Role of the Drosophila BEAF Protein in Chromatin Domain Insulator and Promoter Function

Mukesh Maharjan
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Biochemistry Commons, Molecular Biology Commons, and the Other Biochemistry, Biophysics, and Structural Biology Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/4917

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
ROLE OF THE DROSOPHILA BEAF PROTEIN IN CHROMATIN DOMAIN INSULATOR AND PROMOTER FUNCTION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biological Sciences

by
Mukesh Maharjan
B.S., Purbanchal University, 2010
May 2019
ACKNOWLEDGEMENTS

This is my great opportunity to express my profound gratitude to my advisor, Dr. Craig M. Hart, for giving me the opportunity to learn and work on my Ph.D. project. He has been a real guru (who dispels our darkness or ignorance) in my journey of Ph.D. work. I am sure, his valuable time and thoughtful inputs in our discussions will inspire me also in my future endeavors. I am very much indebted for his understanding and patience even in adverse times. No words would be able to describe my gratitude for him.

I will always remain indebted to Dr. Scott Herke and Ms. Xiao Ying for their timely help and useful advice. I am very thankful to my committee members, Dr. David Donze, Dr. Joomyeong Kim, and Dr. Charles D. Isbell for providing their thoughtful suggestions in meetings and during my project. I am very thankful to Dr. Kim for mentoring my teaching classes and also being a constant support in matters beyond my research. I am very grateful for Dr. DiMario (and students in his lab) for letting me use his facilities and reagents. Also, I am indebted to Dr. William Wischusen, Dr. Jane Reiland, Miss Ann Jolissant, Miss. Brooke Trabona and Dr. Christopher Gregg for helping me with schedules and TA hours. Without Miss Priscilla and Miss Chimene my transition to graduate school, and life in US, would not have been this smooth.

I am very thankful to undergraduate researchers Sunni Lehman and Kyle Patrick for their incessant effort in mapping P-elements in fly genome and various other helps in my project.

To my family, thank you for caring me so much and for constant love. To my past and current lab mates: Satya, Shraddha, Keller, Yuankai and Sunday, thank you for being there and listening to my every idea and thoughts, and for the countless valuable memories. I have learnt a lot from our togetherness. To my friends outside of lab, Baton Rouge peeps, thank you for your support and company all the times. My friends from Nepal who live out of state are also
invaluable to my days of Ph.D. research and breaks. And finally, to my wife and a real better-half, Pratima Maharjan, thank you for your absolute love. Even your being is a blessing for me. I love you and I am excited for our future hereafter.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................... ii

ABSTRACT ............................................................................................................................. v

CHAPTER 1. LITERATURE REVIEW ....................................................................................... 1
  1.1 Genome organization .................................................................................................... 1
  1.2 Large scale chromatin organization and TADs (Topologically associated domains) .... 21
  1.3 Regulatory elements .................................................................................................... 27
  1.4 Boundary Element Associated Factor (BEAF) ........................................................... 34
  1.5 References .................................................................................................................. 36

CHAPTER 2. USING A PHIC31 “DISINTEGRASE” TO MAKE NEW ATTP SITES IN THE DROSOPHILA GENOME AT LOCATIONS SHOWING CHROMOSOMAL POSITION EFFECTS ....................................................................................................................... 53
  2.1 Introduction ................................................................................................................ 53
  2.2 Materials and methods ............................................................................................... 55
  2.3 Results ....................................................................................................................... 58
  2.4 Discussion ................................................................................................................ 64
  2.5 References ................................................................................................................ 66

CHAPTER 3. INSULATOR AND PROMOTER ACTIVITY ANALYSIS OF SCS’ INSULATOR ................................................................................................................................. 68
  3.1 Introduction ................................................................................................................ 68
  3.2 Materials and methods ............................................................................................... 71
  3.3 Results ....................................................................................................................... 73
  3.4 Discussion ................................................................................................................ 79
  3.5 References ................................................................................................................ 81

CHAPTER 4. IDENTIFICATION OF BEAF ASSOCIATED PROTEIN .................................... 86
  4.1 Introduction ................................................................................................................ 86
  4.2 Materials and methods ............................................................................................... 88
  4.3 Result and discussion ............................................................................................... 91
  4.4 Conclusion ................................................................................................................. 97
  4.5 References ................................................................................................................ 98

CHAPTER 5. CONCLUSION AND FUTURE DIRECTIONS .................................................. 101
  5.1 Conclusion ............................................................................................................... 101
  5.2 Potential future direction ........................................................................................ 104
  5.3 References ................................................................................................................ 104

APPENDIX. COMPLETE LIST OF SIGNIFICANT PROTEINS FROM CO-IP WITH BEAF .......................................................................................................................... 106

VITA ..................................................................................................................................... 109
ABSTRACT

Proper folding of eukaryotic genomes is required to allow correct interactions between different parts of chromosomes. Precise and timely interactions among different parts of a chromosome allow proper functioning inside a nucleus, including gene regulation, DNA replication and DNA repair. Eukaryotic regulatory elements that facilitate folding and interactions include enhancers, promoters and insulator elements. Insulator elements and their binding proteins play an important role in regulating correct chromatin structure and function. The *Drosophila melanogaster* special chromatin structure (scs’) is one such insulator. The Boundary Element Associated Factor (BEAF) binds to scs’. BEAF is a 32 kDa protein that has two isoforms, 32A and 32B. Genomic studies have indicated that BEAF binds from 1800 to 3000 sites in the *Drosophila* genome, usually near transcriptions start sites (TSSs) mainly of housekeeping and highly active genes. In this study, we performed a detailed analysis of scs’ to more precisely understand the role of BEAF in insulator and promoter function. We dissected the scs’ insulator to find minimal sequences required for insulator and promoter functions. We found these two functions overlap by 110 base pairs (bp) but can be separated. BEAF is necessary for both, but insulator function requires 50 bp additional downstream DNA sequences while promoter activity requires 50 bp additional upstream sequences. Attempts to identify binding proteins that might work with BEAF have so far been unsuccessful. We also took another approach to getting at BEAF function. Proteins that physically associate with BEAF were identified by co-immunoprecipitation from nuclear protein extracts followed by proteomic mass-spectrometry. These results suggest that in addition to classical insulator function, BEAF might play a more direct role in gene expression. Notably, chromatin remodeling proteins, histone chaperones and transcription factors were identified. This supports the idea that BEAF might
play a role in keeping promoters active by helping to establish or maintain nucleosome depleted regions around TSSs.
CHAPTER 1. LITERATURE REVIEW

1.1 Genome organization

Each human cell contains 2 m of genomic DNA in its nucleus. A fascination question in biology is how is genomic DNA organized into nuclei of a few micro meter (um) size? W. Flemming, around 1880, discovered and named a nuclear substance that was clearly visible on staining using primitive light microscopes as ‘chromatin’, which is believed to be the basic structure of genomic DNA organization (Olins and Olins 2003). Deoxyribonucleic acid (DNA) due to its phosphate backbone is negatively charged and produces electrostatic repulsion between adjacent DNA regions. So, chemically it is difficult for DNA to fold only by itself (Bloomfield 1996). An octamer of histone proteins consists of the proteins H2A, H2B, H3 and H4, helps

![Diagram of DNA organization](http://www.spring8.or.jp/en/news_publications/press_release/2012/120218/)

**Figure 1.1** A) Thin DNA strands with a diameter of 2 nm (top column) wrap around histones to form nucleosome fibers with a diameter of approximately 11 nm (second column). The nucleosome fibers had long been considered to be regularly folded to form 30-nm-diameter chromatin fibers (third column). B) The conventionally proposed model explains that chromatin fibers form a regular helical hierarchical structure (building-block structure), i.e., they are helically wrapped to form 100-nm-diameter fibers, 200- to 250-nm-diameter fibers, then 500- to 750-nm-diameter fibers. Figure adapted from: [http://www.spring8.or.jp/en/news_publications/press_release/2012/120218/](http://www.spring8.or.jp/en/news_publications/press_release/2012/120218/)
DNA to be wrapped around and balances the net charge. This structure is called the Nucleosome (Figure 1.1, 1.2)(Kornberg and Lorch 1999).

**Nucleosome**

In the nucleosome core particle, 147 bp of DNA wraps 1.7 left-handed super helical turns around the histone octamer. Then ‘linker DNA’ connects two nucleosomes. This is why originally the whole structure of chromatin as visualized in electron micrographs was described as ‘beads on a string’ (Olins and Olins 2003). The core histones have tails with positively charged lysine and arginine residues but only neutralize 60% of the negative charges of DNA; consequently, for further folding, the remaining 40% of the DNA charge has to be neutralized by other factors, such as linker histone H1 or cations (Strick *et al.* 2001).

