions and critical evaluation of my research work.

I would like to thank all our lab members, both past and present, who have been like a family to me during my time at UT. I'm especially grateful to Joon, who has been a great friend and a source of constant support. His encouragement, as well as our many discussions, have helped me through some of the most difficult times during my PhD research. I also want to thank Piedad for our many wonderful conversations and for helping to create a lively and energetic atmosphere in the lab. Thanks also to Cherie, Shaofei, Todd, Srilalitha, and Misty for their friendship and the many experiences we shared over the past several years.

I would like to thank my family, all of whom supported me in numerous ways during my lengthy education. Without them, I could not have made it this far. I owe special thanks to my Mother, who has always supported me in whatever path I chose to pursue, and who has worked hard to help keep me on track when my motivation faltered. Finally, I want to thank all my friends, especially Suzanne, who have been so supportive and have shared my ups and downs throughout the last six years.

ABSTRACT

Chromatin insulators are required for proper temporal and spatial expression of genes in metazoans. Insulators are thought to play an important role in the regulation of gene expression through the formation of higher-order chromatin structures. One of the best characterized insulators is the *Drosophila gypsy* insulator, which is located in the *gypsy* retrovirus. Several proteins are required for *gypsy* insulator function, including Su(Hw), Mod(mdg4), and CP190. In addition to the *gypsy* insulator, these proteins are located throughout the genome at sites which are thought to correspond to endogenous insulators. Analysis of the distribution of insulator proteins across a region of chromosome 2R in *Drosophila* polytene chromosomes shows that Su(Hw) is found in three structures differentially associated with insulator proteins: bands, interbands and domains of coexpressed genes. Bands are formed by condensation of chromatin within genes containing one or more Su(Hw) binding sites, while Su(Hw) sites in interbands appear to form structures normally associated with open chromatin. Bands characterized by the lack of CP190 and BEAF-32 insulator proteins are formed by clusters of coexpressed genes, and these bands correlate with the distribution of specific chromatin marks. Conservation of the band interband pattern, as well as the distribution of insulator proteins in nurse cells, suggests that this organization may represent the basic organization of interphasic chromosomes. We also show that, in addition to the *gypsy* insulator, sequence analysis predicts the presence of Su(Hw) binding sites within a number of transposable elements. Su(Hw) binds to predicted sites within *gtwin* and *jockey,* which possesses enhancer-blocking activity. Su(Hw) affects the tissue-specific expression of transposable elements, although this effect is unrelated to

the presence of Su(Hw) binding sites within the element or control of the elements via the piRNA pathway. Additionally, the effect of Su(Hw) on transposable element expression often differs from that of Mod(mdg4). Taken together, these results suggest that insulator proteins associate specifically with, and may help to define, various levels of chromatin organization on polytene chromosomes. Also, *gypsy* insulator proteins may influence the expression of transposable elements in a way that does not depend on Su(Hw) binding sites within the elements themselves.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Gene Regulation and Chromatin Organization

The ability of organisms to precisely regulate gene expression is central to their development. Proper temporal and spatial expression of higher eukaryotic genes involves activation of transcription at the appropriate developmental stage and in the appropriate tissue. Gene regulation is established through the activity of *cis*-regulatory elements including proximal promoters, enhancers, repressors and silencers (Markstein and Levine, 2002; Ptashne and Gann, 1997; Walhout, 2006). Genes are often not located in close proximity to these *cis*-regulatory elements, which may affect their expression. Enhancers, for example, can act on promoters in a manner independent of direction and distance, sometimes acting over distances as large as 50 kb (Kellum and Schedl, 1992). Also, it is not uncommon for genes to be located in an environment surrounded by regions of highly condensed chromatin, which may spread into neighboring regions. For example, heterochromatin propagates by methylation of histone H3 lysine 9 along the chromatin fiber and can silence the expression of nearby genes (Grewal and Moazed, 2003). Despite these influences from their surroundings, genes are precisely regulated to be expressed in specific tissues and at particular times during development. The action of regulatory elements, such as enhancers or promoters, found in the linear DNA sequence is not sufficient in itself to explain the complexity of the regulation of gene expression. It is becoming increasingly clear that

higher-order chromatin structure and epigenetic regulation through chromatin modifications play important regulatory roles in gene expression.

DNA in eukaryotic genomes is packaged into chromatin, the basic unit of which is the nucleosome. Nucleosomes consist of a core histone complex consisting of two histone H2A-H2B dimers and an H3-H4 tetramer (Kornberg and Lorch, 1999). Approximately 146 base pairs of DNA are wrapped around a histone octamer. Repetition of the nucleosome units forms chromatin fibers in which the nucleosomes are arranged as a linear array along the DNA polymer, giving the appearance of 'beads on a string' (Rando and Ahmad, 2007). These nucleosomal arrays can then be further compacted into higher-order chromatin structures. Packaging of the chromatin fiber in this manner creates a repressive chromatin environment that is inaccessible to DNA binding proteins that regulate transcription (Croston and Kadonaga, 1993). Remodeling of chromatin structure can occur at the level of modifications of the amino-terminal histone tails, which protrude from the nucleosome core. These covalent modifications, which include acetylation and methylation, are produced by histone acetyltransferases or histone methylases, respectively, and alter the accessibility of the chromatin (Luo and Dean, 1999). For example, acetylation of lysine residues is thought to neutralize the negative charge of DNA, thereby decreasing the affinity of the histones for the DNA (Hong et al., 1993). Subsequent recruitment of ATP-dependent chromatin remodeling complexes results in disruption of the interactions between DNA and histones, altering nucleosomes in such a way that facilitates binding of transcription factors and assembly of the transcriptional machinery (Strahl and Allis, 2000). Although it is evident that chromatin structure and epigenetic regulation via chromatin modifications are important

factors in transcriptional control, it is becoming increasingly apparent that the spatial organization of chromosomes in the nucleus may also play an important regulatory role in the process of gene expression (Dekker, 2008).

Chromatin is thought to be organized in the nucleus in a non-random manner. The most well-known example of this organization is the distinction between regions of heterochromatin, which is generally inaccessible to DNA binding factors and transcriptionally silent, and euchromatin, which is more accessible and transcriptionally active (Grewal and Moazed, 2003). Fluorescence *in situ* hybridization studies, as well as high-resolution light microscopy and electron microscopy, have shown that chromosomes occupy specific regions of the nucleus known as chromosome territories (Cremer et al., 1982; Pinkel et al., 1986; Schardin et al., 1985; Visser et al., 2000). The nuclear periphery is a region traditionally associated with localization of transcriptionally silent heterochromatin, although a number of transcriptionally active genes have also been associated with this territory (Lanctot et al., 2007). For example, during erythroid differentiation, expression of the β-globin locus is initiated at the nuclear periphery just before it relocates to the nuclear interior (Ragoczy et al., 2006). In contrast, the *interferon-γ* locus is located at the nuclear periphery regardless of its transcriptional state (Hewitt et al., 2004). Positioning of chromosomes within chromosome territories is likely to be maintained by attachment to nuclear structures, such as the nuclear envelope or the nuclear matrix (Parada and Misteli, 2002). Attachment of chromatin fibers to the nuclear matrix is mediated by specific sequences known as matrix attachment regions (MARs) and results in the formation of chromatin loop domains (Galande et al., 2007). High salt extraction of chromatin-associated factors allows for

the visualization of DNA loops attached to the nuclear matrix by MARS, sequences which have been shown to possess gene regulatory properties (Fernandez et al., 2001; Forrester et al., 1999). Chromatin loops have also been observed extending out from the chromosome territories during transcriptional activation. A group of genes in mouse erythroid progenitor cells were shown to loop out into transcription factories (Osborne et al., 2004). Chromosome conformation capture (3C) assays indicate that the formation of chromatin loops allows distant regulatory elements to position close to their target genes for transcriptional regulation (Kleinjan and van Heyningen, 2005; Spilianakis and Flavell, 2004; Tolhuis et al., 2002). These long-range inter- and intra-chromosomal interactions may contribute to efficient gene regulation and expression through the establishment of functional chromatin domains, which represent independent units of transcriptional activity (Fraser and Bickmore, 2007; Lanctot et al., 2007).

The partitioning of chromatin into domains is supported by a number of observations. For example, regulatory elements, such as enhancers, may be separated from their target promoter by large distances (Kellum and Schedl, 1992). Yet despite the distance between a gene and its regulatory elements, and despite the potential promiscuity of enhancers, genes are able to maintain proper spatio-temporal gene expression. In addition to the local regulation of gene transcription by *cis*-regulatory elements, the position and ordering of genes within the eukaryotic chromosomes appears to be non-random and plays an essential role in the coordinated expression of adjacent genes during development (Hurst et al., 2004; Kosak and Groudine, 2004; Sproul et al., 2005). Analysis of transcriptome datasets has revealed that genes in many eukaryotic organisms including *Drosophila*, *Caenorhabditis elegans* and humans are

organized into clusters of coexpressed genes that tend to be transcribed coordinately at specific times throughout development and/or the cell cycle, regardless of whether they are functionally related (Boutanaev et al., 2002; Caron et al., 2001; Cohen et al., 2000; Lee and Sonnhammer, 2003; Lercher et al., 2003; Spellman and Rubin, 2002). A number of well characterized chromatin domains, which influence the coordinated transcription of several neighboring genes, have been described in different species (Talbert and Henikoff, 2006; Tolhuis et al., 2006; Tolhuis et al., 2002). More importantly, a recent study demonstrated that half of the *Drosophila* genome consists of evolutionarily conserved multi-gene chromatin domains defined by the binding of specific chromatin proteins (de Wit et al., 2008).

The existence of independent domains of gene expression implies the existence of regulatory elements that must be able to contribute to and maintain the functional independence of the domains as well as prevent encroachment of differential transcriptional states onto each other (Capelson and Corces, 2004). The underlying molecular mechanisms mediating the organization of chromatin into independent domains are poorly understood, but it has been hypothesized that the partitioning of genomes into functional multi-gene chromatin may be established at the level of higher-order chromatin structure by chromatin insulators (Capelson and Corces, 2004; Kuhn and Geyer, 2003; Labrador and Corces, 2002; West et al., 2002). Insulators are DNA sequences bound by proteins that are defined by two functional characteristics: they can block communication between enhancers and promoters when located between them and they prevent heterochromatin from spreading along the chromatin fiber (Chung et al., 1993; Gaszner and Felsenfeld, 2006; Gerasimova et al., 1995; Geyer and

Corces, 1992; Kellum and Schedl, 1992; Zhao et al., 1995). Insulators are thought to play important roles in the regulation of gene transcription by preventing the inappropriate action of regulatory elements on the expression of genes in neighboring domains, which may be established as a result of their functional properties (Labrador and Corces, 2002; West et al., 2002).

Chromatin Insulators in Eukaryotes

Insulator function is highly conserved in eukaryotes, existing in organisms from yeast to humans. The insulators characterized in yeast have been shown to possess only boundary activity. For example, sequences called subtelomeric anti-silencing regions, or STARs, which are located in the X and Y subtelomeric repeats, contain binding sites for Tbf1p and Reb1p which function to limit telomeric silencing to certain regions (Fourel et al., 1999). Also, the *tRNA*^{Thr} gene located in the heterochromatic HMR locus exhibits boundary activity when bound by the transcription factors TFIIIC and TFIIIB (Donze and Kamakaka, 2001).

The CCCTC binding factor, or CTCF, is the only major protein linked to insulator activity in vertebrates, and CTCF insulators have been described in a number of species. CTCF is associated with the enhancer-blocking activity in the chicken β-globin locus, where it binds to both the 5' and 3' DNase I hypersensitive sites (HSs), protecting this region from nearby enhancers (Bell et al., 1999; Saitoh et al., 2000). In addition to its role in enhancer-blocking, the 5' HS4 exhibits boundary activity independent of CTCF, acting as the 5' boundary of the open β-globin chromatin (Pikaart et al., 1998). In humans and mice, CTCF insulators play important roles in imprinting of the *H19/Igf2* locus which occurs by means of a differentially methylated domain (DMD) located

between the two genes. CTCF binds the DMD on the maternally inherited allele, blocking the activation of the *Igf2* promoter by enhancers downstream of *H19*. On the paternally inherited allele, methylation of the DMD prevents binding of the CTCF protein, thereby preventing enhancer blocking (Bell and Felsenfeld, 2000; Thorvaldsen et al., 1998). In vertebrates, CTCF was recently shown to interact with cohesin, a protein complex involved in sister chromatid cohesion during the S phase of the cell cycle through anaphase of mitosis or meiosis (Michaelis et al., 1997). Cohesin plays an important role in mediating insulator activity at CTCF binding sites such as the *H19/Igf2* locus (Wendt et al., 2008). CTCF is conserved in *Drosophila* as well, where it has been shown to possess insulator properties similar to that in vertebrates. For example, dCTCF binds to domain boundaries within the *Abd-B* gene located in the Bithorax complex (BX-C), including the *Mcp* and *Fab-8* insulators (Holohan et al., 2007; Smith et al., 2009).

The *Drosophila* scs and scs' elements were the first DNA sequences found to possess insulator activity. These sequences are located at the chromatin boundaries of the 87A7 hsp70 locus. Scs and scs' contain binding sites for the Zeste-white 5 (Zw-5) and the Boundary Element Associated Factor of 32 kDa (BEAF-32) proteins, respectively, and binding of these proteins is required for insulator function (Gaszner et al., 1999; Zhao et al., 1995). Other insulators that have been described in *Drosophila* are the *eve* promoter, the SF1 boundary, and the *Mcp* and *Fab-7* boundaries of the BX-C, all of which contain binding sites for the GAGA factor protein (Busturia et al., 2001; Mihaly et al., 1998; Mishra et al., 2001; Ohtsuki and Levine, 1998). One of the best characterized insulators is the *gypsy* insulator of *Drosophila* (Byrd and Corces, 2003;

Cai and Shen, 2001; Gerasimova and Corces, 1998; Gerasimova et al., 1995; Ghosh et al., 2001; Kumaran et al., 2008; Muravyova et al., 2001; Roseman et al., 1993).

Gypsy insulator

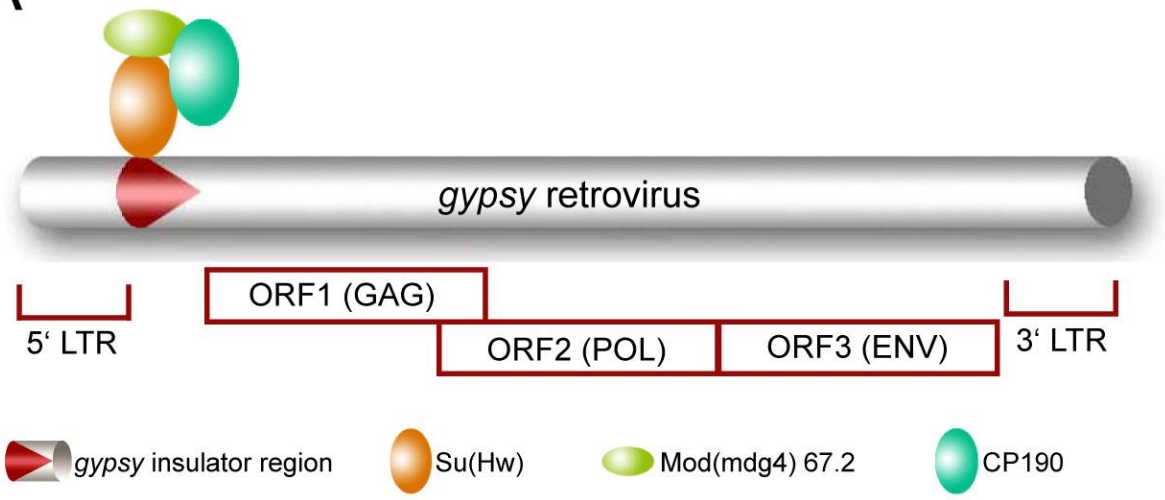
The *Drosophila gypsy* insulator is a 350 bp element located in the 5' untranslated region of the *gypsy* retrovirus upstream of the *gag* open reading frame (Byrd and Corces, 2003; Gerasimova et al., 2000; Gerasimova and Corces, 1998, 2001; Geyer and Corces, 1992). The insulator DNA consists of twelve repeats of a motif sequence (5'-YRYTGCATAYBY-3'), to which the Suppressor of Hairy-wing [Su(Hw)] protein binds, and short AT-rich sequences separating the 12 Su(Hw)-binding motifs (Spana and Corces, 1990; Spana et al., 1988). In addition to Su(Hw), at least two other proteins are required for insulator function: Modifier of mdg4 67.2 [Mod(mdg4) 67.2] and Centrosomal Protein of 190 kD (CP190) (Gerasimova and Corces, 1998; Gerasimova et al., 1995; Pai et al., 2004). Mod(mdg4) and CP190 do not bind directly to the *gypsy* insulator DNA, but interact with Su(Hw) as well as with each other via protein-protein interactions (Figure 1.1) (Gerasimova et al., 1995; Pai et al., 2004).

The Su(Hw) insulator protein is a zinc finger protein which is required for *gypsy* insulator function and binds to insulator DNA via a stretch of 12 zinc finger motifs at its central portion (Figure 1.1B). Four of the 12 zinc finger motifs are essential for recognizing and binding the 5'-YRYTGCATAYBY-3' repeats in the *gypsy* insulator DNA sequence. Su(Hw) possesses both N-terminal and C-terminal acidic and a leucine zipper domain, which is homologous to the helix 2–coiled-coil region of the bHLH-Zip proteins (Harrison et al., 1993; Kim et al., 1996). Three regions at the C-terminal end of

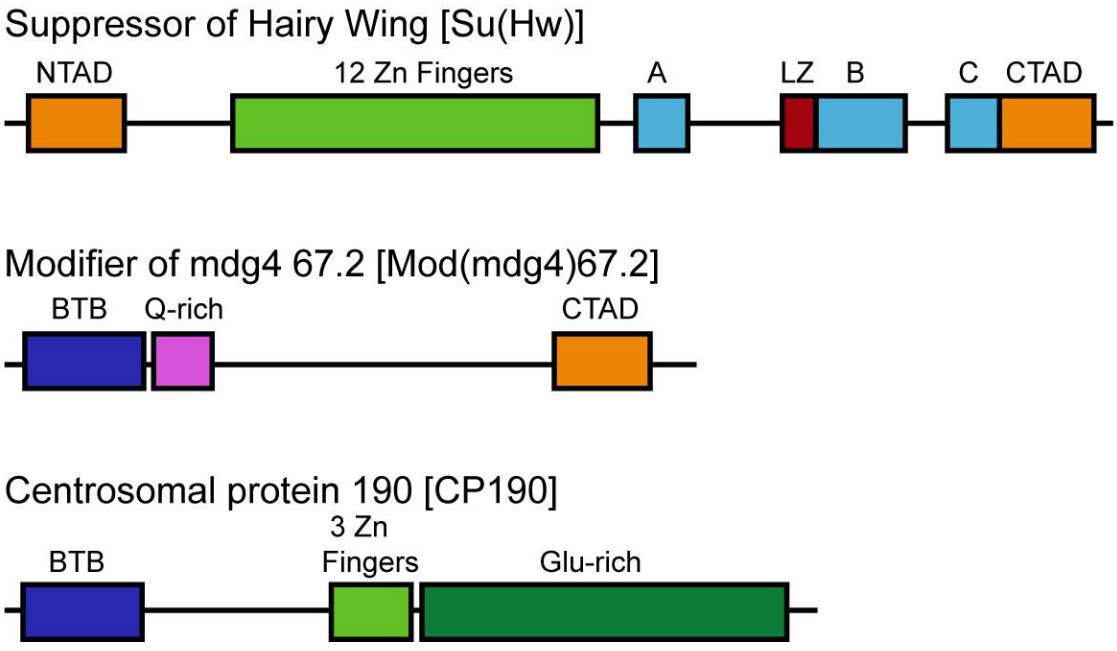
Figure 1.1 The *gypsy* insulator and core insulator proteins.

(A) The *gypsy* insulator is indicated by the red region near the 5' end of the retrovirus. The 3' and 5' long terminal repeats (LTRs) and *gag*, *pol*, and *env* open reading frames (ORFs) are shown below the *gypsy* retrovirus. Su(Hw), Mod(mdg4)67.2 and CP190 are depicted as orange, green, and blue ovals respectively. **(B)** The structures of three core insulator proteins. Domains are depicted as colored boxes; NTAD, amino-terminal acidic domain; LZ, leucine zipper; CTAD, carboxy-terminal acidic domain; BTB, BTB/POZ domain; Q-rich, glutamine rich domain; Glu-rich, glutamate rich domain.

A


gypsy retrovirus
ORF1 (GAG)
5‘ LTR
ORF2 (POL)
ORF3 (ENV)
3‘ LTR
gypsy insulator region
Su(Hw)
Mod(mdg4) 67.2
CP190

B


Suppressor of Hairy Wing [Su(Hw)]
NTAD
12 Zn Fingers
A
LZ
B
C
CTAD
Modifier of mdg4 67.2 [Mod(mdg4)67.2]
BTB
Q-rich
CTAD
Centrosomal protein 190 [CP190]
BTB
3 Zn Fingers
Glu-rich

Su(Hw), termed A, B, and C, are highly conserved among a number of *Drosophila* species but do not show recognizable homology to any known functional domain (Gdula and Corces, 1997). Although region A appears to be dispensable for Su(Hw) function, interaction between Su(Hw) and Mod(mdg4)67.2 are mediated by regions B and C along with the leucine zipper domain (Ghosh et al., 2001).

Mod(mdg4)67.2 is a second protein component of the *gypsy* insulator and is the major isoform encoded by the *mod(mdg4)* gene, which encodes at least 26 different isoforms. The protein is generated by trans-splicing of two independent pre-messenger RNA transcripts to form the final mRNA (Labrador et al., 2001; Mongelard et al., 2002). Mod(mdg4)67.2 contains a BTB/POZ (*b*ric-à-brac, *t*ramtrack and *b*road-complex/pox virus and zinc finger) domain, a highly conserved protein-protein interaction domain, at the N-terminus (Figure 1.1B). BTB domain proteins are able to interact with each other symmetrically to form stable homodimers (Ahmad et al., 1998). The Mod(mdg4)67.2 protein is capable of interacting with itself, and this homodimerization was shown to be mediated by the BTB domain. In addition to the BTB/POZ domain, Mod(mdg4)67.2 also possesses a C-terminal acidic domain which mediates interactions with the Su(Hw) protein (Ghosh et al., 2001).

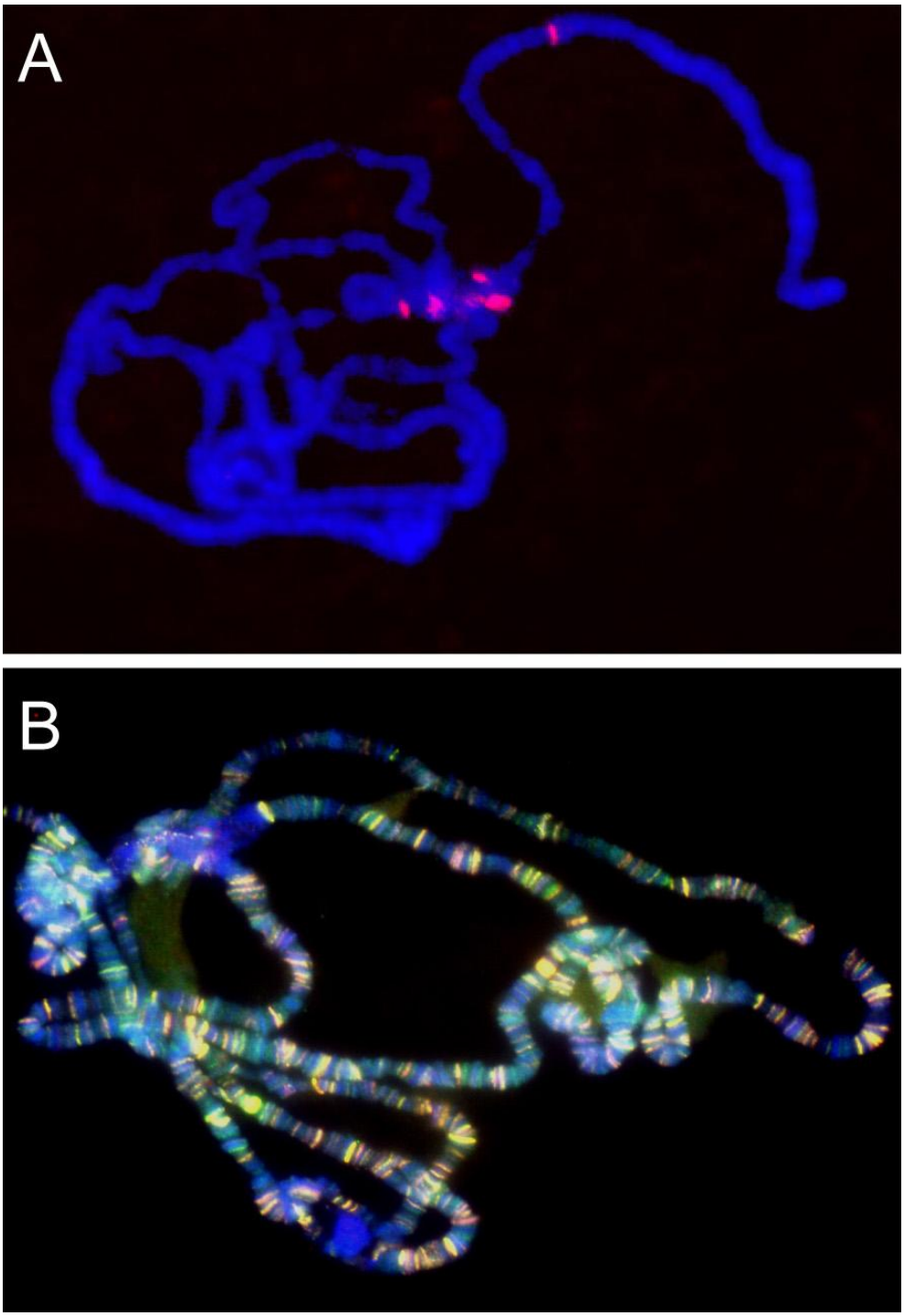
Centrosomal Protein 190 (CP190), a third essential component of the *gypsy* insulator, was initially found to be associated with centrosomes during mitosis. During interphase, CP190 localizes to the nucleus where it binds to numerous sites on polytene chromosomes (Whitfield et al., 1995). CP190 was identified as an essential component of the *gypsy* insulator through a genetic screen for dominant enhancers of *mod(mdg4)*, and its localization on polytene chromosomes was found to overlap significantly with

that of Su(Hw) and Mod(mdg4)67.2. The CP190 protein is able to interact with both Su(Hw) and Mod(mdg4)67.2 proteins, likely through its N-terminal BTB/POZ domain. CP190 also contains three C2H2 zinc-finger motifs, although it is not thought to bind directly to DNA (Figure 1.1B) (Pai et al., 2004).

Endogenous insulators in genome organization

It has been suggested that chromatin insulators may contribute to genome organization by establishing boundaries between different levels of chromatin organization, such as the transitions between interbands and condensed chromatin in bands from polytene chromosomes (Gilbert et al., 2006; Labrador and Corces, 2002; Pai et al., 2004; Spana et al., 1988; Zhao et al., 1995). The pattern of bands and interbands in *Drosophila* polytene chromosomes indicates the presence of an underlying structural organization that divides the chromosomes into domains that may facilitate the regulation of gene expression (Gerasimova, et al, 2000). A number of different insulator proteins have been shown to be present at the boundaries between bands and interbands in polytene chromosomes, in agreement with a putative role for insulators in establishing or maintaining the band/interband domains. The Su(Hw), Mod(mdg4)67.2, and CP190 proteins co-localize at several hundred sites on polytene chromosomes of *Drosophila* larval salivary glands which do not correspond to insertion sites of the *gypsy* retrovirus (Figure 1.2) (Pai et al., 2004). These sites are thought to represent endogenous *gypsy* insulators and the abundance of these sites suggests that they have important functional roles (Labrador and Corces, 2002).

Figure 1.2 Distribution of Su(Hw) and Mod(mdg4) 67.2 on polytene chromosomes. **(A)** *In situ* hybridization of *gypsy* retrovirus on polytene chromosomes is shown is red. DNA is stained with DAPI and is shown in blue. **(B)** Immunostaining of Su(Hw) (shown in red) and Mod(mdg4)67.2 (shown in green) on polytene chromsomes.


A
B

Numerous potential *gypsy* endogenous insulators have been identified by mapping insulator elements at the level of DNA sequence. However, endogenous Su(Hw) binding sites may differ from those in the *gypsy* retrotransposon, which contains 12 tandem repeats of the Su(Hw) binding site, in that most Su(Hw) binding sites are present in the genome as single copies (Parnell et al., 2006; Ramos et al., 2006) suggesting that their properties may differ from the *gypsy* insulator. The first genomic Su(Hw) binding site to be identified was the 1A2 site, which is located in an intergenic region between the *yellow* and *achaete* genes and was shown in transgenic assays to possess enhancer-blocking activity (Golovnin et al., 2003). The properties of endogenous insulators appear to correlate with the number of binding sites, with single binding sites having less enhancer-blocking effect than multiple binding sites (Golovnin et al., 2003; Parnell et al., 2003; Ramos et al., 2006). Bioinformatic approaches using the consensus Su(Hw) binding site sequence obtained from the *gypsy* insulator led to the identification of new binding sites and provided information about the function and genomic distribution of these sites (Parnell et al., 2006; Ramos et al., 2006). For example, enhancer-blocking assays showed that insulator activity of Su(Hw) binding sites depends both on the number of Su(Hw) binding sites as well as the genomic context. Additionally, Su(Hw) binding sites were found to be located predominantly in intergenic regions and within long genes containing at least one intron (Ramos et al., 2006). These findings suggest a functional role for Su(Hw) binding sites in organizing the genome into transcriptional domains.

In addition to the broad distribution of Su(Hw), Mod(mdg4) and CP190, other insulator proteins such as BEAF, dCTCF, GAGA factor and Zw5 are located throughout

the entire *Drosophila* genome as well. Much of what we know about the distribution of insulator proteins in the *Drosophila* genome comes from ChIP-on-chip tiling array data from a number of recent studies which have revealed a large degree of colocalization among these insulator proteins (Adryan et al., 2007; Bartkuhn et al., 2009; Bushey et al., 2009; Jiang et al., 2009; Negre et al., 2010; Smith et al., 2009). For example, CP190 colocalizes with Su(Hw) and Mod(mdg4) on *Drosophila* polytene chromosomes. However, it is also present at numerous sites that lack these proteins. Genome-wide mapping of insulator proteins revealed that CP190 localizes to sites where dCTCF binds, a finding which supports previous data showing interactions between CP190 and dCTCF (Bartkuhn et al., 2009; Bushey et al., 2009; Gerasimova et al., 2007; Mohan et al., 2007). CP190 was found to be associated with 47% of Su(Hw) sites and 62% of dCTCF sites. Additionally, the finding that 71% of BEAF sites colocalized with CP190 identified BEAF as a third subclass of CP190-containing insulators (Bushey et al., 2009). While the fact that these three subclasses of insulator proteins share CP190 as a common functional component may suggest that they function using similar mechanisms, each subclass shows a distinct distribution pattern in relation to gene location. 84% of BEAF sites and 47% of dCTCF binding sites are located within 1 kb of the ends of genes, most often near the 5’ end, with BEAF in close proximity to genes involved in metabolic processes, and dCTCF located near genes with roles in developmental processes. Su(Hw) binding sites, on the other hand do not show a strong association with the 5’ ends of genes and are frequently located at long distances from genes (Bushey et al., 2009; Negre et al., 2010). Also, the majority of dCTCF and BEAF sites located at the 5’ end of genes are associated with genes that

are highly expressed, while Su(Hw) sites are associated with genes that have low expression. The distribution of these insulator proteins in relation to genomic landmarks suggests that they may have differential roles in genome organization and the establishment of regulatory domains. However, the precise role that insulator proteins play in chromosome organization remains largely unknown (Phillips and Corces, 2009).

The mechanism by which insulator proteins are thought to establish regulatory domains is unclear, but evidence suggests that this may occur through intra- and inter-chromosomal interactions among insulator sites throughout the genome**.** For example, while Su(Hw), Mod(mdg4)67.2, dCTCF and CP190 are visible at hundreds of sites in polytene chromosomes, in diploid interphasic cells colocalization of the three proteins is visible as only 25-30 dots per nucleus (Gerasimova et al., 2000; Gerasimova and Corces, 1998; Pai et al., 2004). It has been suggested that these dots correspond to insulator bodies, which result from the coalescence of several endogenous insulators to form independent chromatin loop domains consisting of the intervening DNA sequences (Figure 1.3) (Labrador and Corces, 2002). A number of findings lend support to this idea. For example, the insertion of two *gypsy* retrotransposons between a promoter and a distal enhancer resulted in enhancer bypass of the insulators and activation of the promoter, suggesting that the interaction of two insulators may loop out the sequences between them, thereby bringing the enhancer and promoter into close proximity (Cai and Shen, 2001; Muravyova et al., 2001). In support of this idea, *in situ* hybridization showed that two *gypsy* retrotransposons present in the same fly line at different chromosomal locations result in a colocalization of hybridization signals in diploid cell nuclei. Also, chromatin loops which are anchored by Su(Hw) insulators have been

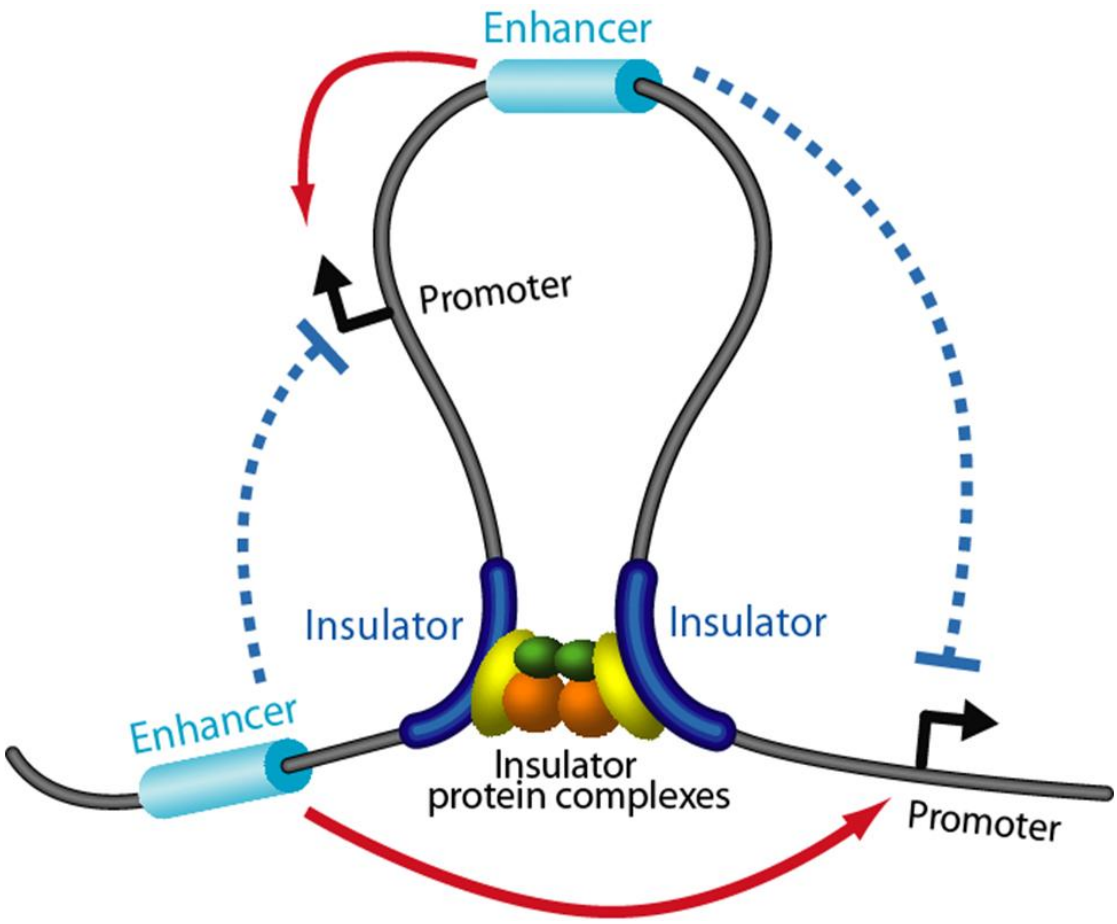


Figure 1.3 Formation of independent chromatin domains mediated by insulators. Insulators separate the chromatin fiber into loop domains. Enhancers located outside the loop domain are only able to activate promoters outside the domain, whereas enhancers within the domain can only activate promoters which are also within the chromatin domain.

visualized by FISH in salt-extracted nuclei (Byrd and Corces, 2003; Gerasimova et al., 2000). Further evidence of the formation of chromatin loops comes from the scs and scs' insulators, which are bound by Zw5 and BEAF, respectively, and can pair with each other both in vitro and in vivo (Blanton et al., 2003). Finally, CTCF insulator sites in vertebrates are able to physically interact in a manner that is dependent on gene activity (Ling et al., 2006; Splinter et al., 2006).

Insulator proteins may not only establish chromatin loops via interactions between themselves, but may also form loops by tethering chromatin to nuclear structures. Mammalian CTCF has been shown to interact with nucleophosmin, a protein which is found at the nucleolus. CTCF binding sites in transgenes frequently localize to the nucleolus, and this localization is abolished when the CTCF binding sites are mutated (Yusufzai et al., 2004). In *Drosophila*, insulator bodies formed in diploid cells by coalescence of insulator proteins largely localize to the nuclear periphery. This localization may be mediated by dTopors, an E3 ubiquitin ligase, which interacts with Su(Hw) and Mod(mdg4)67.2. dTopors also associates with the nuclear lamina, and may therefore serve as an anchor for the attachment of insulator sites to the lamina (Capelson and Corces, 2005). The ability of insulator sequences to form chromatin loops may result in the formation of structures which affect nuclear organization in such a way that may be important for gene expression (Gurudatta and Corces, 2009).

Classes of transposable elements

Transposable elements (TEs) are selfish DNA entities which can use a host genome for survival and propagation, and they are abundant components in the

genomes of most living organisms (Almeida and Allshire, 2005; Kaminker et al., 2002). TEs are divided into two classes, Class I and Class II, based on their structure and mechanism of transposition. Class I elements include the retrotransposons, which transpose via reverse transcription of an RNA intermediate. Class II elements are the DNA transposons, which transpose by a “cut-and-paste” mechanism of DNA excision and repair (Malone and Hannon, 2009; Slotkin and Martienssen, 2007). Class II transposons mainly consist of autonomous elements which encode their own transposase and contain flanking terminal inverted repeats (TIRs) (Kaminker et al., 2002). The transposase recognizes the TIRs and catalyzes excision and reintegration of DNA elements (Malone and Hannon, 2009).

The Class I retrotransposons are grouped into LTR elements, those that have long terminal repeats (LTRs), and non-LTR elements, such as long interspersed elements and short interspersed elements (LINEs and SINEs). Autonomous non-LTR elements contain two open reading frames (ORFs), one which encodes for a DNA binding protein and the other which encodes both an endonuclease and reverse transcriptase (Malone and Hannon, 2009). LTR elements are thought to be derived from retroviruses and contain *gag* and *pol* genes which encode for a viral particle coat, a reverse transcriptase, a ribonuclease H and an integrase, all of which are important for their transposition to new sites (Kazazian, 2004). Some LTR elements, classified as infectious retroviruses, contain an *env* gene which codes for an envelope protein by which movement to other cells may occur. The *Drosophila gypsy* retrotransposon is one such example of an infectious retrovirus (Kim et al., 1994; Song et al., 1994).

Silencing of transposable elements by piRNA pathways

Mobile elements exist in the genome of most organisms and carry with them the potential to generate negative effects on their hosts (Brennecke et al., 2007). Transposable elements are often considered 'selfish' elements, because their success negatively impacts the fitness of the host. TEs frequently insert in protein coding genes, and can result in chromosome breakage and recombination. Additionally, they may affect the expression of nearby genes by changing patterns of splicing and polyadenylation, and they may even function as enhancers or promoters (Girard and Freeling, 1999). The mutagenic potential of TEs is particularly critical in the germline of the developing organism (Vagin et al., 2006). Transpositions in the germline may introduce mutations and chromosomal rearrangements which are then transmitted to offspring. Genomes thus have had to develop mechanisms to regulate transposable elements assuring that the majority of TEs remain silent and inactive in the germline (Zaratiegui et al., 2007).

Transposon resistance and silencing of repetitive sequences has been linked to small RNA regulatory pathways (Brennecke et al., 2007). Core components of the RNAi pathway were shown to be required for TE repression in *C elegans*, *D. melanogaster*, and *M. musculus*. Mutations in known components of the RNAi pathway in *Drosophila* induce transposon mobility in the germline (Kalmykova et al., 2005; Sarot et al., 2004). In the germline of *Drosophila*, as well as most other multicellular animals, a class of small RNAs, referred to as repeat-associated small interfering RNAs (rasiRNAs) or piwi-interacting RNAs (piRNAs), ensure genomic stability by suppression of transposable elements (Aravin et al., 2006; Vagin et al., 2006). piRNAs arise from all known forms of

repetitive elements including retrotransposons and DNA transposons (Sontheimer and Carthew, 2005).

The core components of the piRNA pathway are Piwi family proteins and their associated piRNAs. piRNAs interact with several members of the Piwi family, which is a subclass of the Argonaute family of RNA interference proteins (Carthew, 2006). Piwi proteins, like other members of the Argonaute family, act by binding small RNAs, which are used as guides for mRNA recognition and cleavage (Carmell et al., 2002). The Piwi subfamily consists of Piwi, Aubergine (Aub), and Argonaute 3 (Ago3) proteins which are expressed solely in the germline. Piwi subfamily proteins play integral roles in germline development in *Drosophila*. Aubergine is required for pole cell formation, and *aub* mutations impair the production of functional oocytes. Aub is involved in *TART* transposon silencing (Savitsky et al., 2006) in the female germline as well as Su(Ste) silencing of *Stellate* genes (Aravin et al., 2004). While little is known about the function of Ago3, it is known that *piwi* mutations result in defects in oogenesis and loss of germline cells, indicating its essential role in germline stem cell self-renewal (Cox et al., 1998). Piwi is the only one of these proteins which is expressed in somatic gonadal cells, and it has been recently shown to act alone in a specialzed piRNA pathway in somatic cells (Malone et al., 2009).