More than 30 years ago, Finch and Klug first proposed that the nucleosome helps DNA to be folded into 30-nm chromatin fibers (Figure 1.2) (Finch and Klug 1976). They observed isolated chromatin as a ‘solenoid’ which looked like fibers with a diameter of 30 nm under transmission electron microscopy (TEM). Nucleosomes are located next to each other in the fiber, folding into a simple one-start helix (Figures 1.2a and c). Subsequently, a different model of the ‘two start helix’ was proposed on the basis of microscopic observations nucleosomes (Figure 1.2b and d) (Woodcock *et al.* 1984). That model was also supported by other studies and shows that nucleosomes are arranged in a zigzag manner, where straight linker DNA connects two opposing nucleosome cores, creating the opposing rows of nucleosomes that form so called “two-start” helix. In zigzag model, alternate nucleosomes (for example, N1 and N3) become interacting partners (Figure 1.2b and d) (Bassett *et al.* 2009; Dorigo *et al.* 2004; Woodcock *et al.* 1984).
Shortly after, however, Rhodes and co-workers, using cryo-electron microscopy (cryo-EM) again proposed that the 30-nm chromatin fiber is an interdigitated solenoid (Figure 1.2a and c) (Robinson and Rhodes 2006). In classical solenoid model a nucleosome interacts with its first neighbor. However, the interdigitated solenoid model shows that it contacts with fifth and sixth neighbors along the DNA path (Figure 1.2a and c) (Robinson et al. 2006). Moreover, they updated and supported the one-start helix model for the 30-nm fiber, in 2009 (Kruithof et al. 2009). Their work also suggested that the solenoid or zigzag mode of compactions is, in fact, a function of the length of the linker DNA (Routh et al. 2008). But, interestingly in 2009, a study

![Figure 1.2. Models of a 30-nm chromatin fiber. one-start helix (solenoid) (a) and two-start helix (zigzag) (b). Positions from the first (N1) to eighth (N8) nucleosome are labelled. (c) In the one-start helix the 30-nm chromatin fiber is an interdigitated solenoid, in which a nucleosome in the fiber interacts with its fifth and sixth neighbors. Alternative helical gyres cores are colored blue and orange. (d) In the two-start model, nucleosomes are essentially arranged in a zigzag manner such that a nucleosome in the fiber binds to the second neighbor nucleosome. Alternate nucleosome pairs are colored blue and orange. Figure adapted from Maeshima et al. 2010.]
done by Grigoryev et al. showed that the two-start zigzag and one-start solenoid modes may exist simultaneously in a 30-nm chromatin fiber, suggesting instead that observations made in in vitro experiments might be an isolation artifact due to strictly cationic low-salt environment or chemical cross-linking (e.g., glutaraldehyde fixation) (Grigoryev et al. 2009). More recent studies now question the existence of 30nm fibers in vivo, and chromatin may be more flexible and dynamic than previously thought (Ou et al. 2017). A major problem in studying the higher-order levels of chromatin compaction by electron microscopy is that images of individual chromatin fibers overlap on one another and hinder analysis of individual fibers. Ou et al. developed technique called ChromEMT, which combines electron microscopy tomography (EMT) with a labeling method (ChromEM) that selectively enhances the contrast of DNA. By using this technique, they showed that chromatin forms flexible chains with diameter between 5 to 24 nm. Adding to the story, Maeshima et al. also found that 30nm fiber seems to be conditional to the presence of contaminating ribosomal aggregates and almost no 30nm chromatin structures exist in mitotic chromosomes (Nishino et al. 2012). Therefore, the structural details of the 30-nm chromatin fiber are still controversial.

Factors affecting regulation of chromatin accessibility through modulation of nucleosomes

Histone octamers in the nucleosome core are inseparably related to gene expression and regulation in all eukaryotes. The presence of nucleosomes in the promoter area signifies the gene is inactive while genes being actively transcribed have nucleosomes displaced from the promoter area and the histone proteins of nearby nucleosomes are modified to aid transcription (Henikoff 2008). Post-translational modification of histone N- terminal tails through the addition of functional groups including methylation, acetylation, and phosphorylation can lead to activation or silencing of transcription because of either loosening or
tightening the interaction between histones and DNA, histones in other nucleosome, or other proteins (Wolffe and Hayes 1999). For example, trimethylation of the H3 histone protein at lysine 4 (H3K4me3) or lysine 27 (H3K27me3) are well-studied. H3K4me3 is highly enriched near the 5’ end of actively transcribed genes but the H3K27me3 variant is linked to gene silencing via chromatin condensation and is associated with polycomb group proteins. Transcription factors can also recruit coactivator proteins that acetylate histones and, thus, positively affect the activation of transcription (Strahl and Allis 2000). Histone modifications are taken widely as markers associated with gene activation or repression and also with various genomic features, including promoters, transcribed regions, enhancers and insulators (Figure 1.4). Here we discuss the factors that are responsible for gene expression regulation by changing chromatin accessibility through nucleosome modulation.

**Chromatin remodelers**

Chromatin remodelers are ATP dependent proteins and they function individually or as members of larger multiprotein complexes to maintain proper chromatin structure. It can be divided into four subfamilies: SWI/SNF, CHD, ISWI, and INO80 (Narlikar et al. 2013; Clapier et al. 2017). All four remodeling subfamilies have ATPase domains that drive DNA translocation (shown as Tr in Figure 1.3b) in common and certain DNA recognition domain flanks ATPase domain. Figure 1.3 summarizes properties and domains of all four subfamilies of chromatin remodelers.

**ISWI (Imitation switch) subfamily**

The ATPase domain of ISWI subfamily remodelers, contain two RecA-like lobes (lobe 1 and lobe 2 in Figure 1.3), which are separated by a small insertion sequence. They also have a carboxy-terminal HAND–SANT–SLIDE (HSS) domain that binds the unmodified histone H3.
tail and the linker DNA flanking the nucleosome. Most ISWI subfamily complexes assemble and regularly space nucleosomes to limit chromatin accessibility and gene expression. Moreover, a subset of ISWI subfamily remodelers have accessory subunits that confer access and that promote
Figure 1.3. Consequences and domains of different classes of chromatin remodelers. a) Functional classification of remodelers. The ATPase–translocase subunit of all remodelers is depicted in pink; additional subunits are depicted in green (imitation switch (ISWI) and chromodomain helicase DNA-binding (CHD)), brown (switch/sucrose non-fermentable (SWI/SNF)) and blue (INO80). Nucleosome assembly: particularly ISWI and CHD subfamily remodelers participate in the random deposition of histones, the maturation of nucleosomes and their spacing. Chromatin access: primarily, SWI/SNF subfamily remodelers alter chromatin by repositioning nucleosomes, ejecting octamers or evicting histone dimers. Nucleosome editing: remodelers of the INO80 subfamily (INO80C or Swr1 complex (SWR1C)) change nucleosome composition by exchanging canonical and variant histones, for example, and installing H2A.Z variants (yellow). We note that this functional classification is a simplification, as INO80C, the ISWI remodeler nucleosome remodeling factor (NURF) and certain CHD remodelers can promote chromatin access. b | Domain organization of remodeler subfamilies. The ATPase–translocase domain (Tr) of all the remodelers is sufficient to carry out DNA translocation. It is comprised of two RecA-like
transcription like NURF (nucleosome remodeling factor) complex (Xiao et al. 2001). We can find different versions of ISWI remodelers in eukaryotes. For example, yeast have ISWI, *Drosophila* have NURF, ACF, CHRAC and RSF and mammals have Snf2H and Snf2L (Längst and Manelyte 2015).

**CHD (Chromodomain-helicase-DNA binding) subfamily**

In addition to ATPase domain, CHD subfamily remodelers also contain two tandemly arranged chromodomains at N-terminus and DNA binding domain (DBD) at the C-terminus (Figure 1.3). Unlike in yeast, where CHD is a monomeric protein and can only do chromatin assembly, metazoans can have complexes of protein subunits-like NuRD (nucleosome remodeling deacetylase), and can have all three functions of chromatin assembly, exposing promoters and editing (incorporating histone H3.3). NuRD complex can be recruited by repressors to bind to chromatin and repress gene expression through deacetylase function (Murawska and Brehm 2011).

**SWI/SNF (switch/sucrose nonfermenting) subfamily**

In addition to the ATPase domain, SWI/SNF remodelers contain an N-terminal helicase/SANT-associated (HSA) domain that binds actin and/or actin-related proteins, an adjacent post-HS domain, AT-hooks and a C-terminal bromodomain, which binds to acetylated lysines of histones, (Mohrmann and Verrijzer 2005) (Figure 1.3). SWI/SNF subfamily remodelers typically facilitate access by sliding and evicting nucleosomes, however they are not involved in chromatin assembly. They play role in both gene activation as well as repression (Kasten et al. 2011; Längst and Manelyte 2015). We can find different versions of SWI/SNF remodelers in
eukaryotes. For example, yeast have RSC (Remodels the Structure Chromatin), Drosophila has Brahma containing remodelers, BAP/PBAP and mammals have BAF/PBAF (Becker and Workman 2013). In addition, tissue-specific BAF complexes have also been reported. They interact with a variety of transcription factors in different cell types, allowing the complexes to take on context dependent functions arising from their different interaction partners (Ho and Crabtree 2010).

**INO80 (Inositol requiring 80) subfamily**

There is a long insertion between the two RecA like lobes (ATPase domain) in INO80 remodelers. In yeast it is much shorter (~250 amino acid) while in mammals it is longer (>1000 amino acids). INO80 remodelers have HSA (helicase-SANT associated domain) domain, similar to SWI/SNF modelers, at its N-terminus which binds actin and its related proteins (ARPs). INO80 have subunits like Rvb1, Rvb2 (RuVB like proteins) that are involved in DNA repair and recombination functions as well. That suggests INO80 subfamily remodelers have unique editing functions in addition to chromatin access and nucleosome spacing functions; they replace canonical H2A-H2B diners with H2A.Z histone variant-containing H2A.Z-H2B dimers, while in vertebrates INO80 can also replace H3.1 with variant H3.3 (Pradhan et al. 2016).

**Histone chaperone**

For proper gene expression and regulation, nucleosome provides many forms of flexibilities. It can go through various post-translational modifications, transportation and incorporate histone variants. Additionally, histone proteins are assembled and recycled during DNA replication, DNA repair and transcription. In all these processes, and to prevent its unwanted interactions with DNA, protein complex called histone chaperones play an indispensable role (Laskey et al. 1978).
Histone chaperones have been implicated in histone eviction and deposition during transcription. H2A-H2B dimers and the (H3-H4)_2 tetramer occupy distinct positions in the nucleosome. The external dimers are less tightly bound to DNA than is the tetramer, and they are therefore the main candidates for displacement from DNA (Thiriet and Hayes 2005).