Drosophila piRNAs map to locations of known transposon insertion. Active transposons are located in the euchromatin as well as the pericentric and telomeric heterochromatin. The heterochromatin which borders the centromeres is also enriched in inactive transposons which are partial or defective and are no longer capable of transposition (Hoskins et al., 2002). The majority of piRNAs are generated from

heterochromatic loci referred to as piRNA clusters. piRNAs from these clusters show a tissue-specific expression pattern which correlates with the tissue-specific regulation of the classes of elements to which they correspond. Most clusters produce piRNAs from both DNA strands, and these dual-strand clusters generate piRNAs that correspond to elements targeted by the germline piRNA pathway (Klattenhoff et al., 2009; Malone et al., 2009). For example, the cluster located in the 42AB cytological region of chromosome 2R is the largest piRNA cluster and produces approximately 30% of all piRNAs (Brennecke et al., 2007). The *flamenco* locus, which resides in heterochromatin on the X chromosome, is responsible for silencing of *gypsy, ZAM, and Idefix* in *Drosophila* ovaries (Desset et al., 2008; Mevel-Ninio et al., 2007; Prud'homme et al., 1995). Analysis of Piwi interacting RNAs (piRNAs) which map to this locus indicates that it is enriched for sequences of each of these three transposable elements as well as other *gypsy* family LTR elements (Malone et al., 2009). The production of piRNAs correlating with a biased orientation of transposon fragments indicates that piRNAs may be produced from only one strand and from long precursor transcripts that extend the length of *flamenco* (Brennecke et al., 2007; Malone et al., 2009). Elements represented in the *flamenco* cluster are silenced by the somatic piRNA pathway.

Although mobilization of transposable elements may have deleterious consequences for the host genome, the relationship between TEs and their host is a complicated one. Evidence suggests that TEs may also be beneficial to the host. For example, *Drosophila* telomeres lack the tandem repeat sequences found at the chromosome ends of many eukaryotes. Chromosomal ends are instead maintained by the non-LTR transposons *HeT-A* and *TART* (Levis et al., 1993). Also, reverse

transcriptase encoded by the Ty retrotransposon was found to be capable of repairing double-strand breaks in yeast when it was shown that Ty sequences are located at the break sites in *rad52* mutants when the RT was overexpressed (Moore and Haber, 1996).

Transposable elements in genome organization

TEs may serve as drivers of genome evolution. Transposition events facilitate the translocation of genomic sequences and the shuffling of exons. For example, in the human genome, L1s insert into transcribed genes and retrotranspose flanking sequences. In this manner sequences such as promoters or exons are mobilized into existing genes in a new genomic location thus generating the potential for new genes to evolve (Moran et al., 1999). By inserting in or near coding regions, TEs have the ability to alter sequence composition, as well as temporal and spatial patterns of gene expression. In *Drosophila*, the insertion of the *blood* retrotransposon in the *GPDH* gene alters the expression of several GPDH isozymes (Wilanowski et al., 1995). In plants, many TEs reside in close proximity to genes, where they may contribute regulatory elements that can influence transcription (White et al., 1994). The ability of transposons to shape the genome of their host in this way makes transposable elements an important part of evolution and gene regulation.

The establishment of chromatin domains is important for the temporal and spatial regulation of gene expression. In some cases, retrotransposons can define blocks of transcriptionally active and silent chromatin. For example, in mammals, LINE elements are frequently found within MARs (Purbowasito et al., 2004), which can function as

domain boundaries (Bode and Maass, 1988; Levy-Wilson and Fortier, 1989; Loc and Stratling, 1988). The murine growth hormone locus contains a B2 SINE element whose tissue-specific expression is required for gene activation and which serves as a boundary to block the influence of repressive chromatin modifications (Lunyak et al., 2007). In *Drosophila*, insertion of the *gypsy* retroelement in the *yellow* gene between enhancers and promoter affects gene expression controlled by the distal enhancers without affecting the function of other downstream enhancers (Gdula et al., 1996). The ability of *gypsy* to affect communication between regulatory elements is a result of its insulator activity.

In addition to *gypsy*, at least one other LTR retrotransposon has been shown to possess insulator activity. In transgenic assays, the 5' LTR of *Idefix* was able to block activation of the *white* gene by an enhancer located within *ZAM*, another LTR element (Conte et al., 2002a). *Idefix* not only possesses enhancer-blocking activity, but can function as a boundary as well. Transgenes are protected from chromosomal position effects when flanked with *Idefix* element (Conte et al., 2002a). Recent evidence suggests that *ZAM* may also possess an insulator within its 5' UTR which shows some degree of enhancer-blocking capability (Minervini et al., 2010). It is unclear whether insulator activity is unique to these elements or if it is a common occurrence among TEs. Given the proposed role of insulators in organizing the genome into functional chromatin domains, the existence of insulators within transposable element sequences may be another means by which TEs contribute to the regulatory landscape of the genome.

CHAPTER II

Chromatin insulators specifically associate with different levels of higher-order chromatin structure in *Drosophila*

This chapter is from a paper published in *Chromosoma* (2010 Apr;119(2):177-94). As first author, I contributed the majority of the experimental data to this work. The contributions of the other authors are as follows: Antibodies used for immunostaining and chromatin immunoprecipitation experiment were generated by Maria P. Plata. Hyuck-Joon Kang and Misty Ross contributed to part of the *in situ* hybridizations and some of the immunostaining data.

ABSTRACT

Chromatin insulators are required for proper temporal and spatial expression of genes in metazoans. Here we have analyzed the distribution of insulator proteins on the 56F-58A region of chromosome 2R in *Drosophila* polytene chromosomes, to assess the role of chromatin insulators in shaping genome architecture. Data shows that the Suppressor of Hairy wing protein [Su(Hw)] is found in three structures differentially associated with insulator proteins: bands, interbands and multi-gene domains of coexpressed genes. Results show that bands are generally formed by condensation of chromatin that belongs to genes containing one or more Su(Hw) binding sites, whereas in interbands Su(Hw) sites appear associated with open chromatin. In addition, clusters of coexpressed genes in this region form bands characterized by the lack of CP190 and

BEAF-32 insulator proteins. This pattern correlates with the distribution of specific chromatin marks and is conserved in nurse cells, suggesting that this organization may not be limited to one cell type but represents the basic organization of interphasic chromosomes.

INTRODUCTION

The ability of organisms to precisely regulate gene expression is central to their development. Proper temporal and spatial expression of genes in higher eukaryotes requires activation of transcription during the appropriate developmental stages. Gene regulation is established through the activity of *cis*-regulatory elements including proximal promoters, enhancers, repressors and silencers (Markstein and Levine, 2002; Ptashne and Gann, 1997; Walhout, 2006). However, these processes alone are not sufficient to explain the complexity of gene expression regulation in a chromatin context, given mounting evidence supporting that higher-order chromatin structures and long-range interactions are an important aspect in the process of gene transcription regulation during development and cell differentiation in metazoans (Apostolou and Thanos, 2008; Osborne et al., 2004; Schuettengruber et al., 2007; Spilianakis et al., 2005).

Chromatin insulators are regulatory elements found in *Drosophila* and in vertebrates that are considered to have a major role in higher-order chromatin organization based on their capacity to mediate long range interactions within the chromatin fiber (Bartkuhn and Renkawitz, 2008; Cai and Shen, 2001; Kuhn-Parnell et al., 2008; Kyrchanova et al., 2008; Ling et al., 2006; Maksimenko et al., 2008;

Muravyova et al., 2001; Splinter et al., 2006). Chromatin insulators are DNA sequences that have the ability to block communication between enhancers and promoters when located between them and to prevent heterochromatin spreading (Chung et al., 1993; Gaszner and Felsenfeld, 2006; Gerasimova et al., 1995; Geyer and Corces, 1992; Wallace and Felsenfeld, 2007; Zhao et al., 1995). It is believed that these properties result from *cis*-interactions between insulator proteins, which loop out the intervening DNA sequences to form functionally independent chromatin domains (Capelson and Corces, 2004; Gaszner and Felsenfeld, 2006; Gerasimova and Corces, 2001; Wallace and Felsenfeld, 2007).

In *Drosophila*, Suppressor of Hairy wing [Su(Hw)], Mod(mdg4)67.2 and centrosomal protein 190 (CP190) are proteins associated with the insulator activity initially identified within the *gypsy* retrovirus (Gerasimova and Corces, 1998, 2001; Gerasimova et al., 1995; Geyer and Corces, 1992; Ghosh et al., 2001; Pai et al., 2004; Spana et al., 1988). In addition to the *gypsy* insulator, however, these proteins are present in multiple locations throughout the genome. Similarly, the boundary element associated factor-32 (BEAF-32) and Zw5 were initially identified as components of the scs-scs' insulator at the hsp70 locus, but are also found through the entire genome (Gaszner et al., 1999; Kellum and Schedl, 1991; Zhao et al., 1995). Finally, the insulator activity of the CCCTC-binding factor (CTCF) was initially described in vertebrates but is also found in *Drosophila*, where it has been shown to have insulator properties similar to that in vertebrates, suggesting that insulator function and properties are conserved across species (Bell and Felsenfeld, 2000; Bell et al., 1999; Gerasimova et al., 2007; Moon et al., 2005).

The distribution of insulator protein binding sites in the *Drosophila* genome has been recently revealed by chromatin immunoprecipitation on tiling microarrays (Adryan et al., 2007; Bartkuhn et al., 2009; Bushey et al., 2009; Jiang et al., 2009; Smith et al., 2009) and by the modENCODE consortium. The analysis of these results suggests that insulator sites have the ability to define boundaries of regulatory units through complex interactions. However, the precise role that insulator proteins play in chromosome organization, if any, remains largely speculative (Phillips and Corces, 2009). It has been suggested, for example, that one of the roles for chromatin insulators is providing boundaries between different levels of chromatin organization, such as the transitions between interbands and condensed chromatin in bands from polytene chromosomes (Gilbert et al., 2006; Labrador and Corces, 2002; Pai et al., 2004; Spana et al., 1988; Zhao et al., 1995). In addition, it has been hypothesized that the functional integrity of multi-gene chromatin domains, such as those described in *Drosophila* and other organisms (Boutanaev et al., 2002; Caron et al., 2001; Dorus et al., 2006; Lee and Sonnhammer, 2003; Miller et al., 2004; Roy et al., 2002; Versteeg et al., 2003; Yi et al., 2007), is protected by chromatin insulators to facilitate the coordinate expression of genes without interference from regulatory sequences adjacent to the domains (Capelson and Corces, 2004; Kuhn and Geyer, 2003; Labrador and Corces, 2002; West et al., 2002).

Whereas modern techniques allow for detailed analysis of the local organization of the chromatin fiber, studies of the role that higher-order chromatin organization plays in gene transcription regulation are hindered by our inability to directly visualize chromatin organization within the interphase nucleus. Polytene chromosomes, however,

provide a unique opportunity to examine different levels of chromatin organization during interphase directly under the microscope. Polytene chromosomes form in cells that grow in size without dividing and, therefore, remain permanently in interphase. In a large number of cell types in *Drosophila* and other dipterans, multiple replication rounds without cell division result in more than 1000 DNA strands per chromosome. These strands remain aligned and attached to each other, forming the large structures that we know as polytene chromosomes. This alignment of multiple chromatin fibers amplifies the basic organization of the chromosome and allows differences in chromatin compaction to be seen as a series of bands and interbands extending across the width of the chromosome arms. In addition, because transcription only takes place in the more decondensed chromatin found in interbands, it is clear that these two levels of organization also translate into two different levels of transcriptional activity and chromatin structure (Kaplan et al., 2000; Labrador and Corces, 2003; Lis, 2007; Saunders et al., 2003; Yao et al., 2008).

Here, we take advantage of the recently described genome-wide distribution of insulator sites in *Drosophila* and use this information to devise a series of experiments assessing the role that insulator proteins have in the organization of higher-order chromatin structures in interphase chromosomes. We have analyzed the distribution of the *Drosophila* insulator protein Suppressor of Hairy-wing [Su(Hw)] along the 56F-58A region of chromosome 2R, and have established a correspondence between Su(Hw), CP190, BEAF-32 and the CCCTC-binding factor (CTCF) insulator sites, transcriptional units and the band-interband pattern observed in polytene chromosomes. Our results suggest that most Su(Hw) binding sites are largely associated with compacted

chromatin, and that Su(Hw) binding sites associated with open chromatin in interbands may have properties different from Su(Hw) binding sites found in condensed chromatin. Based on our findings, we propose a model in which different levels of higher-order chromatin organization mediated by insulators and other proteins alternate with each other to generate functional differences in chromatin compaction along the chromatin fiber, which are amplified to form the classic alternate sequence of bands and interbands in polytene chromosomes.

MATERIALS AND METHODS

Analysis of the distribution of Su(Hw) binding sites and gene clusters

ChIP-on-chip data files for Su(Hw), CP190, BEAF-32 and CTCF (accession number GEO GSE16245) from embryos were downloaded from the modENCODE website (http://www.modENCODE.org). Signal intensities were converted to normalized log2 ratios using Tiling Analysis Software (Affymetrix). BED files containing Su(Hw), CP190, BEAF-32 and CTCF peak data were obtained from Supplemental Material included in Bushey, et al (2009). Peak data files were uploaded into Integrated Genome Browser (http://igb.bioviz.org/) for visualization alongside the April 2006 version of the *Drosophila melanogaster* genome. Graph thresholding was set to visualize peaks above the 95th percentile. The distribution of coexpressed genes was analyzed using Excel in which clusters of coexpressed genes from Spellman and Rubin (2002) were merged with the annotated *D. melanogaster* genome project (release 5.5).

In Situ Hybridization and immunostaining of polytene chromosomes

Approximately 500bp DNA fragments corresponding to each endogenous insulator were obtained by PCR (see primers and probe sizes in supplemental table 2). Biotin-labeled DNA was prepared using the Biotin High-Prime random priming kit (Roche). The labeled probe DNA was ethanol precipitated and resuspended in hybridization buffer (4× SSC, 50% formamide, 1× Denhardts, and 0.4 mg/ml of salmon sperm DNA). Polytene chromosomes obtained from salivary glands of third instar larvae were dissected in 0.7% NaCl and fixed in a 1:2:3 mixture of acetic acid/water/lactic acid. Polytene chromosomes from nurse cells were obtained from the ovaries of 3 to 5 day old otu^{7}/otu^{11} females maintained in vials with males and in the presence of dry yeast to stimulate oogenesis. Ovaries were dissected in PBS, fixed and squashed in a 1:2:3 mixture of acetic acid/water/lactic acid. Slides were heated at 65°C in 2×SSC for 30 min, dehydrated in an ethanol series, and denatured in 0.07 M NaOH.

For hybridization of DNA, boiled probes were added to the slide, covered immediately with a coverslip, and incubated at 37°C overnight in a humidified chamber. Following hybridization, coverslips were removed and the slides were washed in 2× SSC at 37°C, then at room temperature in 1× PBS, and finally in antibody dilution buffer. The slides were incubated overnight in dilution buffer containing a 1:300 dilution of Su(Hw) primary antibody. Slides were then washed in antibody dilution buffer and incubated with a 1:300 dilution of FITC- or Texas red-conjugated goat anti-rabbit IgG (Jackson Laboratories) for 2 hr at room temperature. Slides were stained for 30 seconds with 4′,6-diamidino-2-phenylindole (DAPI; 0.5 µg/ml) and mounted in Vectashield mounting medium (Vector Laboratories). Slides were analyzed using a

Leica DM6000B fluorescence microscope. Antibodies specific against Su(Hw) protein were raised using an N-terminal peptide containing the first 218 amino acids of the protein. Antibodies specific against CP190 protein were raised using a C-terminal peptide containing amino acids 488 to 1084. Antibodies specific against Mod(mdg4)67.2 protein were raised using a C-terminal peptide containing amino acids 458 to 610. Antibodies were raised in rats and rabbits by Pocono Rabbit Farm and Laboratory. (Canadensis, PA 18325 USA) and were validated using westerns and by co-immunostainings with previously characterized rat or rabbit antibodies.

Immunostaining of polytene chromosomes

Antibodies used for immunostaining were as follows: rabbit anti-Su(Hw) and anti-CP190 (1:300 dilution) rat anti-Mod(mdg4)67.2 (1:300 dilution); mouse anti-H14 (RNA Pol II) (1:50 dilution) was purchased from Covance (Princeton, New Jersey), mouse anti- heterochromatin protein 1 (HP1) (1:25 dilution). The Hp1 antibody was developed by L.L. Wallrath and was obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the National Institute of Child Health and Human Development and maintained by The University of Iowa, Department of Biological Sciences, Iowa City, IA; The following antibodies were purchased from Upstate (Millipore Corporation, Billerica, MA): rabbit anti-H3K4me3 (dilution1:25), rabbit anti-H3K9ac (dilution 1:50), rabbit anti-H3K27me1 (dilution 1:50), rabbit anti-H3K4ac (dilution 1:50), rabbit anti-H3K27me3 (dilution 1:50); rabbit anti-Polycomb (dilution 1:50)was purchased from Santa Cruz Biotechnology (Santa Cruz, CA.). Secondary antibodies, FITC-conjugated goat anti-rabbit IgG, Texas red goat anti-rabbit IgG, Texas

red donkey anti-mouse IgM, and Texas red donkey anti-rat IgG were purchased from The Jackson Laboratory (Bar Harbor, ME).

Polytene chromosomes obtained from salivary glands of third instar larvae were dissected in 0.7% NaCl and fixed and squashed in fixative solution containing 3.7% formaldehyde and 45% acetic acid. Primary antibodies were diluted in PBS containing 0.1% Igepal and 1% milk and incubated on the slides overnight at 4 °. Slides were then washed in PBS + 0.1% Igepal and incubated with a 1:300 dilution of the appropriate FITC- or Texas red-conjugated secondary antibody for 2 hr at room temperature. Slides were stained for 30 seconds with 4′, 6-diamidino-2-phenylindole (DAPI; 0.5 µg/ml) and mounted in Vectashield mounting medium (Vector Laboratories; Burlingame, CA). Slides were analyzed using a Leica DM6000B fluorescence microscope.

Chromatin immunoprecipitation

Chromatin was prepared from 17-h embryos collected on grape juice agar media. Embryos were homogenized in buffer A1 (60 mM KCl, 15 mM NaCl, 4 mM MgCl2, 15 mM HEPES (pH 7.6), 0.5% Triton X-100, 0.5 mM DTT, 10 mM sodium butyrate, 1X EDTA-free protease inhibitor cocktail (Roche) containing 1.8% formaldehyde. Crosslinking was stopped by adding 225 mM glycine solution. Cells were lysed and chromatin was sheared to an average length of 500-700bp by sonication. In each ChIP experiment, a chromatin solution corresponding to 200 mg of live material was incubated with either Su(Hw) antibody or normal rabbit IgG. Immunoprecipitation and washing were performed as described elsewhere (Cavalli, 1999). The same Su(Hw) antibody used in immunostaining experiments was used in ChIP assays.

Real-time PCR quantification analysis of immunoprecipitated DNA

Real-time PCR quantification of immunoprecipitated DNA was carried out with ABGene (Rockford, IL) SYBR green PCR master mix. For input PCRs, a 1/100 dilution was used as template. Primers were designed to amplify 100-200bp fragments. PCR conditions for each primer pair were tested to determine the efficiency of amplification and to ensure amplification was in the linear range. PCR products for each primer pair were amplified from at least three separate immunoprecipitation products from at least two different chromatin preparations using the BioRad iQ5 Multicolor Real-Time PCR detection system (Primers listed in Supplemental Table 3). Enrichment of immunoprecipitated DNA fragments was calculated using the ΔCt method based on the threshold cycle (Ct) value for each PCR reaction (BioRad real time PCR application guide). Results are presented as percentage of total input. The statistical significance of the results was calculated by Student's t-test.

RESULTS

Only a fraction of the Su(Hw) binding sites identified in embryos correspond to major Su(Hw) bands visible in polytene chromosomes.

We have used fluorescence *in situ* hybridization (FISH) combined with immunostaining to perform an analysis of the distribution of the Su(Hw) protein localized within a total of 2 contiguous Mb along the 56F-58A cytological region on polytene chromosome 2R. This chromosome region has a specific morphology as well as a band-interband pattern that makes it easily identifiable in polytene chromosome spreads. With the help of ChIP-chip data on insulator site distribution in the *Drosophila*

genome from the modENCODE consortium and Bushey et al. (2009), we have used 29 specific probes containing individual Su(Hw) binding sites from this region to perform FISH combined with immunostaining (Table A.1 and Figure A.1). Probes for *in situ* hybridization were designed as 500bp sequences containing a single binding site identified based on ChIP-chip peak data. We used these data with the intention of, first, directly mapping the specific Su(Hw) binding sites in the chromosomes in relation to the band-interband pattern; second, to determine whether the specific sites are associated with a Su(Hw) immunostaining signal; and finally, using the *in situ* hybridized sites as landmarks to determine the relative position of other insulator sites and genes within the same region. One of the arguments against the suitability of immunostaining and *in situ* hybridization on polytene chromosomes is the low resolution of the technique. However, the level of resolution of this technique in our experiments is well below 15kb, even in chromosome regions in which chromatin appears as DAPI condensed bands (Figure A.1). This high level of resolution greatly facilitates the analysis of the distribution of chromatin proteins directly on chromosomes as a complement to analysis performed using ChIP.

Results show that the 29 *in situ* hybridization probes used in these experiments are sufficient to identify all the major Su(Hw) immunostaining signals found within the 56F-58A cytological region (Figures 2.1 and 2.2). In general, there is a noticeable genome-wide variability in the intensity of immunostaining signals when antibodies against Su(Hw) are used to immunostain polytene chromosomes (Gerasimova and Corces, 1998; Gerasimova et al., 1995). The same level of variation is observed among

Fig. 2.1 Su(Hw) binding sites associated with visible bands in polytene chromosomes.

Fifteen Su(Hw) bands are shown in a region of chromosome 2R spanning nucleotides 16,050,000 to 17,500,000. Numbers 1 to 15 indicate Su(Hw) bands as determined by FISH experiments not shown in this Figure (See Figure A.1). Immunostaining images of Su(Hw) bands are shown side by side with ChIP–on-chip peaks from modENCODE data and the corresponding annotated genes viewed using Integrated Genome Browser. Peaks shown in green correspond to Su(Hw) sites. Peaks shown in red correspond to CP190 sites. Arrowheads in polytene chromosomes point to FISH signals, shown in red. Arrowheads in microarray peaks point to sites used as probes in polytene chromosome FISH experiments, shown above, indicated by the corresponding color arrowhead. White arrows (c) point to a Su(Hw) band that corresponds to a *gypsy* retrotransposon insertion.

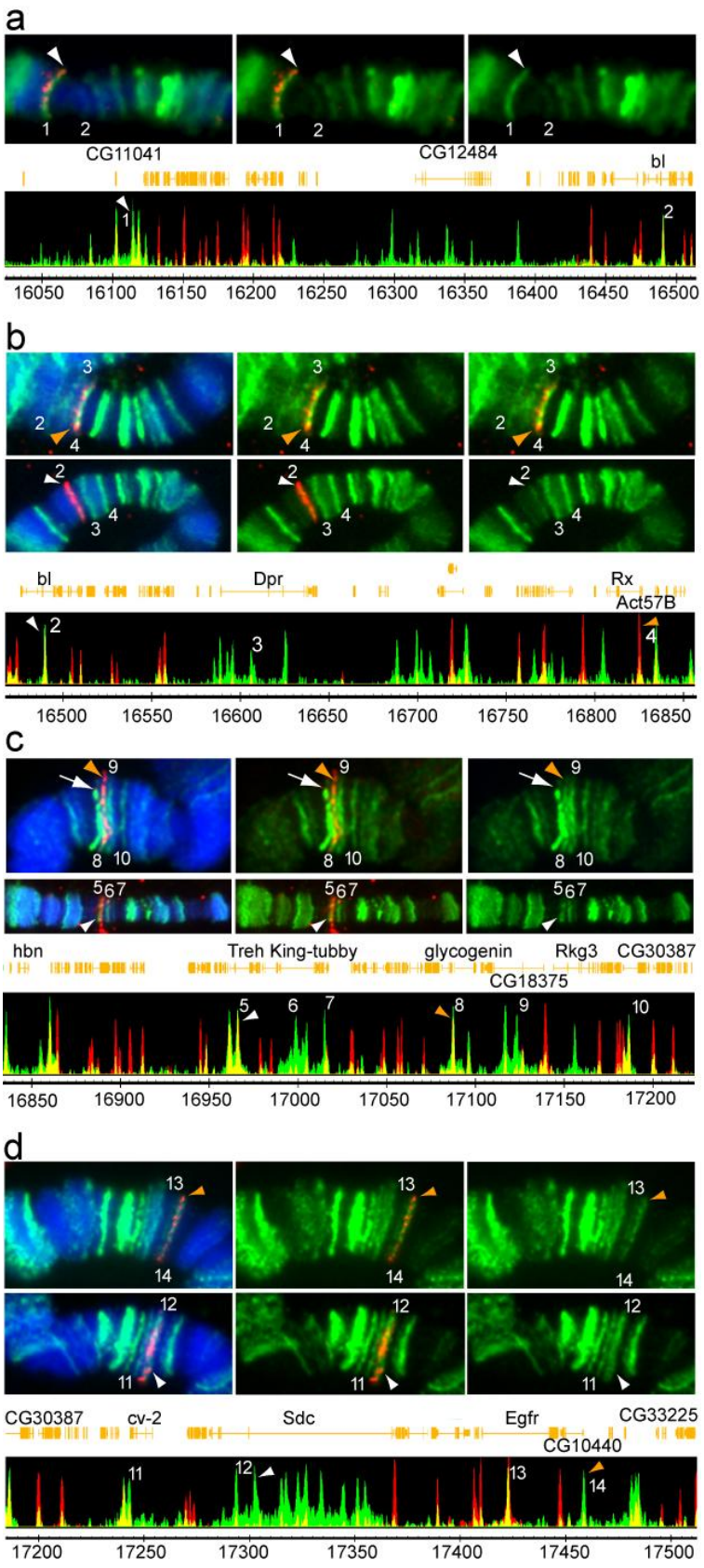
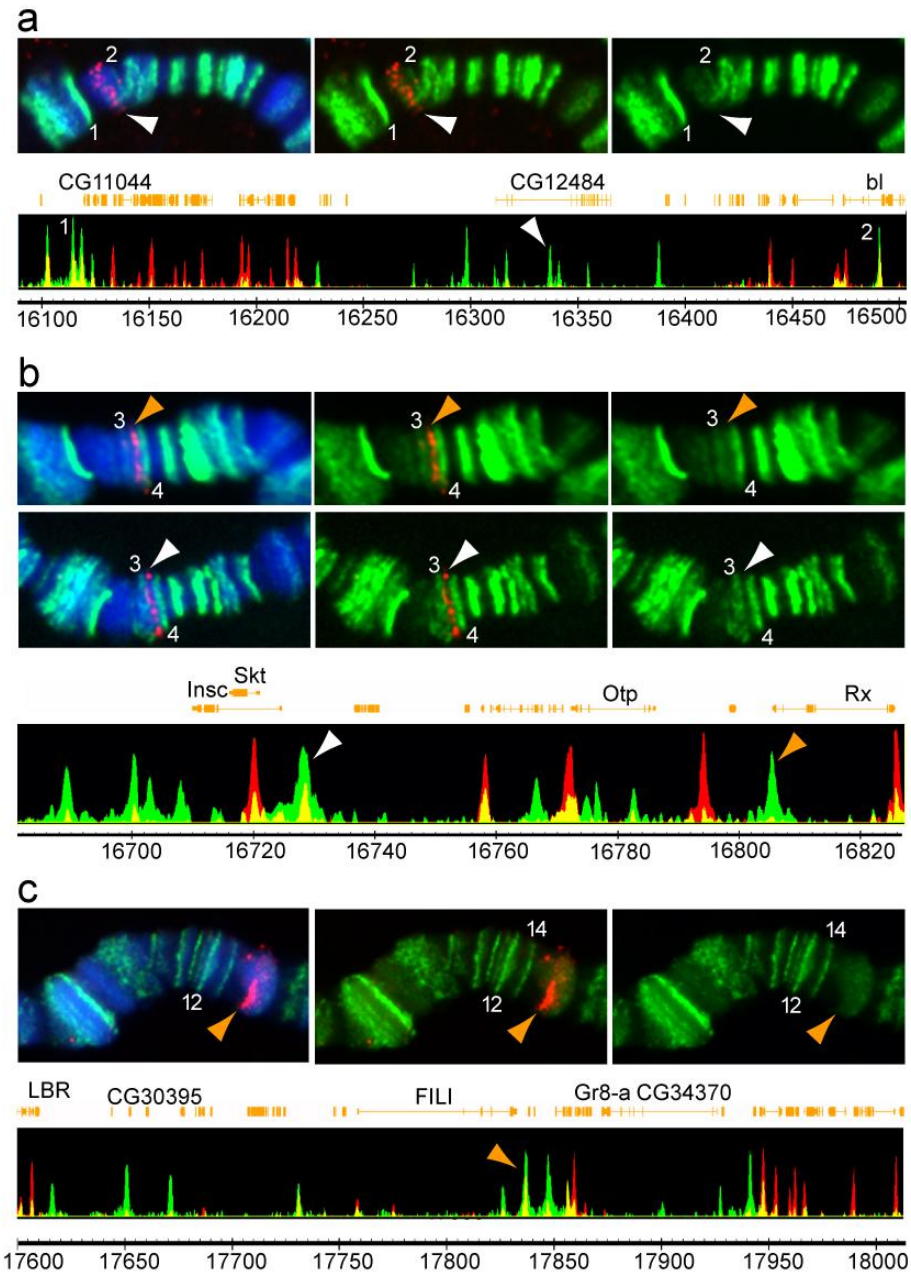

a
1
2
CG11041
CG12484
bl
16050
16100
16150
16200
16250
16300
16350
16400
16450
16500
b
3
2
4
bl
Dpr
Rx
Act57B
16500
16550
16600
16650
16700
16750
16800
16850
c
9
8
10
5
6
7
hbn
Treh
King-tubby
glycogenin
Rkg3
CG30387
CG18375
16850
16900
16950
17000
17050
17100
17150
17200
d
13
14
12
11
CG30387
cv-2
Sdc
Egfr
CG33225
CG10440
17200
17250
17300
17350
17400
17450
17500

Fig. 2.2 A large fraction of Su(Hw) binding sites identified by tiling microarrays are not associated with visible bands in polytene chromosomes.

a to **d**. Three large chromosomal segments spanning nucleotides 16,100,000 – 16,500,000, 16,700,000 – 16,800,000 and 17,600,000 – 18,000,000 are shown. Numbers in polytene chromosomes indicate Su(Hw) bands as determined by FISH experiments (See Figure 2.1). FISH combined with immunostaining using Su(Hw) antibodies are shown side by side with ChIP-on-chip peaks from modENCODE data and the corresponding annotated genes viewed using Integrated Genome Browser. Peaks shown in green correspond to Su(Hw) sites. Peaks shown in red correspond to CP190 sites. Arrowheads in polytene chromosomes point to FISH signals, shown in red. Arrowheads in microarray peaks point to sites used as probes in polytene chromosome FISH experiments, shown above, indicated by the corresponding color arrowhead.


a
2
1
CG11044
CG12484
bl
1
2
16100
16150
16200
16250
16300
16350
16400
16450
16500
b
3
4
Insc
Skt
Otp
Rx
16700
16720
16740
16760
16780
16800
16820
c
12
14
LBR
CG30395
FILI
Gr8-a CG34370
17600
17650
17700
17750
17800
17850
17900
17950
18000

the Su(Hw) immunostaining signals that we have tested along the 56F-58A cytological region. Su(Hw) signals can appear as low, medium, or high intensity bands (Figure 2.1). In addition, a large number of sites show very weak and diffuse or no signal (Figure 2.2). Only 15 Su(Hw) bands are present within the 56F-58A cytological region (Figure 2.1), to which 18 of the probes localize, while the remaining 11 probes are located in regions with a weak or absent Su(Hw) signal. Similarly, using ChIP-chip tiling array data from Bushey, et al (2009) and modENCODE consortium we found that out of the 78 binding sites mapped to this region, only 45 are associated with the 15 immunostaining bands visible in polytene chromosomes, whereas 33 do not localize with any visible Su(Hw) band. Sites with an immunostaining signal that can be identified as a band are listed in Table 2.1, and sites with no identifiable immunostaining signal are listed in Table 2.2.

In order to further understand the differences in signal intensity between sites observed as immunostaining signals in salivary glands, we have compared the distribution of immunostaining signals in polytene chromosomes with that of sites obtained from microarray data using the Integrated Genome Browser (http://igb.bioviz.org/). Fluorescence is very intense at some binding sites in polytene chromosomes (Figure 2.1), whereas it is largely absent or very faint at others (Figure 2.2). Strong bands correspond mostly to intragenic sites, whereas sites with no Su(Hw) immunostaining signal are found mostly in intergenic regions (see Tables 2.1 and 2.2).

Since we observed differences between polytene chromosome results and ChIP-chip data, we performed experiments to validate *in situ* as well as ChIP-chip data using

Table 2.1 Visible Su(Hw) immunostaining bands in polytene chromosome cytological subdivisions 56F-58A.

Reference gene	Su(Hw) modENCODE	Su(Hw) Bushey, et al.	Co-CP190 sites	Nucleotide position (2R)	Location in relation to gene	
CG11044 (1)	4	4	3	16102708 - 16123603	(4) intergenic	
Bl (2)	1	1	1	16490726		(1) intragenic
Dpr (3)	6	5	0	16586135 - 16626367	(1) intergenic	(5) intragenic
Actin57B (4)	4	3	3	16835497 - 16860183	(4) intergenic	
Treh (5)	2	2	2	16961395 - 16965920	(1) intergenic	(1) intragenic
King tubby (6)	4	4	0	16996021 - 17004778		(4) intragenic
CG4050 (7)	1	1		17014809		(1) intragenic
Glycogenin (8)	2	2	1	17087325 - 17096029	(1) intergenic	(1) intragenic
CG18375 (9)	2	2	1	17116585 - 17123159		(2)intragenic
Rgk3 (10.1)	1	1	0	17155651		(1) intragenic
CG30387 (10.2)	1	1	1	17186101		(1) intragenic
cv-2 (11)	2	2	1	17240608 - 17243293	(1) intergenic	(1) intragenic
Sdc (12)	10	9	0	17294077 - 17354975		(10) intragenic
Egfr (13)	1	1	1	17422744		(1) intragenic
CG10440 (14)	4	2	2	17458944 - 17484857	(3) intergenic	(1) intragenic
Total	45	40	16		(15) intergenic	(30) intragenic

Parentheses in the reference genes column indicate the Su(Hw) band number as seen in Figure 1. Reference genes are the genes closest to the sites involved in the formation of specific Su(Hw) immunostaining bands. Numbers in the Su(Hw) modENCODE, Su(Hw) Bushey et al. and Co-CP190 columns indicate the number of Su(Hw) binding sites involved in the formation of each Su(Hw) immunostaining band. Nucleotide position is assigned according to the center of the ChIP peak. Parentheses in the “Location in relation to gene” column indicate number of binding sites in intergenic or intragenic locations for each reference gene.

Table 2.2 Su(Hw) sites that do not associate with a visible immunostaining band in polytene chromosome cytological subdivisions 56F-58A.

Reference gene	Su(Hw) modENCODE	Su(Hw) Bushey, et al.	Co-CP190 sites	Nucleotide position (2R)	Location in relation to gene	
CG16898	2	0	1	15918411 - 15927927	(2) intergenic	
CG11041	1	0	0	16049313 - 16084419	(1) intergenic	
CG12484	7	9	0	16298449 - 16354910	(3) intergenic	(4) intragenic
Obp57c	1	1	0	16387959	(1) intergenic	
CG13430	1	0	1	16429609	(1) intergenic	
Sktl	6	5	2	16689311 - 16728385	(6) intergenic	
Ipk1	2	1	1	16758160 - 16766718	(1) intergenic	(1) intragenic
Otp	3	0	1	16772030 - 16782654		(3) intragenic
Rx	2	0	1	16805279	(1) intergenic	(1) intragenic
LBR	1	0	0	17617230	(1) intergenic	
CG30395	1	1	0	17651973	(1) intergenic	
Lox2	1	1	0	17672334	(1) intergenic	
CG4372	1	1	0	17732356	(1) intergenic	
FILI	1	1	0	17827320		(1) intragenic
CG13488	1	1	1	17837535	(1) intergenic	
CG34369	2	2	1	17848326 - 17857295	(1) intergenic	(1) intragenic
Total	33	23	9		(22) intergenic	(11) intragenic

Reference genes are the genes closest to the sites listed in the Su(Hw) modENCODE, Su(Hw) Bushey et al. and Co-CP190 columns. Parentheses indicate number of binding sites in intergenic or intragenic locations for each reference gene.

Chromatin Immunoprecipitation (ChIP) assays in embryos (Figure A.3). We selected strong, middle and weak or absent polytene signals, as well as the 5 sites within the *Sdc* gene that were used in *in situ* experiments. ChIP assays were performed using chromatin from *Drosophila* embryos and an antibody against Su(Hw). This was followed by quantitative real-time PCR. Primers flanking a subset of Su(Hw) binding sites were designed to assess the level of DNA enrichment after immunoprecipitation. Primers were designed to amplify a target sequence of approximately 150bp containing a Su(Hw) binding site identified using ChIP-chip data. Primers targeting insulator sequences from two genomic copies of the *gypsy* retrotransposon and primers targeting the *rp49* coding sequences were used as positive and negative controls, respectively (Table A.3). Results show that 12 of the 13 tested sequences interact positively with the Su(Hw) protein in vivo ($P<0.0001$). The level of enrichment, however, varies greatly between sites. In general, sites with strong immunostaining signals in polytene chromosomes correlate with a higher level of enrichment whereas weaker or more diffuse signals correlate with a lower level of enrichment (Figure A.3), suggesting that binding sites may have significant differences in the binding affinity of Su(Hw). In addition, data confirmed that sites 58A.8 and 58A.1 are present in embryos and are not occupied by Su(Hw) in polytene chromosomes.

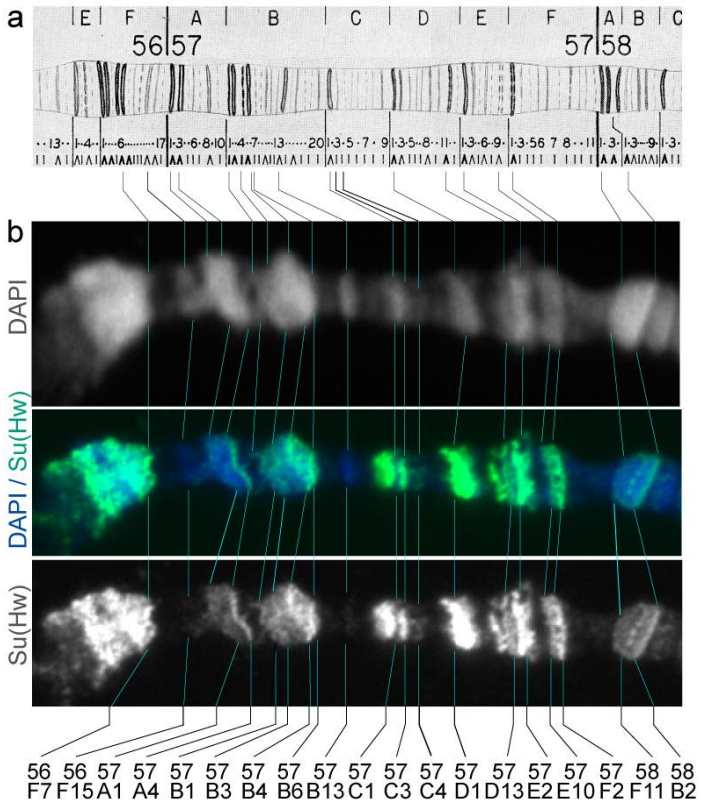
Su(Hw) binding sites are mostly associated with DAPI bands of condensed chromatin and are frequently found within the transcribed sequences of genes.