**FACT (Facilitates Chromatin Transcription)**

FACT was discovered as a protein that allows passage of the transcribing RNA polymerase through H3-H4 in vitro, yet, under certain conditions it can also bind to free DNA (Orphanides et al. 1998). FACT is a hetero-dimer of structure specific recognition protein-1 (SSRP1) and Suppressor of Ty 16 (SPT16) subunits. FACT binds lateral surface of H3-H4 tetramer through SPT16 and this binding enables one H2A–H2B dimer to be lost. FACT is essential for processes such as transcription, DNA replication, and DNA repair (Orphanides et al. 1998; Shimojima et al. 2003; Nakayama et al. 2007). FACT is involved also in elongation by travelling along with the polymerase. It promotes the displacement of H2A-H2B dimers to facilitate RNA Pol II passage (Orphanides et al. 1998; Belotserkovskaya et al. 2003; Yang et al. 2016). A study done by Nakayama et al. in 2012 purposed that GAGA factor recruits FACT and PBAP to certain chromatin boundaries. Then FACT displaces a H2A-H2B pair from a nucleosome with the displaced H2A-H2B being anchored by FACT. This facilitates PBAP to access H3-H4 tetramer and finally displaces H3-H4 (Nakayama et al. 2012). Recently, it was found that FACT complex is also involved in chaperon function of histones in DNA repair. A recent study done by Piquet et al. 2018, has identified FACT as the responsible histone chaperone for new H2A and H2A.X deposition at DNA repair sites. Interestingly, FACT also helps the deposition of another H2A variant, macroH2A1.2, at sites of replication stress in mammalian cells (Kim et al. 2018).
**HIRA (Histone regulatory homolog A)**

HIRA was originally named TUPLE1 because of a sequence similarity to the yeast corepressor Tup1 and Drosophila E(sp1) (Halford et al. 1993). Later it was found that TUPLE1 was more similar to Hir1 and Hir2/Spt1, which are repressors of histone gene transcription in yeast, and it was renamed as HIRA (histone regulatory homolog A) (Lamour et al. 1995). Furthermore, HIRA was shown to have histone chaperone activity (Magnaghi et al. 1998). The HIR complex is functionally related to the SWI/SNF complex. Hir1 and Hir2 repress (Magnaghi et al. 1998) the promoter activity of the histone genes that are specifically expressed in the S phase. The ATP-dependent nucleosome remodeling complex SWI/SNF is required for the expression of the histone genes and is recruited to this locus through its interaction with Hir1 and Hir2 (Dimova et al. 1999). As discussed in previous sub section, FACT complex, with the aid of PBAP, displaces H2A-H2B pair from nucleosome. H2A-H2B is anchored by FACT and PBAP will displace H3-H4. Further, a study from Nakayama et al. 2012, showed that HIRA-ASF1 replaces H3.1 with the H3.3 variant at the chromatin boundaries like d1, Fab-7 and bxd with the help of PBAP complex. The study also showed that PBAP-induced chromatin alteration of chromatin structure is then restored by HIRA.

**CAF 1 (Chromatin assembly factor 1)**

Chromatin assembly factor 1 (CAF-1) is another example of histone chaperone complex. In eukaryotes, it is a conserved heterotrimeric protein complex that promotes histone H3 and H4 deposition onto newly synthesized DNA during replication or DNA repair. In many species the CAF-1 subunits are designated p150, p60, and p48 (Houlard et al. 2006). CAF-1 is also associated with histone H4 acetylated at N-terminal tail residues, namely lysines 5, 8, or 12. These acetylations, which act as markers of newly synthesized histones, can be recognized by
other chromatin factors, the discrimination between newly synthesized and old histone proteins might be utilized in other nuclear events (Eitoku et al.).

**Other Histone Chaperones**

In addition to FACT, HIRA and CAF1 complexes there are many other histone chaperones found in yeast to humans. Spt6, FACT and Asf1 have been implicated in the deposition of histones behind RNA Pol II (Belotserkovskaya et al. 2003). Moreover, recently it is found that FACT binds H3-H4 and cooperates with other histone chaperones, CAF-1 and Rtt106, to participate in replication-coupled nucleosome assembly (Yang et al. 2016). Nucleolin is thought to remove H2A-H2B from assembled nucleosomes in a manner similar to FACT. Histone chaperones such as yeast Nap1 (nucleosome assembly protein-1) and nucleoplasmin are thought to facilitate transcription factor binding by removing an H2A-H2B dimer (Orphanides et al. 1998). In addition, Nap1 and nucleophosmin are also believed to be involved in histone removal during elongation of chromatin templates in vitro (Swaminathan et al. 2005). In yeast, activator-mediated removal of histones H3-H4 from promoter sites seems to be promoted by Asf1 (Korber et al. 2006). This function of Asf1 is further supported by structural studies done by Natsume et al. 2007 who found evidence that tetramer stability is compromised in the presence of ASF1 in vitro.

**Histone variants**

The four core canonical histone proteins H3, H4, H2A and H2B interact in an ordered manner during nucleosome assembly, giving rise to the modular nature of the nucleosome. However, there are variants for the core histones H3, H2A and H2B, and for the linker histone H1. These variants differ from the core histones either by changing a few amino acids or by the addition of larger domains. As discussed briefly also in previous section, histone chaperones are
considered most likely candidates responsible for histone variant deposition in chromatin. Here the different histone variants and their associated chaperones are being discussed.

One of the extensively studied histone variants is H3.3. Although histone H3.3 differs with canonical H3.1 by only four amino acids, they differ in their mechanisms of chromatin deposition. Histone H3.1 is assembled into chromatin during DNA replication, whereas histone H3.3 deposition occurs throughout the cell cycle (Ahmad and Henikoff 2002). Furthermore, the chaperone complex for H3.1 contains CAF-1, p150, p60 and p48, whereas HIRA is the main chaperon complex for H3.3 (Tagami et al. 2004). H3.3 is found to be associated with active histone modifications such as H3K4 methylation and a nucleosome containing H3.3 exhibits instability (McKittrick et al. 2004; Jin and Felsenfeld 2007). Cells lacking HIRA is seen to exhibit reduced H3.3 occupancy at the gene bodies of both active and repressed genes but no effect was seen in the localization of H3.3 at telomeres and other regulatory elements. This suggests that HIRA is required for the assembly and exchange of H3.3 at genic regions, whereas similar study showed that Daxx–ATRX is involved in H3.3 deposition at telomeric regions (Goldberg et al. 2010). Additionally, in Drosophila DEK is likely another H3.3 chaperone which maintains heterochromatin integrity with interactions with HP1alpha (Kappes et al. 2011).

Histone variant of canonical H2A, H2A.Z has a protein sequence that is highly conserved across species, and it is expressed alongside canonical H2A in all organisms. Even though H2A.Z has considerable amino acid changes in comparison to canonical H2A, the change in structure is very subtle. However, the stability of H2A.Z-containing nucleosomes decreases; one of the factors for instability is the steric hinderance between their L1 loops at the opposite end of dyad axis. H2A.Z incorporation affects the interface between the H2A.Z–H2B dimer and the H3–H4 tetramer (Suto et al. 2000). H2A.Z is enriched at +1 and −1 nucleosomes near the
nucleosome-free region (NFR) surrounding the TSSs in budding yeast (Zhang et al. 2005). While in mammalian cells, NFRs may be marked by labile nucleosomes containing H3.3 and H2A.Z. A member of SWI/SNF family of chromatin remodeler, Swr1 (Swi2/Snf2 related ATPase) is essential for the incorporation of H2A.Z–H2B into chromatin. Nap1 mediates H2AZ–H2B nuclear import and Chz1 presents the dimer to Swr1. These studies reveal that incorporation of H2A.Z–H2B into nucleosomes needs both histone chaperones and chromatin remodeling complexes (Mizuguchi et al. 2004; Luk et al. 2007).

H2A.X shares significant homology with canonical H2A. In response to DNA damage the C-terminus of H2A.X is phosphorylated at serine 139, and this phosphorylation recruits downstream factors involved in DNA damage signaling and DNA repair. FACT has been shown to mediate exchange of H2A.X–H2B for canonical H2A–H2B and is regulated by H2A.X phosphorylation (Heo et al. 2008).

MacroH2A is H2A variant which contains a large C-terminal tail and has two paralogs, MacroH2A.1 and MacroH2A.2. MacroH2A is enriched at heterochromatin and gene repression. Recently it has been shown that FACT helps in the deposition of macroH2A1.2, at sites of replication stress in mammalian cells (Kim et al. 2018). Histone variants for H2B and H1 are typically associated with condensed or transcriptionally repressed chromatin. An H1 variant is thought to be involved in differentiation processes while an H2B variant is thought to be involved in packaging of chromatin in sperm cells (Kamakaka and Biggins 2005; Terme et al. 2011)
Histone tail residues and modification marks are shown in Figure 1.4. a) The normal distribution of DNA methylation, DNA hydroxymethylation, and histone marks in the enhancer, promoter, and gene body of actively transcribed genes. Actively transcribed genes typically have chromatin modifications within the gene body to facilitate transcription initiation and elongation. b) Common chromatin modifications found in the enhancer, promoter, and gene body of silenced genes. c) Bivalent/poised genes have both activating and silencing chromatin modifications to facilitate rapid changes in gene expression during development. Figure modified from Layman and Zuo 2015 and Rodríguez-Paredes and Esteller 2011.
As described earlier a nucleosome has a protein core of histone octamers. It includes two copies each of histones H2A, H2B, H3, and H4. Further, the core histones fold as heterodimers, H2A/H2B and H3/H4, and two H3/H4 heterodimers will form an octamer (Luger et al. 1997). A superhelical turn of 145–167 bp length of DNA winds around this histone octamer. However, the N-terminal tails of all histone proteins protrude out from the nucleosome as well as the C-terminal tails of the two H2A proteins (Figure 1.4). These 10 histone tails are main sites of post-translational modifications (PTMs), also known as histone marks (Luger et al. 1997). The addition and removal of these PTMs determine the accessibility of chromatin to other proteins including transcription factors as well as RNA polymerase II. The following discussion will focus on the different types of PTMs found in histone tails and their effects on chromatin dynamics (Figure 1.4).

Interestingly, four of the 10 tails are near the DNA entry/exit of the nucleosome, and others protrude from the flat surface of the histone octamer. The former could affect the winding of DNA around a single nucleosome, whereas the later could impact the packing ability of nucleosomes against each other. Between the nucleosomes are regions of linker DNA that are associated with linker histone H1 (Allan et al. 1980).

There are broadly two states of chromatin in eukaryotes, in terms of transcriptional activity. Heterochromatin is a more condensed structure, transcriptionally inactive, and is found around centromeres and telomeres; euchromatin is a less condensed and transcriptionally active (Grewal and Moazed 2003). However, histones tails are modified in both heterochromatin and euchromatin. Additionally, the boundaries between heterochromatin and euchromatin are also largely regulated by histone PTMs (Hathaway et al. 2012). Hence, different PTMs play very important role in chromatin state, nucleosome stability and ultimately gene expression. Here we
will be discussing the various important PTMs and their effects in chromatin regulation and their cellular outcomes.

**Acetylation**

Two independent discoveries in 1996 of histone acetyl-transferase (HAT) and histone deacetylase (HDAC) activities provided the first direct evidence for PTMs and histone modifying enzymes. These findings suggested reversible histone modification is possible through these enzymes and the modifications can act as on/off switches in regulating transcription (Brownell et al. 1996; Taunton et al. 1996). The HAT enzyme (Gcn5p of yeast) utilize acetyl-CoA as the acetyl group donor to catalyze the acetylation of histone lysines (Figure 1.5).

![Acetylation Diagram](image1.png)

**Figure 1.5.** Histone modification and change in nucleosome stability. a) HAT enzymes catalyze an acetyl group to be added to lysine residue resulting less positive charge to the histone core and destabilizing the chromatin. Whereas, HDAC acts opposite and adds positive charge to histones making compact chromatin. B) The interaction between DNA and histone and resulting chromatin compaction in favor of gene transcription catalyzed by HATs and opposite by HDACs. Figure adapted from Chrun et al. 2017 and [http://www.webbooks.com/MoBio/Free/Ch4G.htm](http://www.webbooks.com/MoBio/Free/Ch4G.htm).
In the acetylation of histones, the addition of acetyl groups neutralizes the positive charge of histones which decreases the strength of nucleosome-DNA interactions or histone-linker DNA interactions, and/or histone-histone interactions. This will result in loosening of chromatin and nucleosome packing, which will increases the accessibility of chromatin to other proteins related to transcription, DNA replication, DNA recombination and DNA repair (Bannister et al. 2002; Chrun et al. 2017). In chromatin different sites of acetylation of histones are found and can be related to different cellular activities. H3K27ac associates with active enhancers, and both H3K9ac and H4K16ac associate with actively transcribed genes. Furthermore, acetylation of K4/K7 residues of H2A, acetylation of K5/K11/K16/K12/K15 residues of H2B, acetylation of K4/K14/K18/K23/K36/K56 residues of H3, and finally acetylation of K5/K8/K12/K16/K19 residues of H4 in yeast and mammalian cells are all associated with transcription activation, DNA replication and DNA repair (Zhou et al. 2011; Layman and Zuo 2015).

Methylation

Unlike histone acetylation, which is associated with active chromatin configurations, histone methylation, depending on both the histone and amino acid residue modified, can contribute to either active or repressive chromatin configurations (Figure 1.6). Although histone methylation is not as well understood as acetylation, histones H3 and H4 are common methylation targets that can be methylated on arginine and lysine residues. Lysine may be mono-, di- or tri- methylated, whereas arginines may be either mono-methylated, or symmetrically or asymmetrically di-methylated. Unlike acetylation, it does not affect the overall charge of the histone residue. However, a very important change it brings is that specific histone methylations can serve as binding sites for additional regulatory proteins such as chromatin remodelers (Bannister et al. 2002; Bannister and Kouzarides 2011; Venkatesh and Workman 2015; Layman
and Zuo 2015). Lysine-specific histone methyltransferases (HMTs) are subdivided into SET (Su(var)3–9, Enhancer of Zeste, Trithorax) domain-containing and non-SET domain-containing proteins. The arginine-specific protein arginine methyltransferases (PRMTs) are responsible for methylating arginine residues on the histones. HMTs and PRMTs together have over 60 different family members all of which use S-Adenosyl methionine (SAM) as a cofactor and methyl donor, releasing S-adenosyl-L-homocysteine (Desjarlais and Tummino 2016).

Histone methylation for many years was thought to be a permanent or irreversible histone modification due to the low turnover rate of methylated histones (Byvoet et al. 1972). However,
after the discovery of lysine specific demethylase 1 (LSD1) and later the JmjC-domain-containing lysine demethylase family the concept was completely changed. LSD1 can catalyze the demethylation of H3K4me1/2 and H3K9me1/2, which means that LSD1 can both silence and activate gene transcription. Moreover, LSD1 can also demethylate nonhistone target proteins such as p53, DNMT1, and E2F1. Unlike the LSD demethylases, the JmjC-domain-containing demethylases can also demethylate trimethylated lysines. (Kooistra and Helin 2012; Helin and Dhanak 2013; Layman and Zuo 2015).

**Other modifications**

Another very important and dynamic histone modification is phosphorylation. It plays important roles in the DNA damage response (DDR), transcriptional regulation, and chromatin compaction. Serine, threonine, and tyrosine histone residues are phosphorylated and dephosphorylated by multiple kinases and phosphatases. Additionally, all histones (H1, H2A, H2B, H3, and H4) are phosphorylated at serine, threonine, and tyrosine residues as a regulatory step in mitosis, meiosis, or transcription (Rossetto et al. 2012). All histone kinases transfer a phosphate group from ATP to the hydroxyl group of the target amino-acid side chain and the negative charge added to the histone then influences the chromatin structure and also provides potential protein binding sites (Bannister and Kouzarides 2011).

The primary function of non-histone protein polyubiquitylation, which is the association of multiple 76-amino acid ubiquitin groups through the ε-amino group of lysine, is to mark proteins for degradation by the 26S proteasome. There is no evidence to date that histone proteins are ubiquitylated for this purpose. However, C- and N-terminal lysine monoubiquitylation serves as a histone mark and play a role in transcriptional regulation. Monoubiquitylation on K 13, 15 and 119 of H2A and K 34, 120 and 125 of H2B are the most
prominent ones and have been shown to play important biological roles (Desjarlais and Tummino 2016). H2AK119ub1 is involved in gene silencing, whereas H2BK123ub1 plays an important role in transcriptional initiation and elongation (Bannister and Kouzarides 2011). Sumoylation involves the covalent attachment of small ubiquitin-like molecules to histone lysines via the action of E1, E2 and E3 enzymes. It has been detected on all four core histones and seems to prevent acetylation and ubiquitylation in the same lysine side chain. Consequently, it has mainly been associated with repressive functions (Bannister and Kouzarides 2011). Mono-ADP-ribosyltransferase and poly-ADP-ribose polymerase catalyze the mono- and poly- ADP-ribosylation of histones, respectively. ADP-ribosylation adds negative charge to the histones and hence are linked with relatively relaxed chromatin state. Mono-ADP-ribosylation is also detected in linker histone H1. ADP-ribosylation is significantly increased upon DNA damage (Hassa et al. 2006; Cohen-Armon et al. 2007).

Since the histone PTMs do not change the DNA sequence itself, but changes the environment of chromatin and gene expression, it is known as epigenetic regulator. Understanding epigenetic codes and mechanisms has become one of the major focuses for research in nuclear function including the study of diseases.

1.2 Large scale chromatin organization and TADs (Topologically associated domains)

Although for simplicity, we visualize DNA as a linear structure, in fact, it is highly folded and organized. Since the discovery of electron microscopes in 1930s, the double helix structure described by Watson and Crick in 1953 was first visualized as a fiber structure with nucleosomes as regular beads in the fiber. The initial visualization of the spatial organization of chromosomes by FISH (Fluorescent In-Situ Hybridization) confirmed that individual chromosomes are spatially organized as distinct chromosome territories (CTs), in interphase nuclei and is true for
all kinds of cells (Figure 1.7). Additionally, inactive regions of chromatin are often found in proximity to the nuclear envelope whereas active chromatin generally has a more internal position within the nucleus (Croft et al. 1999).