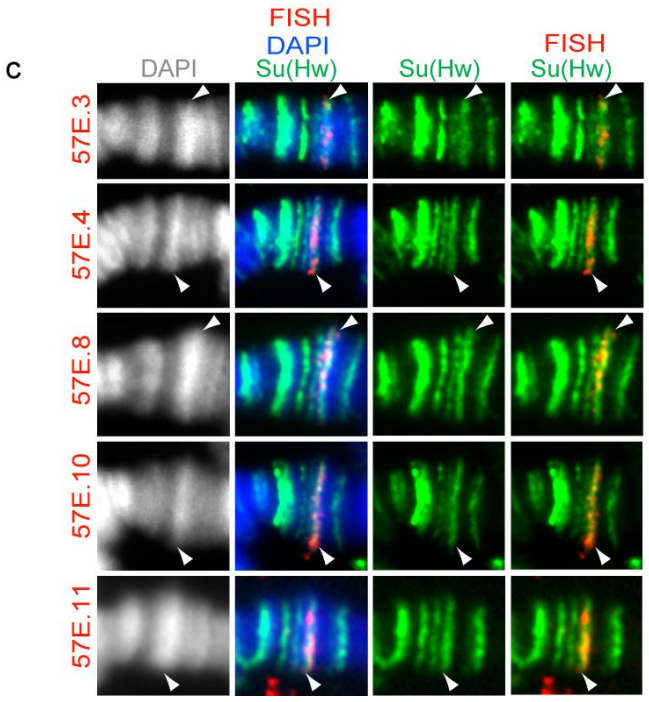
It has been already noted that Su(Hw) is frequently found within the transcribed

regions of genes (Adryan et al., 2007; Bushey et al., 2009). In the chromosome segment spanning 56F-58A subdivisions, 41 out of 78 sites identified in ChIP-chip assays are found within gene sequences (Table A.1). Out of 27 genes longer than 10kb, 16 are decorated with one or more internal Su(Hw) binding sites, and all genes spanning more than 20kb are associated with Su(Hw) binding sites in their transcribed sequences (Table A.1). Interestingly, visible bands of Su(Hw) in polytene chromosomes associate most often with insulator sites that localize within genes. Table 2.1 shows that, out of 45 Su(Hw) sites, 30 are intragenic. These Su(Hw) sites originate from the association of more than one binding site (frequently two sites) forming clusters that may span more than 50kb. On the other hand, Su(Hw) sites that appear as weak or absent Su(Hw) signals form no clusters and localize mostly with intergenic regions. Of the 33 sites in this category, only 11 are intragenic (Table 2.2). Statistical analysis using a contingency table shows that differences in the distribution of sites between intragenic and intergenic sites are significant (Chi-square=8.48; P=0.004). Together, these observations indicate that an important fraction of the condensed chromatin found in polytene chromosome bands corresponds to highly condensed DNA from large genes bound by Su(Hw) proteins. Accordingly, cytogenetic analysis of the 56F-58A subdivisions has revealed that all but 1 of the 15 strong Su(Hw) immunostaining signals visible in this region corresponds to a band distinguishable by DAPI staining, and all are traceable to specific binding sites mapping within genes (See Table 2.1 and Figure 2.3). In support of this observation we found that the only two intense DAPI bands (56F15

Fig. 2.3 Su(Hw) bands extensively colocalize with DAPI bands.

a. Cytological subdivisions 56F-58A, as depicted by Bridges (1939). **b.** The same cytological regions stained with DAPI and immunostained with Su(Hw) antibody. Correspondence between Bridges bands, DAPI bands and immunostaining bands is established. **c.** *In situ* hybridization combined with immunostaining using 5 different probes containing Su(Hw) binding sites shows that the *Sdc* gene completely colocalizes with a DAPI band and that all insulator signals coalesce in a single Su(Hw) band.


a
E F A B C D E F A B C
56 57
57 58
b
DAPI
DAPI / Su(Hw)
Su(Hw)
56 F7
56 F15
57 A1
57 A4
57 B1
57 B3
57 B4
57 B6
57 B13
57 C1
57 C3
57 C4
57 D1
57 D13
57 E2
57 E10
57 F2
58 F11
58 B2


c
DAPI
FISH
DAPI
Su(Hw)
Su(Hw)
FISH
Su(Hw)
57E.3
57E.4
57E.8
57E.10
57E.11

and 57B13-14) that completely lack Su(Hw) correspond to large stretches of DNA devoid of genes (Figure 2.3 and data not shown).

The *Sdc* gene is an example of a long gene containing multiple Su(Hw) binding sites within the transcriptional unit and is organized in condensed chromatin forming an intense DAPI band in polytene chromosomes. *Sdc* contains 10 Su(Hw) binding sites distributed along the approximately 90kb length of the transcriptional unit. *In situ* hybridization using probes specific for each of 5 sites shows a complete colocalization of all sites with the same Su(Hw) band (Figure 2.3C), supporting the idea that strong Su(Hw) immunostaining signals are comprised of clusters of binding sites located within genes. All 5 sites were also analyzed using ChIP to confirm that each of the sites were true Su(Hw) binding sites (see Figure A.3). ChIP results show that in chromatin isolated from embryos, each one of the sites found within the *Sdc* gene also interacts positively with Su(Hw) ($P<0.0001$). Results also show that all binding sites are located within a unique strong DAPI band. This DAPI band corresponds to cytological bands 57E2 to 57E6 in Bridges map, which normally are seen as a single band that appears to be generated by the condensation of the DNA of the *Sdc* gene (Figure 2.3C).

Su(Hw) sites in interbands display different properties than sites within condensed chromatin in bands.

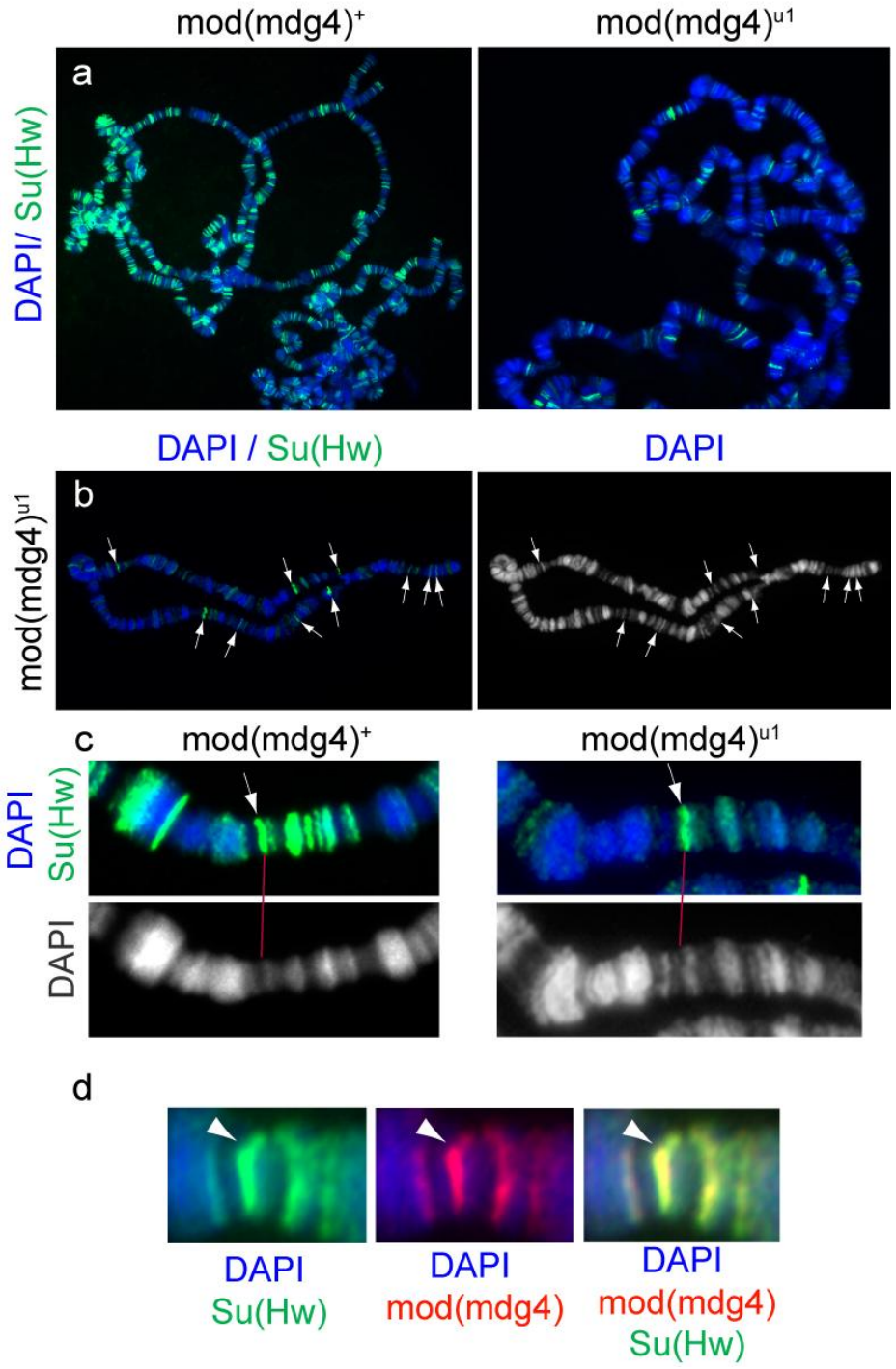
Since Su(Hw) interacts with the Modifier of mdg4 protein (Mod(mdg4)67.2) and mutations of *mod(mdg)67.2* have an effect on *gypsy* insulator function (Gerasimova and Corces, 1998), we examined whether mutations in the *mod(mdg4)* gene could also

influence the distribution of Su(Hw) in polytene chromosomes. It has been noted before that the amount of Su(Hw) protein in chromosomes is significantly reduced in *mod(mdg4)*^{u1} mutants (Gerasimova and Corces, 1998). An interpretation of this observation is that since Mod(mdg4)67.2 is part of the Su(Hw) insulator complex, the lack of this protein in *mod(mdg4)*^{u1} mutant flies could reduce the DNA binding affinity of Su(Hw). We have performed immunostaining experiments using Su(Hw) antibodies in *mod(mdg4)*^{u1} mutant chromosomes and observed the same global reduction in Su(Hw) immunostaining signal intensity (Figure 2.4A). However, a closer look at the distribution of the Su(Hw) proteins reveals that the amount of Su(Hw) protein associated with DAPI bands is significantly reduced in the mutants, whereas the Su(Hw) protein that localizes to interbands remains practically the same (Figure 2.4B). We have confirmed this general observation by comparing the distribution of Su(Hw) in the 56F-58A subdivisions in wild-type and *mod(mdg4)*^{u1} mutants. Results show that the levels of Su(Hw) protein associated with all DAPI bands is significantly reduced in mutants, whereas the signal levels in binding site 57B.26, the only site found in an interband, remain unchanged (Figure 2.4C). We have shown that Mod(mdg4)67.2 is also found at this site in wild-type flies by using immunostaining with antibodies raised specifically against Mod(mdg4)67.2 (Figure 2.4D).

Additional evidence suggesting that insulator proteins organize chromatin in bands in a different manner than chromatin in interbands comes from *in situ* hybridization results obtained using a 1kb DNA probe spanning Su(Hw) binding site 57B.26. Images revealed an *in situ* hybridization signal that extends outward from the

Fig. 2.4 Su(Hw) sites localizing to DAPI bands have different properties than sites localizing to interbands.

a. The amount of Su(Hw) in polytene chromosomes decreases significantly in *mod(mdg4)*^{u1} mutants. Only certain Su(Hw) signals remain as intense Su(Hw) bands. **b.** Intense Su(Hw) bands in *mod(mdg4)*^{u1} mutants map to interbands. **c**. Within the cytological subdivisions 56F-58A, only the Su(Hw) band that contains insulator sites 57B.25 and 57B.26, at the *Treh* gene, is clearly found in an interband and remains as a strong immunostaining signal in *mod(mdg4)*^{u1} mutants. **d**. Su(Hw) sites 57B.25 and 57B.26 at the *Treh* gene also contain Mod(mdg4) proteins in polytene chromosomes.

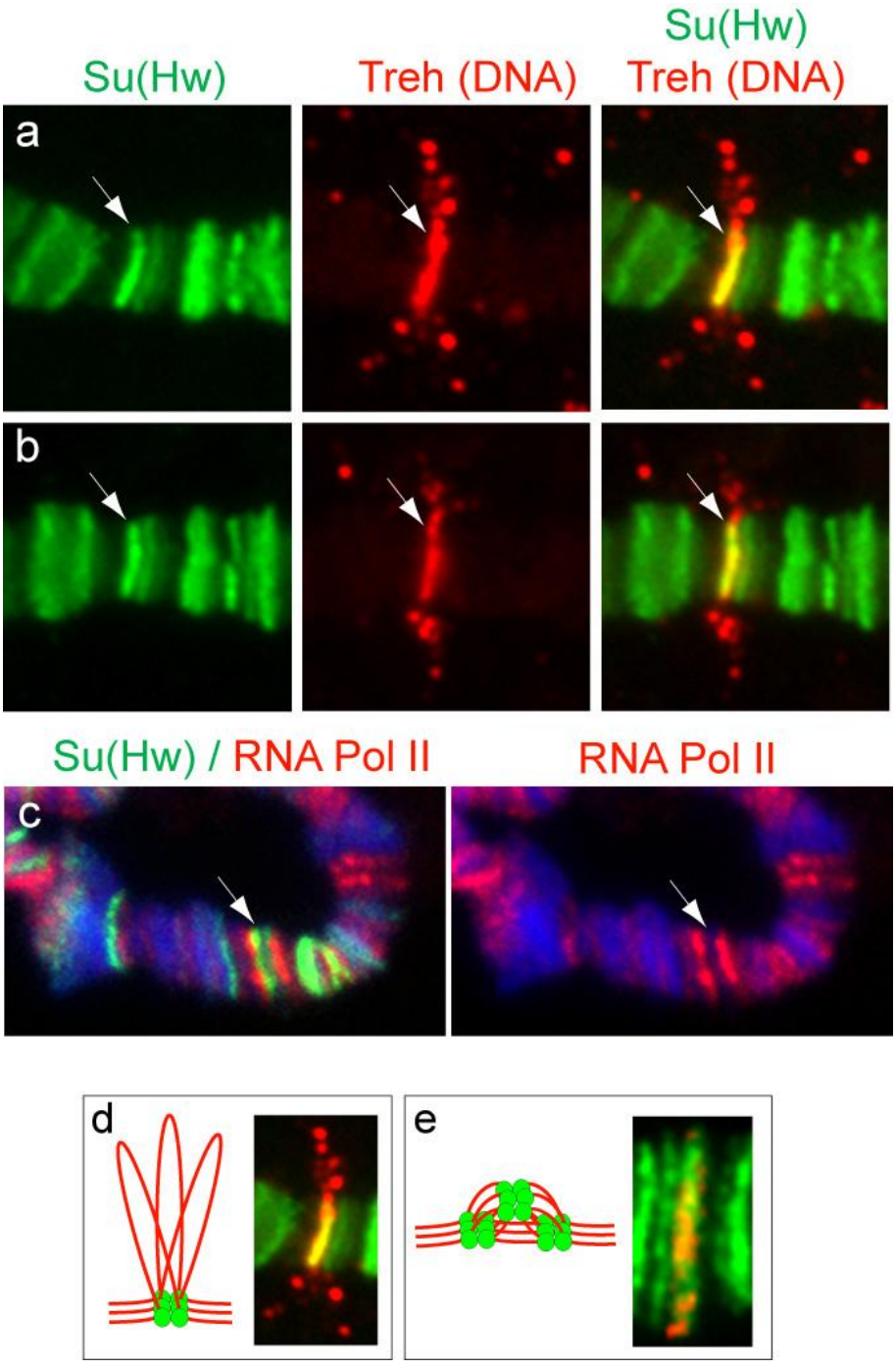

mod(mdg4)+
mod(mdg4)u1
a
DAPI/ Su(Hw)
DAPI / Su(Hw)
DAPI
b
mod(mdg4)u1
c
mod(mdg4)+
mod(mdg4)u1
DAPI
Su(Hw)
DAPI
d
DAPI
Su(Hw)
DAPI
mod(mdg4)
DAPI
mod(mdg4)
Su(Hw)

polytene chromosome (Figure 2.5). The *in situ* hybridization images shown in Figure 2.5 have been reproduced multiple times, originating from different larvae and different nuclei, suggesting that the signal observed may be a reflection of the more open DNA organization in this interband region. Experiments using RNA Pol II specific antibodies show that *Treh* is not highly transcribed in salivary glands during the third instar larval stage, suggesting that this region of open chromatin does not correspond to actively transcribing DNA forming a puff (Figure 2.5C).

To gain insight into whether highly condensed DAPI bands associated with Su(Hw) protein are irreversibly silenced or, alternatively, can be transcriptionally activated, we used the GAL4 binary system to ectopically activate transcription at the promoter of the *Sdc* gene, which is condensed forming a large DAPI band and contains 10 Su(Hw) binding sites (see Table A.1 and Figure 2.3C). To activate transcription at the *Sdc* promoter we have used the EP line $Sdc^{EY04602}$, which has a P element integrated 5' to the promoter of *Sdc* (Rorth, 1996). This EP line provides an *Hsp70* promoter with an upstream GAL4-UAS that can be used to activate transcription of downstream DNA sequences in the presence of the transcriptional activator GAL4. We have used a P(Act5C-GAL4) transgene as a GAL4 source and analyzed changes in the chromatin organization of the gene in polytene chromosomes using immunostaining. Figure A.4 shows how the structure of the polytene chromosome undergoes a dramatic reorganization at the *Sdc* DAPI band. This reorganization affects an area that spans more than 350 kb, well beyond the size of the 87.5kb *Sdc* gene (Figure A.4C). Changes

Fig. 2.5 Chromatin associated with Su(Hw) sites localizing to bands have a different organization than chromatin associated with sites localizing to interbands.

a and **b.** Two examples of FISH combined with immunostaining revealing that DNA associated with insulator site 57B.25, at the *Treh* gene, extends outward from the chromosome reflecting the underlying open chromatin organization in this region. **c**. Immunostaining using RNA pol II antibodies shows that transcription occurs to the left of the Su(Hw) site associated with *Treh*, suggesting that transcription is not taking place at the gene. Arrow is pointing to the *Treh* gene. **d.** and **e**. Proposed model illustrating the organization of DNA at Su(Hw) sites in interbands and in bands, respectively. Green circles indicate Su(Hw) protein at insulator sites. Red lines correspond to DNA.


Su(Hw)
Treh (DNA)
Su(Hw)
Treh (DNA)
a
b
Su(Hw) / RNA Pol II
RNA Pol II
c
d
e

in the structure of the polytene chromosome at the *Sdc* locus correlate with the disappearance of the DAPI bands as well as Su(Hw) and CP190, another insulator protein that interacts with Su(Hw), in the same region. To confirm that Gal4 is activating transcription, and that these changes are mediated by RNA polymerase II activity, we performed immunostaining experiments using the phosphoserine 5 specific H14 antibody, which recognizes the active form of RNA Pol II. Results show that RNA Pol II accumulates at a large disorganized puff formed at the site corresponding to *Sdc* (Figure 2.6). In addition, it appears that RNA Pol II never colocalizes with Su(Hw), suggesting that during passage of RNA Pol II Su(Hw) proteins are removed from the DNA (Figure 2.6B).

Some large DAPI bands contain clusters of genes that are devoid of insulator proteins.

Despite the high level of association of Su(Hw) with chromosome bands, there are three large DAPI bands in which the presence of Su(Hw) is limited to a diffuse or completely absent immunostaining signal (see Figure 2.2 and Table 2.2). These bands map approximately at sites 57A1-4, 57B3-B4 and 57F11-58B2 in Bridges map (Figure 2.3). In order to test whether this distribution is unique to Su(Hw) or is common to other insulator proteins, we performed immunostaining of polytene chromosomes using antibodies specific against the CP190 protein. Results show that the distribution of CP190 proteins is similar to that of Su(Hw) and the same DAPI bands in which Su(Hw) appears diffuse, completely lack CP190 protein (Figure 2.7A). This result suggests the

Fig. 2.6 Phosphorylated RNA Pol II decorates transcriptionally activated *Sdc* DNA. Activation of *Sdc* by Gal4 in $Sdc^{EY04602}$ mutants can be monitored by immunostaining of polytene chromosomes using an antibody specific against phosphorylated RNA Pol II (**a**). Co-immunostaining of Su(Hw) and phosphorylated Pol II shows that after transcriptional activation, Su(Hw) remains associated with dense DAPI stained areas forming broken bands. In the puffed region, Su(Hw) is observed as small grains that normally do not colocalize with the RNA Pol II dots (**b**) suggesting that passage of RNA Pol II evicts Su(Hw) proteins from their binding sites.

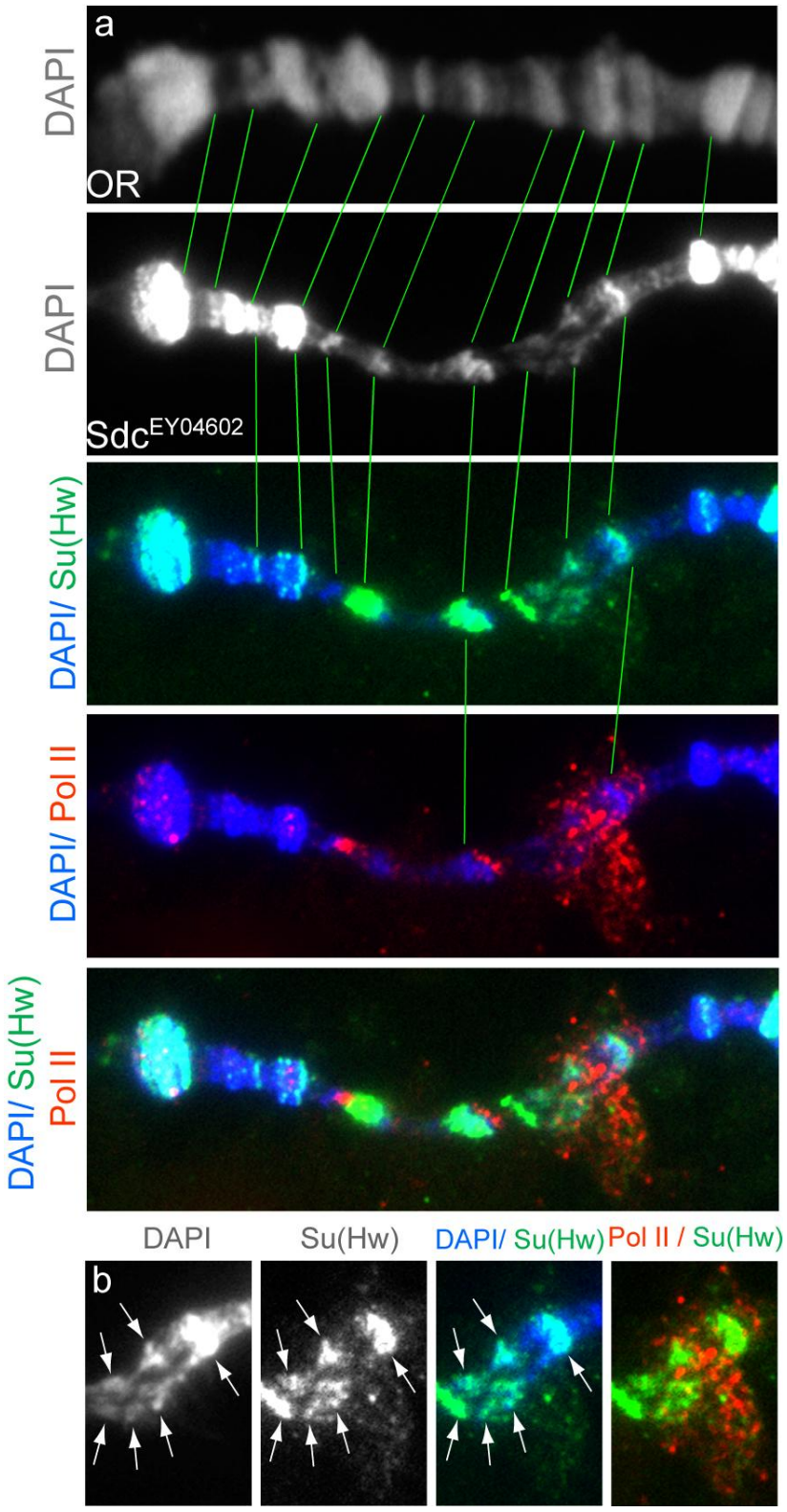
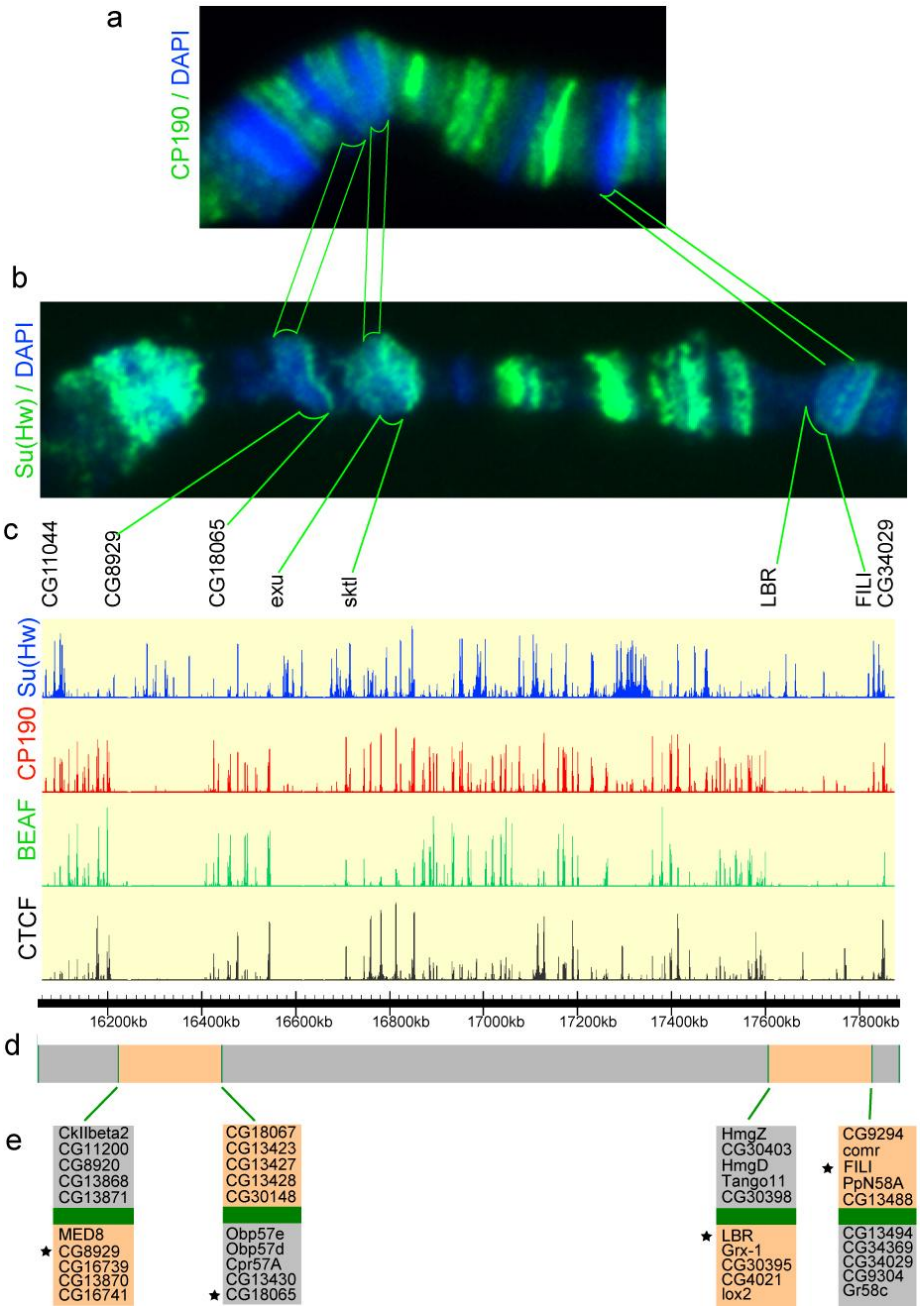

a
DAPI
OR
DAPI
Sdc^EY04602
DAPI/ Su(Hw)
DAPI/ Pol II
DAPI/ Su(Hw)
Pol II
DAPI
Su(Hw)
DAPI/ Su(Hw)
Pol II / Su(Hw)
b

Fig. 2.7 Domains of coexpressed genes overlap with regions with a low concentration of insulator proteins in cytological subdivisions 56F-58A.

Distribution of insulator proteins Su(Hw) and CP190 on polytene chromosomes (**a** and **b**) and Su(Hw), CP190, BEAF, and CTCF using tiling microarray data (**c**) is compared with the distribution of clusters of coexpressed genes defined by Spellman and Rubin (2002). **d**. Graphical representation of the distribution of clusters of coexpressed genes (in orange) and clusters of non-coexpressed genes (grey). Boundaries between clusters are indicated in green (see Table A.2). Boxes in E show genes found close to boundaries between clusters. Genes flanking the clusters as defined by CP190 insulator proteins are indicated in E by stars.


a
CP190 / DAPI
b
Su(Hw) / DAPI
c
CG11044
CG8929
CG18065
exu
sktl
LBR
FILI
CG34029
Su(Hw)
CP190
BEAF
CTCF
16200kb
16400kb
16600kb
16800kb
17000kb
17200kb
17400kb
17600kb
17800kb
d
e
CkIIbeta2
CG11200
CG8920
CG13868
CG13871
MED8
CG8929
CG16739
CG13870
CG16741
CG18067
CG13423
CG13427
CG13428
CG30148
Obp57e
Obp57d
Cpr57A
CG13430
CG18065
HmgZ
CG30403
HmgD
Tango11
CG30398
LBR
Grx-1
CG30395
CG4021
lox2
CG9294
comr
FILI
PpN58A
CG13488
CG13494
CG34369
CG34029
CG9304
Gr58c

possibility that chromatin organization associated with these DAPI bands in polytene chromosomes is different from that of the DAPI bands that colocalize with intense Su(Hw) immunostaining signals.

To further test the nature of these large DAPI bands that do not colocalize with Su(Hw) immunostaining signals and to determine the distribution of other insulator proteins in such bands, we have analyzed insulator proteins Su(Hw), CP190, BEAF-32, and CCCTC-binding factor (CTCF) using microarray data from modENCODE and Bushey et al. (2009). Whereas the analysis of the Su(Hw) binding sites shows that there is not a significant difference in Su(Hw) site distribution along the 56F-58A cytological subdivisions, binding sites for BEAF-32, CP190 and to a lesser extent CTCF, are largely absent from DNA sequences spanning each one of the three DAPI bands mentioned above (Figure 2.7). The general lack of insulator proteins led us to ask whether the genes found within these DAPI bands are clustered into domains of coregulated gene expression.

The most important reference to clusters of genes that appear to be coexpressed in *Drosophila* came from experiments using multiple sets of expression microarrays by Spellman and Rubin (Spellman and Rubin, 2002). These experiments suggested that genes tend to be organized in the *Drosophila* genome by clustering into groups of adjacent genes that are coexpressed. These clusters of coexpressed genes are flanked by generally larger sets of genes that are not coexpressed. Using Spellman and Rubin data (Spellman and Rubin, 2002), we have mapped all gene clusters within the 56F-58A cytological region (Table A.2). Remarkably, two of the large DAPI bands (57A1-4 and

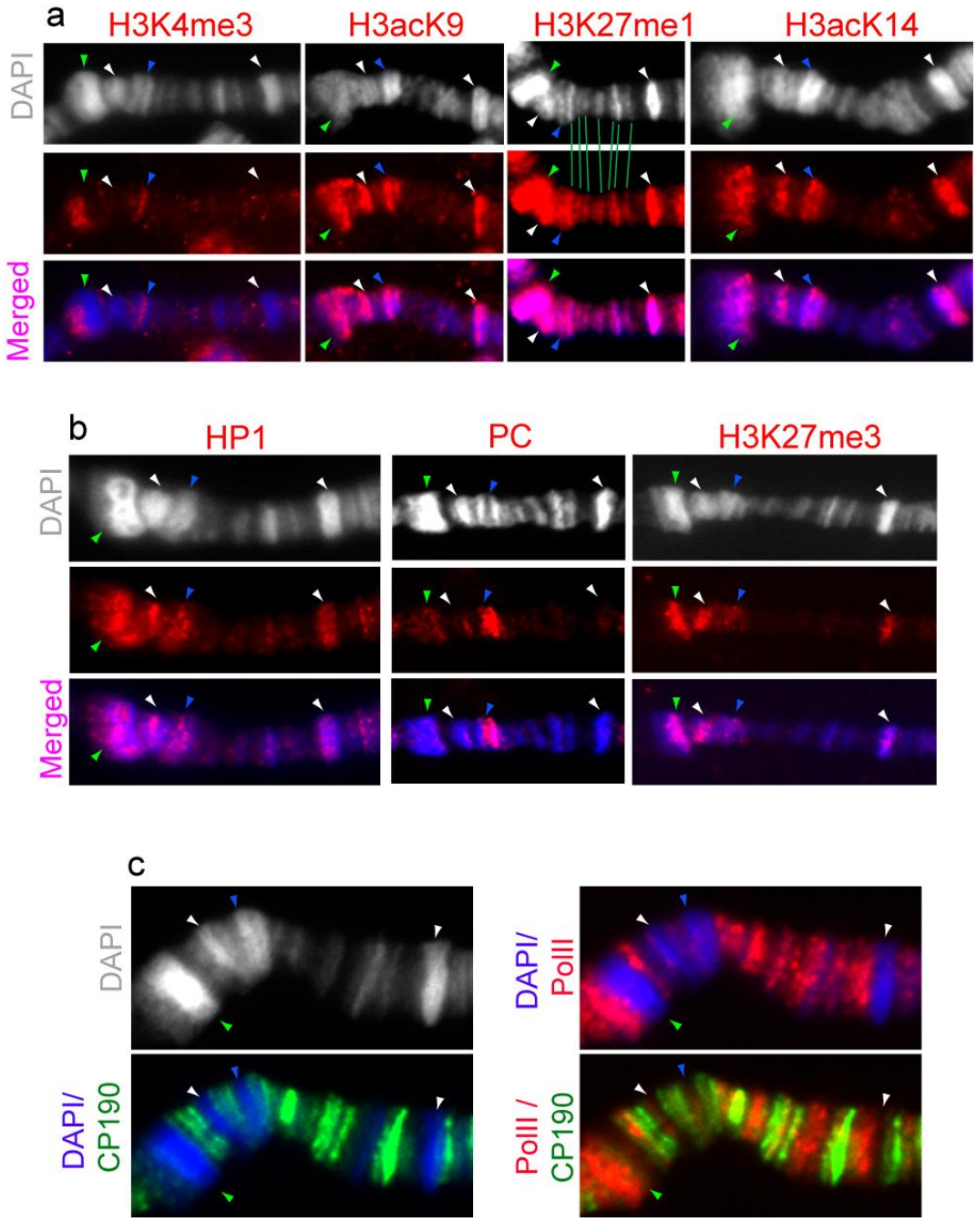
57F11-58B2) shown in (Figure 2.7B) precisely map to domains previously defined as clusters of coexpressed genes (Figure 2.7D). In contrast, the sharp and well defined Su(Hw) immunostaining signals coincide with regular DAPI bands and are located within the clusters of genes that are not coexpressed (Figure 2.7B and D).

It is important to mention that inaccuracies in the description of the boundaries of clusters or even in the description of complete clusters are expected in the data obtained by Spellman and Rubin, given that experiments were realized using a very early annotated version of the *Drosophila* genome (Manak et al., 2006). If there is a true correlation between lack of insulator proteins and clusters of coexpressed genes, these inaccuracies could explain why the third DAPI region (57B3-B4), which also shows no binding of insulator proteins in polytene chromosomes and lacks binding sites of insulator proteins in embryos as well, does not appear as a cluster of coexpressed genes in Spellman and Rubin data (Spellman and Rubin, 2002).

Both CP190 and BEAF-32, and to some extent CTCF, have been largely associated with gene promoters (Bushey et al., 2009; Jiang et al., 2009) suggesting that the absence of these proteins from these gene clusters indicates that these chromosome segments are enriched in genes that lack insulator proteins at their promoters. Comparison of polytene chromosome data with tiling microarray data obtained from embryos suggests that the lack of the Su(Hw) protein in sites associated with clusters of coexpressed genes in salivary glands is tissue specific, since these sites are occupied by Su(Hw) in embryos. Given that immunostaining experiments do not show significant levels of RNA Pol II associated with these clusters (Figure 2.8), it can

Fig. 2.8 Clusters of coexpressed genes appear in polytene chromosomes as highly condensed chromatin and intense DAPI bands, poor in insulator proteins and decorated with both repressive as well as active chromatin marks.

a and **b.** Immunostaining of cytological region 56F-58A using antibodies against chromatin proteins HP1, PC and histone modifications H3K4me3, H3acK9, HeK27me1, H3acK14, and H3k27me3. White arrowheads point to DAPI bands corresponding to clusters of coexpressed genes included within the cytological regions 56F-58A. Green arrowheads point to a cluster of coexpressed genes adjacent to 56E but not included in this study (Table A.2). Blue arrowheads point to a dense DAPI band that colocalize with a region devoid of CP190 and BEAF proteins. This region was not defined by Spellman and Rubin (2002) as belonging to a cluster of coexpressed genes. Only HeK27me1 binds to all polytene bands. Green bars connect DAPI bands from the DAPI channel to HeK27me1 bands. **c.** CP190 immunostaining and RNA Pol II are not observed within the domains of coexpressed genes (Green and white arrowheads).


a
H3K4me3
H3acK9
H3K27me1
H3acK14
DAPI
Merged
b
HP1
PC
H3K27me3
DAPI
Merged
c
DAPI
DAPI/
CP190
DAPI/
PolII
PolII /
CP190

be speculated that without transcription, Su(Hw) does not bind to sites localized to coexpressed gene clusters. Therefore, the presence of Su(Hw) associated with these sites in embryos would be an indication that these genes are actively transcribed in only a fraction of embryo tissues, as it is expected if the genes are developmentally regulated. On the other hand, the lack of CP190, BEAF-32 and CTCF proteins from promoters of genes forming coexpressed gene clusters could be interpreted as a structural characteristic that can possibly define these clusters.

Domains with low abundance of insulator proteins are enriched with repressive as well as with active chromatin marks.

To acquire further understanding of the chromatin organization associated with bands, interbands and with clusters of genes associated to large DAPI bands containing no-insulator proteins, we have performed immunostaining of polytene chromosomes using antibodies against specific histone modifications and chromatin proteins. We have used antibodies raised against chromatin proteins that associate with repressive chromatin such as Polycomb (PC) and Heterochromatin protein 1 (HP1) (Eissenberg and Elgin, 2000; Fanti and Pimpinelli, 2008; Schwartz and Pirrotta, 2007). We have also used antibodies against histone H3 tri-methylated at lysine 27 (H3K27me3), a histone post-translational modification linked to transcriptional repression (Cao et al., 2002; Plath et al., 2003; Sarma et al., 2008), as well as antibodies against histone modifications H3 acetylated at lysine 9 (H3K9ac), H3 acetylated at Lysine 14 (H3K14ac) and H3 tri-methylated at Lysine 4 (H3K4me3), which are normally

associated with transcriptional activation (Barski et al., 2007; Jenuwein and Allis, 2001; Ng et al., 2003; Santos-Rosa et al., 2002; Wang et al., 2008). In addition, we have also used H3 mono-methylated at lysine 27 (H3K27me1), which was previously identified because of its association with heterochromatin but also with DAPI bands as well as with active transcription (Barski et al., 2007; Labrador et al., 2008).

Results confirm that H3K27me1 is generally associated with DAPI bands since it is found in all DAPI bands within the 56F-58A cytological region, including bands that correspond to clusters of coexpressed genes (Figure 2.8A). H3K27me1 is the only chromatin mark with a broad distribution within the cytological region 56E-58A. All other marks showed a much more restricted distribution, limited to only a few sites within the same region (Figure 2.8A and B). PC is enriched in a single region, which coincides with one of the large DAPI bands (57B3-B4) lacking insulator proteins, but did not correspond to a previously identified cluster of coexpressed genes (Figure 2.7, Table A.2). This region also contains Su(Hw) binding sites described in embryos that are not visible in polytene chromosomes (Figure A.3, Figure 2.8, Table 2.2). Surprisingly, we found that this same region is the only site with a significant H3K4me3 immunostaining signal (Figure 2.8A). Both H3K4me3 and PC have opposite functions and are not expected to colocalize. However, the resolution of the immunostaining at this point is not sufficient to rule out the possibility that both PC and H3K4me3 actually map to the same sequences. Antibodies against RNA Pol II show that this region is not actively transcribing (Figure 2.8C).

All other markers, including H3K27me1, H3K9ac, H3K14ac, H3K27me3 and HP1, strongly associate with the three DAPI bands lacking insulator proteins that could correspond to putative domains of coexpressed genes present in this region (Figure 2. 8 and Table A.2). In addition to these domains, a third region spanning genes CG8654 to CG11007 (nucleotides 15390584 to 15872262) is strongly stained with all antibodies with exception of H3K4me3 and also corresponds to a cluster of coexpressed genes that is equally poor in insulator proteins (See Table A.2). These results suggest that the chromosome regions previously identified as containing clusters of coexpressed genes are largely defined by a low abundance of insulator proteins and are associated with a specific subset of histone modifications that include both transcriptionally active as well as repressive chromatin marks. The role that this chromatin structure and absence of insulator proteins play in these domains is still unclear but intriguingly is reminiscent of the chromatin structure associated with bivalent domains in vertebrates (Bernstein et al., 2006; Ku et al., 2008).

The band-interband structure and their association with insulator proteins are largely conserved between different tissues.

To determine whether the distribution of insulator proteins as well as the pattern of bands and interbands observed in polytene chromosomes changes significantly between tissues, we have compared the 56F-58A cytological region from salivary gland chromosomes with the corresponding region in ovarian nurse cell chromosomes. Polytene chromosomes from nurse cells were obtained from 3 to 5 days old otu^{7}/otu^{11}

females, which have polytene chromosomes that are comparable in size to those of salivary glands (Mal'ceva et al., 1997). The band-interband morphology as well as *in situ* hybridization combined with immunostaining was used to adequately locate the 56F-58A cytological region in nurse cell chromosomes. Results show that the band-interband pattern is largely conserved between both tissues (Figure 2.9). These results suggest that the well defined band-interband pattern of polytene chromosomes is maintained between different tissues. In addition to morphological conservation, the distribution of the Su(Hw) protein is largely conserved as well.

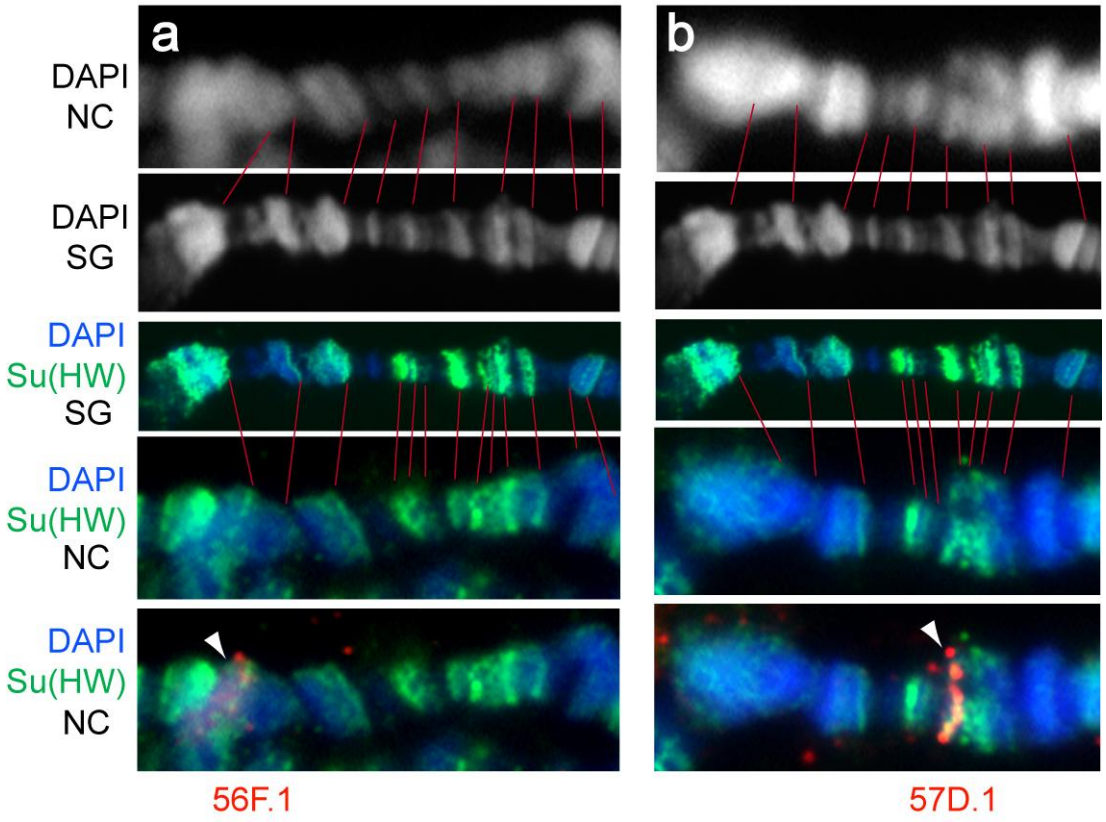


Fig. 2.9 The DAPI and Su(Hw) banding patterns are conserved between salivary gland cells (SG) and Nurse Cells (NC) in cytological subdivisions 56F-58A.

Su(Hw) immunostaining combined with FISH using specific probes identify subdivisions 56F-58A in nurse cells (**a** and **b**). Arrowheads indicate *in situ* hybridization signal. Lines between NC and SG chromosomes indicate the correspondence of DAPI and Su(Hw) staining between NC and SG cells.

DISCUSSION

To further understand the role that insulator proteins have in chromosome structure and function, we have used polytene chromosomes to study the distribution of insulator protein sites. In this work, we have identified all Su(Hw) sites, mapped by ChIP-chip tiling array data, in a 2Mb region of polytene chromosome 2R. Our data shows that the Su(Hw) binding sites within the cytological region 56F-58A map to three different categories of chromatin organization: bands, interbands, and multi-gene domains of coexpressed genes. We propose a model in which the association of insulator proteins defines the structure and the properties of each of the different levels of chromatin organization (see Figure 2.10).

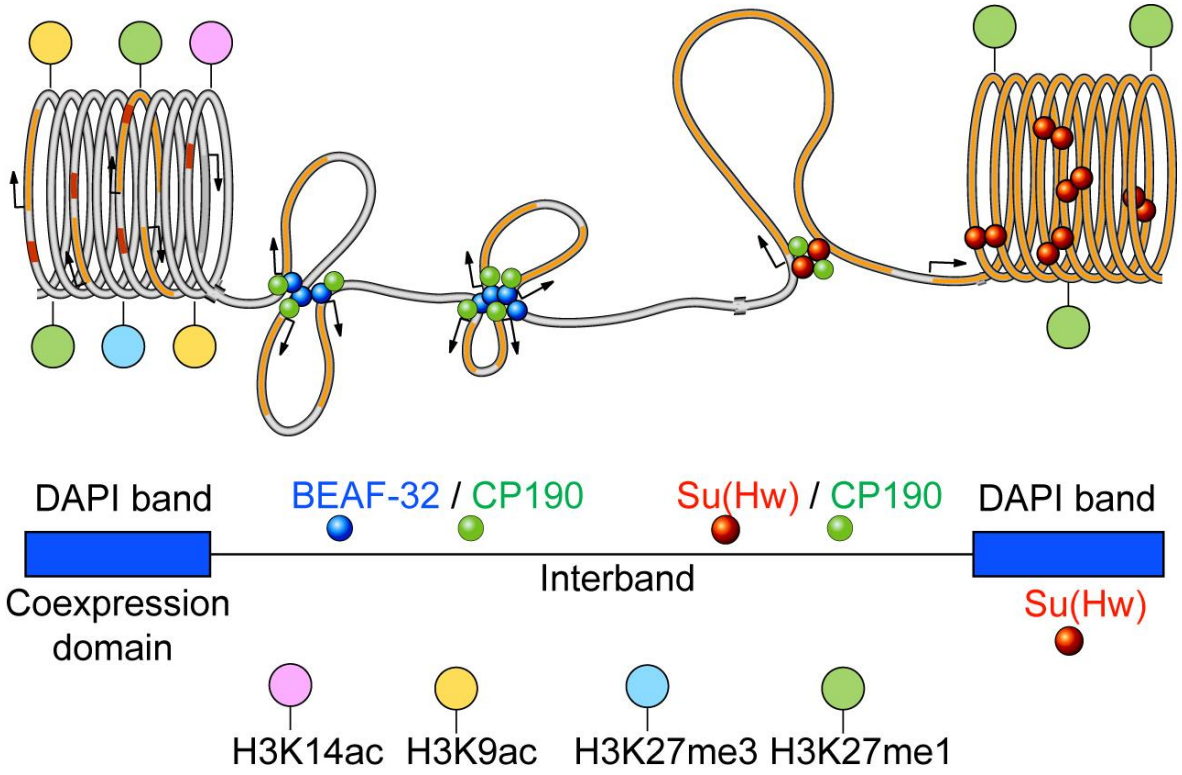


Fig. 2.10 A model integrating the band-interband pattern in polytene chromosomes with the distribution of insulator proteins and gene organization. Chromatin fiber is indicated as a grey cord. Orange fragments within the chromatin fiber correspond to transcriptional units. Broken arrows correspond to gene promoters. Different insulator proteins are indicated by spheres and are color encoded. Short red segments correspond to Su(Hw) binding sites unoccupied by the Su(Hw) protein. Histone modifications are indicated by color encoded flags. Condensed chromatin is indicated by spirals, which map to bands, whereas non-condensed chromatin localizes to interbands. Su(Hw) localizes to all three types of chromatin organization.

DAPI bands frequently correspond to condensed chromatin pertaining to silent gene sequences bound to Su(Hw) insulator proteins.

It is remarkable that all Su(Hw) bands identified in this study, except for two sites at the *Treh* gene, map within a DAPI band. In fact all the DAPI bands in the region (except for 2 that are devoid of coding sequences) either are associated with Su(Hw) immunostaining signals or have functional binding sites for Su(Hw) uncovered in embryos. The significance of this association remains unclear, but our data suggests that Su(Hw) insulator proteins in these sites can have an active role in the process of transcription, by facilitating chromatin decondensation during the process of transcriptional activation or, alternatively, a repressive role by actively participating in the process of chromatin condensation that leads to the formation of bands. We have shown that condensed chromatin associated with Su(Hw) remains potentially active, since it can be efficiently decondensed by passage of RNA pol II. When the *Sdc* gene is ectopically activated by Gal4, the RNA pol II is capable of inducing the formation of large loops of DNA, generating a prominent puff at the site that was originally occupied by a large DAPI band. DNA in the induced puff is highly enriched in RNA Pol II, which rarely colocalizes with Su(Hw) proteins. This observation suggests that in large genes containing Su(Hw) sites, the interaction of Su(Hw) with the DNA can be disrupted after transcriptional activation. The eviction of Su(Hw) would also disrupt higher order interactions between Su(Hw) sites, which in turn may help in the process of chromatin decondensation.

Su(Hw) in interbands associates to open chromatin.

Experiments in Figure 2.4 show that in the absence of Mod(mdg4)67.2, Su(Hw) proteins remain stably bound to DNA in interbands, whereas the binding to DNA in bands becomes ostensibly weaker, suggesting that Su(Hw) may have alternative functions depending on whether it is targeted to chromosome bands or to interbands. We have shown that the Su(Hw) site 57B.26 associated with the gene *Treh* unequivocally maps to an interband, and that *in situ* hybridization images show a signal which may be indicative of the open chromatin structure of interbands. Interestingly, microarray data shows that there is an additional Su(Hw) binding site, 57B.25, approximately 5kb upstream of the 57B.26 site, which maps within the transcribed region of the *Trehalase* (*Treh*) gene (Table A.2). These two sites cannot be resolved as two independent immunostaining signals (see also Figure A.2 and Figure 2.1). It is possible that the distance between these two sites is too small to allow resolution of the two independent Su(Hw) bands using our technique. However, an interesting possibility is that the signal observed by *in situ* hybridization actually corresponds to DNA loops that are anchored to the two interacting Su(Hw) binding sites (see Figure 2.5 panels a,b and d). Previous reports have shown that insulator proteins differentially localize to bands or to the band-interband boundaries in polytene chromosomes (Capelson and Corces, 2004; Pai et al., 2004; Zhao et al., 1995), but to our knowledge this is the first time that the distribution of insulator proteins has suggested that they may have differentiated roles in bands and interbands.

Insulator proteins may help organize the genome into differentiated transcriptional domains.

In situ hybridization combined with immunostaining revealed that three chromatin domains in the cytological region 56F-58A appear as large DAPI bands that lack Su(Hw) and CP190 proteins. Analysis of tiling microarray data from embryos reveals that insulator proteins CP190, BEAF-32 and CTCF are also largely absent from the same domains ((Bushey et al., 2009) and modENCODE consortium). In contrast, the same data shows that Su(Hw) binding sites are found both within the domains and outside the domains. Therefore, true Su(Hw) binding sites within the domains as determined by microarray experiments in embryos, are not bound by Su(Hw) in polytene chromosomes where genes are transcriptionally silent (Figure 2.7).

We also show that 2 of these DAPI bands map exactly to the only 2 clusters of coexpressed genes described by Spellman and Rubin in this region (Spellman and Rubin, 2002). Clusters of coexpressed genes are common to eukaryotic organisms including *Drosophila*, *Caenorhabditis elegans* and humans. These clusters are ordered non-randomly in the genome and consists of similarly expressed multi gene domains (Boutanaev et al., 2002; Caron et al., 2001; Dorus et al., 2006; Lee and Sonnhammer, 2003; Miller et al., 2004; Roy et al., 2002; Versteeg et al., 2003; Yi et al., 2007). This result opens the possibility that the domains of coexpressed genes could actually be defined by a low abundance or a complete absence of insulator proteins such as CP190, BEAF-32 and CTCF. The lack of all insulator proteins would explain the coexpression of all the genes in the domain, and the activity of insulator proteins within

the clusters of non-coexpressed genes would explain the lack of coexpression, but this possibility raises the question of the role that Su(Hw) plays in the domains.

A rapid inspection of the classes of genes found in each type of multi-gene domain defined by their coexpression level suggests that genes found in non-coexpressed multi gene domains frequently have CP190 and BEAF-32 at their promoters, are highly expressed, show very short intergenic distances between them and encode proteins with essential cellular functions. For example, it has been shown that most genes regulated by BEAF-32 are involved in basic functions, such as cell cycle regulation or chromosome segregation (Emberly et al., 2008). Genes found in the domains of coexpressed genes, however, lack CP190 and BEAF-32, are generally tissue specific, are developmentally regulated and show large intergenic distances.

Our data shows that the organization described above is largely conserved between tissues, given the minimal differences found when salivary gland cells and nurse cells are compared (Figure 2.9). In both tissues the band-interband pattern as well as the distribution of Su(Hw) protein is practically identical. It is tempting to speculate that this organization is not specific to salivary glands or nurse cells, but instead may largely represent how *Drosophila* chromosomes are generally organized regardless of cell lineage. If this hypothesis is correct, it would be expected that major differences in the band-interband pattern as well as in the distribution of Su(Hw) protein between cell lineages would only occur upon induction or tissue-specific activation of genes, such as *Sdc*, residing within the bands, whereas there would be genes that will always localize to interbands regardless of the cell type. Such genes will correspond to

genes encoding proteins required for basic cell functions such as those having BEAF-32 at their promoters or inducible genes such as *Treh*. Inducible genes have to remain in a chromatin state that will allow rapid activation but would stay off in the absence of induction in a variety of tissues and cell types (see proposed model in Figure 2.10). The specific role of Su(Hw) in the formation of these higher order structures or domains remains obscure, since mutations at the *su(Hw)* gene alone do not significantly change the band-interband pattern or the morphology of polytene chromosomes, nor do they have severe effects on viability (Smith and Corces, 1995). However, since most bands and interbands are conserved between very unrelated tissues it is expected that the condensed chromatin associated to bands or to domains of coexpressed genes would form before cell differentiation, when there is still a supply of maternally deposited Su(Hw) and would be transmitted epigenetically during cell division.

The fact that histone modifications normally associated with transcriptional activation are found within the same domains as histone modifications normally associated with transcriptional repression is reminiscent of the bivalent domains initially found in stem cells and deserves additional characterization. In depth analysis assessing the chromatin properties of bands, interbands, and clusters of coexpressed genes should provide further insights into the role of insulators in the organization of these chromatin domains and whether they play an active role in gene expression regulation during development and cell differentiation.

CHAPTER III

Analysis of the role of Su(Hw) protein binding sites in transposable elements

ABSTRACT

Chromatin insulators are thought to play an important role in higher-order chromatin organization and the regulation of gene expression. The *gypsy* insulator is a 350-bp element which is located in the *Drosophila gypsy* retrovirus, and contains twelve binding sites for the Suppressor of Hairy-wing [Su(Hw)] protein. At least one other retrotransposon has been shown to possess insulator activity, but it remains unclear whether insulators might be a common occurrence among transposable elements. Here we used a Su(Hw) consensus sequence to predict Su(Hw) binding sites within the canonical sequences of a number of *Drosophila* transposable elements. Data shows that Su(Hw) is enriched at predicted binding sites within *jockey* and *gtwin elements,* and that *jockey* possesses enhancer-blocking activity. Su(Hw) appears to have a tissue-specific effect on expression of these elements in a manner that appears to be unrelated to the presence of Su(Hw) binding sites. Results suggest that the effect of Su(Hw) on TE expression is not related to control of these elements via the piRNA pathway. Finally, mutations in the *mod(mdg4)* and *su(Hw)* genes have different effects on the transcript levels of TEs in the same tissue.

INTRODUCTION

Transposable elements (TEs) are selfish DNA entities which can use a host genome for survival and propagation, and they are abundant components in the genomes of most living organisms (Almeida and Allshire, 2005; Kaminker et al., 2002). TEs are divided into two classes based on their mechanism of transposition. Class I elements consist of the retrotransposons, which transpose by an RNA intermediate. The Class I retrotransposons are divided into LTR elements, which contain long terminal repeats (LTRs), and non-LTR elements, such as long interspersed elements and short interspersed elements (LINEs and SINEs). Class II elements consist of the transposons, which transpose by DNA excision and integration into a new site (Slotkin and Martienssen, 2007).

Mobile elements, which exist in the genome of most organisms, have the potential to have deleterious effects on the fitness of the host (Brennecke et al., 2007). The mutagenic activity of TEs is particularly critical in the germline, where transpositions have the potential to introduce heritable mutations or chromosomal rearrangements (Vagin et al., 2006). For this reason, organisms have developed mechanisms to regulate transposable elements to assure that most TEs remain silent and inactive in the germline (Zaratiegui et al., 2007). For example, mobility of a number or elements, including *gypsy*, is controlled by the *flamenco* locus located on the X chromosome (Desset et al., 2008; Mevel-Ninio et al., 2007; Prud'homme et al., 1995). Accumulation of transcripts of the elements controlled by this locus is repressed in the germline by a small interfering RNA pathway which depends on the Argonaute proteins Piwi,

Argonaute3 (Ago3) and Aubergine (Aub), and Piwi interacting short RNAs (piRNAs) (Pelisson et al., 2007; Sarot et al., 2004).

It has recently been found that piRNAs are involved in two distinct pathways, composed of different components, which function in ovarian germline and somatic cells (Malone et al., 2009). In the germline, Aub and Ago3 associate with piRNAs derived from clusters that produce small RNAs from both genomic strands which are involved in suppression of a broad spectrum of elements. Piwi was found to act alone in the soma in a pathway that is driven by the *flamenco* cluster. The somatic piRNA pathway functions to silence retroviral elements that may propagate by infecting neighboring germ cells. piRNAs derived from the *flamenco* piRNA cluster are enriched for sequences of the *gypsy* family of LTR elements (Malone et al., 2009).

Although it was once thought that TEs contributed little or none to the fitness of their host, it is clear that some TEs provide functions that are beneficial. For instance, the *Het-A* and *TART* retrotransposons are involved in the maintenance of telomere ends in *Drosophila* (Pardue and DeBaryshe, 2003). TEs may help to drive evolution of new genes based on their ability to frequently insert in protein coding genes and alter both sequence composition and the expression pattern of these or other nearby genes, sometimes even functioning as enhancers or promoters (Girard and Freeling, 1999). Retrotransposons have also been shown to act as boundaries, defining blocks of transcriptionally active and silent chromatin. (Purbowasito et al., 2004). The ability of transposons to shape the genome of their host in this way makes transposable elements an important source of genetic variation during evolution.

The best known example of a retrotransposon which can function as a boundary comes from the *Drosophila gypsy* retrovirus, in which both enhancer-blocking and boundary functions are found within an insulator located in its 5' LTR. The *gypsy* insulator requires several proteins for its function, including Suppressor of Hairy-wing [Su(Hw)] Modifier of mdg4 67.2 [Mod(mdg4) 67.2] and Centrosomal Protein of 190 kD (Pai et al., 2004; Spana and Corces, 1990; Spana et al., 1988). The Su(Hw) protein contains a stretch of 12 zinc finger motifs at its central portion. A number of the 12 zinc finger motifs are involved in recognizing and binding the 5'-YRYTGCATAYBY-3' repeats in the *gypsy* insulator DNA sequence (Harrison et al., 1993; Kim et al., 1996). The C-terminal acidic and leucine zipper domains of Su(Hw) are involved in its interaction with the Mod(mdg4)67.2 protein (Gdula and Corces, 1997). Null mutations in *su(Hw)* cause female sterility due to degeneration of the egg chambers before completion of oogenesis (Harrison, 1993).

In addition to *gypsy*, at least one other LTR retrotransposon has been shown to possess insulator activity. Transgenic assays revealed that the 5' LTR of *Idefix* possesses both enhancer-blocking as well as boundary function (Conte et al., 2002a). Additionally, *ZAM* may also possess an insulator within its 5' UTR which shows some degree of enhancer-blocking capability (Minervini et al., 2010). Evidence that *ZAM* can act as a transcriptional enhancer, as well as the finding that both of these elements possess insulator activity, indicates that transposable elements are capable of influencing transcriptional regulation of the host genome (Conte et al., 2002a, b). Perhaps one way in which TEs might have influenced genome organization is by

contributing to the spread of chromatin insulators. However, it is unclear whether insulator activity is unique to these elements or if it is a common occurrence among TEs.

We sought to determine whether other transposable elements might contain binding sites for some insulator proteins. By searching for Su(Hw) consensus binding sites within the canonical sequences of *Drosophila* transposable elements, we identify a number of TEs with predicted Su(Hw) sites. ChIP assays show that Su(Hw) binds significantly to only a few elements containing predicted sites, including *jockey*, which we show possesses enhancer-blocking activity in transgenic assays. Analysis of transcripts from a number of elements indicates that Su(Hw) may have a tissue-specific effect on expression of these elements in a manner that is independent of the presence of Su(Hw) binding sites. This effect of Su(Hw) on TE expression does not appear to be related to control of these elements via the piRNA pathway. Finally, mutations in *mod(mdg4)* and *su(Hw)* affect transcript levels of TEs differently in the same tissue.

MATERIALS AND METHODS

Analysis of Su(Hw) binding sites in TEs

ChIP-on-chip data files for Su(Hw) (accession number GEO GSE16245) from embryos were downloaded from the modENCODE website (http://www.modENCODE.org). Signal intensities were converted to normalized log2 ratios using Tiling Analysis Software (Affymetrix). The CisGenome software (Ji et al., 2008) was used to obtain sequence data corresponding to the genome regions enriched

for the Su(Hw) protein. Sequence data was submitted to the MEME motif discovery program (http://meme.sdsc.edu/) (Bailey and Elkan, 1994) in order to generate the consensus sequence and position weight matrix for Su(Hw) binding sites. The position weight matrix was submitted to FIMO, part of the MEME sequence analysis suite, along with the canonical sequences of all *Drosophila* transposable elements, in order to search for Su(Hw) sites within the TEs. Canonical transposable element sequence files were downloaded from the Berkeley *Drosophila* Genome project website (http://www.fruitfly.org/p_disrupt/TE.html).

In Situ Hybridization and immunostaining of polytene chromosomes

Approximately 500bp DNA fragments corresponding to each endogenous insulator were obtained by PCR (see primers and probe sizes in table A.4). Biotin-labeled DNA was prepared using the Biotin High-Prime random priming kit (Roche). The labeled probe DNA was ethanol precipitated and resuspended in hybridization buffer (4× SSC, 50% formamide, 1× Denhardts, and 0.4 mg/ml of salmon sperm DNA). Polytene chromosomes obtained from salivary glands of third instar larvae were dissected in 0.7% NaCl and fixed in a 1:2:3 mixture of acetic acid/water/lactic acid. Slides were heated at 65°C in 2×SSC for 30 min, dehydrated in an ethanol series, and denatured in 0.07 M NaOH.

For hybridization of DNA, boiled probes were added to the slide, covered immediately with a coverslip, and incubated at 37°C overnight in a humidified chamber. Following hybridization, coverslips were removed and the slides were washed in 2×

SSC at 37°C, then at room temperature in 1× PBS, and finally in antibody dilution buffer. The slides were incubated overnight in dilution buffer containing a 1:300 dilution of Su(Hw) primary antibody. Slides were then washed in antibody dilution buffer and incubated with a 1:300 dilution of FITC- or Texas red-conjugated goat anti-rabbit IgG (Jackson Laboratories) for 2 hr at room temperature. Slides were stained for 30 seconds with 4′,6-diamidino-2-phenylindole (DAPI; 0.5 µg/ml) and mounted in Vectashield mounting medium (Vector Laboratories). Slides were analyzed using a Leica DM6000B fluorescence microscope. Antibodies specific against Su(Hw) protein were raised using an N-terminal peptide containing the first 218 amino acids of the protein. Antibodies were raised in rats and rabbits by Pocono Rabbit Farm and Laboratory. (Canadensis, PA 18325 USA) and were validated using westerns and by co-immunostainings with previously characterized rat or rabbit antibodies.

Chromatin immunoprecipitation

Chromatin was prepared from 17-h embryos collected on grape juice agar media. Embryos were homogenized in buffer A1 (60 mM KCl, 15 mM NaCl, 4 mM MgCl2, 15 mM HEPES (pH 7.6), 0.5% Triton X-100, 0.5 mM DTT, 10 mM sodium butyrate, 1X EDTA-free protease inhibitor cocktail (Roche) containing 1.8% formaldehyde. Crosslinking was stopped by adding 225 mM glycine solution. Cells were lysed and chromatin was sheared to an average length of 500-700bp by sonication. In each ChIP experiment, a chromatin solution corresponding to 200 mg of live material was incubated with either Su(Hw) antibody or normal rabbit IgG. Immunoprecipitation and

washing were performed as described elsewhere (Cavalli, 1999). The same Su(Hw) antibody used in immunostaining experiments was used in ChIP assays.

Real-time PCR quantification analysis of immunoprecipitated DNA

Real-time PCR quantification of immunoprecipitated DNA was carried out with ABGene (Rockford, IL) SYBR green PCR master mix. For input PCRs, a 1/100 dilution was used as template. Primers were designed to amplify 100-200bp fragments. PCR conditions for each primer pair were tested to determine the efficiency of amplification and to ensure amplification was in the linear range. PCR products for each primer pair were amplified from at least three separate immunoprecipitation products from at least two different chromatin preparations using the BioRad iQ5 Multicolor Real-Time PCR detection system (Primers listed in Supplemental Table 3). Enrichment of immunoprecipitated DNA fragments was calculated using the ΔCt method based on the threshold cycle (Ct) value for each PCR reaction (BioRad real time PCR application guide). Results are presented as percentage of total input. The statistical significance of the results was calculated by Student's t-test.

Enhancer-blocking assay

The *yellow* gene was used as a reporter for the enhancer-blocking capability of the predicted Su(Hw) binding site located in *jockey*. The pUASY vector, constructed from the *Drosophila* P element transformation vector pUAST, contains the promoter, coding and intron regions of the *yellow* gene and contains unique restriction sites *Age*I-

*Asc*I-*Avr*II between the wing enhancer and body enhancer and *BsiW*I-*Nhe*I-*Fse*I between the body enhancer and promoter. *Jockey* was amplified from Oregon R genomic DNA by PCR using primers designed with terminal restriction sites or *Fse*I and *Nhe*I (Primers listed in Table A.2). For enhancer blocking assays, the *jockey* element was cloned between the wing and body enhancer or between the body enhancer and promoter, generating pUASY-JY. The construct was microinjected into w^{1118} flies (Bestgene Inc., CA). Phenotypes were determined by crossing transgenic flies into the yw^{67c} background and scoring pigmentation in the wing and body cuticle of the offspring. Flies were raised on standard cornmeal and agar medium at 25°C. Insulator activity was assessed by comparing pigmentation with transgenic flies containing only empty pUASY vector. Phenotypes were analyzed using a Leica MZ16FA stereomicroscope and pictures were taken using a Leica DFC420 digital camera.

RNA isolation and Reverse Transcriptase PCR

Salivary gland RNA was obtained from 60 pairs of salivary glands dissected from 3rd instar larvae. Ovary RNA was obtained from 10 pairs of ovaries, dissected from 3-5 day old adult females. Embryonic RNA was isolated from approximately 200 0-2 hour embryos. Tissue samples were homogenized in TRIzol reagent and incubated at RT for 5 min. The mixture was centrifuged at 10,000 g for 10 min and the supernatant was transferred into a fresh eppendorf tube. Chloroform was added (200 ul/ml TRIzol) and the sample was vortexed for 15 sec. followed by a 3 min incubation at RT. Samples were then centrifuged at 12,000 g for 15 min at 4°C and the aqueous phase was

transferred to a fresh tube. Samples were mixed with Isopropyl alcohol (500ul/1ml TRIzol) and incubated for 10 minutes at RT. Samples were centrifuged (12,000 g for 10 min) and the RNA pellet was washed with 75% ethanol and centrifuged for 5 minutes at 7500 g. RNA was dissolved in RNase-free water and the sample was DNase treated using TURBO DNA-free DNAse kit to remove genomic DNA (Applied Biosystems). Approximate concentration of total RNA was calculated using the NanoDrop spectrophotometer (Thermo Scientific). RT-PCR was performed using the SuperScript First-strand cDNA synthesis kit (Invitrogen) according to manufacturer's instruction. Primers used are listed in table A.4.

Real-time PCR quantification analysis of TE expression

Real-time PCR quantification of TE expression was carried out with ABGene (Rockford, IL) SYBR green PCR master mix. For input PCRs, a 1/100 dilution was used as template. Primers were designed to amplify 100-200bp fragments. PCR conditions for each primer pair were tested to determine the efficiency of amplification and to ensure amplification was in the linear range. PCR products for each primer pair were amplified from cDNA using the BioRad iQ5 Multicolor Real-Time PCR detection system (Primers listed in Table A.4). cDNA was reverse transcribed from at least three different RNA samples, with the exception of the analysis of expression in y^2ct^6;*mod(mdg4)*^{u1}/*Tb* and y^2ct^6;*mod(mdg4)*^{u1}/ *mod(mdg4)*^{u1} ovaries (Figure 3.10). In this case, only a single biological replicate was used for *stalker, gtwin, baggins, beagle, 412, Idefix,* and *Tabor*. Ct values were normalized to those for the housekeeping gene *rp49*. Change in

expression level was calculated using the ΔΔCt method based on the threshold cycle (Ct) value for each PCR reaction (BioRad real time PCR application guide). Results are presented as fold change in mutant relative to wild-type. The statistical significance of the results was calculated by Student's t-test.

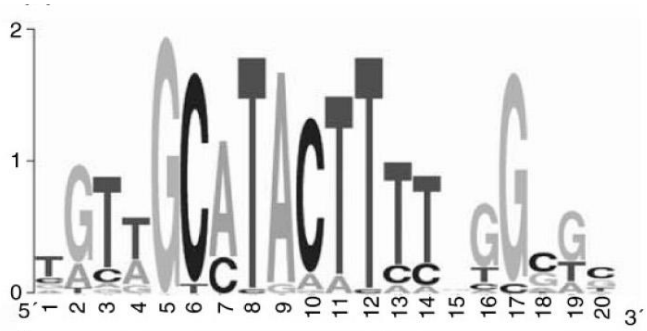
RESULTS

The DNA sequence of a large fraction of *Drosophila* TEs contain Su(Hw) binding sites.

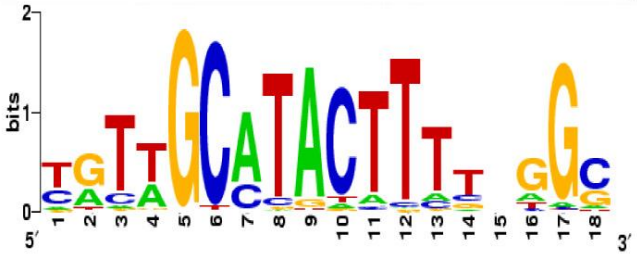
Data on endogenous *gypsy* insulators suggests that the insulator found in the 5'untranslated region of the *gypsy* retrotransposon is exceptional in that it contains 12 tandem repeats of the Su(Hw) binding site, whereas the great majority of endogenous Su(Hw) insulators only contain one binding site (Ramos et al., 2006). These differences raise the question of whether the properties of endogenous insulators with only one binding site are equivalent to the properties observed in *gypsy* insulators. In addition, the role that the insulator plays in the biology of the *gypsy* retrovirus is still unknown. Although other transposable elements appear to have insulator properties (Brasset et al., 2010; Conte et al., 2002a; Minervini et al., 2010), it is unknown whether *gypsy* insulator sequences are common among TEs. We investigated whether the incorporation of insulator sequences into *gypsy* was a unique event among transposable elements or, alternatively, whether the Su(Hw) sites provide an advantage to *gypsy* that could also be utilized by other elements.

We have used ChIP-on-chip tiling array data from modENCODE to generate a Su(Hw) binding site consensus sequence which we used to predict putative Su(Hw)-binding sites in transposable elements. Regions enriched for Su(Hw) were obtained from the modENCODE project website (http://www.modencode.org). CisGenome (Ji et al., 2008) was used to obtain sequence data corresponding to the genomic regions enriched for the Su(Hw) protein, and these sequences were submitted to the MEME motif discovery program (Bailey and Elkan, 1994). Because MEME only accepts up to 60000 characters at a time, the 5109 sequences enriched for Su(Hw) were partitioned into 12 groups so that sequences in each group did not exceed 60 kb. A similar genomic Su(Hw) binding consensus was found in all 12 groups submitted to MEME, and letter frequencies of Su(Hw) binding motifs, shown as a sequence logo, were very similar in all groups (Figure 3.1B). Additionally, the consensus found was in agreement with consensus sequences from recently published studies using ChIP-on-chip analysis to identify endogenous Su(Hw) binding sites (Figure 3.1A) (Adryan et al., 2007; Bushey et al., 2009). A position weight matrix of one group obtained from MEME was generated (Figure 3.1C) and submitted to FIMO to search for Su(Hw) binding sites within the sequences of all canonical transposable elements. 166 different consensus sequences were found in 95 different transposons. These TEs include LTR and non-LTR retroelements as well as several DNA elements, suggesting that the role of the insulator may not be limited to elements that transpose by an RNA intermediate. Some of these candidate sequences in the transposons were selected for further analysis.

A



B



C

	A	C	G	T
1	0.079810	0.320741	0.054041	0.545409
2	0.319990	0.031524	0.571178	0.077308
3	0.054291	0.134351	0.034276	0.777083
4	0.335752	0.009507	0.062047	0.592695
5	0.010008	0.005004	0.980235	0.004754
6	0.003252	0.961471	0.003252	0.032024
7	0.745059	0.235176	0.009757	0.010008
8	0.015261	0.060045	0.017763	0.906930
9	0.913935	0.008256	0.063297	0.014511
10	0.032775	0.876657	0.029522	0.061046
11	0.073805	0.043533	0.009257	0.873405
12	0.014261	0.034276	0.013510	0.937953
13	0.090568	0.088566	0.032274	0.788591
14	0.074056	0.190393	0.114086	0.621466
15	0.298724	0.207906	0.323493	0.169877
16	0.157118	0.040030	0.684764	0.118089
17	0.054041	0.016262	0.915937	0.013760
18	0.074556	0.633225	0.241181	0.051038

background letter frequency in Drosophila genome

A	C	G	T
0.288000	0.212000	0.212000	0.288000

Fig. 3.1 Search for Su(Hw) binding sites within sequences of transposable elements.

(A) WebLogo of the published consensus sequence from ChIP-on-chip analysis of endogenous Su(Hw) binding sites (Adryan et al., 2007). (B) Sequence logo of consensus obtained from MEME using modENCODE ChIP-on-chip data. (C) Position weight matrix corresponding to the sequence logo obtained from MEME.

A number of Su(Hw) binding sites correspond to Su(Hw) signals in polytene chromosomes.

To determine whether sequences found in TEs correspond to Su(Hw) binding sites, we first performed fluorescence *in situ* hybridization in salivary gland polytene chromosomes from Oregon R flies. DNA probes of approximately 1 kb were generated using primers against the region in the canonical sequence of these TEs which contains the predicted Su(Hw) binding sites. Hybridization of these probes to polytene chromosomes was combined with immunostaining using an antibody against Su(Hw). TEs analyzed included the non-LTR elements *jockey*, *X-element*, and *Doc3* as well as the LTR elements *nomad* and *rooA*. Each element has multiple insertions in the genome, therefore multiple *in situ* hybridization signals are visible for each TE. Each of the transposable elements tested shows various degrees of colocalization with the Su(Hw) protein (Figures 3.2 and 3.3). *Jockey* and *nomad* clearly show the highest number of sites which correspond to Su(Hw) immunostaining signals. *Jockey* is found at approximately 50 sites, mainly along the chromosome arms, with Su(Hw) protein visible at about 40 of these (Figure 3.2A). Probes for *nomad* shows 15 copies along the chromosome arms, and Su(Hw) can be seen at 13 (Figure 3.2B). *RooA* and *Doc3* elements show a slightly lower correlation between *in situ* hybridization signal and Su(Hw) immunostaining signals. *RooA* is the most abundant of the elements we tested, visible at approximately 100 sites, however Su(Hw) is only visible at about 50% of these (Figure 3.3A). *Doc3* exhibits a similar ratio, with Su(Hw) colocalizing at 11 out of 20

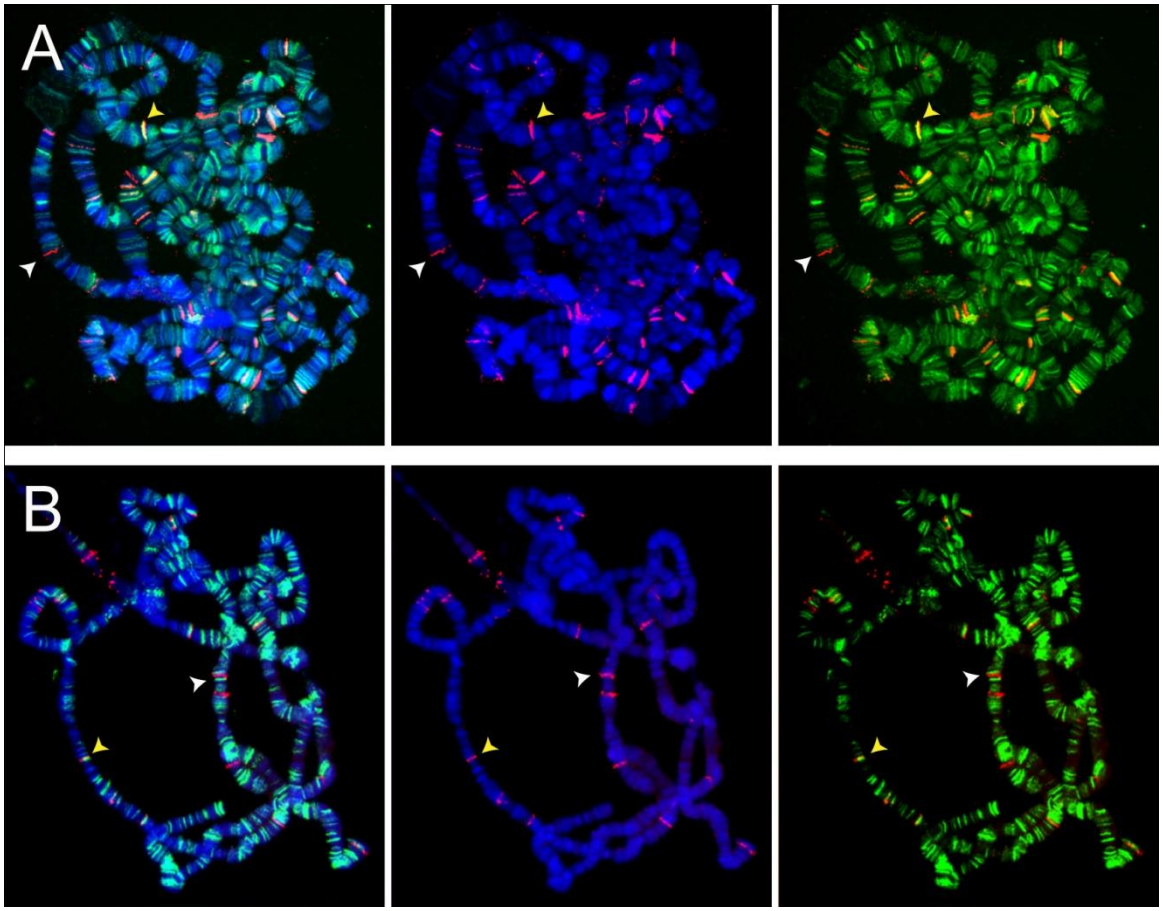


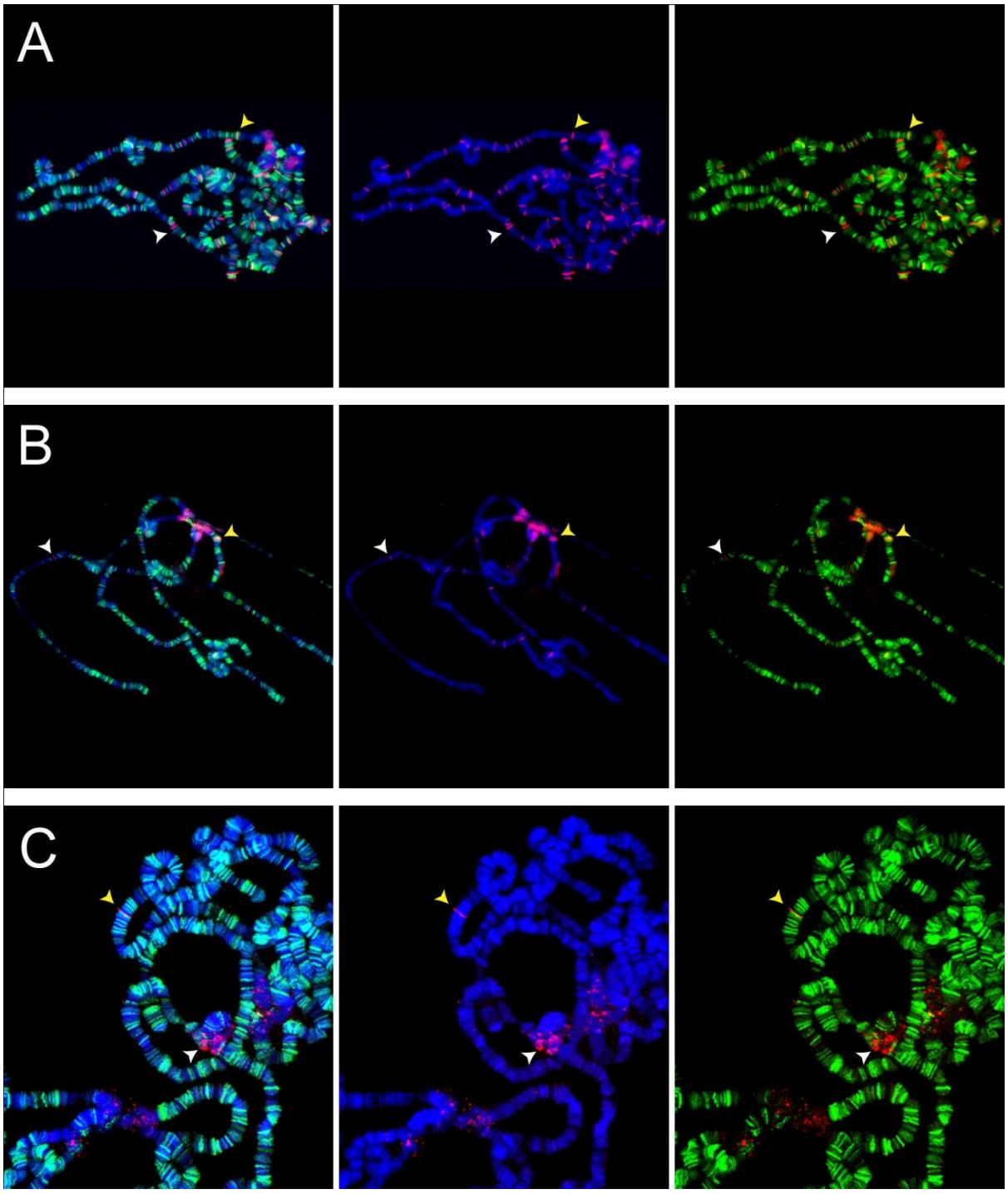
Fig. 3.2 Predicted Su(Hw) binding sites within *jockey* and *nomad* extensively colocalize with Su(Hw) on polytene chromosomes.