Furthermore, techniques like FISH gave more ideas about the chromosomal organization (Langer-Safer et al. 1982). More recently, methods like chromosomal conformation capture (3C), which detects interactions between a single pair of genomic loci (Dekker et al. 2002), played a key role in understanding the conformation of chromosome. It also helped to develop

**Figure 1.7.** Large-scale structure of the genome. A) 3D structure of a haploid mouse ES genome with expanded views of the separate chromosome territories. B) The spatial distribution of the A (blue) and B (red) compartments. C) Cross-sections through five superimposed 3D structures colored according to: whether the sequence is in the A or B compartment (left); whether the sequence is part of a constitutive Lamina Associated Domains (cLAD) (yellow) or contains highly expressed genes (blue) (center); and chromosome identity (right). Figure adapted from Stevens et al. 2017.
new methods like 4C, which captures interactions between one locus and all other genomic loci (Zhao et al. 2006), 5C, which detects all interactions within a given region (Dostie and Dekker 2007), ChIA-PET, which detects all the associated interactions mediated by a protein of interest (Fullwood et al. 2009) and Hi-C, which detects all interactions of whole genome (Lieberman-Aiden et al. 2009). Lately, Hi-C technique has brought a more detailed understanding of folding of chromosomes and has led us to map chromosomes and specific loci. It has suggested that chromosomes are in fact organized in chromosomal territories and the proximity of the chromosomes (Lieberman-Aiden et al. 2009) and their parts depends on their state (Ramírez et al. 2017). These studies found that the more active and gene rich chromosomes and regions tend to associate, and they situate farther from gene poor chromosomes and inactive regions of chromosomes. Using Hi-C also revealed two major types of structural domains, termed A and B compartments (Lieberman-Aiden and Berkum 2009). The A compartment is active chromatin (denoted by transcriptional activity, higher chromatin accessibility and H3K36me3 deposition) while the B compartment, more compacted, is inactive chromatin (denoted by low transcriptional activity, association with the nuclear lamina and H3K27me3 deposition) (Lieberman-Aiden and Berkum 2009; Rao et al. 2014). Consistent with previous microscopy studies, a recent Hi-C study conducted on single mammalian cells has provided striking views of the spatial arrangements of A and B compartments (Stevens et al. 2017). In modeling the arrangement of all chromosomes within the nucleus, it was shown that DNA from the A compartment is organized in an inner ring-shaped structure, while DNA from the B compartment preferentially associates with the lamina and the edges of nucleoli (Figure 1.7).
These compartments can be rearranged by histone modifications and completely reshuffled in mitosis (Wijchers et al. 2016). Some recent research also suggest that liquid-liquid phase separation can also result in these kinds of non-membrane bound compartments in cells (Maeshima et al. 2016; Larson et al. 2017; Strom et al. 2017). These studies found the nucleus is a phase separated compartment containing several different immiscible liquid-like sub-compartments. HP1 containing heterochromatin has liquid-like properties and appears to form by

**Figure 1.8.** Hierarchical genome organization. Hi-C heatmaps for different scales: whole genome (a), whole chromosome (b), megabase (c, d) and hundred kilobases (e), and a model of genome folding at these scales (f–h) is shown. Whole-genome contact maps show that chromosomes occupy separate chromosomal territories and rarely interact with each other (a, f). Megabase level heatmaps with clear square formations along the diagonal are indicative of topological domains (c, d, g). Plaid-like pattern corresponding to compartments A and B is also visible (b, c, g). Individual peaks corresponding to chromatin loops are clearly seen on the high-resolution heatmaps (e, h). Figure replicated from Szalaj and Plewczynski 2018.
phase separation, which are dissolved or formed by specific ligands on the basis of nuclear context (Larson et al. 2017).

At finer resolution, in more small scale, interactions among chromatin in Drosophila and mammals have shown the separation of the genome into physical domains ranging from kilobases to megabases, that generally contain a small number of genes, 10 or fewer. These domains are known as Topologically Associating Domains, TADs (Dixon et al. 2012) (Figure 1.8d). TAD boundaries interact more frequently with each other and the region in between the TAD boundaries interact inside this local TAD more preferentially than across the TADs (Dixon et al. 2012; Rao et al. 2014). Interestingly, many factors/proteins are thought to be responsible for maintaining these TADs. In mammals, CTCF and cohesin loop anchors at many TAD boundaries, however, in Drosophila, in addition to dCTCF, CP190, BEAF, M1BP and Pita and other architectural proteins are found to be associated (Ali et al. 2016; Cubeñas-Potts et al. 2017; Rowley et al. 2017; Ramírez et al. 2018). Boundaries containing BEAF-32 were stronger when present together with motifs like motif-6, Pita, or Zipic motif, and weaker with motif-8 (Ramírez et al. 2018). Studies done separately by Beagan and Weintraub and colleagues also found that factors like CTCF are responsible for maintaining the TAD boundaries, inside a TAD other proteins like transcription factor Ying Yang 1 (YY1) might play an indispensable role in activating transcription by making loops to bring enhancers and promoters together (Beagan et al. 2017; Weintraub et al. 2017). Consistent with this, mutation studies of CTCF and YY1 binding sites showed disruption of chromatin loops and domain structures. These studies also showed disrupting CTCF and YY1 sites may lead to novel enhancer-promoter interactions and mis-expression of genes (Lupíañez et al. 2015; Guo et al. 2015) (Figure 1.9). Another interesting finding was from Sarah Rennie and colleagues which revealed that genes with similar temporal
expression tend to stay together. For example, more housekeeping gene promoters are in between TADs and developmental promoters are inside the TADs (Rennie et al. 2018).

Figure 1.9. A hypothetical example of a genome reorganization after a TAD boundary disruption, shown using 3 different perspectives: contact maps, genomic diagram, and a chromatin looping model (top, central and bottom row, respectively). a A sample region with three domains (marked with green bars and labeled I, II, and III) separated by TAD boundary elements (black rectangles) is presented. The domains are further divided into sub-domains (blue bars) separated by sub-TAD boundary elements (gray rectangles). Interactions between genes and enhancers are restricted to domains (E1-G1, E2-G3, E4-G4), but they can bypass the subdomain boundaries (E1-G1). b After the boundary disruption (marked with red arrow), former domains II and III merge together allowing for contacts between previously separated loci, as indicated by increased interaction frequency between the domains observed in the heatmap. Without the insulating barrier, enhancer E4 changes its target from G4 to G3, which disrupts prior interactions. In this example, G4 loses its enhancer while E2 gains a new target gene. Figure replicated from Szalaj and Plewczynski 2018.

In conclusion, studies suggest chromatin loops and interactions play key roles in gene regulation. Even so, many questions are still unanswered in the field of chromatin organization and gene expression. However, with the help of new technologies like Hi-C, single cell techniques, START-seq, STARR-seq, high throughput computational techniques etc. exciting mysteries are being unfolded.
1.3 Regulatory elements

Bacteria as well as multicellular organisms have levels of genome organization and are rigorously regulated for the expression of genes at the right time. Especially in the case of multicellular organisms, the organization and expression of different genes depends on the cell type as well as in developmental stages. Various factors affect gene expression and regulation from transcriptional to post transcriptional and post translational stages. Most regulation of the gene expression is at the transcriptional initiation level. Many gene regulatory elements play roles in this complex yet highly managed process. Enhancers, promoters and insulator elements are major players in this regulation.

**Enhancers and promoters**

Enhancers and promoters are two distinct classes of functional cis elements. Enhancers can be distally positioned and regulate transcription from promoters at different stages of development (temporal) and in different tissue types (spatial) (Kim and Shiekhattar 2015). Enhancers were first described as nucleosome-depleted regions (NDRs) with many short sequence motifs recognized by DNA-binding transcription factors (Small et al. 1992; Spitz and Furlong 2012), which in turn recruit coactivators (Collis et al. 1990), such as p300/CBP, which acetylate TFs and histone H3 Lys27 [H3K27ac]. However, how enhancer sequences ultimately give enhancer activity is still not fully answered even at this stage of the genomics era. With the advancement of genomic studies, more characteristics of enhancers have been recognized providing more mechanistic detail (Calo and Wysocka 2013; Shlyueva et al. 2014).

The promoter commonly is referred to a DNA region that allows accurate transcription initiation of a gene at transcriptional start sites (TSSs) by recruiting RNA polymerase II (Smale and Kadonaga 2003). The core promoter is a segment of DNA around a TSS which has specific DNA motifs or sequences (e.g., the TATA box, initiator, downstream core promoter element,
motif ten element, TCT, TFIIB recognition element [BRE]) that recruits basal transcription machinery, including RNA polymerase II (RNAPII). Although the DNA sequences (Summarized in Table 1) can vary significantly depending upon core promoter region for particular genes, its general role is to drive precise transcription initiation (Smale and Kadonaga 2003). Transcription factors can facilitate the recruitment of the basal transcription machinery onto the core promoter or mediate the recruitment of specific distal enhancers to the core promoter (Akbari et al. 2008).

**Table 1.1. Consensus sequences of some core promoter elements**

<table>
<thead>
<tr>
<th>Motif</th>
<th>Location</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>TATA box</td>
<td>Upstream T at −32 to −28</td>
<td>TATAWR</td>
</tr>
<tr>
<td>BRE&lt;sup&gt;u&lt;/sup&gt;</td>
<td>Upstream of TATA box</td>
<td>SSRCGCC</td>
</tr>
<tr>
<td>BRE&lt;sup&gt;d&lt;/sup&gt;</td>
<td>−23 to −17</td>
<td>RTDKKKK</td>
</tr>
<tr>
<td>Inr</td>
<td>−2 to +4</td>
<td>TCA+1GTY (<em>Drosophila</em>)</td>
</tr>
<tr>
<td></td>
<td>−3 to +3</td>
<td>BBCA+1BW (human)</td>
</tr>
<tr>
<td>TCT</td>
<td>−2 to +6</td>
<td>YYC+1TTTY (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>−1 to +6</td>
<td>YC+1TY (human)</td>
</tr>
<tr>
<td>XCPE1</td>
<td>−8 to +2</td>
<td>DSGYGRAS+1M</td>
</tr>
<tr>
<td>XCPE2</td>
<td>−9 to +2</td>
<td>VCYCRTTRCM+1Y</td>
</tr>
<tr>
<td>MTE</td>
<td>+18 to +22</td>
<td>CGANC</td>
</tr>
<tr>
<td></td>
<td>+27 to +29</td>
<td>CGG</td>
</tr>
<tr>
<td>DPE</td>
<td>+28 to +32</td>
<td>RGWYV</td>
</tr>
<tr>
<td>DCE</td>
<td>Box I: +6 to +11</td>
<td>CTTC</td>
</tr>
<tr>
<td></td>
<td>Box II: +16 to +21</td>
<td>CTGT</td>
</tr>
<tr>
<td></td>
<td>Box III: +30 to +34</td>
<td>AGC</td>
</tr>
<tr>
<td>DTIE</td>
<td>+23 to +31</td>
<td>GSGRDHG</td>
</tr>
</tbody>
</table>

(W) A or T; (R) A or G; (S) G or C; (D) A, G, or T (not C); (K) G or T; (Y) C or T; (B) C, G, or T (not A); (M) A or C; (V) A, C, or G (not T); (N) A, C, G, or T (any base); (H) A, C, or T (not G). (BRE) TFIIB recognition element upstream (u) or downstream (d); (Inr) initiator; (XCPE1) X core promoter element 1; (MTE) motif ten element; (DPE) downstream core promoter element; (DCE) downstream core element; (DTIE) downstream transcription initiation element. Table adapted from (Vo Ngoc et al. 2017).