Fluorescence *in situ* hybridization signals corresponding to labeled genomic DNA probes from *jockey* **(A)** and *nomad* **(B)**, which contain predicted Su(Hw)-binding sites are shown in red. Chromosomes were labeled with antibodies against Su(Hw), shown in green. DNA, shown in blue, was visualized with DAPI. Yellow arrowheads indicate regions in which the *in situ* hybridization signals colocalize with Su(Hw) immunostaining signals, while white arrowheads indicate regions in which no colocalization is observed.

sites (Figure 3.3B). *X-element* contained the fewest number of copies and is located predominantly in centromeric heterochromatin with 1 or 2 sites along the chromosome visible at approximately 100 sites, however Su(Hw) is only visible at about 50% of these (Figure 3.3A). *Doc3* exhibits a similar ratio, with Su(Hw) colocalizing at 11 out of 20 sites (Figure 3.3B). *X-element* contained the fewest number of copies and is located predominantly in centromeric heterochromatin with 1 or 2 sites along the chromosome arms. Only the *X-element* site located outside of the centromere colocalized with Su(Hw), and no colocalization was observed with the Su(Hw) immunostaining signals in both centromeric sites (Figure 3.3B). These results suggest that a number of the sequences identified by computer searches belong to regions of TEs that interact with the Su(Hw) protein in vivo. Nevertheless, low amounts of colocalization observed in some elements may suggest that binding of Su(Hw) protein to the sequences may be dependent on genome context.

In order to determine whether there is a correlation between *in situ* hybridization and immunostaining data in polytene chromosomes, and to verify binding of the Su(Hw) protein to these sites in vivo, we have performed Chromatin Immunoprecipitation (ChIP) assays. Chromatin was isolated from *Drosophila* embryos and was immunoprecipitated with an antibody against Su(Hw) to pull down protein-DNA complexes containing Su(Hw) binding sites. To assess the level of DNA enrichment after immunoprecipitation, this procedure was followed by quantitative real-time PCR using primers flanking the predicted sites in each element. The *gypsy* retrotransposon and *rp49* coding sequences were used as positive and negative controls, respectively. Results from ChIP assays

Fig. 3.3 Predicted Su(Hw) binding sites within with *rooA, Doc3,* and *X-element* show varying degrees of colocalization with Su(Hw) on polytene chromosomes. Fluorescence *in situ* hybridization signals corresponding to labeled genomic DNA probes from *rooA* **(A)**, *Doc3* **(B)**, and *X-element* **(C)**, which contain predicted Su(Hw)-binding sites, are shown in red. Chromosomes were labeled with antibodies against Su(Hw), shown in green. DNA, shown in blue, was visualized with DAPI. Yellow arrowheads indicate regions in which the *in situ* hybridization signals colocalize with Su(Hw) immunostaining signals, while white arrowheads indicate regions in which no colocalization is observed.


A
B
C

show that only *jockey* exhibits a significant enrichment for the Su(Hw) protein in vivo ($P<0.0001$), with a level of enrichment that is half of that seen in *gypsy* (Figure 3.4). The level of enrichment in ChIP assays does not appear to correlate with observations from *in situ* hybridization and immunostaining experiments. For example, although *jockey* and *nomad* show a similar amount of colocalization with the Su(Hw) protein in polytene chromosomes, only *jockey* exhibits a significant level of enrichment in ChIP experiments. These data indicate that the majority of predicted sites within the TEs are not bound by the Su(Hw) protein in embryos.

A *jockey* insulator has enhancer-blocking activity.

To analyze whether the predicted Su(Hw) binding site sequence within *jockey* might possess enhancer-blocking properties, we have performed a transgenic enhancer-blocking assay. We used a pUAST-Yellow reporter vector, containing the *yellow* (*y*) gene coding region as well as the corresponding upstream regulatory sequences, which is designed to test interactions between insulators cloned into a pUAST transformation vector. The *y* gene is involved in the production of pigmentation in the larval and adult cuticle structures, and its regulatory sequences include several tissue-specific enhancers required for expression in various tissues, including wing blades, body cuticle, and bristles. Mutations at the *yellow* locus affect pigment production, changing coloration of cuticle structure from brownish-black to yellow (Geyer et al., 1986). Several tissue-specific enhancers, including the upstream

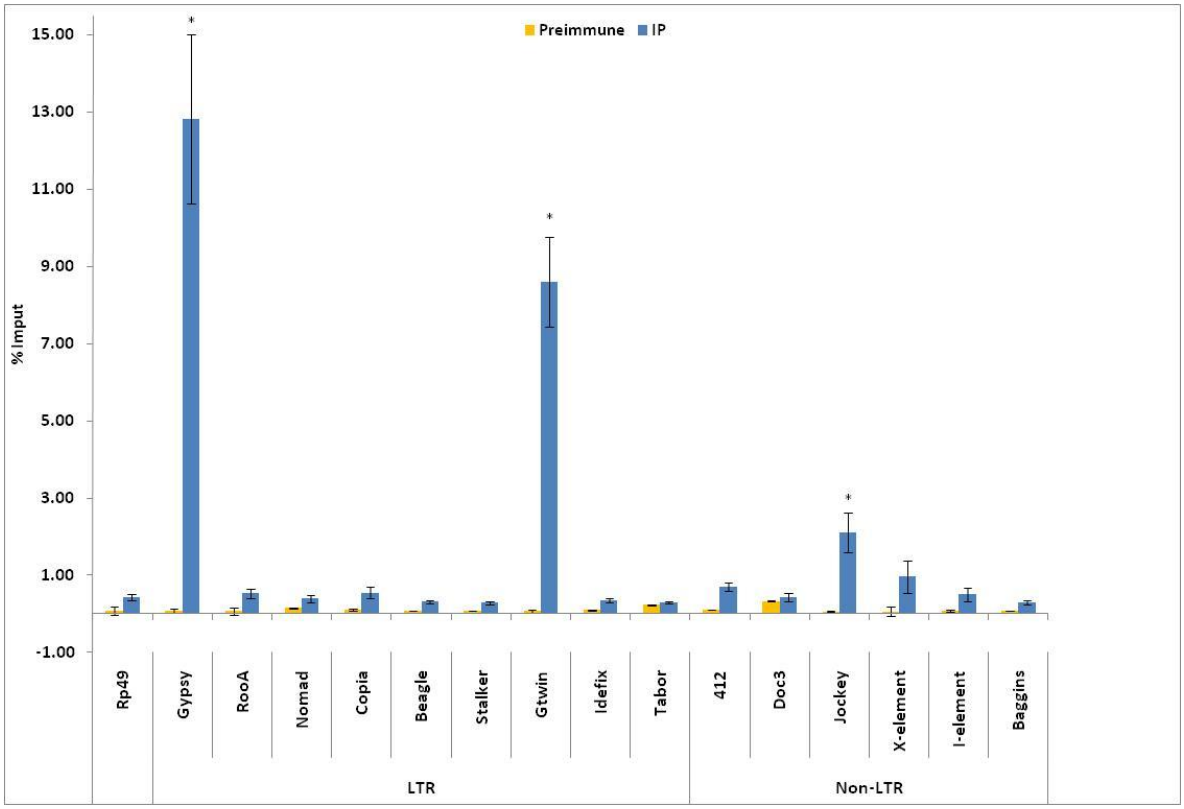


Fig. 3.4 Chromatin immunoprecipitation analysis of predicted Su(Hw) binding sites. Percentage of input obtained from ChIP experiments using predicted binding sites in 14 different transposable elements. Chromatin from Drosophila embryos was immunoprecipitated with an antibody against. Su(Hw)The *gypsy* insulator was used as positive control (*gypsy*) and the *RpL32* gene (Rp49*)* was used as a negative control. Results are shown as percentage of input.

wing and body enhancers and the bristle enhancer, are required for *yellow* transcription (Geyer et al., 1986). The pUAST-Yellow vector contains unique restriction sites which allow a DNA sequence to be cloned between wing and body enhancers, between the body enhancer and promoter or in both sites in the same clone.

P element transposons were generated by cloning an approximately 500 bp fragment of *jockey* flanking the predicted Su(Hw) site into the pUAST-Yellow vector between the body enhancer and the promoter. If the predicted sequences function as insulators, the enhancers located upstream of the insulator would be prevented from activating transcription at the promoter and therefore the pigmentation would be decreased in the corresponding tissue. Enhancer-blocking was evaluated by comparing pigmentation levels from flies carrying the transgenes containing *jockey* sequences with those carrying transgenes lacking the insertion (Figure 3.5). Five independent lines were analyzed and compared for each insulator in order to eliminate the possibility of differences in expression level due to position effects. The resulting phenotype of transgenic flies containing single insertions of *jockey* located between body enhancer and promoter was a y^2 phenotype with yellow wings, yellow pigmentation in the body cuticle and black bristles (Figure 3.5B). The levels of pigmentation in transgenic flies were similar to those seen in control transgenic flies containing the *gypsy* insulator cloned into the same site. These results suggest that the predicted Su(Hw) binding site in *jockey* has insulator activity as shown by its ability to block body and wing enhancers from activating the promoter of the *y* gene.

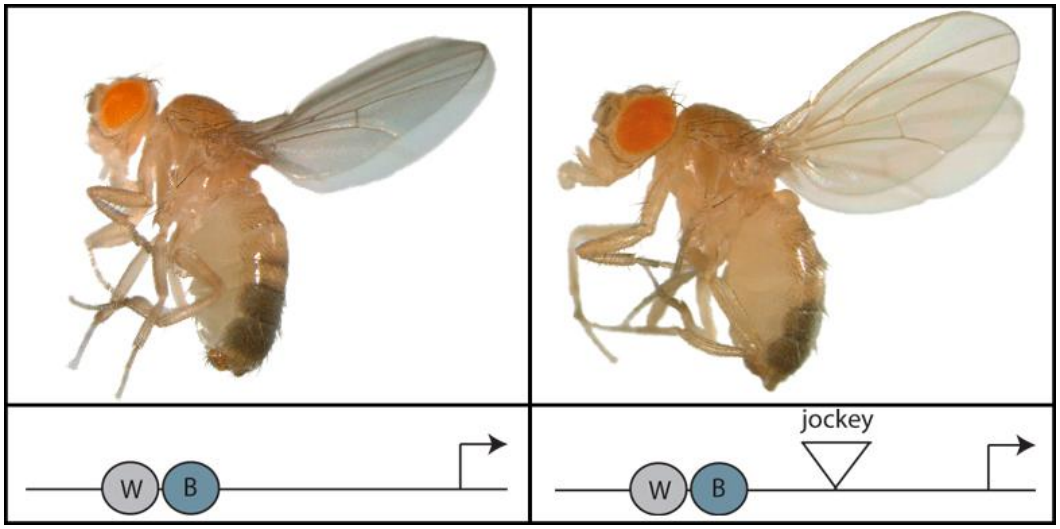


Fig. 3.5 A Su(Hw) binding site in *jockey* has enhancer-blocking activity.

Changes in yellow phenotype in pUASy-JY transgenic flies in the absence (**A**) and presence (**B**) of a *jockey* element inserted between the body enhancer and promoter. The triangle indicates *jockey* insertion site. Broken arrows correspond to the *yellow* gene promoter. Circles represent wing blades and body cuticle enhancers (W and B, respectively).

Su(Hw) ^{e04061} **mutant affects TE expression differentially in ovary and salivary gland tissues.**

It has been proposed that Su(Hw) regulates the transcription of the *gypsy* element, and *gypsy* RNA levels have been shown to decrease in Su(Hw) homozygous mutants (Parkhurst and Corces, 1986). Therefore, Su(Hw) appears to be important for regulation of *gypsy* transcription. In order to investigate the function of Su(Hw) binding sites within these transposons, we asked whether the Su(Hw) protein plays a role in the expression of these elements. To answer this, we performed quantitative RT-PCR using RNA extracted from salivary glands of third instar larvae and ovaries from wild-type and *su(Hw)*^{e04061} mutant homozygotes. The *su(Hw)*^{e04061} mutation is the result of a P element integrated into the coding region at the 5' end of the gene and reduces the expression of the *su(Hw)* gene by 40 fold (manuscript in preparation). Primers used for ChIP experiments were again used for qRT-PCR, and *rp49* primers were designed and used as an internal control. *Copia* and *I-element* were used as negative controls for the retroviral elements and the non-LTR elements, respectively, since they do not contain any predicted Su(Hw) binding site in their canonical sequence.

As expected, based on previous findings, expression of *gypsy* decreased significantly in *su(Hw)* ^{e04061} mutants in both tissues, although the magnitude of the decrease was much greater in salivary glands when compared with ovaries (Figure 3.6). Expression changes of other elements differed considerably depending on tissue. In salivary glands, aside from *gypsy,* only the expression of *jockey* decreased significantly,

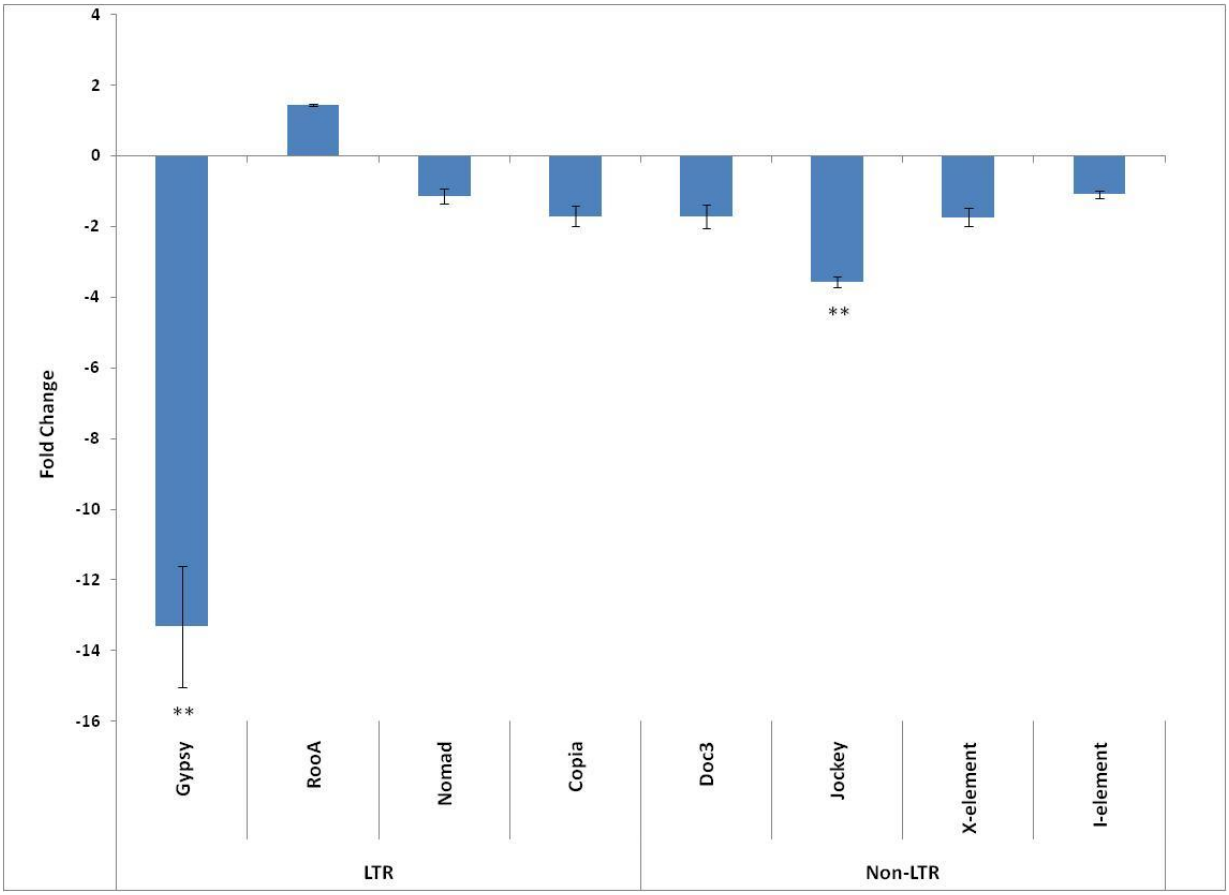


Fig. 3.6 Expression of transposable elements in *su(Hw)*^{e04061} mutant salivary glands. Transcript levels were quantified by real-time PCR and were normalized to *rp49*. Fold change values represent the relative expression of mRNA in *su(Hw)*^{e04061} homozygotes compared with *su(Hw)*^{e04061} / *Tb* heterozygotes. Two asterisks indicate P < 0.001.

while expression of the other elements showed little change. Since *jockey* was the only element which showed a significant enrichment for the Su(Hw) protein in ChIP experiments, the observed change in expression level suggests that, similar to *gypsy*, *jockey* expression requires the presence of the Su(Hw) protein.

In ovaries, a dramatically different pattern of expression change was observed (Figure 3.7). Unlike *gypsy*, whose expression decreased in mutants, the other elements tested exhibited at least a 2-fold increase in expression in mutants relative to wild-type. This increase was observed for *Copia* and *I-element* as well, although these elements contain no predicted Su(Hw) binding sites. This result suggests that the Su(Hw) protein may have a different functional role in the ovary, and its effect on TE expression may be unrelated to the presence of the Su(Hw) binding site within the element. Considering that transposons are suppressed by the piRNA pathway in the germline (Brennecke et al., 2007), the increase in TE expression in the mutant ovaries suggests that Su(Hw) may be involved in this suppression, either directly or indirectly perhaps by regulating a protein involved in TE regulation.

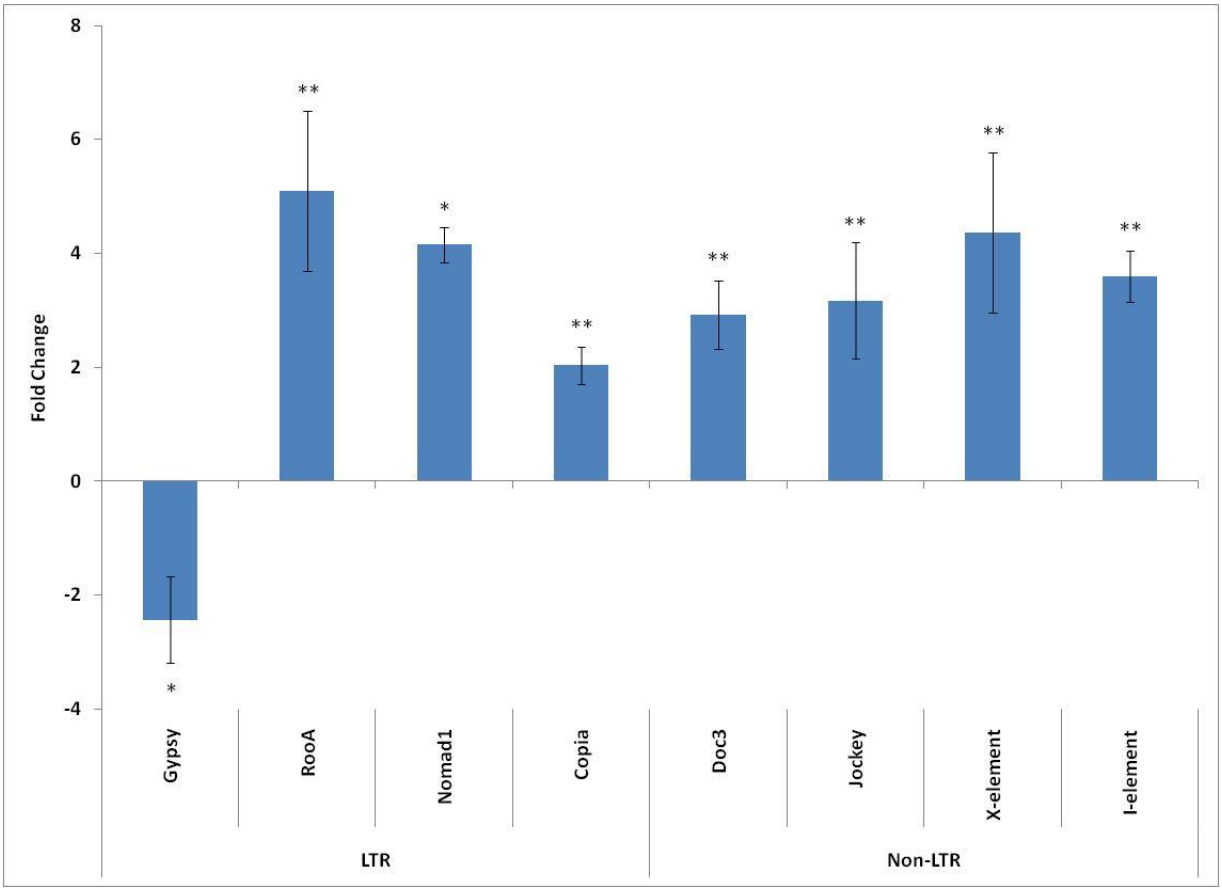


Fig. 3.7 Expression of transposable elements in *su(Hw)*^{e04061} mutant ovaries.

Transcript levels were quantified by real-time PCR and were normalized to *rp49*. Fold change values represent the relative expression of mRNA in *su(Hw)*^{e04061} homozygotes compared with *su(Hw)*^{e04061} / *Tb* heterozygotes. Single asterisks indicate $P < 0.05$ and two asterisks indicate $P < 0.001$.

The regulation of TE transcripts by *gypsy* insulator protein may not be related to piRNA pathways.

Recently, it has been found that in somatic tissues, a separate piRNA pathway exists which suppresses elements distinct from those controlled by the germline pathway. The existence of separate somatic and germline piRNA pathways was determined by analysis of piRNA pools in whole ovaries and early stage embryos, in which the follicular epithelium has been shed, prior to activation of the zygotic genome. Existence of piRNAs in both ovaries and early embryos suggested germline control, while those underrepresented in early embryo pools appear to be under somatic control (Malone et al., 2009). Unlike the other elements we tested by qRT-PCR, which are under germline control, *gypsy* is controlled by the somatic piRNA pathway (Malone et al., 2009). This observation may provide a possible clue to account for the fact that only this element shows a decrease in expression in *su(Hw)*^{e04061} mutant ovaries. In order to determine whether Su(Hw) might differentially affect the expression of elements suppressed by the somatic and germline pathways, we selected a larger number of transposable elements with predicted Su(Hw) sites, which included both somatically and germline controlled elements, according to Malone, et al, 2009.

First, to test whether the predicted Su(Hw) sites in the newly selected transposable elements were bound by the protein in vivo, we again performed ChIP assays in *Drosophila* embryos using the Su(Hw) antibody. Primers for quantitative real-time PCR were designed flanking the predicted sites in each element. The *gypsy* retrotransposon and the *rp49* coding sequences were used as positive and negative

controls, respectively. Results show that among the elements tested, only *gtwin* exhibited a significant level of enrichment of the Su(Hw) protein (Figure 3.4). This finding is not surprising, given that the *gtwin* element is mostly closely related to *gypsy* and there is strong degree of sequence similarity between the two elements (Bowen and McDonald, 2001; Ludwig and Loreto, 2007).

In order to address whether the differential effect of Su(Hw) on *gyspy* expression and expression of other transposons in the ovaries is related to whether these elements are controlled by somatic or germline pathways, we again performed quantitative RT-PCR using RNA extracted from 0-2 hour old embryos and ovaries of y^2ct^6; +/+ and y^2ct^6;$mod(mdg4)^{u1}$/ $mod(mdg4)^{u1}$ mutant flies. The use of these stocks allowed for differentiation between mutant and wild-type embryos, which is not possible with $su(Hw)^{e04061}$ mutants, which are maintained as heterozygotes since homozygote females are infertile (Harrison, 1993). The $mod(mdg4)^{u1}$ mutant allele is caused by the insertion of the *Stalker* retrotransposon into the sequences encoding the carboxyl-terminal end of the Mod(mdg4)67.2 protein (Georgiev and Gerasimova, 1989). Mutations in *mod(mdg)67.2* affect *gypsy* insulator function (Gerasimova and Corces, 1998). Because Su(Hw) interacts with the Modifier of mdg4 protein (Mod(mdg4)67.2), and in $mod(mdg4)^{u1}$ mutants the amount of Su(Hw) protein bound to chromosomes is significantly reduced (Gerasimova and Corces, 1998), we expected that the effect of the $mod(mdg4)^{u1}$ mutant may indirectly reflect the effect of $su(Hw)^{e04061}$ mutant.

For quantitative RT-PCR, primers used for ChIP experiments were again used for qRT-PCR, and *rp49* primers were designed and used as an internal control. In

ovaries, only a few elements showed a significant change in expression in *mod(mdg4)*^{u1}mutants when compared with wild-type (Figure 3.8). As expected, *gypsy* transcript levels decreased in a manner similar to that observed in *su(Hw)* ^{e04061} mutant ovaries. *Tabor*, another element under somatic control by *flamenco*, decreased as well. *I element, gtwin*, *412*, and *stalker* all showed significant increase in transcript levels in mutants, with *stalker* showing the largest overall fold change. Considering the structural and sequence similarities between *gypsy* and *gtwin*, as well as significant enrichment of Su(Hw) in *gtwin*, the opposite pattern of transcript level change between *gypsy* and *gtwin* in mutants was somewhat surprising.

In contrast to the ovary, in early embryos, the majority of TEs exhibited a significant increase in transcript levels in *mod(mdg4)*^{u1} mutants (Figure 3.9). This was true of TEs under both somatic and germline pathway control. Interestingly, unlike in other tissues where transcript levels decreased, *gypsy* transcript levels increased almost 7 fold in the mutant. Since *gypsy* is controlled by the somatic piRNA pathway (Malone et al., 2009), the significant increase in *gypsy* transcript in *mod(mdg4)*^{u1} mutant early embryos indicates that there is not any clear correlation between TE expression regulation by insulator proteins and the piRNA pathway.

Homozygote *mod(mdg4)*^{u1}mutants used for comparison of TE expression levels in *mod(mdg4)*^{u1}mutant and wild-type embryos were obtained from the *y*^{2}*ct*^{6};*mod(mdg4)*^{u1}/*Tb* stock. Because the *Tb* selection marker is only useful for larval or pupal stages, we used the *y*^{2}*ct*^{6}*;*+/+ line which was created by crossing

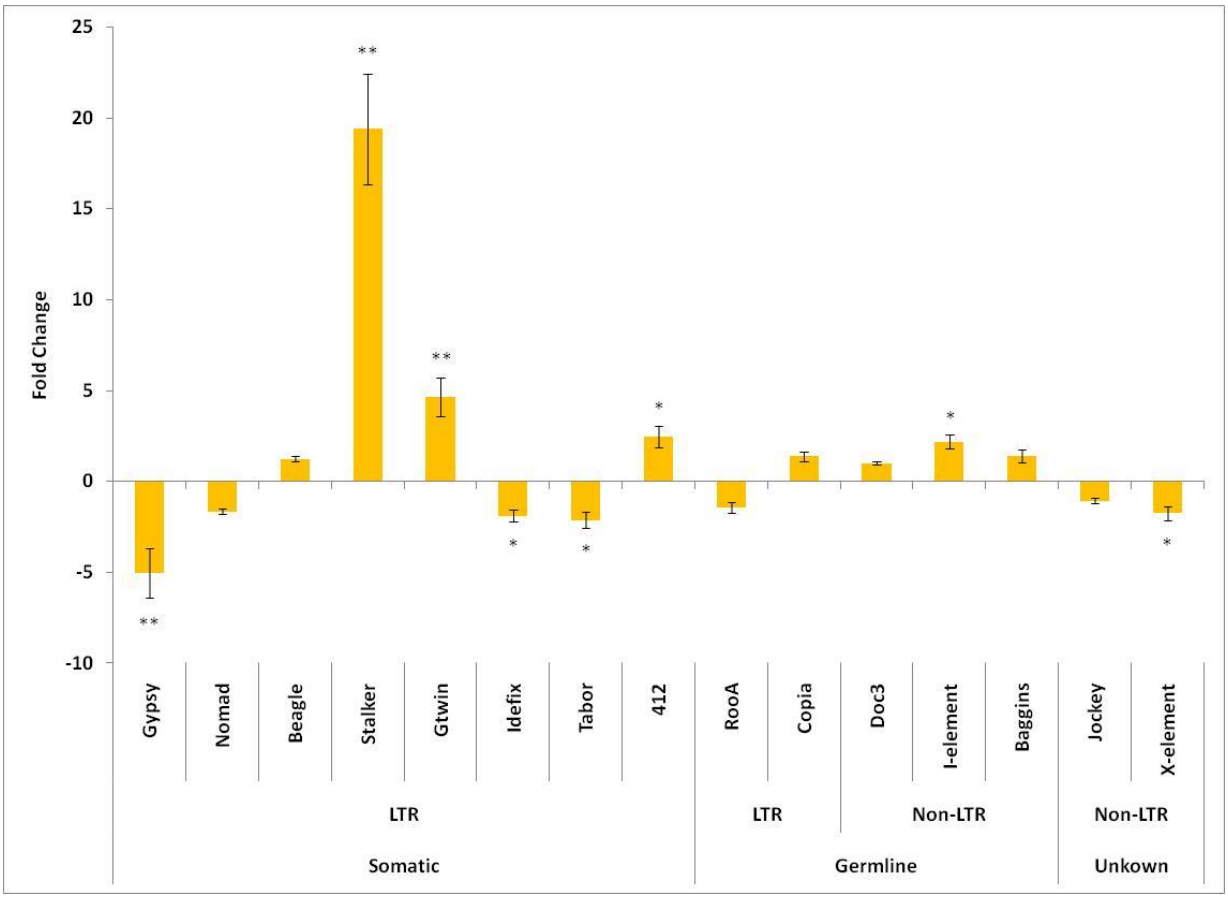


Fig. 3.8 Expression of transposable elements in *mod(mdg4)*^{u1} mutant ovaries.

Transcript levels were quantified by real-time PCR and were normalized to *rp49*. Fold change values represent the relative expression of mRNA in ovaries from *y*^{2}*ct*^{6};*mod(mdg4)*^{u1} homozygotes compared with ovaries from *y*^{2}*ct*^{6}*;* +/+. Single asterisks indicate P < 0.05 and two asterisks indicate P < 0.001. Elements are grouped according to piRNA pathway by which they are controlled.

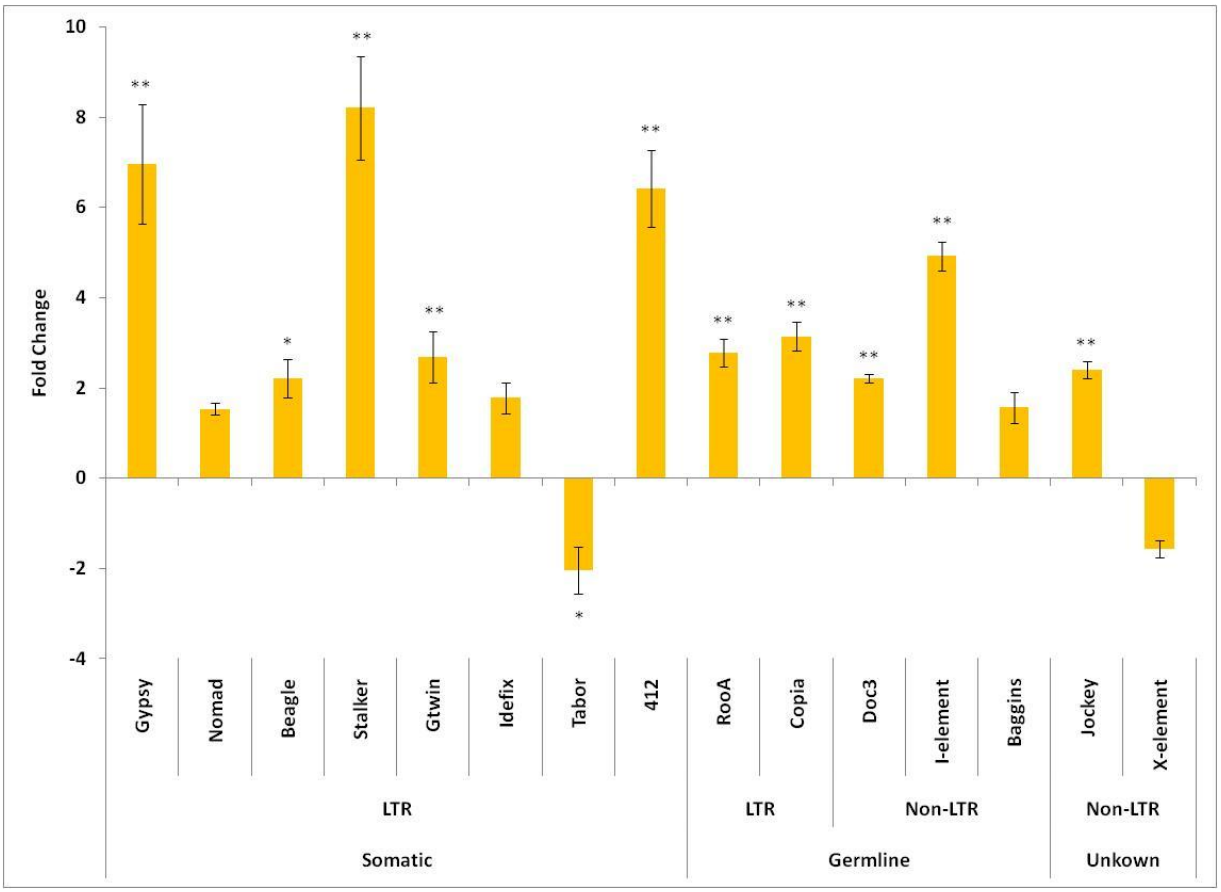


Fig. 3.9 Expression of transposable elements in *mod(mdg4)*^{u1} mutant embryos.

Transcript levels were quantified by real-time PCR and were normalized to *rp49*. Fold change values represent the relative expression of mRNA y^2ct^6;*mod(mdg4)*^{u1} homozygous embryos compared with y^2ct^6; +/+ embryos. Single asterisks indicate P < 0.05 and two asterisks indicate P < 0.001. Elements are grouped according to piRNA pathway by which they are controlled.

y^2ct^6;*mod(mdg4)*^{u1}/*Tb* with a wild-type stock for comparison of TE expression levels in *mod(mdg4)*^{u1}mutant and wild-type embryos. However, crossing with other stocks with heterogeneous genomic backgrounds may result in variation in number and location of TEs between the stocks, which could have an effect on the overall fold changes. To test this possibility, we repeated the quantitative RT-PCR using ovaries from y^2ct^6; *mod(mdg4)*^{u1}/*Tb* and y^2ct^6;*mod(mdg4)*^{u1}/*mod(mdg4)*^{u1} flies. Comparison of this data with that from the experiment using y^2ct^6;+/+ and y^2ct^6;*mod(mdg4)*^{u1}/*mod(mdg4)*^{u1} stocks show that changes in transcript levels of most of the TEs tested remains consistent, with only a few exceptions (Figure 3.10). For example, *gtwin*, which showed a significant increase in y^2ct^6;*mod(mdg4)*^{u1} when compared with y^2ct^6;+/+ ovaries, showed an almost 2-fold decrease in mutants when compared with y^2ct^6; *mod(mdg4)*^{u1}/*Tb*. Additionally, the large magnitude of the increase in *stalker* transcripts seen initially was not observed in y^2ct^6;*mod(mdg4)*^{u1} when compared with *mod(mdg4)*^{u1}/*Tb* ovaries, although the increase of transcript in mutants remained significant. The data suggests that, with the exception perhaps of *gtwin* and *stalker*, the use of separately maintained wild-type and mutant stocks did not have a significant effect on the transcript level changes observed.

Several mutant alleles of Su(Hw) are female sterile. In these mutants, egg chamber development arrests at stage 10, and is followed by eventual deterioration, suggesting that Su(Hw) may play an important role in oogenesis (Harrison et al., 1993; Harrison et al., 1992). Our initial quantitative RT-PCR experiments

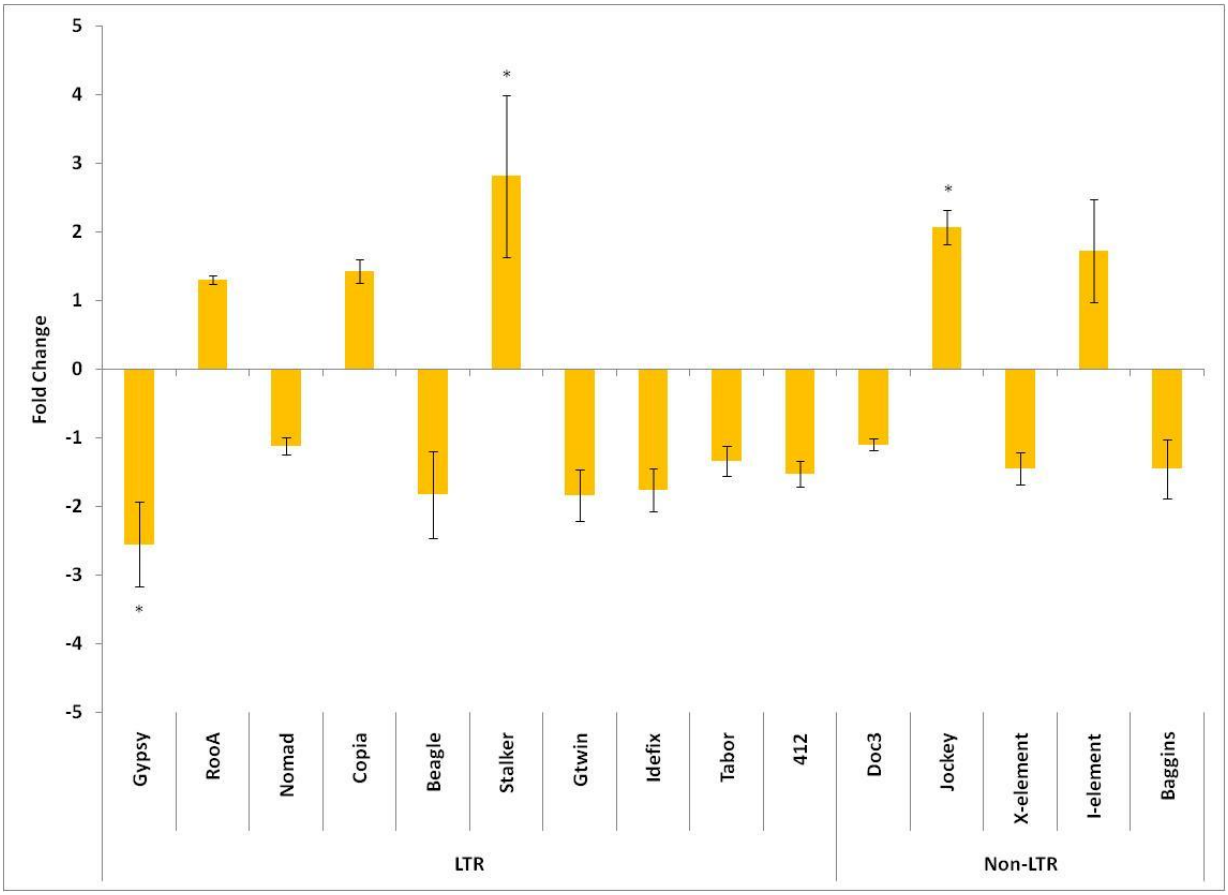


Fig. 3.10 Expression of transposable elements in *mod(mdg4)*^{u1} mutant ovaries.

Transcript levels were quantified by real-time PCR and were normalized to *rp49*. Fold change values represent the relative expression of mRNA in ovaries from *y*^{2}*ct*^{6};*mod(mdg4)*^{u1} homozygotes compared with ovaries from *y*^{2}*ct*^{6}*; mod(mdg4)*^{u1}*/Tb*. Single asterisks indicate P < 0.05.

measuring transposon expression in *su(Hw)* ^{e04061} mutants were performed using RNA from whole ovaries, which include some later stage egg chambers. It is possible that the striking changes observed in transposon transcript levels in *su(Hw)* ^{e04061} mutants might result from a global misregulation of transposons due to defects in oogenesis, rather than from specific action of Su(Hw) on their expression. Therefore, when we repeated quantitative RT-PCR experiments using RNA extracted from wild-type and *su(Hw)*^{e04061} mutant ovaries, all stage 9 or later egg chambers had been removed from wild-type and *su(Hw)*^{e04061} mutant ovaries. When compared with data from whole ovaries, the direction of transcript level change remained the same for most elements, although in general the changes were on a much smaller scale (Figures 3.7 and 3.11). In whole ovaries, for the majority of transposons transcript levels increased by more than 3 fold (Figure 3.7). When egg chambers later than stage 9 were removed, however, the level of fold change was less than 3 for all elements (Figure 3.11). On the other hand, while the decrease in *gypsy* transcripts was relatively modest in whole ovaries, these results showed a 21 fold reduction in *gypsy* transcripts. Based on these results, it appears that the large transcript level increase observed previously in *su(Hw)* ^{e04061} mutant ovaries might be due, in part, to the arrested later stage egg chambers.

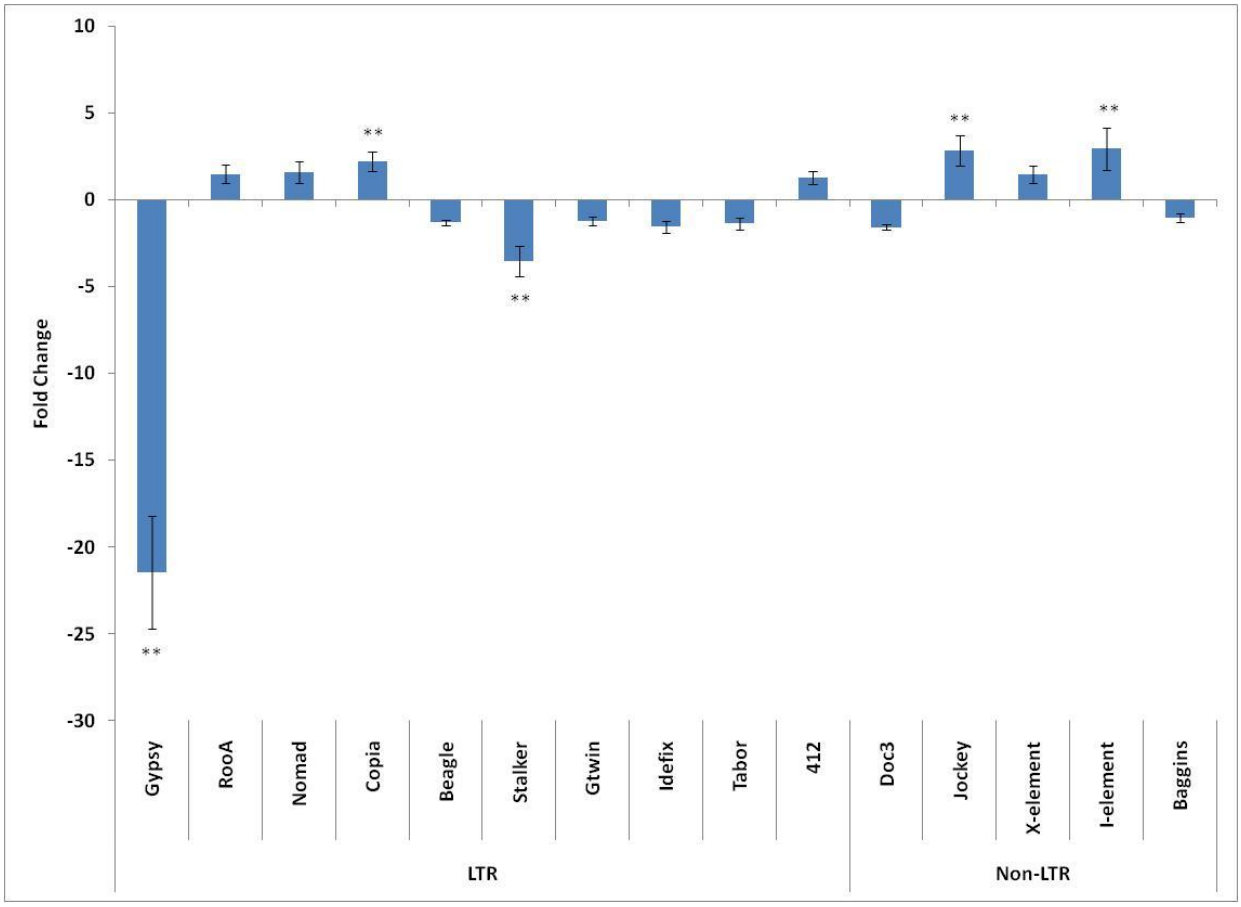


Fig. 3.11 Expression of transposable elements in *su(Hw)*^{e04061} mutant ovaries with later stage egg chambers removed. Transcript levels were quantified by real-time PCR and were normalized to *rp49*. Fold change values represent the relative expression of mRNA in ovaries from *y*^{2}*ct*^{6};*mod(mdg4)*^{u1} homozygotes compared with ovaries from *y*^{2}*ct*^{6}*; mod(mdg4)*^{u1}*/Tb*. Two asterisks indicate P < 0.001.

Mod(mdg4)^{u1} mutant and su(Hw) ^{e04061} mutant affect TE transcript levels differently in ovary.

In order to investigate whether these effects of the *mod(mdg4)^{u1}* mutant reflect the effects of the *su(Hw) ^{e04061}* mutant in the ovary, the new quantitative RT-PCR results from *su(Hw)^{e04061}* mutant ovaries, from which all stage 9 or later egg chambers had been removed, were compared with results of *mod(mdg4)^{u1}* mutant ovaries. Although *gypsy* transcript levels decreased in *mod(mdg4)^{u1}* mutants (Figure 3.10), the magnitude of the decrease is considerably reduced when compared with that seen in the *su(Hw)^{e04061}* mutant (Figure 3.11). It is possible that this is due to the fact that, although the *mod(mdg4)^{u1}* mutation results in a decrease of binding of Su(Hw) to DNA, Su(Hw) binding is not completely eliminated and so the effect of the *mod(mdg4)^{u1}* mutant might be reduced. However, this explanation does not account for the opposite effects observed for many other elements when comparing *mod(mdg4)^{u1}* and *su(Hw)^{e04061}* mutants. For example, in *stalker*, the transcript levels decrease in *su(Hw)^{e04061}* mutant ovaries but significantly increase in *mod(mdg4)^{u1}* mutant ovaries. This result may indicate that Su(Hw) may be involved in the expression of TE transcripts without direct binding to transposable elements. This idea is supported by the observation that transcripts of *copia* and *I-element*, which contain no predicted Su(Hw) binding sites, were increased in *su(Hw)^{e04061}* mutant ovaries.

DISCUSSION

Previous studies using computational approaches for identifying Su(Hw) binding sites in the *Drosophila* genome have provided evidence in support of the validity of this approach. Sites identified by *in silico* methods have been shown experimentally to correspond to sites that are occupied in vivo by Su(Hw) (Ramos et al., 2006). Using a consensus binding site sequence derived from genome wide ChIP-chip analysis of Su(Hw) binding sites, we have searched for Su(Hw) binding sites within the sequences of 179 *Drosophila* transposable elements and have identified 95 transposable elements with predicted sites that match the Su(Hw) consensus. Using *in situ* hybridization combined with immunostaining on a subset of these elements containing predicted sites, we have found that sites within each element show varying degrees of colocalization with the Su(Hw) protein. Some elements, such as *jockey*, show Su(Hw) localizing to a larger proportion of the regions containing *in situ* hybridization signals than others, and no element showed 100% colocalization. This observation may be to some extent related to the genomic context in which each copy of the TEs reside. It has been shown that Su(Hw) associates differently with different levels of chromatin organization in polytene chromosomes, and location within the genome appears to be correlated with appearance and intensity of Su(Hw) immunostaining signals (Wallace et al., 2010).

Recent ChIP-on-chip studies suggest that localization patterns of insulator proteins may be cell type-specific (Bushey et al., 2009; Ramos et al., 2006). Additionally, comparison of immunostaining data from polytene chromosomes with

ChIP-on-chip data obtained from embryos suggests that association of the Su(Hw) protein with some binding sites may be tissue specific. In general, however, a strong correlation could be observed between the intensity of Su(Hw) immunostaining signals in polytene chromosomes and enrichment in ChIP experiments performed in embryos (Ramos et al., 2006; Wallace et al., 2010). Based on these findings, it is surprising that while 4 elements showed colocalization between the Su(Hw) immunostaining signal and *in situ* hybridization at approximately 50% or more of the sites on polytene chromosomes, *jockey* was the only element which showed a significant level of Su(Hw) binding. In total, out of 16 elements tested by ChIP in our analysis, in addition to *gypsy*, only *jockey* and *gtwin* were significantly enriched for the Su(Hw) protein. *Gtwin* is referred to as a *gypsy*-twin element, and shows a high degree of sequence similarity with *gypsy.* The *gag*, *pol*, and *env* genes of *gtwin* show 53%, 76%, 77% percent amino acid identity, respectively, when compared with *gypsy* (Ludwig and Loreto, 2007). The arrangement of predicted Su(Hw) sites in *gtwin* resembles that of the 12 binding sites in the 5' UTR of *gypsy*. In *gtwin*, 4 predicted Su(Hw) binding sites are clustered within a span of 200 bp next to the 5' LTR, and 2 of the sites are identical in sequence to binding sites found in the *gypsy* insulator region. Given this sequence and structural similarity, as well as the clustering of multiple binding sites in close proximity, it is not surprising that we see a large amount of enrichment of the Su(Hw) protein.

It is possible that the level of enrichment of Su(Hw) at the sites tested in *gtwin* is a result of the number of binding sites present. It has been shown through competitive EMSA studies that multiple binding sites were able to outcompete single binding sites

for Su(Hw) binding (Ramos et al., 2006). This indicates that the strength of interaction of Su(Hw) with DNA correlates with the number of binding sites and may indicate some level of cooperativity in the binding of Su(Hw) to DNA. This observation may explain the difference in enrichment levels for Su(Hw) in *gypsy* and *gtwin* compared with *jockey*. While *jockey* has 4 predicted binding sites dispersed throughout its 5 kb length, the enrichment for Su(Hw) is approximately 87% less than that of *gypsy* and 25% less than that of *gtwin*. Based on this finding, it appears that the amount of Su(Hw) protein bound to the DNA is correlated with not only the number, but also the proximity of the Su(Hw) binding sites to each other.

The Su(Hw) binding region of the *gypsy* retrovirus, when bound by functional Su(Hw) protein, was shown to be able to induce expression of β-galactosidase from a *lacZ* reporter gene in tissues where *gypsy* is normally expressed (Smith and Corces, 1995). This data indicates that Su(Hw) acts as a transcriptional activator of *gypsy* expression. Based on this finding, as well as the observation that a number of transposable elements were predicted to contain Su(Hw) binding sites, and that among them *gtwin* and *jockey* showed significant enrichment for the Su(Hw) protein at the predicted sites in ChIP experiments, we wondered whether the presence of the predicted insulator sites might affect expression of these elements in a manner similar to that observed for *gypsy*. This idea is supported by our quantitative RT-PCR data from salivary glands in which, like *gypsy*, expression of *jockey* decreases significantly in *su(Hw)*^{e04061} mutants. However, in *su(Hw)*^{e04061} mutant ovaries, in contrast to *gypsy* transcript levels, which show a large decrease, *jockey* transcript levels increase while

gtwin transcript levels remain virtually unchanged. Additionally, we observed significant changes in expression of elements such as *stalker*, which contain predicted binding sites but were not found to bind Su(Hw) in ChIP assays, as well as elements, such as *I-element* and *copia*, which do not contain any predicted Su(Hw) sites. These results indicate that, at least in ovaries, the effect of Su(Hw) on expression of transposable elements is not related to the presence of Su(Hw) binding sites. The ability of Su(Hw) to affect TE expression in a manner that is independent of its ability to bind insulator sequences is reminiscent of the effect of certain *su(Hw)* mutations on female fertility. Interestingly, zinc finger 10 of the Su(Hw) protein was shown to be essential for binding to *gypsy* insulator DNA sequences, but is not required for female fertility (Kim et al., 1996), suggesting that the regions involved in binding *gypsy* insulator DNA sequences are not required for oogenesis and that the role of Su(Hw) in oogenesis might involve binding to sequences other than those in the insulator (Kim et al., 1996).

Gypsy is expressed in a tissue-specific pattern during *Drosophila* development, and *gypsy* RNA is mainly found in the gonads, salivary glands and fat body. This complex pattern of expression is thought to be under the control of transcription factors which interact with sequences located in the 5' LTR and 5'UTR (Smith and Corces, 1995). Like *gypsy*, other transposons that bind Su(Hw) may have their own tissue-specific expression patterns for which the Su(Hw) protein may be necessary. In addition to Su(Hw), other tissue-specific proteins may control the pattern of temporal and spatial expression of these transposons. For example, the leucine zipper domain of Su(Hw) was found to mediate transcriptional repression of *gypsy* in larval tissues and

transcriptional activation of *gypsy* in ovaries, suggesting that this domain may be involved in interaction with proteins that regulate *gypsy* expression differently in the two tissues (Smith and Corces, 1995). This system of tissue-specific regulation may provide a possible explanation for our results which show the opposing effects of *su(Hw)*^{e04061} mutation on the expression of *jockey* in salivary glands and ovaries, as well as *gypsy* in ovaries and embryos.

The pattern of transcript level changes initially observed in *su(Hw)*^{e04061} mutant ovaries led us to the hypothesis that Su(Hw) may play a role in regulation of transposable element expression via the piRNA pathway. A recent study found that two distinct piRNA pathways exist to control TE expression; one that functions in somatic tissues and another that functions in the germline. Elements with piRNAs which originate from the *flamenco* cluster are predominately members of the *gypsy* family of LTR retrotransposons and are controlled by Piwi in the somatic pathway. On the other hand, elements with corresponding piRNAs which are produced from other clusters, such as *42AB,* are under control of Aub and Ago3 in the germline piRNA pathway (Malone et al., 2009). Our initial quantitative RT-PCR results in ovaries showed a large increase in transcript levels of each element, with the exception of *gypsy*. Among the elements tested, only *gypsy* had been shown to be regulated in the somatic follicle cells, while the other elements are controlled in the germline, an observation which suggested that the different effect of the *su(Hw)*^{e04061} mutation on *gypsy* versus the other elements might be related to the mechanism by which their expression is regulated. When we expanded our analysis to include a larger number of transposable elements regulated

by each pathway, however, the results did not remain consistent. For example, in ovaries unlike *gypsy*, expression of *412*, which is also regulated by *flamenco*, increased in *su(Hw)*^{e04061} mutants while expression of *baggins*, which is under germline control, decreased. Additionally, comparison of TE expression changes in *mod(mdg4)*^{u1} mutants in ovaries and early embryos revealed a pattern which was not consistent with our hypothesis. For example, if Su(Hw) affected the expression of TEs controlled by the two different piRNA pathways, we would expect that in early embryos, which represent only germline tissues, *mod(mdg4)* mutations, in which Su(Hw) binding is significantly reduced, would affect only those elements which are controlled by that piRNA pathway. Instead, we see large changes in expression of almost all the elements irrespective of the pathway by which they are controlled. While the piRNA pathways are involved in the suppression of transposable elements, we observed that the absence of the Su(Hw) protein may result in either an increase or decrease in TE expression, depending on the tissue, indicating that Su(Hw) protein has a less straightforward role in regulating TE expression. Taken together, these findings indicate that Su(Hw) may not be directly involved in the regulation of transposable elements via the piRNA pathways.

Data from *mod(mdg4)*^{u1} mutants itself is not sufficient enough to make conclusions about the role of Su(Hw) in the expression of transposable elements. While we expected, based on the reduction of Su(Hw) binding to DNA in these mutants, that Mod(mdg4) might have a similar effect on TE expression to that of Su(Hw), comparison of data from ovaries of the two mutants indicates that this is not always the case. Even for *gypsy*, whose expression decreased in both *su(Hw)*^{e04061} and *mod(mdg4)*^{u1} mutants,

the decrease observed was approximately 10 times less in *mod(mdg4)*^{u1} than in *su(Hw)*^{e04061} mutants. In many cases, transcript changes showed opposite patterns for the same elements in the two mutants. It is possible that, in much the same way as Su(Hw) is thought to interact with different proteins to regulate expression of *gypsy* in various tissues, in the absence of Mod(mdg4), Su(Hw) may be able to interact with other binding partners which might alter its effects on TE expression. This possibility was suggested by results of a functional study of insulator proteins in the *gypsy* retrovirus (Gerasimova et al., 1995). The *y2* allele, which is caused by the insertion of a single *gypsy* retrovirus between the body enhancer and the promoter in the *yellow* gene, showed a variegated body cuticle and bristle phenotype in homozygous *mod(mdg4)*^{u1} alleles. It was suggested that the absence of Mod(mdg4) results in free leucine zipper and acidic domains of Su(Hw), and these free domains may interact with other proteins which may be involved in the condensation or formation of heterochromatin (Gerasimova et al., 1995). This may be one explanation as to why the absence of Mod(mdg4) may result in a different effect than the absence of Su(Hw) on the expression of many transposable elements.

Taken together, our data shows that Su(Hw) may influence the expression of a number of TEs, although, with the exception of *gypsy*, it does not appear to have a major effect. Since our data suggests that Su(Hw) may not be involved in the regulation of TE expression by the piRNA pathway, it is possible that the action of the piRNA pathway in suppressing TEs is masking the effect of Su(Hw) on TE expression. For instance, Su(Hw) may influence transcription of TEs, but the piRNA pathway may act at

a higher level of regulation to recognize and degrade transcripts from these elements. In other words, under comprehensive suppression of TE by the piRNA pathways, Su(Hw) may only be able to produce slight alterations in TE expression. Analysis of the effect of Su(Hw) on TE expression in *flamenco* mutant lines, for example, where piRNA production is inhibited, may help to determine whether Su(Hw) plays major role in TE expression.

LIST OF REFERENCES

Adryan, B., Woerfel, G., Birch-Machin, I., Gao, S., Quick, M., Meadows, L., Russell, S., and White, R. (2007). Genomic mapping of Suppressor of Hairy-wing binding sites in *Drosophila*. Genome Biol *8*, R167.

Ahmad, K.F., Engel, C.K., and Prive, G.G. (1998). Crystal structure of the BTB domain from PLZF. Proc Natl Acad Sci U S A *95*, 12123-12128.

Almeida, R., and Allshire, R.C. (2005). RNA silencing and genome regulation. Trends Cell Biol *15*, 251-258.

Apostolou, E., and Thanos, D. (2008). Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. Cell *134*, 85-96.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T.*, et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Klenov, M.S., Vagin, V.V., Bantignies, F., Cavalli, G., and Gvozdev, V.A. (2004). Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. Molecular and Cellular Biology *24*, 6742-6750.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bartkuhn, M., and Renkawitz, R. (2008). Long range chromatin interactions involved in gene regulation. Biochim Biophys Acta *1783*, 2161-2166.

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

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

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

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

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

Bode, J., and Maass, K. (1988). Chromatin domain surrounding the human interferon-beta gene as defined by scaffold-attached regions. Biochemistry *27*, 4706-4711.

Boutanaev, A.M., Kalmykova, A.I., Shevelyov, Y.Y., and Nurminsky, D.I. (2002). Large clusters of co-expressed genes in the *Drosophila* genome. Nature *420*, 666-669.

Bowen, N.J., and McDonald, J.F. (2001). *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. Genome Res *11*, 1527-1540.

Brasset, E., Hermant, C., Jensen, S., and Vaury, C. (2010). The *Idefix* enhancer-blocking insulator also harbors barrier activity. Gene *450*, 25-31.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell *128*, 1089-1103.

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

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

Byrd, K., and Corces, V.G. (2003). Visualization of chromatin domains created by the *gypsy* insulator of *Drosophila*. The Journal of Cell Biology *162*, 565-574.

Cai, H.N., and Shen, P. (2001). Effects of *cis* arrangement of chromatin insulators on enhancer-blocking activity. Science (New York, NY) *291*, 493-495.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Capelson, M., and Corces, V.G. (2004). Boundary elements and nuclear organization. Biol Cell *96*, 617-629.

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

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes & Development *16*, 2733-2742.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A*., et al.* (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science (New York, NY) *291*, 1289-1292.

Carthew, R.W. (2006). Molecular biology. A new RNA dimension to genome control. Science (New York, NY) *313*, 305-306.

Cavalli, G., Orlando, V., and Paro, R. (1999). Mapping DNA target sites of chromatin-associated proteins by formaldehyde cross-linking in *Drosophila* embryos. In Chromosome Structural Analysis: A Practical Approach, pp20-37, W.A. Bickmore, ed. (Oxford University Press UK).

Chung, J.H., Whiteley, M., and Felsenfeld, G. (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.

Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nature Genetics *26*, 183-186.

Conte, C., Dastugue, B., and Vaury, C. (2002a). Coupling of enhancer and insulator properties identified in two retrotransposons modulates their mutagenic impact on nearby genes. Molecular and Cellular Biology *22*, 1767-1777.

Conte, C., Dastugue, B., and Vaury, C. (2002b). Promoter competition as a mechanism of transcriptional interference mediated by retrotransposons. The EMBO Journal *21*, 3908-3916.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes & Development *12*, 3715-3727.

Cremer, T., Cremer, C., Baumann, H., Luedtke, E.K., Sperling, K., Teuber, V., and Zorn, C. (1982). Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. Hum Genet *60*, 46-56.

Croston, G.E., and Kadonaga, J.T. (1993). Role of chromatin structure in the regulation of transcription by RNA polymerase II. Current opinion in cell biology *5*, 417-423.
de Wit, E., Braunschweig, U., Greil, F., Bussemaker, H.J., and van Steensel, B. (2008). Global chromatin domain organization of the *Drosophila* genome. PLoS Genetics *4*, e1000045.

Dekker, J. (2008). Gene regulation in the third dimension. Science *319*, 1793-1794.
Desset, S., Buchon, N., Meignin, C., Coiffet, M., and Vaury, C. (2008). In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. PloS One *3*, e1526.

Donze, D., and Kamakaka, R.T. (2001). RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. The EMBO Journal *20*, 520-531.

Dorus, S., Busby, S.A., Gerike, U., Shabanowitz, J., Hunt, D.F., and Karr, T.L. (2006). Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. Nature Genetics *38*, 1440-1445.

Eissenberg, J.C., and Elgin, S.C. (2000). The HP1 protein family: getting a grip on chromatin. Current Opinion in Genetics & Development *10*, 204-210.

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

Fanti, L., and Pimpinelli, S. (2008). HP1: a functionally multifaceted protein. Current Oin Genetics & Development *18*, 169-174.

Fernandez, L.A., Winkler, M., and Grosschedl, R. (2001). Matrix attachment region-dependent function of the immunoglobulin mu enhancer involves histone acetylation at a distance without changes in enhancer occupancy. Molecular and Cellular Biology *21*, 196-208.

Forrester, W.C., Fernandez, L.A., and Grosschedl, R. (1999). Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. Genes & Development *13*, 3003-3014.

Fourel, G., Revardel, E., Koering, C.E., and Gilson, E. (1999). Cohabitation of insulators and silencing elements in yeast subtelomeric regions. The EMBO Journal *18*, 2522-2537.

Fraser, P., and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. Nature *447*, 413-417.

Galande, S., Purbey, P.K., Notani, D., and Kumar, P.P. (2007). The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. Current Opinion in Genetics & Development *17*, 408-414.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet *7*, 703-713.

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

Gdula, D.A., and Corces, V.G. (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., Gerasimova, T.I., and Corces, V.G. (1996). Genetic and molecular analysis of the *gypsy* chromatin insulator of *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America *93*, 9378-9383.

Georgiev, P.G., and Gerasimova, T.I. (1989). Novel genes influencing the expression of the *yellow* locus and *mdg4* (*gypsy*) in *Drosophila melanogaster*. Mol Gen Genet *220*, 121-126.

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

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

Gerasimova, T.I., and Corces, V.G. (2001). Chromatin insulators and boundaries: effects on transcription and nuclear organization. Annual Review of Genetics *35*, 193-208.

Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O., and Corces, V.G. (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., Lei, E.P., Bushey, A.M., and Corces, V.G. (2007). Coordinated control of dCTCF and *gypsy* chromatin insulators in *Drosophila*. Molecular Cell *28*, 761-772.

Geyer, P.K., and Corces, V.G. (1992). DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. Genes & Development *6*, 1865-1873.

Geyer, P.K., Spana, C., and Corces, V.G. (1986). On the molecular mechanism of *gypsy*-induced mutations at the *yellow* locus of *Drosophila melanogaster*. The EMBO Journal *5*, 2657-2662.

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

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

Girard, L., and Freeling, M. (1999). Regulatory changes as a consequence of transposon insertion. Dev Genet *25*, 291-296.

Golovnin, A., Biryukova, I., Romanova, O., Silicheva, M., Parshikov, A., Savitskaya, E., Pirrotta, V., and Georgiev, P. (2003). An endogenous Su(Hw) insulator separates the yellow gene from the Achaete-scute gene complex in *Drosophila*. Development (Cambridge, England) *130*, 3249-3258.

Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science (New York, NY) *301*, 798-802.

Gurudatta, B.V., and Corces, V.G. (2009). Chromatin insulators: lessons from the fly. Brief Funct Genomic Proteomic *8*, 276-282.

Harrison, D.A., Gdula, D.A., Coyne, R.S., and Corces, V.G. (1993). A leucine zipper domain of the suppressor of Hairy-wing protein mediates its repressive effect on enhancer function. Genes & Development *7*, 1966-1978.

Harrison, D.A., Mortin, M.A., and Corces, V.G. (1992). The RNA polymerase II 15-kilodalton subunit is essential for viability in *Drosophila melanogaster*. Molecular and Cellular Biology *12*, 928-935.

Hewitt, S.L., High, F.A., Reiner, S.L., Fisher, A.G., and Merkenschlager, M. (2004). Nuclear repositioning marks the selective exclusion of lineage-inappropriate transcription factor loci during T helper cell differentiation. Eur J Immunol *34*, 3604-3613.

Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., and White, R. (2007). CTCF genomic binding sites in *Drosophila* and the organisation of the bithorax complex. PLoS Genetics *3*, e112.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. The Journal of Biological Chemistry *268*, 305-314.

Hoskins, R.A., Smith, C.D., Carlson, J.W., Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G.*, et al.* (2002). Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. Genome Biol *3*, RESEARCH0085.

Hurst, L.D., Pal, C., and Lercher, M.J. (2004). The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet *5*, 299-310.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science (New York, NY) *293*, 1074-1080.

Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., and Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nature Biotechnology *26*, 1293-1300.

Jiang, N., Emberly, E., Cuvier, O., and Hart, C.M. (2009). Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in *Drosophila melanogaster* links BEAF to transcription. Molecular and Cellular Biology *29*, 3556-3568.

Kalmykova, A.I., Klenov, M.S., and Gvozdev, V.A. (2005). Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. Nucleic Acids Res *33*, 2052-2059.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M.*, et al.* (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol *3*, RESEARCH0084.

Kaplan, C.D., Morris, J.R., Wu, C., and Winston, F. (2000). Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. *melanogaster*. Genes & Development *14*, 2623-2634.

Kazazian, H.H., Jr. (2004). Mobile elements: drivers of genome evolution. Science (New York, NY) *303*, 1626-1632.

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

Kellum, R., and Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. Molecular and Cellular Biology *12*, 2424-2431.

Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Purd'homme, N., and Bucheton, A. (1994). Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America *91*, 1285-1289.

Kim, J., Shen, B., Rosen, C., and Dorsett, D. (1996). The DNA-binding and enhancer-blocking domains of the *Drosophila* suppressor of Hairy-wing protein. Molecular and Cellular Biology *16*, 3381-3392.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A.*, et al.* (2009). The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.

Kleinjan, D.A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet *76*, 8-32.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.

Kosak, S.T., and Groudine, M. (2004). Gene order and dynamic domains. Science (New York, NY) *306*, 644-647.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S.*, et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet *4*, e1000242.

Kuhn-Parnell, E.J., Helou, C., Marion, D.J., Gilmore, B.L., Parnell, T.J., Wold, M.S., and Geyer, P.K. (2008). Investigation of the properties of non-*gypsy* suppressor of hairy-wing-binding sites. Genetics *179*, 1263-1273.

Kuhn, E.J., and Geyer, P.K. (2003). Genomic insulators: connecting properties to mechanism. Current Opinion in Cell Biology *15*, 259-265.

Kumaran, R.I., Thakar, R., and Spector, D.L. (2008). Chromatin dynamics and gene positioning. Cell *132*, 929-934.

Kyrchanova, O., Chetverina, D., Maksimenko, O., Kullyev, A., and Georgiev, P. (2008). Orientation-dependent interaction between *Drosophila* insulators is a property of this class of regulatory elements. Nucleic Acids Res *36*, 7019-7028.

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

Labrador, M., and Corces, V.G. (2003). Phosphorylation of histone H3 during transcriptional activation depends on promoter structure. Genes & Development *17*, 43-48.

Labrador, M., Mongelard, F., Plata-Rengifo, P., Baxter, E.M., Corces, V.G., and Gerasimova, T.I. (2001). Protein encoding by both DNA strands. Nature *409*, 1000.

Labrador, M., Sha, K., Li, A., and Corces, V.G. (2008). Insulator and Ovo Proteins Determine the Frequency and Specificity of Insertion of the *gypsy* Retrotransposon in *Drosophila melanogaster*. Genetics.

Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nat Rev Genet *8*, 104-115.

Lee, J.M., and Sonnhammer, E.L. (2003). Genomic gene clustering analysis of pathways in eukaryotes. Genome Res *13*, 875-882.

Lercher, M.J., Blumenthal, T., and Hurst, L.D. (2003). Coexpression of neighboring genes in Caenorhabditis elegans is mostly due to operons and duplicate genes. Genome Res *13*, 238-243.

Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., and Sheen, F.M. (1993). Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell *75*, 1083-1093.

Levy-Wilson, B., and Fortier, C. (1989). The limits of the DNase I-sensitive domain of the human apolipoprotein B gene coincide with the locations of chromosomal anchorage loops and define the 5' and 3' boundaries of the gene. The Journal of Biological Chemistry *264*, 21196-21204.

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

Lis, J.T. (2007). Imaging *Drosophila* gene activation and polymerase pausing in vivo. Nature *450*, 198-202.

Loc, P.V., and Stratling, W.H. (1988). The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. The EMBO Journal *7*, 655-664.

Ludwig, A., and Loreto, E.L. (2007). Evolutionary pattern of the *gtwin* retrotransposon in the *Drosophila melanogaster* subgroup. Genetica *130*, 161-168.

Lunyak, V.V., Prefontaine, G.G., Nunez, E., Cramer, T., Ju, B.G., Ohgi, K.A., Hutt, K., Roy, R., Garcia-Diaz, A., Zhu, X.*, et al.* (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science (New York, NY) *317*, 248-251.

Luo, R.X., and Dean, D.C. (1999). Chromatin remodeling and transcriptional regulation. J Natl Cancer Inst *91*, 1288-1294.

Maksimenko, O., Golovnin, A., and Georgiev, P. (2008). Enhancer-promoter communication is regulated by insulator pairing in a *Drosophila* model bigenic locus. Molecular and Cellular Biology *28*, 5469-5477.

Mal'ceva, N.I., Belyaeva, E.S., King, R.C., and Zhimulev, I.F. (1997). Nurse cell polytene chromosomes of *Drosophila melanogaster* otu mutants: morphological changes accompanying interallelic complementation and position effect variegation. Dev Genet *20*, 163-174.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. Cell *137*, 522-535.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656-668.

Manak, J.R., Dike, S., Sementchenko, V., Kapranov, P., Biemar, F., Long, J., Cheng, J., Bell, I., Ghosh, S., Piccolboni, A.*, et al.* (2006). Biological function of unannotated transcription during the early development of *Drosophila melanogaster*. Nature Genetics *38*, 1151-1158.

Markstein, M., and Levine, M. (2002). Decoding *cis*-regulatory DNAs in the *Drosophila* genome. Current Opinion in Genetics & Development *12*, 601-606.

Mevel-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The *flamenco* locus controls the *gypsy* and *ZAM* retroviruses and is required for *Drosophila* oogenesis. Genetics *175*, 1615-1624.

Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell *91*, 35-45.

Mihaly, J., Hogga, I., Barges, S., Galloni, M., Mishra, R.K., Hagstrom, K., Muller, M., Schedl, P., Sipos, L., Gausz, J.*, et al.* (1998). Chromatin domain boundaries in the Bithorax complex. Cell Mol Life Sci *54*, 60-70.

Miller, M.A., Cutter, A.D., Yamamoto, I., Ward, S., and Greenstein, D. (2004). Clustered organization of reproductive genes in the C. elegans genome. Curr Biol *14*, 1284-1290.

Minervini, C.F., Ruggieri, S., Traversa, M., D'Aiuto, L., Marsano, R.M., Leronni, D., Centomani, I., De Giovanni, C., and Viggiano, L. (2010). Evidences for insulator activity of the 5'UTR of the *Drosophila melanogaster* LTR-retrotransposon ZAM. Mol Genet Genomics *283*, 503-509.

Mishra, R.K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S.E., and Schedl, P. (2001). The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. Molecular and Cellular Biology *21*, 1311-1318.

Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinl, N., Bardenhagen, I., Leers, J., White, R.A., Renkawitz-Pohl, R., Saumweber, H.*, et al.* (2007). The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. The EMBO Journal *26*, 4203-4214.

Mongelard, F., Labrador, M., Baxter, E.M., Gerasimova, T.I., and Corces, V.G. (2002). Trans-splicing as a novel mechanism to explain interallelic complementation in *Drosophila*. Genetics *160*, 1481-1487.

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R.*, et al.* (2005). CTCF is conserved from

Drosophila to humans and confers enhancer blocking of the *Fab-8* insulator. EMBO Reports *6*, 165-170.

Moore, J.K., and Haber, J.E. (1996). Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature *383*, 644-646.

Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. Science (New York, NY) *283*, 1530-1534.

Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V., and Georgiev, P. (2001). Loss of insulator activity by paired Su(Hw) chromatin insulators. Science (New York, NY) *291*, 495-498.

Negre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A.*, et al.* (2010). A comprehensive map of insulator elements for the *Drosophila* genome. PLoS Genet *6*, e1000814.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Molecular Cell *11*, 709-719.

Ohtsuki, S., and Levine, M. (1998). GAGA mediates the enhancer blocking activity of the eve promoter in the *Drosophila* embryo. Genes Dev *12*, 3325-3330.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W.*, et al.* (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nature Genetics *36*, 1065-1071.

Pai, C.Y., Lei, E.P., Ghosh, D., and Corces, V.G. (2004). The centrosomal protein CP190 is a component of the *gypsy* chromatin insulator. Molecular Cell *16*, 737-748.

Parada, L., and Misteli, T. (2002). Chromosome positioning in the interphase nucleus. Trends Cell Biol *12*, 425-432.

Pardue, M.L., and DeBaryshe, P.G. (2003). Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. Annual Review of Genetics *37*, 485-511.

Parkhurst, S.M., and Corces, V.G. (1986). Mutations at the suppressor of forked locus increase the accumulation of *gypsy*-encoded transcripts in *Drosophila melanogaster*. Molecular and Cellular Biology *6*, 2271-2274.

Parnell, T.J., Kuhn, E.J., Gilmore, B.L., Helou, C., Wold, M.S., and Geyer, P.K. (2006). Identification of genomic sites that bind the *Drosophila* suppressor of Hairy-wing insulator protein. Molecular and Cellular Biology *26*, 5983-5993.

Parnell, T.J., Viering, M.M., Skjesol, A., Helou, C., Kuhn, E.J., and Geyer, P.K. (2003). An endogenous suppressor of hairy-wing insulator separates regulatory domains in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America *100*, 13436-13441.

Pelisson, A., Payen-Groschene, G., Terzian, C., and Bucheton, A. (2007). Restrictive *flamenco* alleles are maintained in *Drosophila melanogaster* population cages, despite the absence of their endogenous *gypsy* retroviral targets. Molecular Biology and Evolution *24*, 498-504.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194-1211.

Pikaart, M.J., Recillas-Targa, F., and Felsenfeld, G. (1998). Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. Genes & Development *12*, 2852-2862.

Pinkel, D., Straume, T., and Gray, J.W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proceedings of the National Academy of Sciences of the United States of America *83*, 2934-2938.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science (New York, NY) *300*, 131-135.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). *Flamenco*, a gene controlling the *gypsy* retrovirus of *Drosophila melanogaster*. Genetics *139*, 697-711.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Purbowasito, W., Suda, C., Yokomine, T., Zubair, M., Sado, T., Tsutsui, K., and Sasaki, H. (2004). Large-scale identification and mapping of nuclear matrix-attachment regions in the distal imprinted domain of mouse chromosome 7. DNA Res *11*, 391-407.

Ragoczy, T., Bender, M.A., Telling, A., Byron, R., and Groudine, M. (2006). The locus control region is required for association of the murine beta-globin locus with engaged

transcription factories during erythroid maturation. Genes & Development *20*, 1447-1457.

Ramos, E., Ghosh, D., Baxter, E., and Corces, V.G. (2006). Genomic organization of *gypsy* chromatin insulators in *Drosophila melanogaster*. Genetics *172*, 2337-2349.

Rando, O.J., and Ahmad, K. (2007). Rules and regulation in the primary structure of chromatin. Current Opinion in Cell Biology *19*, 250-256.

Rorth, P. (1996). A modular misexpression screen in *Drosophila* detecting tissue-specific phenotypes. Proceedings of the National Academy of Sciences of the United States of America *93*, 12418-12422.

Roseman, R.R., Pirrotta, V., and Geyer, P.K. (1993). The su(Hw) protein insulates expression of the *Drosophila melanogaster* white gene from chromosomal position-effects. The EMBO Journal *12*, 435-442.

Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. (2002). Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature *418*, 975-979.

Saitoh, N., Bell, A.C., Recillas-Targa, F., West, A.G., Simpson, M., Pikaart, M., and Felsenfeld, G. (2000). Structural and functional conservation at the boundaries of the chicken beta-globin domain. The EMBO Journal *19*, 2315-2322.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407-411.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Molecular and Cellular Biology *28*, 2718-2731.

Sarot, E., Payen-Groschene, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the *gypsy* endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. Genetics *166*, 1313-1321.

Saunders, A., Werner, J., Andrulis, E.D., Nakayama, T., Hirose, S., Reinberg, D., and Lis, J.T. (2003). Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. Science (New York, NY) *301*, 1094-1096.

Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., and Gvozdev, V. (2006). Telomere elongation is under the control of the RNAi-based mechanism in the *Drosophila* germline. Genes & Development *20*, 345-354.

Schardin, M., Cremer, T., Hager, H.D., and Lang, M. (1985). Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. Hum Genet *71*, 281-287.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet *8*, 9-22.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet *8*, 272-285.

Smith, P.A., and Corces, V.G. (1995). The suppressor of Hairy-wing protein regulates the tissue-specific expression of the *Drosophila gypsy* retrotransposon. Genetics *139*, 215-228.

Smith, S.T., Wickramasinghe, P., Olson, A., Loukinov, D., Lin, L., Deng, J., Xiong, Y., Rux, J., Sachidanandam, R., Sun, H.*, et al.* (2009). Genome wide ChIP-chip analyses reveal important roles for CTCF in *Drosophila* genome organization. Dev Biol *328*, 518-528.

Song, S.U., Gerasimova, T., Kurkulos, M., Boeke, J.D., and Corces, V.G. (1994). An env-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. Genes & Development *8*, 2046-2057.

Sontheimer, E.J., and Carthew, R.W. (2005). Silence from within: endogenous siRNAs and miRNAs. Cell *122*, 9-12.

Spana, C., and Corces, V.G. (1990). DNA bending is a determinant of binding specificity for a *Drosophila* zinc finger protein. Genes & Development *4*, 1505-1515.

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

Spellman, P.T., and Rubin, G.M. (2002). Evidence for large domains of similarly expressed genes in the *Drosophila* genome. J Biol *1*, 5.

Spilianakis, C.G., and Flavell, R.A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nature Immunology *5*, 1017-1027.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature *435*, 637-645.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes & Development *20*, 2349-2354.

Sproul, D., Gilbert, N., and Bickmore, W.A. (2005). The role of chromatin structure in regulating the expression of clustered genes. Nat Rev Genet *6*, 775-781.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Talbert, P.B., and Henikoff, S. (2006). Spreading of silent chromatin: inaction at a distance. Nat Rev Genet *7*, 793-803.

Thorvaldsen, J.L., Duran, K.L., and Bartolomei, M.S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes & Development *12*, 3693-3702.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. Nature Genetics *38*, 694-699.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular Cell *10*, 1453-1465.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science (New York, NY) *313*, 320-324.

Versteeg, R., van Schaik, B.D., van Batenburg, M.F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H.J., and van Kampen, A.H. (2003). The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res *13*, 1998-2004.

Visser, A.E., Jaunin, F., Fakan, S., and Aten, J.A. (2000). High resolution analysis of interphase chromosome domains. Journal of Cell Science *113 (Pt 14)*, 2585-2593.

Walhout, A.J. (2006). Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. Genome Res *16*, 1445-1454.

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

Wallace, J.A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. Current Opinion in Genetics & Development *17*, 400-407.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nature genetics *40*, 897-903.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T.*, et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796-801.

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

White, S.E., Habera, L.F., and Wessler, S.R. (1994). Retrotransposons in the flanking regions of normal plant genes: a role for *copia*-like elements in the evolution of gene structure and expression. Proceedings of the National Academy of Sciences of the United States of America *91*, 11792-11796.

Whitfield, W.G., Chaplin, M.A., Oegema, K., Parry, H., and Glover, D.M. (1995). The 190 kDa centrosome-associated protein of *Drosophila melanogaster* contains four zinc finger motifs and binds to specific sites on polytene chromosomes. Journal of Cell Science *108 (Pt 11)*, 3377-3387.

Wilanowski, T.M., Gibson, J.B., and Symonds, J.E. (1995). Retrotransposon insertion induces an isozyme of sn-glycerol-3-phosphate dehydrogenase in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America *92*, 12065-12069.

Yao, J., Zobeck, K.L., Lis, J.T., and Webb, W.W. (2008). Imaging transcription dynamics at endogenous genes in living *Drosophila* tissues. Methods (San Diego, Calif *45*, 233-241.

Yi, G., Sze, S.H., and Thon, M.R. (2007). Identifying clusters of functionally related genes in genomes. Bioinformatics *23*, 1053-1060.

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Molecular Cell *13*, 291-298.

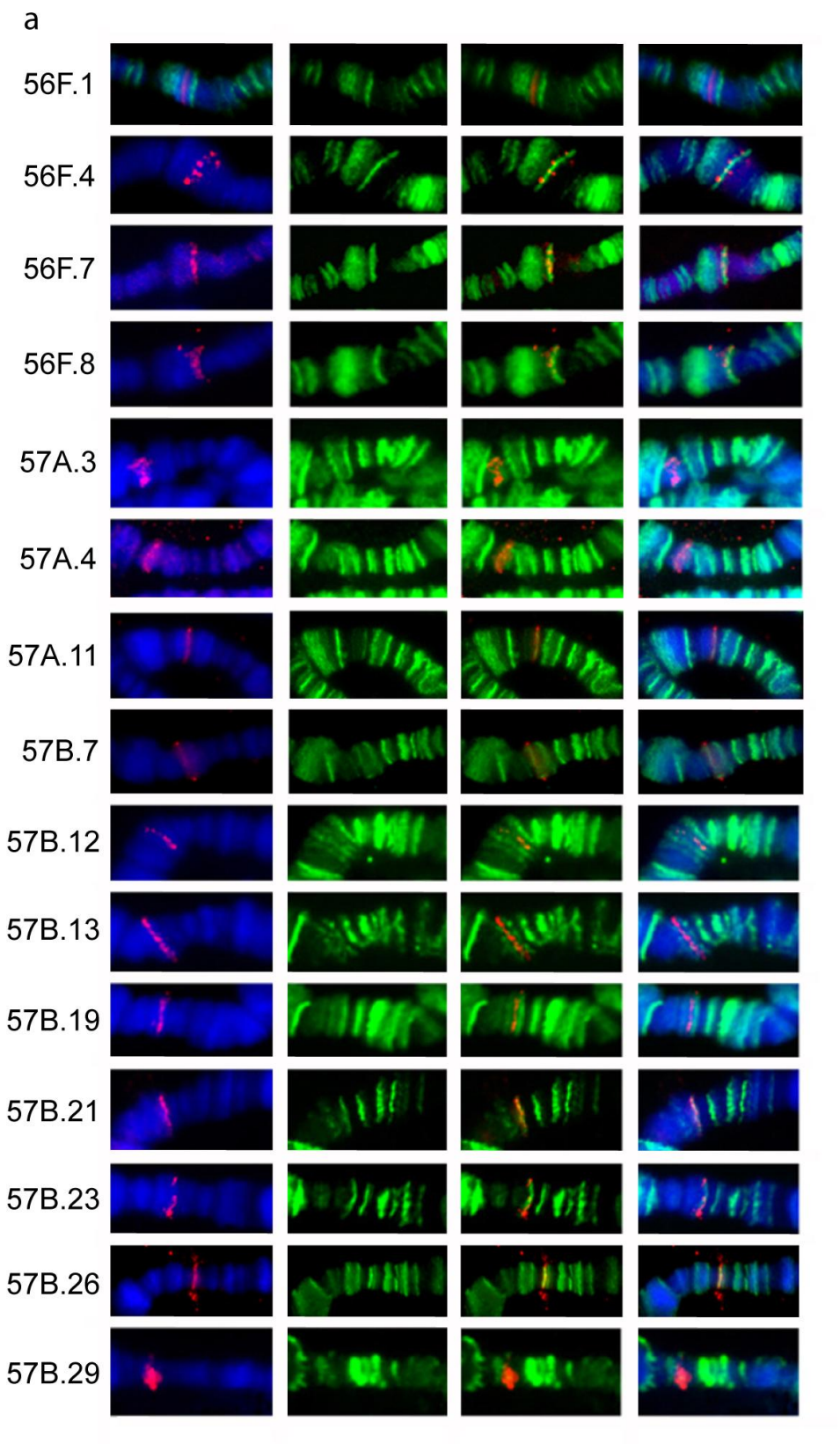
Zaratiegui, M., Irvine, D.V., and Martienssen, R.A. (2007). Noncoding RNAs and gene silencing. Cell *128*, 763-776.