Although these functional definitions remain mostly correct, there have been findings that challenge these classical concepts of gene regulation. Enhancers can act as promoters in certain cases (Kowalczyk et al. 2012) and vice versa (Dao et al. 2017; Diao et al. 2017). One of the factors determining this role is the directionality of the transcription of the element (Mikhaylichenko et al. 2018). Intergenic enhancers that are transcribed bidirectionally can
generally function as weak bidirectional promoters whereas unidirectional enhancers lack promoter activity. Also, bidirectionally transcribed promoters were found to act as strong enhancers, but unidirectional promoters lacked enhancer functions.

Monomethylation of histone H3 Lys4 (H3K4me1) was found to be enriched at enhancers, whereas, gene promoters exhibit trimethylated histone H3K4 (H3K4me3) (Heintzman et al. 2007; Rada-Iglesias et al. 2011). As a hallmark for identification of enhancers, binding of CBP/p300, presence of enriched H3K27ac, accompanied by high levels of H3K4me1, occupancy by cohesin and low H3K4me3 has been generally used to identify active enhancers (Consortium 2012; Bose et al. 2017; Henriques et al. 2018). Nonetheless, recent work has demonstrated that the presence of H3K4me1 is not a requirement for enhancer function (Rickels et al. 2017; Henriques et al. 2018) and that H3K4me2 and H3K4me3 of local histones are fully compatible with enhancer activity. Hence, the difference between enhancers and promoters is becoming narrower than it was thought. DNA sequence features, RNA Polymerase II (PolII) recruitment, chromatin marks and bidirectional transcription are proving not to be enough to distinguish them (Mikhaylichenko et al. 2018; Rennie et al. 2018; Henriques et al. 2018a; b). Therefore, these findings suggest more work needs to be done to fully understand enhancer and promoter dichotomy or to know if there is a dichotomy.

Nevertheless, to understand mechanisms of gene regulation, we need a clear picture of how enhancers and promoters function. The interaction between enhancers and core-promoters (CP) have been a very fascinating topic for biologists lately. A very interesting finding comes from the Stark Lab in 2015, about the specificity of some transcriptional enhancers with CPs. Previously, before the advent of genomics techniques, it was believed that some enhancers are choosy about which promoters they prefer to interact with, and that DNA sequences associated
near the promoter are responsible for these preferences (Choi and Engel 1988). This hypothesis was supported by data from the Stark Lab using STARR-seq (Self-Transcribing Active Regulatory Region-sequencing), a technique developed by the Stark group previously (Arnold et al. 2013). Zabidi, Arnold and colleagues tested the enhancer–promoter specificity hypothesis on a genome-wide scale, by high-throughput genome-wide screening of enhancer activity using housekeeping (hk) and developmental (d) core promoters (Zabidi et al. 2015). Their study revealed an enrichment of DREF-binding DRE motifs in the housekeeping enhancers, while Trithorax-like (Trl)-binding GAGA motifs were enriched in developmental enhancers (Figure 1.10). BEAF-32B binding site (CGATA) is highly related to DREF binding site (TATCGATA) and BEAF binds near hk TSSs (Jiang et al. 2009a). Although Zabidi et al. did not state anything about BEAF, these facts about BEAF suggest it might also be enriched in their hk enhancers as well.

**Figure 1.10.** Summary of key differences between dCP- and hkCP-preferring enhancers and their target genes as found by Zabidi, Arnold and colleagues. Figure adapted from Lorberbaum and Barolo 2015.
Developmental and housekeeping core promoter enhancer elements has proximity bias. STARR-seq sequences activating the dCP (developmental core promoter) enhancer appear to be gene promoter-distal while sequences activating the hkCP (housekeeping core promoter) enhancers are generally gene promoter-proximal (Arnold et al. 2013). In fact, hkCP enhancers often overlap promoter regions but can function from a distance. Another study also gave an interesting finding in the enhancer realm, based on DNase I hypersensitive sites (DHSs). Unstable DHSs are associated with low level of transcription that is exosome sensitive and are enriched for dCP enhancers. In contrast, large fraction of stable DHSs have high transcription levels and are enriched for hkCP enhancers (Rennie et al. 2018). These findings in Drosophila mirror results from humans (Andersson et al. 2014).

**Insulators and insulator proteins/architectural proteins**

Eukaryotic chromosomes are subdivided into functionally autonomous domains by special elements called chromatin boundaries or insulators (Gaszner and Felsenfeld 2006). Insulators or boundary elements are gene regulatory elements, in addition to enhancers and promoters that control gene activity. One of the classic ways that insulators regulate gene expression is by blocking the interaction between an upstream enhancer and downstream promoter (Kellum and Schedl 1992). Another is to block the spreading of heterochromatic gene silencing, thus acting as domain barriers (Udvardy et al. 1985; Kellum and Schedl 1991). Insulators have been found to be a common feature in eukaryotic genomes ranging from yeast to humans. Several insulator binding proteins, also referred to as architectural proteins, have been described in Drosophila: Suppressor of Hair-wing [Su (Hw)], Drosophila CTCF (dCTCF), boundary element-associated factor (BEAF), Zeste-white 5 (Zw5), GAGA factor (GAF) (Bushey et al. 2009), Ibf1 and Ibf2 (Cuartero et al. 2014), Elba1 and Elba2 (Aoki et al. 2012), Pita and
Early studies of insulators

The gypsy retro-transposon is the most extensively studied and one of the first discovered Drosophila insulator elements. It contains binding sequences for ‘suppressor of Hairy wing’ [su(Hw) (Spana et al. 1988)]. The transposon (or an isolated fragment with su(Hw)-binding sites) was shown to mediate enhancer blocking within the yellow locus (Spana et al. 1988; Adryan et al. 2007). Enhancer-blocking function by su(Hw) was later found to be facilitated by additional factors, like Mod(mdg4) (modifier of mdg4), CP190 (centrosomal protein 190), the ubiquitin ligase dTopors and a putative RNA helicase Rm62 (Gerasimova et al. 1995; Capelson and Corces 2005; Lei and Corces 2006), and found to be working forming chromatin loop-domains. One interesting finding in the study of insulators is the insulator-bypass phenomenon. When two copies of an insulator are placed between an enhancer and promoter of a reporter gene the upstream enhancer can activate the reporter gene. The model proposed for this process is that the two insulators pair with each other forming a loop and brings sequences on the far sides of the two insulators in proximity (Maeda and Karch 2007; Kyrchanova and Georgiev 2014; Kyrchanova et al. 2015). However, insulator bypassing depends upon the relative orientation of these two insulators with respect to each other (Kyrchanova et al. 2011). If the orientation of one of the insulators is reversed, the insulators can still pair, but because of the topology of the loop is different so the enhancer remains far from the promoter, preventing strong activation of the reporter gene (Kyrchanova et al. 2016).

Another well studied insulator system is the scs/scs’ paired elements in Drosophila, which flank the Hsp70 (heat-shock protein 70) locus at cytological position 87A7. The study of
scs/scs’ elements also provides support for the importance of loop domains (Kellum and Schedl 1991, 1992). Proteins Zw5 (Zeste-white 5) and BEAF32 (boundary-element-associated factor of 32 kD) bind to scs and scs’ respectively (Zhao et al. 1995a; Gaszner et al. 1999). The interaction between BEAF32 and Zw5 has been shown to stabilize loop-domain formation at opposite ends of the 87A7 hsp70 locus in vivo (Blanton et al. 2003).

Further examples of insulator function can be found in the bithorax complex (BX-C), which encompasses three homeotic genes, Ultrabithorax (Ubx), abdominal-A (abd-A), and Abdominal-B (Abd-B), and nine distinct cis-regulatory domains, abx/bx, bxd/pbx, iab-2—iab-9, (Sánchez-Herrero et al.; Lewis 1978; Maeda and Karch 2015). The highly conserved insulator factor CTCF is associated with six of the eight boundaries separating the cis-regulatory domains (Holohan et al. 2007). dCTCF is the only one of the known Drosophila insulator factors with a conserved counterpart in vertebrates.

The first vertebrate enhancer-blocking insulator to be identified was HS4, a complex element that combines enhancer-blocking and barrier activity and lies at the 5′ end of the chicken β-globin locus (Chung et al. 1993). The powerful enhancer-blocking activity of this element is associated with a strong binding site for CTCF, which a zinc- finger protein expressed ubiquitously in vertebrates. Like the chicken β-globin insulator, CTCF binding sites are also found in mouse insulators. Imprinting of the Igf2/H19 locus in mouse endodermal cells results in the expression of H19 from the maternal allele and Igf2 from the paternal allele. Transcription of these genes is regulated by an imprinted enhancer located proximal to the H19 gene. An imprinting control region (ICR) is located between the Igf2 and H19 genes (Robinson et al. 2006)(Oki and Kamakaka 2002). The ICR element contains binding sites for the mouse homologue of CTCF, and the DNA-binding activity of CTCF is blocked by methylation of the
ICR. The CpG residues at the ICR are methylated on the paternal allele, which is thought to prevent CTCF from binding, thus inactivating the insulator and allowing the enhancer to activate the \textit{Igf2} gene. On the maternal allele, the CTCF-binding sites are not methylated and the consequent binding of CTCF blocks the enhancer from activating the \textit{Igf2} gene (Oki and Kamakaka 2002).