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

APPENDIX

Fig. A.1 *In situ* hybridization combined with immunostaining in cytological subdivisions 56F-58A from polytene chromosomes.

A total of 29 probes containing Su(Hw) binding sites are shown. DAPI is shown in blue. *In situ* hybridization signal is shown in red and Su(Hw) immunostaining is shown in green. **a**. Sites 56F.1to 57B.29. **b**. Sites 57D.1 to 58A.8.


a
56F.1
56F.4
56F.7
56F.8
57A.3
57A.4
57A.11
57B.7
57B.12
57B.13
57B.19
57B.21
57B.23
57B.26
57B.29

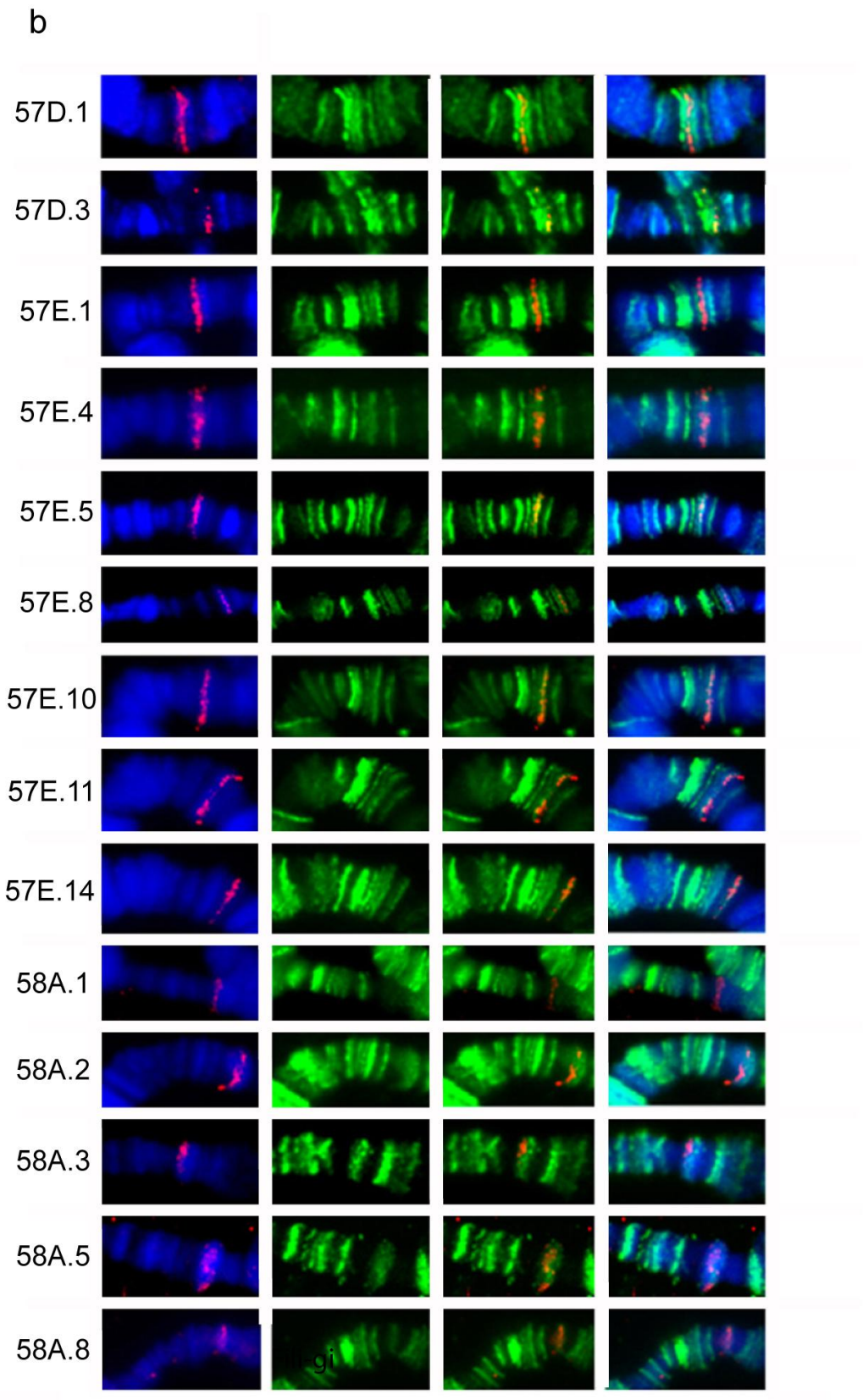

b
57D.1
57D.3
57E.1
57E.4
57E.5
57E.8
57E.10
57E.11
57E.14
58A.1
58A.2
58A.3
58A.5
58A.8

Fig. A2 High resolution of Su(Hw) immunostaining in polytene chromosomes reveals independent binding sites at distances as low as 11 kb. Su(Hw) binding sites obtained by tiling microarray (**a**) are compared with the distribution of Su(Hw) immunostaining sites in polytene chromosomes (**b**). Sites separated by distances less than 10 kb, such as those, *Glycogenin* (57D.1 – 57D.2) or *CG18375* (57D.3 – 57D.4) never appear as independent sites. Sites 57D.5 and 57D.6 are separated by 30kb and may appear as a single or as a double band. (**c**) Distances of 10kb are clearly resolved between *king tubby* (57B.30) and CG4050 (57B.31) sites. Microarray data from modENCODE is shown using Integrated Genome Browser.

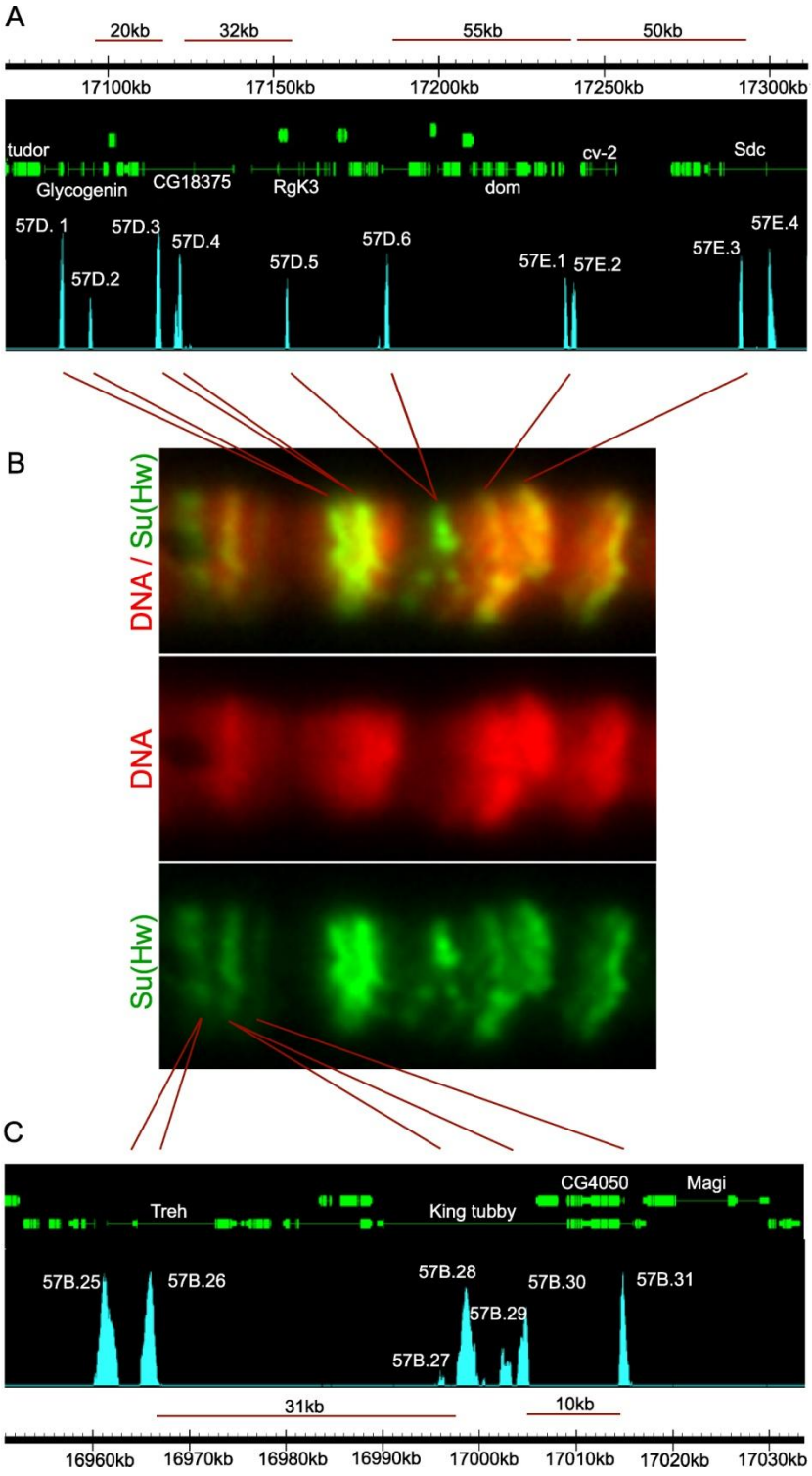

A
20kb
32kb
55kb
50kb
17100kb
17150kb
17200kb
17250kb
17300kb
tudor
Glycogenin
CG18375
RgK3
dom
cv-2
Sdc
57D. 1
57D.2
57D.3
57D.4
57D.5
57D.6
57E.1
57E.2
57E.3
57E.4
B
DNA / Su(Hw)
DNA
Su(Hw)
C
CG4050
Magi
Treh
King tubby
57B.25
57B.26
57B.27
57B.28
57B.29
57B.30
57B.31
31kb
10kb
16960kb
16970kb
16980kb
16990kb
17000kb
17010kb
17020kb
17030kb

Fig. A.3 Chromatin immunoprecipitation results are highly variable when comparing individual Su(Hw) binding sites along the cytological subdivisions 56F-58A.

a. Percentage of input obtained from ChIP experiments using 13 independent Su(Hw) binding sites. An endogenous copy of the *gypsy* retrotransposon was used as positive control (*gypsy*) and the *RpL32* gene (*rp49)* was used as a negative control. **b.** Polytene chromosome immunostaining, using an antibody against Su(Hw), spanning the same cytological region. Although there is a certain correlation between the intensity of polytene chromosome immunostaining and ChIP assays, results also show that relatively intense signals in polytene chromosomes may produce a low level of enrichment in ChIP assays using embryonic chromatin (56F.7 or 57E.1) whereas fainter signals in polytene chromosomes may yield a high enrichment in ChIP assays (58A.1 or 58A.8).

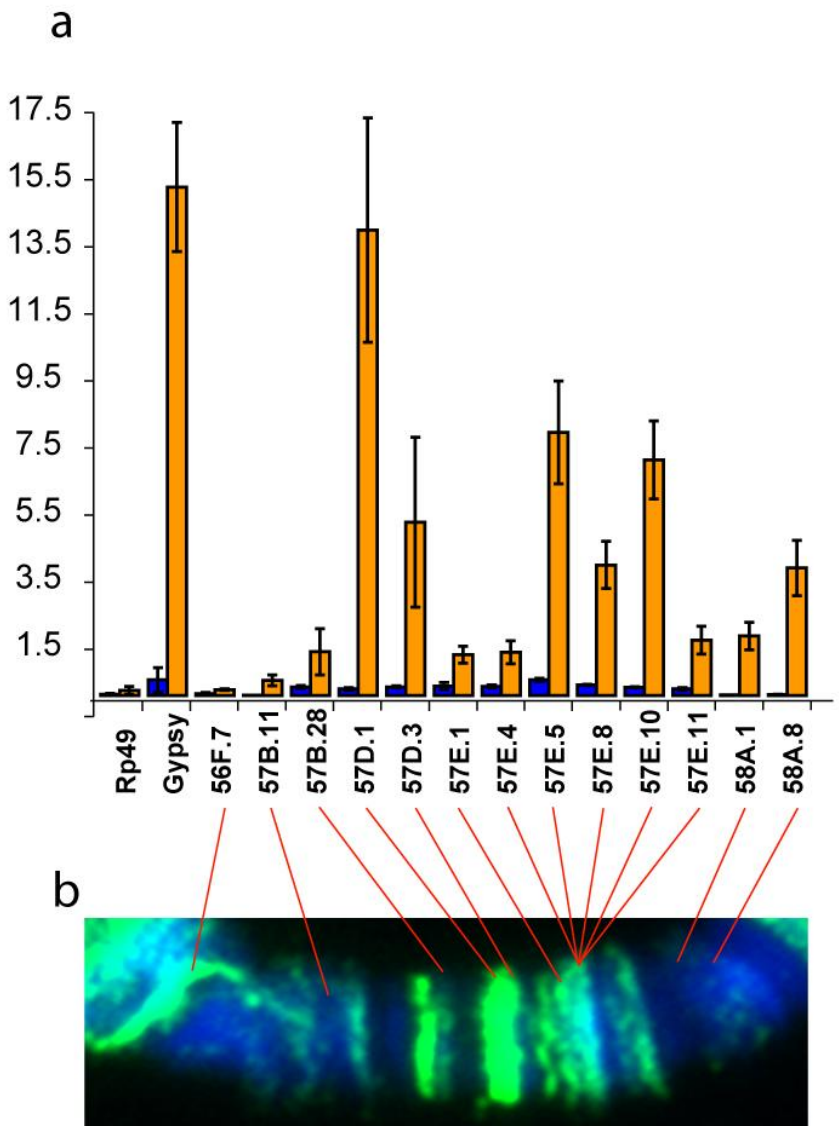

a
17.5
15.5
13.5
11.5
9.5
7.5
5.5
3.5
1.5
Rp49
Gypsy
56F.7
57B.11
57B.28
57D.1
57D.3
57E.1
57E.4
57E.5
57E.8
57E.10
57E.11
58A.1
58A.8
b

Fig. A.4 Ectopic activation of *Sdc* by Gal4 disorganizes the flanking polytene chromosome structure across more than 380 kb.

The banding pattern revealed by DAPI in polytene chromosomes in the Oregon R (OR) stock (**a**) is compared with the same pattern in the $Sdc^{EY04602}$ stock after ectopic activation of transcription at the *Sdc* promoter by Gal4 (**b**). The dense DAPI staining normally associated with *Sdc* disappears and is substituted by dark patches that span more than 300kb in both directions. Su(Hw) and CP190 immunostaining is also altered in the sites flanking *Sdc*, and these changes map hundreds of kb away from the *Sdc* gene (**c**).

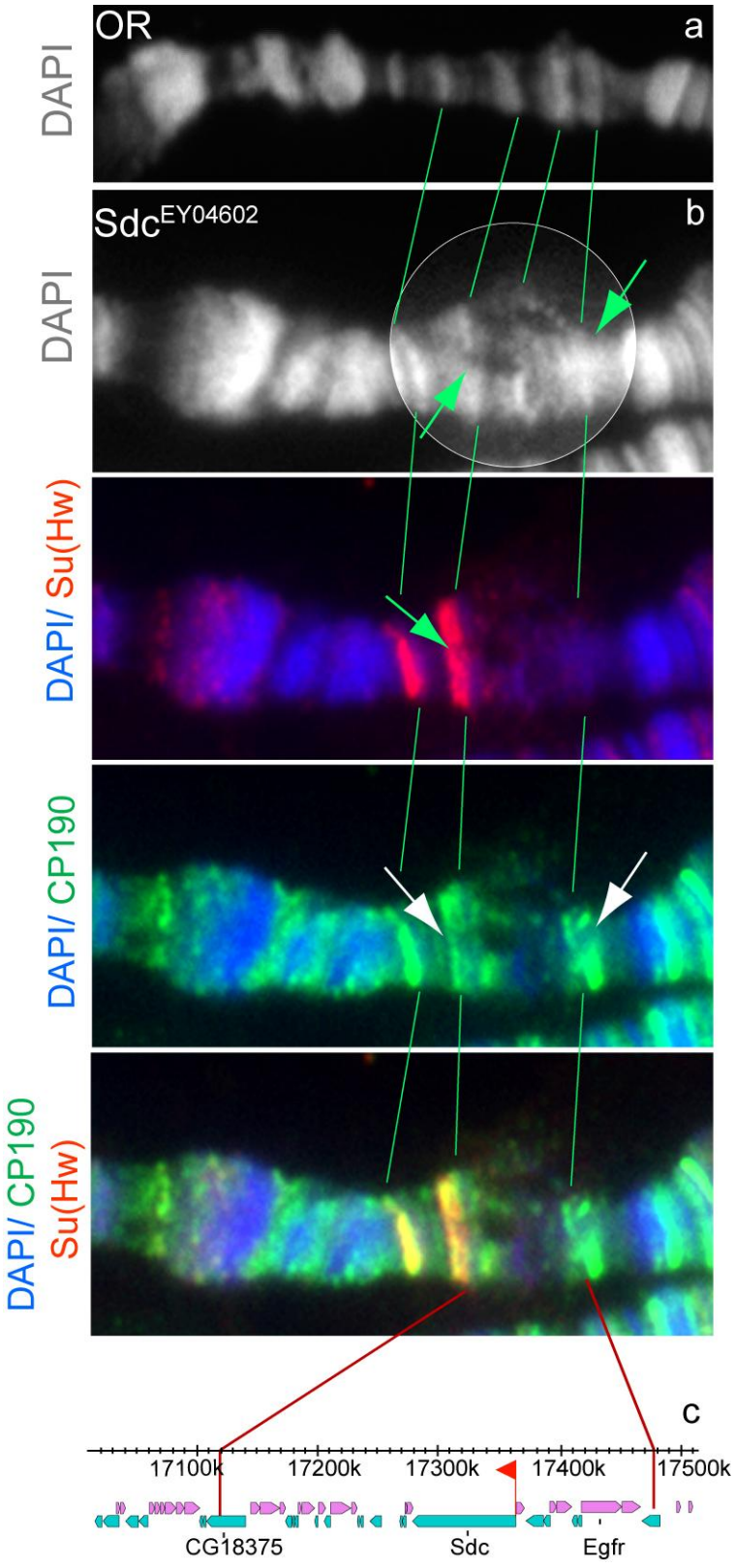

OR
a
DAPI
Sdc^EY04602
b
DAPI
DAPI/ Su(Hw)
DAPI/ CP190
DAPI/ CP190
Su(Hw)
c
17100k
17200k
17300k
17400k
17500k
CG18375
Sdc
Egfr

Table A.1 Distribution of Su(Hw) binding sites in cytological subdivisions 56E-58A from polytene chromosomes.

Su(Hw) binding sites in this region were named sequentially following Bridges Cytological subdivisions. Cytological subdivisions are show in alternative background colors. Sites from modENCODE and Bushey et al (2009) are compared. Only 11 sites where found in data from modENCODE that were not found in Bushey et al (2009) whereas 4 sites were found in Bushey et al (2009) and were missing in modENCODE. The position of binding sites in relation to genes is also indicated. Red characters are used for sites localized within transcribed regions of genes. Black characters are used to indicate sites localized in intergenic DNA. 37 sites map in intergenic DNA whereas 41 sites map in transcribed DNA. Nucleotide positions are given for sites as well as for the probe used in *in situ* hybridizations. Sites selected for *in situ* hybridization that are indicated with a "+" sign correspond to a visible immunostaining band, while those that do not are indicated with a "–" sign. Sites selected for ChIP are indicated with a "+" sign if significantly enriched for the Su(Hw) protein or a "–" sign if the level of enrichment was not significant.

Supplementary table 1: Distribution of Su(Hw) binding sites in cytological subdivisions 56E-58A from polytene chromosomes

Gene (size in kb)	Su(Hw) modENCODE	Su(Hw) Bushey, et al	Nucleotide position (2R)	FISH	FISH Probe Position	ChIP	Gene (size in kb)	Bridges Cytological division
CG10822 (0.4)	**56F.1**		15,863,341	+	15,868,574 - 15,869,594		CG16898 (1.4)	
CG16898 (1.4)	**56F.2** **56F.3**		15,918,234 - 15,927,927				18w (5.4)	
CG11041 (1.1)	**56F.4**		16,084,522	-	16,084,226- 16,084,705		CG16894 (1)	
CG16894 (1)	**56F.5** **56F.6** **56F.7**	**56F.5** **56F.6** **56F.7**	16,102,708 - 16,118,763	+	16,118,252 - 16118727	-	CG11044 (3.1)	**56F**
CG11044 (3.1)	**56F.8**	**56F.8**	16,123,603	+	16,131,453 -16,131,949		CG11048 (2.9)	
CG8929 (3.4)	**56F.9**	**56F.9**	16,228,969				CG16739 (0.7)	
CG11192 (0.9)	**57A.3**	**57A.1** **57A.2** **57A.3**	16,298,449	-	16,298,025 - 16,298,744		CG12484 (54.6)	
CG12484 (54.6)	**57A.4** **57A.5** **57A.7** **57A.8**	**57A.4** **57A.5** **57A.6** **57A.7** **57A.8**	16,316,842 - 16,354,910	-	16,316,699 - 16,317,289			**57A**
CG12484 (54.6)	**57A.9**	**57A.9**	16,387,959				obp57c (1.2)	

Supplementary table 1: Distribution of Su(Hw) binding sites in cytological subdivisions 56E-58A from polytene chromosomes

Gene (size in kb)	Su(Hw) modENCODE	Su(Hw) Bushey, et al	Nucleotide position (2R)	FISH	FISH Probe Position	ChIP	Gene (size in kb)	Bridges Cytological division
CG18065 (1.8)	**57A.10**		16,429,609				CG13424 (3.2)	
bl (19.4)	**57A.11**	**57A.11**	16,490,726	**+**	16.490,201-16,491,317			
CG13438 (0.7)	**57B.1**		16,586,135				dpr (51.6)	
dpr (51.6)	**57B.2** **57B.3** **57B.4** **57B.5** **57B.6** **57B.7**	**57B.3** **57B.4** **57B.5** **57B.6** **57B.7**	16,589,469 - 16,626,367	**+**	16,625,642 - 16,626,788			
CG15225 (0.7)	**57B.8** **57B.9** **57B.10** **57B.11**	**57B.8** **57B.9** **57B.10** **57B.11**	16,689,311 - 16,708,017				Insc (14.8)	**57B**
Insc (14.8)	**57B.12** **57B.13**	**57B.13**	16,720,342 - 16,728,385	- -	16,723,323 - 16,723,870 16,727,864 - 16,728,868	+	CG17999 (2)	
IM14 (0.2)	**57B.14**		16,758,160				Ipk1 (8.3)	
Ipk1 (8.3) **otp (19.8)**	**57B.15** **57B.16**	**57B.15**	16,766,718 16,772,030					

Supplementary table 1: Distribution of Su(Hw) binding sites in cytological subdivisions 56E-58A from polytene chromosomes

Gene (size in kb)	Su(Hw) modENCODE	Su(Hw) Bushey, et al	Nucleotide position (2R)	FISH	FISH Probe Position	ChIP	Gene (size in kb)	Bridges Cytological division
	57B.17		- 16,782,754					
	57B.18							
Rx (20.2)	**57B.19**		16,805,279	**-**	16,805,398 - 16,805,918			
Rx (20.2)	**57B.20**		16,825,919				Actin57B (2.4)	
Actin57B (2.4)	**57B.21**	**57B.21**	16,835,497	+	16,835,321 - 16,835,839		CG33704 (1)	
hbn (6.2)	**57B.22**	**57B.22**	16,854,827		16,859,899 - 16,860,446		CG15649 (0.8)	
	57B.23	**57B.23**	- 16,860,183	+				
CG15653 (0.6)	**57B.24**		16,948,242				mRpL54 (0.8)	
CG10527 (2.7)	**57B.25**	**57B.25**	16,961,395				Treh (14)	
Treh (14)	**57B.26**	**57B.26**	16,965,920	**+**	16,965,221 - 16,966,264			
	57B.27	**57B.27**						
king tubby (22)	**57B.28**	**57B.28**	16,996,021					
	57B.29	**57B.29**	- 17,004,778					
	57B.30	**57B.30**						
CG4050 (4.7)	**57B.31**	**57B.31**	17,014,809	**+**	17,010,117 - 17,010,596	**+**		
CG15658 (5.9)	**57D.1**	**57D.1**	17,087,325	+	17,086,807 - 17,087,417	+	Glycogenin (12.5)	**57D**

Supplementary table 1: Distribution of Su(Hw) binding sites in cytological subdivisions 56E-58A from polytene chromosomes

Gene (size in kb)	Su(Hw) modENCODE	Su(Hw) Bushey, et al	Nucleotide position (2R)	FISH	FISH Probe Position	ChIP	Gene (size in kb)	Bridges Cytological division
Glycogenin (12.5)	**57D.2**	**57D.2**	17.096.029					
CG18375 (33)	**57D.3**	**57D.3**	17,116,585	**+**	17,116,278 -	**+**		
	57D.4	**57D.4**	- 17,123,159		17,116,817			
Rgk3 (23.8)	**57D.5**	**57D.5**	17,155,651					
CG30387 (14.7)	**57D.6**	**57D.6**	17,186,101					
CG17974 (2.7)	**57E.1**	**57E.1**	17,240,608	+	17,239,242 - 17,239,777	+	cv-2 (11)	
cv-2 (11)	**57E.2**	**57E.2**	17,243,293					
	57E.3	**57E.3**			17,299,134 - 17,298,550			
	57E.4	**57E.4**		**+**	17,302,918 -	**+**		
	57E.5	**57E.5**		**+**	17,302,415	**+**		**57E**
	57E.6	**57E.6**						
Sdc (87.5)	**57E.7**	**57E.7**	17,294,077 - 17,354,975		17,315,224 -			
	57E.8	**57E.8**		**+**	17,315,734	**+**		
	57E.9	**57E.9**						
	57E.10	**57E.10**		**+**	17,326,681 -	**+**		
	57E.11			**+**	17,327,295	**+**		
	57E.12	**57E.12**			17,351,113 -			

Supplementary table 1: Distribution of Su(Hw) binding sites in cytological subdivisions 56E-58A from polytene chromosomes

Gene (size in kb)	Su(Hw) modENCODE	Su(Hw) Bushey, et al	Nucleotide position (2R)	FISH	FISH Probe Position	ChIP	Gene (size in kb)	Bridges Cytological division
					17,351,674			
Egfr (36.4)	**57E.13**	**57E.13**	17,422,744					
CG10440 (11.2)	**57E.14**	**57E.14**	17,458,944	**+**	17,458,692 - 17,459,210			
CG33225 (1.2)	**57F.1 57F.2 57F.3**	**57F.3**	17,481,358 -17,484,857				CG10433 (1.3)	**57F**
LBR (3.6)	**58A.1**		17,617,230	-	17,616,860 - 17,617,896	+	Grx-1 (0.5)	
Grx-1 (0.5)	**58A.2**	**58A.2**	17,651,973	-	17,651,142 - 17,652,369		CG30395 (4.7)	
CG4021 (1.5)	**58A.3**	**58A.3**	17,672,334	-	17,672,257 - 17,672,811		lox2 (2)	
CG4372 (1.2)	**58A.4**	**58A.4**	17,732,356				CG9294 (1.1)	**58A**
Fili(74)	**58A.5**	**58A.5**	17,827,320	**-**	17,826,970 - 17,827,523			
Fili (74)	**58A.6**	**58A.6**	17,837,535				CG13488 (0.9)	
CG13494 (0.6)	**58A.7**	**58A.7**	17,848,326				CG34369 (6.5)	
CG34369 (6.5)	**58A.8**	**58A.8**	17,857,295	-	17,848,012 - 17,848,523	+		

Table A.2 Distribution of clusters of coexpressed and non-coexpressed genes found in cytological subdivisions 56F-58A. Clusters of coexpressed genes were obtained from Spellman and Rubin (2002) and merged with the *Drosophila* annotated genome (release 5.5). Clusters of coexpressed genes are indicated by an orange background color whereas clusters of non-coexpressed genes are indicated by a grey color. Boundaries are indicated in green. Insulator sites are annotated in column H. Sites in red correspond to intragenic sites and sites in black correspond to intergenic sites. Sites are named as in Table A.1.