1.4 \textbf{Boundary Element Associated Factor (BEAF)}

Research in our lab is mainly focused on understanding the role of BEAF insulator protein as a tool to better understand chromatin organization and dynamics and how this influences gene regulation. This study is particularly focused on understanding the role of BEAF by dissecting the scs’ insulator element and its promoter activity. We also are trying to identify proteins that work with BEAF. As mentioned earlier, the scs’ element is located downstream of

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure11.png}
\caption{Two BEAF proteins are encoded by a single gene. A) Schematic of a 5 kb genomic fragment containing the \textit{BEAF} gene. Black: UTR’s; green and red: unique 32A and 32B coding sequences, respectively; white and yellow: shared coding sequences. B) Schematic of the two BEAF proteins, colored as in (A). DNA binding and BEAF interaction domains of BEAF (Hart \textit{et al.} 1997). C) schematic showing scs’ element and M Fragment (dimer M2). The promoter of CG3281 and \textit{Aur} gene is shown. The exact start sites of those genes are not known. Dotted lines highlight the monomer M fragment. L: low affinity BEAF binding site; H: high affinity BEAF binding site.}
\end{figure}
the Hsp70 locus at 87A7. Two clusters of three CGATA motifs in scs’ are bound by BEAF-32B, and these binding sites are essential for insulator activity (Zhao et al. 1995a; Hart et al. 1997; Cuvier et al. 1998). BEAF-32A and -32B are 32 kDa proteins that come from same gene, presumably by alternative promoters. They differ only at their N-terminal ends of around 80 amino acids (Figure 1.11A and B). Study of the scs’ element will provide insight into how a boundary element works. scs’ has two BEAF binding sites, a low affinity binding site and high binding site (Figure 1.11 C). Two divergent promoters are located in scs’ at each end (Glover et al. 1995). However, their significance to scs’ function has never been addressed previously. A small version of scs’, 225 bp M fragment, has been identified that retains full insulator activity as a dimer (Cuvier et al. 1998). Because of this small size it is an excellent model system for studying the BEAF binding insulator element.

Previously, immunostaining of polytene chromosomes showed that there are several hundred BEAF binding sites in the Drosophila genome. Other binding sites that have been tested have insulator activity (Cuvier et al. 1998). This finding was extended by genome wide mapping of BEAF binding sites (Jiang et al. 2009b). One interesting finding from genomic data is that 85% of 1820 BEAF peaks had their centers within 300 bp of a transcription start site (TSS). Similarly, over 85% of the genes with a TSS within 300 bp of the center of a BEAF peak are on a list of housekeeping genes (Lam et al. 2012; Ulianov et al. 2016). Quantitative reverse transcription (qRT)-PCR also showed that the expression levels of most tested BEAF-associated genes decrease in embryos and cultured cells lacking BEAF (Jiang 2009).

Together, these results suggest that BEAF may play a role at promoters, particularly of housekeeping genes. It may help to maintain an architecture favorable for transcription, facilitating the recruitment of RNA polymerase II (RNA pol II). In addition to that, BEAF is also
found to co-localize with other insulator proteins at hundreds of sites on DNA. This suggests that it may form insulator protein complexes, implicated in structural and functional demarcation of the genome (Nègre et al. 2010a; Vogelmann et al. 2014). A study of BEAF function and dynamics reported that it might play a role as a component of the mitotic spindle matrix, supporting the report that BEAF physically interacts with Chromator (Vogelmann et al. 2014; Avva and Hart 2016; Yao et al. 2018). Thus, the diversity of interactions between BEAF and other architectural proteins could explain different functions of BEAF in chromatin barrier, architectural and transcriptional regulation functions.

1.5 References


Avva S. V. S. P., Hart C. M., 2016 Characterization of the Drosophila BEAF-32A and BEAF-


Dostie J., Dekker J., 2007 Mapping networks of physical interactions between genomic elements

Eitoku M., Sato L., Senda T., Horikoshi M., Review Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly.


Helin K., Dhanak D., 2013 Chromatin proteins and modifications as drug targets. Nature 502:


Jiang N., 2009 GENOME-WIDE MAPPING OF BEAF BINDING SITES IN DROSOPHILA LINKS BEAF TO TRANSCRIPTION.


Kellner W. A., Bortle K. Van, Li L., Ramos E., Takenaka N., et al., 2013 Distinct isoforms of
the Drosophila Brd4 homologue are present at enhancers, promoters and insulator sites. Nucleic Acids Res. 41: 9274–9283.


McKittrick E., Gafken P. R., Ahmad K., Henikoff S., 2004 From The Cover: Histone H3.3 is


Nishino Y., Eltsov M., Joti Y., Ito K., Takata H., et al., 2012 Human mitotic chromosomes
consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. EMBO J. 31: 1644–1653.


49


Thorpe H. M., Smith M. C. M., 1998 In vitro site-specific integration of bacteriophage DNA


Wolle D., Cleard F., Aoki T., Deshpande G., Schedl P., et al., 2015 Functional Requirements for Fab-7 Boundary Activity in the Bithorax Complex.


CHAPTER 2. USING A PHIC31 “DISINTEGRASE” TO MAKE NEW ATTP SITES IN THE DROSOPHILA GENOME AT LOCATIONS SHOWING CHROMOSOMAL POSITION EFFECTS

2.1 Introduction

The use of transgenic organisms for purposes such as expressing normal or mutant proteins or studying potential regulatory elements is an important tool in basic and applied research. In *Drosophila*, the use of modified P-element DNA transposons has historically been the most common method for generating transgenic flies (Rubin and Spradling 1982). P-element integration into the genome is fairly random, so each integration event needs to be individually characterized as to which chromosome is affected and, if desired, exactly where the insertion occurred. In addition, different insertion events are subject to different chromosomal position effects (CPE) depending on local regulatory elements and chromatin states. Randomness can be advantageous for generating insertion mutations or enhancer-trapping, but can complicate the analysis of transgenes and regulatory elements. One solution is to use site-specific integration to standardize CPE. The integration mechanism of the bacteriophage phiC31 (Thorpe and Smith 1998) has been adapted for this purpose in *Drosophila*, creating a powerful method for integrating different DNA sequences at the same genetic locus. This requires placing an attP integration site into the fly genome, an attB site into the plasmid containing the DNA to be integrated into the genome, and a source of the phiC31 Integrase (Groth *et al.* 2004; Venken *et al.* 2006; Bischof *et al.* 2007; Markstein *et al.* 2008). Alternatively, cassette exchange can be done using a pair of attP and attB sites (Bateman *et al.* 2006).

While many attP landing platforms exist, they are not always suitable. For instance, we are interested in chromatin domain insulator elements. One assay for insulator activity is their ability to shield a bracketed transgene from CPE, resulting in position-independent expression
(Kellum and Schedl 1991; Cuvier et al. 1998). However, when using random integration mediated by P-element insertion, this requires sampling multiple genomic locations so a statistically significant conclusion can be reached as to whether a test element has insulator activity or not. Ideally, ten or more genomic locations should be sampled for each test element. This assay would be simpler if all test constructs could be tested at the same chromosomal location. This requires that the location is subject to CPE, so there is an effect for the test element to block. We tested 13 attP sites available from the Bloomington *Drosophila* Stock Center (BDSC), and none were suitable for our purpose. They either did not exhibit CPE, or activation of the mini-\(w\) reporter gene was not blocked by an insulator. Therefore we wanted to generate new attP landing platforms that suit our needs. We wanted to place attP sites in genomic locations that show CPE activation of a mini-\(w\) reporter gene that can be blocked by an insulator element.

Here we describe the development of a “Drosophilized” transgene encoding a phiC31 “Disintegrase” (Dint) that can be used to collapse an attR-attL pair to an attP site. This is based on the pioneering work of the M.C. Smith lab (Rowley et al. 2008), and is similar to a report that was published while this work was in progress (Knapp et al. 2015). This tool could be useful for purposes other than the one that we use it for here. We also developed a transposon in which an insulator element can be removed by FLP recombinase to test for CPE, and then the mini-\(w\) gene and insulator at its 3’ end can be “dis-integrated” using Dint to leave an attP site. Since this removes the mini-\(w\) marker gene, the transposon also has a yellow\(^+\) gene to mark the presence of the attP site in the remnant transposon. In addition, the y\(^+\) gene is bracketed by loxP sites so that it can be removed by Cre recombinase if desired. We use this system to generate several attP
sites at locations that show varying degrees of CPE. We also show that CPE is still evident when we reintegrate insulated or uninsulated mini-\(w\) transgenes into 3 of the new \(attP\) platforms.