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0034455	CG11007	15390584	15391773	CG11007	CG11007	
FBgn0050223	transfer RNA:CR30223	15393844	15393915	CR30223	tRNA:CR30223	
FBgn0034456	Ionotropic receptor 56b	15395672	15396949	CG15121	Ir56b	
FBgn0034457	Ionotropic receptor 56c	15397273	15398976	CG15122	Ir56c	
FBgn0034458	Ionotropic receptor 56d	15399178	15401076	CG15904	Ir56d	
FBgn0003435	smooth	15417414	15519007	CG9218	sm	
FBgn0053535	transfer RNA:CR33535	15452378	15452449	CR33535	tRNA:CR33535	
FBgn0034459	CG16716	15481295	15484371	CG16716	CG16716	
FBgn0034460	CG18367	15524165	15524803	CG18367	CG18367	
FBgn0034461	CG15124	15530543	15531118	CG15124	CG15124	
FBgn0034462	CG15905	15534257	15535073	CG15905	CG15905	
FBgn0034463	CG15125	15545288	15547092	CG15125	CG15125	
FBgn0046817	mir-6-3	15548227	15548248	CR33039	mir-6-3	
FBgn0046818	mir-6-2	15548379	15548400	CR33040	mir-6-2	
FBgn0046819	mir-6-1	15548517	15548538	CR33004	mir-6-1	
FBgn0046820	mir-5	15548705	15548727	CR33036	mir-5	
FBgn0046821	mir-4	15548806	15548826	CR33038	mir-4	
FBgn0082173	mir-4S	15548838	15548861	CR33561	mir-4S	
FBgn0067706	mir-286	15548940	15548962	CR33602	mir-286	
FBgn0046822	mir-3	15549105	15549126	CR33037	mir-3	
FBgn0067695	mir-309	15549215	15549236	CR33613	mir-309	
FBgn0034464	CG11018	15549941	15551243	CG11018	CG11018	
FBgn0015949	hiiragi	15551422	15557326	CG9854	hrg	
FBgn0028372	isopeptidase-T-3	15558124	15560982	CG11025	isopeptidase-T-3	
FBgn0040730	CG15127	15568272	15568878	CG15127	CG15127	
FBgn0085227		15570029	15570478	CG34198	CG34198	
FBgn0034467	CG15128	15573111	15575583	CG15128	CG15128	
FBgn0040729	CG15126	15578350	15578511	CG15126	CG15126	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0034468	**Odorant-binding protein 56a**	**15585228**	**15586016**	**CG11797**	**Obp56a**	
FBgn0046880	**Odorant-binding protein 56b**	**15586347**	**15586822**	**CG30129**	**Obp56b**	
FBgn0046879	**Odorant-binding protein 56c**	**15587678**	**15588573**	**CG30128**	**Obp56c**	
FBgn0034470	**Odorant-binding protein 56d**	**15590635**	**15591300**	**CG11218**	**Obp56d**	
FBgn0034471	**Odorant-binding protein 56e**	**15599850**	**15600466**	**CG8462**	**Obp56e**	
FBgn0043533	**Odorant-binding protein 56f**	**15600924**	**15601355**	**CG30450**	**Obp56f**	
FBgn0034472	**CG8517**	**15601581**	**15602378**	**CG8517**	**CG8517**	
FBgn0050215	**transfer RNA:CR30215**	**15603070**	**15603143**	**CR30215**	**tRNA:CR30215**	
FBgn0050452	**transfer RNA:CR30452**	**15613125**	**15613196**	**CR30452**	**tRNA:CR30452**	
FBgn0050218	**transfer RNA:CR30218**	**15613410**	**15613481**	**CR30218**	**tRNA:CR30218**	
FBgn0011850	**transfer RNA:glu4:56Fc**	**15613647**	**15613718**	**CR30451**	**tRNA:E4:56Fc**	
FBgn0065076	**snoRNA:185**	**15614603**	**15614657**	**CR33930**	**snoRNA:185**	
FBgn0011849	**transfer RNA:glu4:56Fb**	**15614965**	**15615036**	**CR30455**	**tRNA:E4:56Fb**	
FBgn0011848	**transfer RNA:glu4:56Fa**	**15615486**	**15615557**	**CR30453**	**tRNA:E4:56Fa**	
FBgn0050454	**transfer RNA:CR30454**	**15615693**	**15615764**	**CR30454**	**tRNA:CR30454**	
FBgn0050220	**transfer RNA:CR30220**	**15616392**	**15616463**	**CR30220**	**tRNA:CR30220**	
FBgn0050449	**transfer RNA:CR30449**	**15616787**	**15616858**	**CR30449**	**tRNA:CR30449**	
FBgn0053452	**5SrRNA:CR33452**	**15617067**	**15617201**	**CR33452**	**5SrRNA:CR33452**	
FBgn0053451	**5SrRNA:CR33451**	**15617426**	**15617560**	**CR33451**	**5SrRNA:CR33451**	
FBgn0053450	**5SrRNA:CR33450**	**15617809**	**15617943**	**CR33450**	**5SrRNA:CR33450**	
FBgn0053449	**5SrRNA:CR33449**	**15618175**	**15618309**	**CR33449**	**5SrRNA:CR33449**	
FBgn0053448	**5SrRNA:CR33448**	**15618548**	**15618682**	**CR33448**	**5SrRNA:CR33448**	
FBgn0053447	**5SrRNA:CR33447**	**15618914**	**15619048**	**CR33447**	**5SrRNA:CR33447**	
FBgn0053446	**5SrRNA:CR33446**	**15619273**	**15619407**	**CR33446**	**5SrRNA:CR33446**	
FBgn0053445	**5SrRNA:CR33445**	**15619675**	**15619809**	**CR33445**	**5SrRNA:CR33445**	
FBgn0053444	**5SrRNA:CR33444**	**15620055**	**15620189**	**CR33444**	**5SrRNA:CR33444**	
FBgn0053443	**5SrRNA:CR33443**	**15620431**	**15620565**	**CR33443**	**5SrRNA:CR33443**	
FBgn0053442	**5SrRNA:CR33442**	**15620793**	**15620927**	**CR33442**	**5SrRNA:CR33442**	
FBgn0053441	**5SrRNA:CR33441**	**15621176**	**15621310**	**CR33441**	**5SrRNA:CR33441**	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0053440	**5SrRNA:CR33440**	**15621542**	**15621676**	**CR33440**	**5SrRNA:CR33440**	
FBgn0053439	**5SrRNA:CR33439**	**15621901**	**15622035**	**CR33439**	**5SrRNA:CR33439**	
FBgn0053438	**5SrRNA:CR33438**	**15622277**	**15622411**	**CR33438**	**5SrRNA:CR33438**	
FBgn0053437	**5SrRNA:CR33437**	**15622646**	**15622780**	**CR33437**	**5SrRNA:CR33437**	
FBgn0053436	**5SrRNA:CR33436**	**15623015**	**15623149**	**CR33436**	**5SrRNA:CR33436**	
FBgn0053435	**5SrRNA:CR33435**	**15623384**	**15623518**	**CR33435**	**5SrRNA:CR33435**	
FBgn0053434	**5SrRNA:CR33434**	**15623753**	**15623887**	**CR33434**	**5SrRNA:CR33434**	
FBgn0053433	**5SrRNA:CR33433**	**15624122**	**15624256**	**CR33433**	**5SrRNA:CR33433**	
FBgn0053432	**5SrRNA:CR33432**	**15624491**	**15624625**	**CR33432**	**5SrRNA:CR33432**	
FBgn0053431	**5SrRNA:CR33431**	**15624860**	**15624994**	**CR33431**	**5SrRNA:CR33431**	
FBgn0053430	**5SrRNA:CR33430**	**15625229**	**15625363**	**CR33430**	**5SrRNA:CR33430**	
FBgn0053429	**5SrRNA:CR33429**	**15625591**	**15625725**	**CR33429**	**5SrRNA:CR33429**	
FBgn0053428	**5SrRNA:CR33428**	**15625957**	**15626091**	**CR33428**	**5SrRNA:CR33428**	
FBgn0053427	**5SrRNA:CR33427**	**15626323**	**15626457**	**CR33427**	**5SrRNA:CR33427**	
FBgn0053426	**5SrRNA:CR33426**	**15626699**	**15626833**	**CR33426**	**5SrRNA:CR33426**	
FBgn0053425	**5SrRNA:CR33425**	**15627075**	**15627209**	**CR33425**	**5SrRNA:CR33425**	
FBgn0053424	**5SrRNA:CR33424**	**15627441**	**15627575**	**CR33424**	**5SrRNA:CR33424**	
FBgn0053423	**5SrRNA:CR33423**	**15627810**	**15627944**	**CR33423**	**5SrRNA:CR33423**	
FBgn0053422	**5SrRNA:CR33422**	**15628169**	**15628303**	**CR33422**	**5SrRNA:CR33422**	
FBgn0053421	**5SrRNA:CR33421**	**15628545**	**15628679**	**CR33421**	**5SrRNA:CR33421**	
FBgn0053420	**5SrRNA:CR33420**	**15628911**	**15629045**	**CR33420**	**5SrRNA:CR33420**	
FBgn0053419	**5SrRNA:CR33419**	**15629273**	**15629407**	**CR33419**	**5SrRNA:CR33419**	
FBgn0053418	**5SrRNA:CR33418**	**15629642**	**15629776**	**CR33418**	**5SrRNA:CR33418**	
FBgn0053417	**5SrRNA:CR33417**	**15630008**	**15630142**	**CR33417**	**5SrRNA:CR33417**	
FBgn0053416	**5SrRNA-Psi:CR33416**	**15630381**	**15630523**	**CR33416**	**5SrRNA-Psi:CR33416**	
FBgn0053415	**5SrRNA:CR33415**	**15630758**	**15630892**	**CR33415**	**5SrRNA:CR33415**	
FBgn0053414	**5SrRNA:CR33414**	**15631134**	**15631268**	**CR33414**	**5SrRNA:CR33414**	
FBgn0053413	**5SrRNA:CR33413**	**15631500**	**15631634**	**CR33413**	**5SrRNA:CR33413**	
FBgn0053412	**5SrRNA:CR33412**	**15631866**	**15632000**	**CR33412**	**5SrRNA:CR33412**	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0053410	**5SrRNA:CR33410**	**15632601**	**15632735**	**CR33410**	**5SrRNA:CR33410**	
FBgn0053409	**5SrRNA:CR33409**	**15632967**	**15633101**	**CR33409**	**5SrRNA:CR33409**	
FBgn0053408	**5SrRNA:CR33408**	**15633333**	**15633467**	**CR33408**	**5SrRNA:CR33408**	
FBgn0053407	**5SrRNA:CR33407**	**15633706**	**15633840**	**CR33407**	**5SrRNA:CR33407**	
FBgn0053406	**5SrRNA:CR33406**	**15634079**	**15634213**	**CR33406**	**5SrRNA:CR33406**	
FBgn0053405	**5SrRNA:CR33405**	**15634448**	**15634581**	**CR33405**	**5SrRNA:CR33405**	
FBgn0053404	**5SrRNA:CR33404**	**15634823**	**15634957**	**CR33404**	**5SrRNA:CR33404**	
FBgn0053403	**5SrRNA:CR33403**	**15635182**	**15635316**	**CR33403**	**5SrRNA:CR33403**	
FBgn0053402	**5SrRNA:CR33402**	**15635548**	**15635682**	**CR33402**	**5SrRNA:CR33402**	
FBgn0053401	**5SrRNA:CR33401**	**15635910**	**15636044**	**CR33401**	**5SrRNA:CR33401**	
FBgn0053400	**5SrRNA:CR33400**	**15636279**	**15636413**	**CR33400**	**5SrRNA:CR33400**	
FBgn0053399	**5SrRNA:CR33399**	**15636648**	**15636782**	**CR33399**	**5SrRNA:CR33399**	
FBgn0053398	**5SrRNA:CR33398**	**15637017**	**15637151**	**CR33398**	**5SrRNA:CR33398**	
FBgn0053397	**5SrRNA:CR33397**	**15637383**	**15637517**	**CR33397**	**5SrRNA:CR33397**	
FBgn0053396	**5SrRNA:CR33396**	**15637756**	**15637890**	**CR33396**	**5SrRNA:CR33396**	
FBgn0053395	**5SrRNA:CR33395**	**15638125**	**15638259**	**CR33395**	**5SrRNA:CR33395**	
FBgn0053394	**5SrRNA:CR33394**	**15638494**	**15638628**	**CR33394**	**5SrRNA:CR33394**	
FBgn0053393	**5SrRNA:CR33393**	**15638863**	**15638997**	**CR33393**	**5SrRNA:CR33393**	
FBgn0053392	**5SrRNA:CR33392**	**15639232**	**15639366**	**CR33392**	**5SrRNA:CR33392**	
FBgn0053391	**5SrRNA:CR33391**	**15639601**	**15639734**	**CR33391**	**5SrRNA:CR33391**	
FBgn0053390	**5SrRNA:CR33390**	**15639962**	**15640096**	**CR33390**	**5SrRNA:CR33390**	
FBgn0053389	**5SrRNA:CR33389**	**15640321**	**15640455**	**CR33389**	**5SrRNA:CR33389**	
FBgn0053388	**5SrRNA:CR33388**	**15640694**	**15640828**	**CR33388**	**5SrRNA:CR33388**	
FBgn0053387	**5SrRNA:CR33387**	**15641063**	**15641197**	**CR33387**	**5SrRNA:CR33387**	
FBgn0053386	**5SrRNA:CR33386**	**15641432**	**15641566**	**CR33386**	**5SrRNA:CR33386**	
FBgn0053385	**5SrRNA:CR33385**	**15641801**	**15641935**	**CR33385**	**5SrRNA:CR33385**	
FBgn0053384	**5SrRNA:CR33384**	**15642170**	**15642304**	**CR33384**	**5SrRNA:CR33384**	
FBgn0053383	**5SrRNA:CR33383**	**15642539**	**15642673**	**CR33383**	**5SrRNA:CR33383**	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0053381	**5SrRNA:CR33381**	**15643277**	**15643411**	**CR33381**	**5SrRNA:CR33381**	
FBgn0053380	**5SrRNA:CR33380**	**15643646**	**15643780**	**CR33380**	**5SrRNA:CR33380**	
FBgn0053379	**5SrRNA:CR33379**	**15644015**	**15644149**	**CR33379**	**5SrRNA:CR33379**	
FBgn0053378	**5SrRNA:CR33378**	**15644384**	**15644518**	**CR33378**	**5SrRNA:CR33378**	
FBgn0053377	**5SrRNA:CR33377**	**15644760**	**15644894**	**CR33377**	**5SrRNA:CR33377**	
FBgn0053376	**5SrRNA:CR33376**	**15645129**	**15645263**	**CR33376**	**5SrRNA:CR33376**	
FBgn0053375	**5SrRNA:CR33375**	**15645498**	**15645632**	**CR33375**	**5SrRNA:CR33375**	
FBgn0053374	**5SrRNA:CR33374**	**15645867**	**15646001**	**CR33374**	**5SrRNA:CR33374**	
FBgn0053373	**5SrRNA:CR33373**	**15646240**	**15646374**	**CR33373**	**5SrRNA:CR33373**	
FBgn0053372	**5SrRNA:CR33372**	**15646606**	**15646740**	**CR33372**	**5SrRNA:CR33372**	
FBgn0053371	**5SrRNA-Psi:CR33371**	**15646979**	**15647115**	**CR33371**	**5SrRNA-Psi:CR33371**	
FBgn0053370	**5SrRNA:CR33370**	**15647347**	**15647481**	**CR33370**	**5SrRNA:CR33370**	
FBgn0053369	**5SrRNA:CR33369**	**15647720**	**15647854**	**CR33369**	**5SrRNA:CR33369**	
FBgn0053368	**5SrRNA:CR33368**	**15648096**	**15648230**	**CR33368**	**5SrRNA:CR33368**	
FBgn0053367	**5SrRNA:CR33367**	**15648472**	**15648606**	**CR33367**	**5SrRNA:CR33367**	
FBgn0053366	**5SrRNA:CR33366**	**15648848**	**15648982**	**CR33366**	**5SrRNA:CR33366**	
FBgn0053365	**5SrRNA:CR33365**	**15649207**	**15649341**	**CR33365**	**5SrRNA:CR33365**	
FBgn0053364	**5SrRNA:CR33364**	**15649573**	**15649707**	**CR33364**	**5SrRNA:CR33364**	
FBgn0053363	**5SrRNA-Psi:CR33363**	**15649932**	**15650075**	**CR33363**	**5SrRNA-Psi:CR33363**	
FBgn0053362	**5SrRNA:CR33362**	**15650307**	**15650441**	**CR33362**	**5SrRNA:CR33362**	
FBgn0053361	**5SrRNA:CR33361**	**15650680**	**15650814**	**CR33361**	**5SrRNA:CR33361**	
FBgn0053360	**5SrRNA:CR33360**	**15651049**	**15651183**	**CR33360**	**5SrRNA:CR33360**	
FBgn0053359	**5SrRNA:CR33359**	**15651418**	**15651552**	**CR33359**	**5SrRNA:CR33359**	
FBgn0053358	**5SrRNA:CR33358**	**15651794**	**15651928**	**CR33358**	**5SrRNA:CR33358**	
FBgn0053357	**5SrRNA:CR33357**	**15652153**	**15652287**	**CR33357**	**5SrRNA:CR33357**	
FBgn0053356	**5SrRNA-Psi:CR33356**	**15652526**	**15652668**	**CR33356**	**5SrRNA-Psi:CR33356**	
FBgn0053355	**5SrRNA:CR33355**	**15652914**	**15653048**	**CR33355**	**5SrRNA:CR33355**	
FBgn0053354	**5SrRNA:CR33354**	**15653280**	**15653414**	**CR33354**	**5SrRNA:CR33354**	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0053353	5SrRNA:CR33353	15653649	15653783	CR33353	5SrRNA:CR33353	
FBgn0034473	Odorant receptor 56a	15656966	15658738	CG12501	Or56a	
FBgn0034474	Odorant-binding protein 56g	15671062	15671525	CG13873	Obp56g	
FBgn0011897	transfer RNA:lys2:56EF	15690096	15690168	CR30520	tRNA:K2:56EF	
FBgn0034475	Odorant-binding protein 56h	15703082	15703544	CG13874	Obp56h	
FBgn0034476	Toll-7	15714410	15718750	CG8595	Toll-7	
FBgn0043532	Odorant-binding protein 56i	15755621	15756101	CG30448	Obp56i	
FBgn0034477	CG13872	15766333	15768425	CG13872	CG13872	
FBgn0050447	CG30447	15834726	15835588	CG30447	CG30447	
FBgn0034478	CG10822	15838938	15839456	CG10822	CG10822	
FBgn0034479	CG8654	15868574	15872262	CG8654	CG8654	
FBgn0050212	transfer RNA:CR30212	15887801	15887873	CR30212	tRNA:CR30212	
FBgn0053538	transfer RNA:CR33538	15888324	15888396	CR33538	tRNA:CR33538	
FBgn0034480	CG16898	15891124	15892492	CG16898	CG16898	56F.1
FBgn0004364	18 wheeler	15999016	16004437	CG8896	18w	56F.2 56F.3
FBgn0034481	CG11041	16033638	16034329	CG11041	CG11041	56F.4
FBgn0034483	CG16894	16099109	16100124	CG16894	CG16894	56F.5 56F.6 56F.7
FBgn0011869	transfer RNA:gly3:56EFa	16116986	16117056	CR30214	tRNA:G3:56EFa	
FBgn0034484	CG11044	16119336	16122438	CG11044	CG11044	56F.8
FBgn0011870	transfer RNA:gly3:56EFb	16120302	16120372	CR30138	tRNA:G3:56EFb	
FBgn0034485	CG11099	16123740	16126706	CG11099	CG11099	
FBgn0034486	CG13869	16126075	16126904	CG13869	CG13869	
FBgn0034487	CG11048	16127492	16130387	CG11048	CG11048	
FBgn0034488	CG11208	16130256	16132904	CG11208	CG11208	
FBgn0034489	pickpocket 6	16133276	16135078	CG11209	ppk6	
FBgn0034490	CG9864	16135635	16141444	CG9864	CG9864	
FBgn0025720	Ate1	16141297	16144886	CG9204	Ate1	
FBgn0034491	CG11055	16145157	16149241	CG11055	CG11055	
FBgn0005655	mutagen-sensitive 209	16149257	16150362	CG9193	mus209	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0003114	plutonium	16150463	16151224	CG9183	plu	
FBgn0010411	Ribosomal protein S18	16151477	16152399	CG8900	RpS18	
FBgn0034493	CG8908	16152437	16158290	CG8908	CG8908	
FBgn0034494	CG10444	16159156	16162030	CG10444	CG10444	
FBgn0034495	CG11788	16162191	16163648	CG11788	CG11788	
FBgn0034496	CG9143	16163475	16166501	CG9143	CG9143	
FBgn0034497	CG9090	16166561	16168629	CG9090	CG9090	
FBgn0034498	CG16868	16169134	16174584	CG16868	CG16868	
FBgn0085228		16175120	16175851	CG34199	CG34199	
FBgn0034499	Cuticular protein 56F	16175860	16179448	CG9036	Cpr56F	
FBgn0026136	Casein kinase II beta2 subunit	16177132	16178056	CG8914	CkIIbeta2	
FBgn0034500	CG11200	16192022	16196170	CG11200	CG11200	
FBgn0027529	CG8920	16193796	16213478	CG8920	CG8920	
FBgn0034501	CG13868	16196217	16204415	CG13868	CG13868	
FBgn0034502	CG13871	16204828	16206544	CG13871	CG13871	
FBgn0034503	Mediator complex subunit 8	16213346	16214311	CG13867	MED8	
FBgn0034504	CG8929	16214772	16218226	CG8929	CG8929	56F.9
FBgn0034505	CG16739	16229533	16230279	CG16739	CG16739	
FBgn0034506	CG13870	16231350	16233384	CG13870	CG13870	
FBgn0042198	CG16741	16234541	16235101	CG16741	CG16741	
FBgn0034507	CG11192	16241610	16242558	CG11192	CG11192	57A. 1 57A.2 57A. 3
FBgn0086604	CG12484	16311840	16366459	CG12484	CG12484	57A.4 57A.5 57A.6 57A.7 57A.8
FBgn0034509	Odorant-binding protein 57c	16391061	16392296	CG13421	Obp57c	57A.9
FBgn0043534	Odorant-binding protein 57b	16391755	16392322	CG30142	Obp57b	
FBgn0043535	Odorant-binding protein 57a	16392672	16393265	CG30141	Obp57a	
FBgn0034510	CG13426	16400199	16400674	CG13426	CG13426	
FBgn0034511	CG13422	16413832	16414331	CG13422	CG13422	
FBgn0050151	CG30151	16416009	16416969	CG30151	CG30151	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0050154	**CG30154**	**16420879**	**16421441**	**CG30154**	**CG30154**	
FBgn0034512	**CG18067**	**16422015**	**16422860**	**CG18067**	**CG18067**	
FBgn0034513	**CG13423**	**16424083**	**16425767**	**CG13423**	**CG13423**	
FBgn0034514	**CG13427**	**16426382**	**16426819**	**CG13427**	**CG13427**	
FBgn0034515	**CG13428**	**16427016**	**16427953**	**CG13428**	**CG13428**	
FBgn0050148	**CG30148**	**16434535**	**16435186**	**CG30148**	**CG30148**	
FBgn0050145	**Odorant-binding protein 57e**	**16435321**	**16435796**	**CG30145**	**Obp57e**	
FBgn0034517	**Cuticular protein 57A**	**16435872**	**16437597**	**CG18066**	**Cpr57A**	
FBgn0043536	**Odorant-binding protein 57d**	**16435872**	**16437597**	**CG30150**	**Obp57d**	
FBgn0034518	**CG13430**	**16437838**	**16439591**	**CG13430**	**CG13430**	
FBgn0034519	**CG18065**	**16437838**	**16439591**	**CG18065**	**CG18065**	**57A.10**
FBgn0034520	**CG13424**	**16444364**	**16447458**	**CG13424**	**CG13424**	
FBgn0034521	**UDP-GlcNAc:a-3-D-mannoside-beta-1,2-N-acetylglucosaminyltransferase I**	**16446661**	**16449832**	**CG13431**	**Mgat1**	
FBgn0028622	**lethal (2) 05510**	**16450540**	**16469633**	**CG13432**	**l(2)05510**	
FBgn0067694	**mir-310**	**16471271**	**16471292**	**CR33614**	**mir-310**	
FBgn0067693	**mir-311**	**16471393**	**16471414**	**CR33615**	**mir-311**	
FBgn0067692	**mir-312**	**16471563**	**16471584**	**CR33616**	**mir-312**	
FBgn0067691	**mir-313**	**16471699**	**16471720**	**CR33617**	**mir-313**	
FBgn0034523	**Nnf1a**	**16473727**	**16474712**	**CG13434**	**Nnf1a**	
FBgn0015907	**bancal**	**16475010**	**16494359**	**CG13425**	**bl**	**57A.11**
FBgn0046816	**mir-7**	**16493586**	**16493608**	**CR33042**	**mir-7**	
FBgn0050147	**Hillarin**	**16494386**	**16505719**	**CG30147**	**Hil**	
FBgn0034527	**CG9945**	**16506008**	**16508586**	**CG9945**	**CG9945**	
FBgn0034528	**CG11180**	**16508419**	**16510808**	**CG11180**	**CG11180**	
FBgn0034529	**CG16742**	**16511079**	**16515820**	**CG16742**	**CG16742**	
FBgn0034530	**Reduction in Cnn dots 6**	**16515691**	**16521007**	**CG11175**	**Rcd6**	
FBgn0250850	**rigor mortis**	**16521381**	**16528191**	**CG30149**	**rig**	
FBgn0034532	**CG13436**	**16523897**	**16525928**	**CG13436**	**CG13436**	
FBgn0034534	**maf-S**	**16528476**	**16529179**	**CG9954**	**maf-S**	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0034535	CG11110	16529307	16529922	CG11110	CG11110	
FBgn0061362	CG33785	16529965	16531285	CG33785	CG33785	
FBgn0061361	CG33786	16529966	16531285	CG33786	CG33786	
FBgn0034537	DMAP1	16531543	16533376	CG11132	DMAP1	
FBgn0034538	CG16799	16540720	16542596	CG16799	CG16799	
FBgn0034539	CG11159	16543359	16543997	CG11159	CG11159	
FBgn0034540	CG11136	16544745	16548012	CG11136	CG11136	
FBgn0050155	transfer RNA:CR30155	16545396	16545467	CR30155	tRNA:CR30155	
FBgn0050211	transfer RNA:CR30211	16545609	16545680	CR30211	tRNA:CR30211	
FBgn0034541	CG13437	16549254	16549932	CG13437	CG13437	
FBgn0034542	Fem-1	16549979	16554790	CG9025	Fem-1	
FBgn0000615	exuperantia	16554930	16558650	CG8994	exu	
FBgn0034543	CG30152	16558708	16560059	CG30152	CG30152	
FBgn0041240	Gustatory receptor 57a	16573110	16574415	CG13441	Gr57a	
FBgn0034545	CG13438	16580375	16581114	CG13438	CG13438	57B.1
FBgn0040726	defective proboscis extension response	16586355	16637949	CG13439	dpr	57B.2 57B.3 57B.4 57B.5 57B.6 57B.6
FBgn0034546	CG13442	16615329	16618115	CG13442	CG13442	
FBgn0085230		16638452	16639628	CG34201	CG34201	
FBgn0085231		16640110	16641244	CG34202	CG34202	
FBgn0034548	CG13443	16661234	16662618	CG13443	CG13443	
FBgn0034550	CG15226	16675822	16679643	CG15226	CG15226	
FBgn0034551	CG15225	16680472	16681191	CG15225	CG15225	57B.8 57B.9 57B.10 57B.11
FBgn0011674	inscuteable	16708771	16723569	CG11312	insc	57B.12 57B.13
FBgn0016984	skittles	16714916	16719919	CG9985	sktl	
FBgn0034552	CG17999	16735485	16737424	CG17999	CG17999	
FBgn0034553	CG9993	16737651	16739626	CG9993	CG9993	
FBgn0034554	CG15227	16753606	16754481	CG15227	CG15227	
FBgn0040653	Immune induced molecule 4	16756341	16756826	CG15231	IM4	
FBgn0067905	Immune induced molecule 14	16757930	16758134	CG33990	IM14	57B.14

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0050295	Ipk1	16758311	16766632	CG30295	Ipk1	57B.15
FBgn0034558	CG9236	16767554	16769910	CG9236	CG9236	
FBgn0015524	orthopedia	16771116	16790900	CG10036	otp	57B.16 57B.17 57B.18
FBgn0034560	CG9235	16797209	16798313	CG9235	CG9235	
FBgn0020617	Retinal Homeobox	16804285	16824537	CG10052	Rx	57B.19 57B.20
FBgn0000044	Actin 57B	16831533	16833945	CG10067	Act57B	57B.21
FBgn0065059	snoRNA:660	16833475	16833570	CR33780	snoRNA:660	
FBgn0053704	CG33704	16837607	16838599	CG33704	CG33704	
FBgn0008636	homeobrain	16842169	16848414	CG33152	hbn	57B.22 57B.23
FBgn0034563	CG15649	16860421	16861502	CG15649	CG15649	
FBgn0083951	CG34115	16863005	16864439	CG34115	CG34115	
FBgn0034564	CG9344	16864561	16865105	CG9344	CG9344	
FBgn0034565	CG15650	16865233	16865742	CG15650	CG15650	
FBgn0034566	CG9313	16866849	16871548	CG9313	CG9313	
FBgn0034567	CG15651	16871339	16874000	CG15651	CG15651	
FBgn0034568	CG3216	16874338	16878676	CG3216	CG3216	
FBgn0034569	dim gamma-tubulin 3	16879755	16881841	CG3221	dgt3	
FBgn0003150	Proteasome 29kD subunit	16882407	16883623	CG9327	Pros29	
FBgn0034570	CG10543	16884146	16894148	CG10543	CG10543	
FBgn0050291	CG30291	16895260	16897078	CG30291	CG30291	
FBgn0001133	grauzone	16897271	16899424	CG33133	grau	
FBgn0034572	CG9346	16899477	16903133	CG9346	CG9346	
FBgn0034573	CG3295	16903092	16904951	CG3295	CG3295	
FBgn0050296	CG30296	16905290	16912415	CG30296	CG30296	
FBgn0034576	CG9350	16912585	16913595	CG9350	CG9350	
FBgn0003391	shotgun	16938107	16944664	CG3722	shg	
FBgn0034577	capping protein alpha	16945296	16947388	CG10540	cpa	
FBgn0034578	CG15653	16947404	16947973	CG15653	CG15653	57B.24
FBgn0034579	mitochondrial ribosomal protein L54	16948329	16949096	CG9353	mRpL54	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0034580	Cht8	16949151	16950717	CG9357	Cht8	
FBgn0050293	Cht12	16950683	16952468	CG30293	Cht12	
FBgn0022700	Chitinase 4	16952885	16954592	CG3986	Cht4	
FBgn0034582	Cht9	16955487	16956813	CG10531	Cht9	
FBgn0034583	CG10527	16957568	16960290	CG10527	CG10527	57B.25
FBgn0003748	Trehalase	16961530	16975734	CG9364	Treh	57B.26
FBgn0034585	CG4030	16975888	16978507	CG4030	CG4030	
FBgn0085425		16979743	16988961	CG34396	CG34396	
FBgn0011824	CG4038	16983447	16984808	CG4038	CG4038	
FBgn0015721	king tubby	16989435	17011220	CG9398	king-tubby	57B.27 57B.28 57B.29 57B.30
FBgn0034588	CG9394	17005899	17008214	CG9394	CG9394	
FBgn0020312	CG4050	17011233	17015995	CG4050	CG4050	57B.31
FBgn0002736	mago nashi	17016084	17017245	CG9401	mago	
FBgn0034590	Magi	17016968	17030067	CG30388	Magi	
FBgn0034592	CG9406	17029948	17030702	CG9406	CG9406	
FBgn0021872	X box binding protein-1	17031050	17033255	CG9415	Xbp1	
FBgn0026582	CG9418	17033713	17035113	CG9418	CG9418	
FBgn0034595	CG15657	17034983	17035658	CG15657	CG15657	
FBgn0050389	CG30389	17036238	17048050	CG30389	CG30389	
FBgn0050206	transfer RNA:CR30206	17048568	17048639	CR30206	tRNA:CR30206	
FBgn0011871	transfer RNA:gly3:57BCa	17048861	17048931	CR30207	tRNA:G3:57BCa	
FBgn0050208	transfer RNA:CR30208	17049031	17049102	CR30208	tRNA:CR30208	
FBgn0034598	CG4266	17049218	17056163	CG4266	CG4266	
FBgn0034599	CG9437	17056465	17057518	CG9437	CG9437	
FBgn0034600	CG4279	17057473	17058150	CG4279	CG4279	
FBgn0015844	Xeroderma pigmentosum D	17058508	17062009	CG9433	Xpd	
FBgn0003162	Punch	17062822	17070121	CG9441	Pu	
FBgn0034601	CG4286	17065251	17066141	CG4286	CG4286	
FBgn0053539	transfer RNA:CR33539	17068789	17068860	CR33539	tRNA:CR33539	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0050209	transfer RNA:CR30209	17070183	17070254	CR30209	tRNA:CR30209	
FBgn0011872	transfer RNA:gly3:57BCb	17070361	17070431	CR30210	tRNA:G3:57BCb	
FBgn0003891	tudor	17070937	17079903	CG9450	tud	
FBgn0034602	CG15658	17081078	17086941	CG15658	CG15658	57D.1
FBgn0034603	Glycogenin	17087992	17100475	CG9480	Glycogenin	57D.2
FBgn0027073	CG4302	17100383	17102473	CG4302	CG4302	
FBgn0034605	CG15661	17102939	17105258	CG15661	CG15661	
FBgn0034606	CG18375	17105361	17138458	CG18375	CG18375	57D.3 57D.4
FBgn0085426	Rgk3	17143745	17167538	CG34397	Rgk3	57D.5
FBgn0050391	CG30391	17152344	17153847	CG30391	CG30391	
FBgn0050393	CG30393	17153869	17155084	CG30393	CG30393	
FBgn0085232		17168568	17169161	CG34203	CG34203	
FBgn0034611	CG10069	17169805	17173174	CG10069	CG10069	
FBgn0034612	CG10505	17173225	17177833	CG10505	CG10505	
FBgn0050392	CG30392	17178419	17179683	CG30392	CG30392	
FBgn0050390	CG30390	17179744	17180981	CG30390	CG30390	
FBgn0016726	Ribosomal protein L29	17181275	17181909	CG10071	RpL29	
FBgn0034614	CG9752	17181907	17183092	CG9752	CG9752	
FBgn0050387	CG30387	17183178	17197916	CG30387	CG30387	57D.6
FBgn0050386	CG30386	17187409	17187978	CG30386	CG30386	
FBgn0034617	CG9754	17197873	17199853	CG9754	CG9754	
FBgn0034618	CG9485	17200185	17207074	CG9485	CG9485	
FBgn0250824		17206941	17207522	CG33655	CG33655	
FBgn0050394	CG30394	17207543	17211056	CG30394	CG30394	
FBgn0020306	domino	17210949	17229352	CG9696	dom	
FBgn0086023		17212723	17212828	CR34652	snoRNA:Me28S-A2589a	
FBgn0082972	snoRNA:Psi28S-3316a	17214244	17214375	CR34547	snoRNA:Psi28S-3316a	
FBgn0082971	snoRNA:Psi28S-3316b	17214387	17214527	CR34548	snoRNA:Psi28S-3316b	
FBgn0086669	snoRNA:Psi18S-841a	17214561	17214704	CR33765	snoRNA:Psi18S-841a	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0086024		17215768	17215876	CR34653	snoRNA:Me28S-A2589b	
FBgn0082963	snoRNA:Psi28S-3378	17217203	17217352	CR34549	snoRNA:Psi28S-3378	
FBgn0082970	snoRNA:Psi28S-3316c	17217632	17217764	CR34550	snoRNA:Psi28S-3316c	
FBgn0083018	snoRNA:Psi18S-841b	17217781	17217918	CR34551	snoRNA:Psi18S-841b	
FBgn0082969	snoRNA:Psi28S-3316d	17217939	17218070	CR34552	snoRNA:Psi28S-3316d	
FBgn0083017	snoRNA:Psi18S-841c	17218088	17218220	CR34553	snoRNA:Psi18S-841c	
FBgn0082968	snoRNA:Psi28S-3316e	17218241	17218372	CR34554	snoRNA:Psi28S-3316e	
FBgn0083016	snoRNA:Psi18S-841d	17218411	17218548	CR34555	snoRNA:Psi18S-841d	
FBgn0086025		17219078	17219184	CR34654	snoRNA:Me28S-A2589c	
FBgn0083046	snoRNA:Psi18S-1389a	17221439	17221597	CR34556	snoRNA:Psi18S-1389a	
FBgn0083045	snoRNA:Psi18S-1389b	17221650	17221804	CR34557	snoRNA:Psi18S-1389b	
FBgn0086026		17222888	17222961	CR34655	snoRNA:Me28S-G2596	
FBgn0065102	small non-messenger RNA 184	17223015	17223049	CR33916	snmRNA:184	
FBgn0034622	CG15666	17229490	17233113	CG15666	CG15666	
FBgn0034623	CG9822	17234380	17235356	CG9822	CG9822	
FBgn0034624	CG17974	17235692	17238349	CG17974	CG17974	57E.1
FBgn0000395	crossveinless 2	17242520	17266967	CG15671	cv-2	57E.2
FBgn0034626	CG10795	17270732	17271846	CG10795	CG10795	
FBgn0034627	EfSec	17272060	17273720	CG9841	EfSec	
FBgn0034628	acyl-Coenzyme A oxidase at 57D proximal	17273859	17276788	CG9707	Acox57D-p	
FBgn0034629	acyl-Coenzyme A oxidase at 57D distal	17276901	17279944	CG9709	Acox57D-d	
FBgn0010415	Syndecan	17281122	17368680	CG10497	Sdc	57E.3 57E.4 57E.5 57E.6 57E.7 57E.8 57E.9 57E.10 57E.11 57E.12
FBgn0026369	Smad anchor for receptor activation	17369424	17374618	CG15667	Sara	
FBgn0010470	Fkbp13	17374670	17385431	CG9847	Fkbp13	
FBgn0034631	CG10496	17386405	17389475	CG10496	CG10496	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0043070	Misexpression suppressor of KSR 2	17389591	17404747	CG15669	MESK2	
FBgn0034634	CG10494	17403601	17406611	CG10494	CG10494	
FBgn0050289	CG30289	17407960	17409338	CG30289	CG30289	
FBgn0003731	Epidermal growth factor receptor	17410510	17446932	CG10079	Egfr	57E.13
FBgn0050286	CG30286	17423221	17424300	CG30286	CG30286	
FBgn0050287	CG30287	17424399	17425520	CG30287	CG30287	
FBgn0069056	CG33226	17425755	17426798	CG33226	CG33226	
FBgn0260477	CG30283	17427089	17428078	CG30283	CG30283	
FBgn0034636	CG10440	17447751	17458957	CG10440	CG10440	57E.14
FBgn0050222	CG30222	17470997	17472873	CG30222	CG30222	
FBgn0053225	CG33225	17477891	17479089	CG33225	CG33225	57F.1 57F.2 57F.3
FBgn0034638	CG10433	17493181	17494458	CG10433	CG10433	
FBgn0034639	CG15673	17495898	17500594	CG15673	CG15673	
FBgn0042180	CG18870	17502268	17504398	CG18870	CG18870	
FBgn0000326	cricklet	17504566	17506564	CG9858	clt	
FBgn0016641	Protein tyrosine phosphatase-ERK/Enhancer of Ras1	17506889	17511947	CG9856	PTP-ER	
FBgn0034641	CG10080	17512540	17518766	CG10080	CG10080	
FBgn0260720		17518926	17521545	CR42547	CR42547	
FBgn0034642	CG15674	17519791	17521675	CG15674	CG15674	
FBgn0034643	CG10321	17522173	17525906	CG10321	CG10321	
FBgn0034644	CG10082	17533683	17544944	CG10082	CG10082	
FBgn0050284	CG30284	17538330	17539504	CG30284	CG30284	
FBgn0034645	CG10320	17545580	17546282	CG10320	CG10320	
FBgn0034646	Rae1	17546261	17547508	CG9862	Rae1	
FBgn0034647	poor Imd response upon knock-in	17548472	17549749	CG15678	pirk	
FBgn0259708		17551938	17553613	CG42362	CG42362	
FBgn0259709		17551938	17553613	CG42363	CG42363	
FBgn0259710		17553456	17554084	CG42364	CG42364	
FBgn0259711		17554252	17555099	CG42365	CG42365	
FBgn0034649	CG9865	17555046	17557560	CG9865	CG9865	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0259725	CG42379	17555046	17557560	CG42379	CG42379	
FBgn0259726	CG42380	17555046	17557560	CG42380	CG42380	
FBgn0259727	CG42381	17555046	17557560	CG42381	CG42381	
FBgn0034650	NC2alpha	17557673	17559098	CG10318	NC2alpha	
FBgn0034651	CG15676	17559205	17559850	CG15676	CG15676	
FBgn0050263	CG30263	17563826	17572505	CG30263	CG30263	
FBgn0034654	CG10306	17572775	17573709	CG10306	CG10306	
FBgn0003687	TATA binding protein	17573705	17575145	CG9874	Tbp	
FBgn0050285	CG30285	17575408	17576026	CG30285	CG30285	
FBgn0034655	CG10307	17576417	17577758	CG10307	CG10307	
FBgn0027360	Tim10	17577720	17578643	CG9878	Tim10	
FBgn0260223	CG42497	17577720	17578643	CG42497	CG42497	
FBgn0050290	CG30290	17579095	17580269	CG30290	CG30290	
FBgn0260222	CG42496	17579095	17580269	CG42496	CG42496	
FBgn0034656	CG17922	17580191	17584657	CG17922	CG17922	
FBgn0010228	HMG protein Z	17585056	17591993	CG17921	HmgZ	
FBgn0050403	CG30403	17596955	17599114	CG30403	CG30403	
FBgn0004362	High mobility group protein D	17600836	17604230	CG17950	HmgD	
FBgn0050404	Transport and Golgi organization	17604285	17607466	CG30404	Tango11	
FBgn0050398	CG30398	17605338	17606337	CG30398	CG30398	
FBgn0034657	Lamin B receptor	17608002	17611564	CG17952	LBR	58A.1
FBgn0034658	Grx-1	17644932	17645483	CG7975	Grx-1	58A.2
FBgn0050395	CG30395	17653030	17657701	CG30395	CG30395	
FBgn0034659	CG4021	17660852	17662381	CG4021	CG4021	58A.3
FBgn0034660	lysyl oxidase-like 2	17677121	17679160	CG4402	lox2	
FBgn0085233		17684045	17684357	CG34204	CG34204	
FBgn0042098	CG18735	17685577	17686671	CG18735	CG18735	
FBgn0034661	CG4386	17686832	17688405	CG4386	CG4386	
FBgn0046297	CG9284	17690929	17691846	CG9284	CG9284	
FBgn0034662	CG13492	17708349	17717715	CG13492	CG13492	

Flybase number	Associated Function	End Nucleotide	End Nucleotide	Gene Name	Alternative Gene Name	Su(Hw) Binding Sites
FBgn0054040	CG34040	17720096	17721012	CG34040	CG34040	
FBgn0034663	CG4363	17721725	17722534	CG4363	CG4363	
FBgn0034664	CG4377	17722951	17723835	CG4377	CG4377	
FBgn0034665	CG4372	17725110	17726303	CG4372	CG4372	58A.4
FBgn0050407	transfer RNA:CR30407	17730602	17730673	CR30407	tRNA:CR30407	
FBgn0050406	transfer RNA:CR30406	17735912	17735983	CR30406	tRNA:CR30406	
FBgn0034666	CG9294	17748356	17749469	CG9294	CG9294	
FBgn0034667	cookie monster	17752372	17754364	CG13493	comr	
FBgn0085397	Fish-lips	17759537	17833769	CG34368	Fili	58A.5
FBgn0025573	Protein phosphatase N at 58A	17768857	17770038	CG3245	PpN58A	58A.6
FBgn0034670	CG13488	17839196	17840068	CG13488	CG13488	
FBgn0034671	CG13494	17841743	17842350	CG13494	CG13494	58A.7
FBgn0085398		17851678	17858169	CG34369	CG34369	57A.8
FBgn0054029	CG34029	17858333	17859215	CG34029	CG34029	
FBgn0034674	CG9304	17860631	17863710	CG9304	CG9304	
FBgn0041237	Gustatory receptor 58c	17864167	17865462	CG13491	Gr58c	
FBgn0041238	Gustatory receptor 58b	17865720	17867190	CG13495	Gr58b	
FBgn0041239	Gustatory receptor 58a	17867365	17868604	CG30396	Gr58a	
FBgn0050401	CG30401	17873267	17874149	CG30401	CG30401	
FBgn0085399		17874573	17927063	CG34370	CG34370	
FBgn0034681	CG9308	17922349	17923173	CG9308	CG9308	
FBgn0005778	Protein phosphatase D5	17929172	17930387	CG10138	PpD5	
FBgn0034683	CG13500	17943972	17945186	CG13500	CG13500	
FBgn0034684	CG13501	17946408	17948215	CG13501	CG13501	
FBgn0259142	CG42257	17948458	17957262	CG42257	CG42257	
FBgn0034687	CG11475	17957228	17958791	CG11475	CG11475	
FBgn0034688	CG11474	17958918	17960948	CG11474	CG11474	
FBgn0034689	CG2921	17961042	17962842	CG2921	CG2921	
FBgn0067102	GlcT-1	17963547	17965177	CG6437	GlcT-1	

FBgn0034691	synaptojanin	17967676	17973272	CG6562	synj
FBgn0034692	CG13502	17973174	17976084	CG13502	CG13502
FBgn0034693	CG11073	17978994	17987231	CG11073	CG11073
FBgn0034694	CG6613	17990693	17993294	CG6613	CG6613
FBgn0243516	Verprolin 1	17994243	18010053	CG13503	Vrp1

Table A.3 Oligonucleotide sequences used in used in *in situ* hybridization and Real-Time PCR after chromatin Immunoprecipitation

***In Situ* Hybridization Probes**

Name	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product Size(bp)
56F.1	CATTGATAGTAGAGCGTCGA	GCAGATGGAGATTTCGTTTG	1020
56F.4	GCTGTGAAATATGCCGTCA	CAATTGCGGAGGAGATGAC	493
56F.7	CACGCCGACACACCAGCT	GCTCTGTGTGCATATGTGTG	476
56F.8	TCCTCACTAATCCCTCTCAC	GAGTGGGGAGGAAGGGCA	497
57A.3	TTACTCCGCGGATCAGGCATGTGGGATCTGAATC	ATATCACGCGTGACGCAACCGAAACGTAGCTACA	723
57A.4	GATCACCGCGGACGTGTTTGTGTTTCGGAGAGTGC	ATATCACGCGTGACTCGGCACTCACCATTCTC	714
57A.11	CAGCAGCCTGCGTTTCCT	CCAAAGGCTGCGTCATGTT	1116
57B.7	GACCCAAATCGCATGCAAAT	TCCTGGCAAACAAAGTCCTT	1147
57B.12	GATCACCGCGGACGTATGTGTAGTGGGTGGCTGG	ATATCACGCGTGACATGTTGATAACACACCCCGTG	548
57B.13	TACACGCGTTCTAGTTCGCCACATAGCCATCTCAC	AGTCCGCGGCATGAACTTTGAATGGGGGGGTTCTG	1011
57B.19	TAGCCACGCCGTAGCTTCCGGACTGATCGCCAAATG	TAGCTCCGCGGACGAGAGAGCCGAGTTAAAAGCC	519
57B.21	ACGACACGCGTCATAAAGCGAACATTGTCAGCCC	GATCACCGCGGTGCAAGCGGAAAGTGTTGGAAGC	521
57B.23	TAGCCACGCGTAGCAACGTTATTAGCCGACTGG	TAGCTCCGCGGACGTTGTTCGCTTAATCCCCGTG	550
57B.26	CGCCTATCACAAATGAATAAG	CAGTAACTTGCAAGTCCATAT	1045
57B.29	GGCAAACTGCGCTCTAATGT	GGAGTCGTGGACCAACTGAA	493
57D.1	CGCCATCAAAGAATATTGCCT	GTCATCAAAAGCGAGCATTGA	612
57D.3	TGTTTGGCCAACAATCACG	CAGCCTGACAGTCACTTAAT	540
57E.1	AGTCTGATGGATGCCAACCA	TAGTTATCAAGTGCCGACAC	536
57E.4	ATATCACGCGTGACAAGCGTCAATTGTTGGCCCAC	TTACTCCGCGGATCAACGAAATCCGCTTGCCAGG	613
57E.5	ATCGTACGCGTAGCAAGGATGGTGAACTCAGCATG	CATAACCGCGGATCTTCAAATCGCTCAAACGAGGG	532
57E.8	GATCACCGCGGACGTTTGCGCCGACTTTTGCAC	ATATCACGCGTGACTTCAGCTGGAGTCAAGTAGG	523
57E.10	ATCGTACGCGTAGCAATGAGAGCGAGCAGCAGG	CATAACCGCGGATCTACAAAATGCTCGGCCATAG	615
57E.11	GATCACCGCGGTGCTCAAGGCACGAATTGGTTG	ACGACACGCGTCATTGTTGCATATTATCCCCAGGC	542
57E.14	ACGACACGCGTCATTTGTCCGATCTTCAGTGACC	CATAACCGCGGATCACAAAGGCATCATATGGTGG	519
58A.1	CAGACGCGTCATGCAATGCCCAGGAACTTGCCC	GTACCGCGGACTGCAAGCCTTTGAGCCAACTTACC	1030
58A.2	TACACGCGTAGCTGATGAAGCCATCAAGTTGTGC	ATCCCGCGGATGCATCGAACGCAAAATGTCGTCG	809
58A.3	AGTGCGGGTAAAATGTGACC	GCGTTGGGAGAATTTTACGA	511
58A.5	GTCTGTGGAAAATGCATATG	GCCATTAAATTGTCCTCTCA	511
58A.8	GTACGATCCCTTACAAAAGTC	TGTTTGTGTGTTCGCAGATG	512

Real-Time PCR

Name	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product Size(bp)
Rp49	TGTCCTTCCAGCTTCAAGATGACCATC	CTTGGGCTTGCGCCATTTGTG	194
Gypsy	GGTTTCTCTAAAAAGTATGCAGC	CTGGCCACGTAATAAGTGTGC	120
56F.7	TCAGCGAAAACGACAGC	CGCCACTCAGCTGGGTTC	136
57B.12	GGGTGGGAGAGCAGTAAGC	CACATAATTCAAAGTATGCACT	139
57B.29	CCACGATGTTGTCCATGTTC	GCGGATCCCAAACAGCAGG	122
57D.1	GAACTGAATGCTCTAAGCCTG	GTGCATCAGCCTGAAAGTAGG	111
57D.3	AGCTCCTGCCCGGAAATCC	GGAAGTTCTCAATTTGTGGGC	126
57E.1	CATTCTGACACACGCCATCG	GGTCCATCTAGCATGCTCCG	114
57E.4	GAATTCAATCAGCGAACTCC	GTTCCACTTCGTTGCCATGC	127
57E.5	CTGCGTATACTTGTTCTGGC	GCTAAAGTGCGCAACGAAGC	117
57E.8	CACCGGCACACGCAAATTGC	GCAACTTTTCTGACGGAAGC	126
57E.10	GTTGAGATACCAGAAGCCAAC	GGCTCCTCCACTCAACTTG	119
57E.11	CAATGTTAATACCGTCGGGAG	CAGCAATGCATCGCTCTTCA	108
58A.1	CTAGCTGCCCATAGAGCCG	CATCCTCGCCTGGCATGCA	138
58A.8	CATTGGCAGCGGGAAAGTGC	GTTGTCATGAGCAGCCGAC	134

Table A.4 Oligonucleotide sequences used in used in *in situ* hybridization, and real-time PCR after chromatin Immunoprecipitation, and real-time RT-PCR

Name	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product Size(kb)
RooA	TGGGCCAACTTCCTTTGCAT	GCAATATATCTAGACCCGCT	1032
Jockey	TTCAGACTGACACCCGGAC	ACCCTATAACTATGCACCGG	995
Doc3	ATCTTCCCTTTTACTTAAGAGACTAAC	TGGTAATTGGCAAGAAGACTGCTTACA	2535
Nomad	AGTGATACGAAAAAATTGTGG	AGGGTCTCTTCCATTTTGA	1035
X-element	ACCTGACAAAGTGGATTAAG	AGCGGTGTGTGAATATCG	985
Real Time PCR			
Rp49	TGTCCTTCCAGCTTCAAGATGACCATC	CTTGGGCTTGCGCCATTTGTG	194
Gypsy (ChIP)	GGTTTCTCTAAAAAGTATGCAGC	CTGGCCACGTAATAAGTGTGC	120
Gypsy (RT PCR)	TGGAAGCACCGCAAATCAAG	TCCAGGCCACATACTCGTC	129
Jockey	GCAGCACGGTACTCCTGAG	CAGGGTGCCAGACTCTGTC	128
Doc3	CTTCATGACCTTCATGCAAG	GCCATTAGCGTTCCAGGTA	129
Nomad	CAACGCCTCTCCAGTGTAC	GAGAAGGGTTTACGGACTGT	143
X-element	CCTTCGGCTACAGAACCTAG	GCAGCTTGATGACTGGTACTG	134
RooA	CAGAAGATGTTAACTCCAATTT	TCAATGAGTGTAGCTGTTTCG	135
Baggins	GGACTGTGTACCGATCGTG	GTGTTCAGCCAGTGCAGTG	121
Beagle	CTGACCATCAGCCTTTGAC	CAGAGCGTCGGCTACAGTA	140
Stalker	GTAGCAGACGCACTCTCAC	CCTAGGCAATAGTTCCTTG	132
Gtwin	ATGAAGTCACTCGGCAACCT	ACGCTTGGTAAAAGTATGCAATTG	184
Tabor	GGACCGACAACAAAGAAACATG	GAGAACTTTCGATACCTGAG	123
412	CCGTGTGATGGAATAATCGG	GGACAACTTGGGATCTTGCT	181
Idefix	GTACGGTACTGATCAACTG	GAATACTACTTTCACGTAGATTC	120
Copia	CCCTATTTGAAGCCGTGAGA	GACATGAGGGGTTGTTTGCT	135
I-Element	GCTCTTTCACCTCAACCATC	GCTAGCCAATGTAGTCTCGT	140
Transgene Constructs			
Jockey	AACAGCTAGCGAAAGCTTGGAAATCGGCGA	TGTAGGCCGGCCAACTGAGTATAGGGTTGGGC	491

VITA

Heather Anne Wallace was born in Chattanooga, Tennessee. She graduated from Notre Dame High School in 1996. She received a Bachelor of Science from The University of Tennessee, Chattanooga in 2002. In August 2004, she joined the Graduate Program in Biochemistry, Cellular, and Molecular Biology. Heather was awarded the Cokkinias Graduate Student Excellence Award in 2010.