### 2.2 Materials and Methods

**Plasmids and DNA**

All plasmids used here for testing insulator activity were made from pC4scs, a previously described derivative of pCaSpeR4 (Pirrotta 1988) containing a 990 bp scs insulator sequence inserted downstream of the mini-\(w\) gene (Cuvier et al. 1998). A 50 bp \(attB\) site (CGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCAC) was placed into this plasmid using XhoI and XbaI restriction sites to make pC4-attB-scs. The previously described 215 bp M fragment from scs’ including the high affinity BEAF binding site was inserted into the BamHI site of this plasmid as a monomer (M) or dimer (M2) (Cuvier et al. 1998). The plasmid pRLY, which was used to generate new \(attP\) sites in the \textit{Drosophila} genome, was constructed from pC4scs as follows. A 770 bp fragment containing a 50 bp \(attR\) site (GTAGTGGCCCAACTGGGGTAAACCTTTGGGCTCCCCGGGCGCGTACTCCAC) followed by \textit{FRT} sites bracketing the M2 insulator was placed upstream of mini-\(w\) using XhoI and EcoRI. The PstI site at the 3’ end of scs was used to insert a 50 bp \(attL\) site (CGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCAC) followed by \textit{loxP} sites bracketing a 5 kb intronless \(y^+\) gene. The \(y^+\) gene (Geyer and Corces 1992) was from pC4-yellow (Sigrist and Pirrotta 1997). This allows FLP recombinase-mediated removal of the M2 insulator to test for CPE on mini-\(w\) expression (Golic and Lindquist 1989). The mini-\(w\) gene and scs insulator can be removed by the phiC31 “Disintegrase”, hereafter referred to as Dint, leaving behind an \(attP\) site whose presence is marked by the \(y^+\) gene. If desired, the \(y^+\) gene can be removed by Cre recombinase.
The Dint transgene was made by overlap PCR using the nuclear-targeted *Drosophila* codon-optimized integrase as a template (Bischof et al. 2007). The following primers were used for the overlap PCR: NdeI dPhiC31-1S (AAAACATATGGACACGTATGCGGTCG), dPhiC31-E449KS (CGACGCTTC GGCAAGCTCACTAAGGCGCCAGAGTCCCGGCG), dPhiC31-E449KAS (CGCCGGACTT CTCTGGCGCCTAGTGGCTTGCCGAAGCGTGC) and RI-nls-dPhiC31-1818AS (TTTT GAATTCTTACACCTTGCCTTCTTCTTGGGGGCAGCTACGTTTTCGAGCGTGC). After the overlap PCR reaction, the resulting PCR fragment was cloned into pGEMT-easy (Promega, Wisconsin, USA) and sequenced. To add the *nanos* promoter and UTRs, Dint was excised from pGEMT-easy with NdeI and EcoRI and subcloned into the pHSXnosN vector (Gavis and Lehmann 1992) cut with the same enzymes. A NotI fragment containing the *nanos* promoter, *nanos* 5’UTR, Dint, and *nanos* 3’UTR was then isolated from the resulting plasmid and cloned into a C4 yellow plasmid cut with NotI.

**Germline transformation and transposon hopping**

Thirteen BDSC fly stocks with *attP* sites were tested for CPE. None were suitable (BDSC stock number and location in parentheses): ZH-22A (24481; 2L:22A2); VK37 (28472; 2L:22A3); ZH-51C (24482; 2R:51C1); ZH-51D (24483; 2R:51D9); ZH-58A (24484; 2R:58A3); VK31 (24870; 3L:62E1); VK33 (24871; 3L:65B2); ZH-68E (24485; 3L:68E1); ZH-86Fa (24486; 3R:86E18); ZH-86Fb (24749; 3R:86F8); ZH-96E (24487; 3R:96E10); VK20 (24867; 3R:99F8); ZH-102D (24488; 4:102F4). We designed the strategy described below to obtain suitable stocks.

Injections of pre-blastoderm embryos to generate transgenic flies were done by GenetiVision (Houston, TX) or in the Karch lab using standard techniques. The pRLY P-element plasmid was injected into a y w stock and resulted in one transgenic line, identified by eye color and body
pigmentation. Three \textit{attP} lines generated in this study were used to generate transgenic stocks using phiC31 Integrase (Groth \textit{et al.} 2004; Bischof \textit{et al.} 2007). Two on the X chromosome were combined with \textit{M\{vas-int.B\}ZH-102D} (from BDSC 23649), and one on chromosome 2 was combined with \textit{y\textsuperscript{1} M\{vas-int.Dm\}ZH-2A w\textsuperscript{*}} (from BDSC 24486). The \textit{Integrase} transgenes are marked with \textit{3xP3-RFP}, so their presence can be confirmed by pink eye color in flies more than 4 days old. Stocks were sent to GenetiVision for injections with \textit{attB} plasmids. The \textit{Dint} construct was injected into a \textit{y w} stock and the F1 were screened for inserts mapping to the X chromosome.

 Injections with pRLY resulted in one fly stock with the transposon on chromosome 2. To generate additional insertion sites, the \textit{P\{RLY\}} transposon was placed over a \textit{CyO} balancer chromosome and combined with \textit{ry\textsuperscript{506} Sb\textsuperscript{1} P\{ry\textsuperscript{+}\textit{i7.2} = Delta2-3\}99B} as a marked chromosome 3 transposase source (BDSC 3664). Among the hopped transposons recovered were two hops onto the \textit{CyO} balancer. These were used in subsequent crosses to isolate additional hops, using the same transposase source. Chromosomes containing the new \textit{P\{RLY\}} insertions were mapped using standard genetic methods with lab stocks of \textit{y w} flies, \textit{FM7/Df(1)JA52} flies, \textit{CyO/wg\textsuperscript{Sp-1}} flies and \textit{TM3/Scm\textsuperscript{ET50}} flies.

\textbf{Testing for Chromosomal Position Effects (CPE)}

To test for CPE, \textit{P\{RLY\}} flies were crossed with \textit{P\{70FLP\}10} flies (BDSC 6938) and third instar larvae were heat-shocked in a 37\textdegree C water bath for 1 hour to remove the M2 insulator. Resulting flies were crossed to flies with appropriate balancer chromosomes, and individual progeny were selected, made homozygous for the \textit{P\{RLY\}} chromosome, and checked for removal of the M2 insulator by PCR. Homozygous \textit{P\{RLY\}} and \textit{P\{RLYdelM2\}} flies were crossed to \textit{y w} flies, and eye color of 2 to 3 days old allelic heterozygous females was compared.
Generating \textit{attP} sites

Lines, especially those showing CPE, were selected for generating \textit{attP} landing sites. Selected stocks were crossed to males with a \textit{y w P[nos-Dint1.5 y]} \textit{X} chromosome, and \textit{F1} female virgins were crossed to \textit{y w} males with balancers as appropriate. Individual \textit{F2 y+ w} males were crossed to \textit{y w} females with balancers as appropriate, followed by crosses to assure removal of the \textit{y w P[nos-Dint1.5 y]} chromosome and to make the \textit{P[attP y]} chromosome homozygous. The presence of the \textit{attP} site was verified by PCR and sequencing of the PCR product.

\textbf{Mapping P-element integration sites in genomic DNA}

Mapping of integration sites of P-elements was done by performing TAIL-PCR (Liu and Chen 2007), inverse PCR (Huang \textit{et al}. 2009), or splinkerette PCR (Potter and Luo 2010). Products were sequenced and aligned to the genome using BLAST. Genomic primers were then designed and used with P-element primers to verify the genomic locations.

\textbf{2.3 Results}

The use of site-specific integration allows different transgenes to be tested in the context of identical CPE. It also eliminates the need to map the site of transgene integration. The bacteriophage phiC31 Integrase together with \textit{attP} sites has been adapted for this purpose in \textit{Drosophila}, and is commonly used (Groth \textit{et al}. 2004; Bischof \textit{et al}. 2007). In many cases it is desirable to minimize CPE. In our case, we are interested in chromatin domain insulator element function so we want a CPE that can be blocked by candidate insulator sequences. We tested 13 \textit{attP} landing platforms available from the BDSC using a mini-\textit{w} transgene with an scs insulator downstream and with or without the scs’-derived M2 insulator upstream (Cuvier \textit{et al}. 1998). None were suitable for our purpose. In order to generate \textit{attP} landing sites at genomic locations subject to CPE that can be blocked by insulators, we developed two tools. One is a mutagenized phiC31 Integrase protein capable of mediating recombination between \textit{attR} and \textit{attL} sites to yield
an *attP* site. The other is a P-element-based plasmid in which an insulator can be removed to test for CPE, and that has *attR* and *attL* sites to allow for creation of an *attP* site.

**Design of the phiC31 Disintegrase gene**

In 2008, the lab of M.C. Smith reported a phiC31 integrase mutant capable of catalyzing recombination between phiC31 *attR* and *attL* sites. This phiC31 integrase variant replaces a Glu at position 449 with a Lys (Rowley *et al.* 2008). To recreate this variant in the fruit fly, we replaced the equivalent Glu with a Lys in a nuclear-localized, *Drosophila* codon-optimized integrase (Bischof *et al.* 2007) by overlap PCR. The gene for this mutant version the phiC31 integrase was then cloned into a P-element transposon vector where its expression and localization is controlled by the *nanos* promoter and UTRs. This results in a germline-expressing phiC31 E449K variant whose mRNA localizes to the posterior region of early embryos, where germ cell development initiates. We call this Drosophilized E449K version of the integrase, Disintegrase or Dint (Figure 2.1). We only determined the efficiency of dis-integration between an *attR* and *attL* site to yield an *attP* site for one fly line (see below) and found it to be over 90% efficient (129/136 flies).

![Figure 2.1](image.png)

**Figure 2.1.** C4 yellow Disintegrase transposon. Shown in yellow is the yellow gene, which includes its promoter, 5’ UTR and 3’ UTR. Downstream of this is the nanos promoter with its 5’ and 3’ UTRs (green). Inserted between the nanos 5’ and 3’ UTRs is the Disintegrase sequence (light orange) fused to a nuclear localization sequence (nls in red). The 3’ and 5’ P inverted repeats are labeled in dark orange.
Design and use of a P-transposon to detect CPE

We designed a P-element construct to allow us to identify genomic sites that exhibit CPE affecting \( w \) expression, and then to remove the \( w^+ \) gene leaving behind an \( \text{attP} \) site to allow testing other candidate insulator sequences at the same location (Figure 2.2). We placed the BEAF-dependent scs’-derived M2 insulator (Cuvier et al. 1998), bracketed by \( \text{FRT} \) sites, upstream of a mini-\( w \) transgene with an scs insulator located downstream. An \( \text{attR} \) site was upstream, and an \( \text{attL} \) site was downstream of this assembly. Further downstream of this, the transposon has a \( y^+ \) transgene bracketed by \( \text{loxp} \) sites to serve as a marker for the presence of the \( \text{attP} \) site after removal of mini-\( w \).

![Diagram](image)

**Figure 2.2.** The \( P\{RLY\} \) transposon and the strategy for testing for CPE and making an \( \text{attP} \) landing site. The top shows the part of the pRLY plasmid between the P-element ends (orange rectangles). The mini-\( w \) gene is flanked by M2 (light purple rectangle) and scs (purple rectangle) insulator sequences. M2 is also flanked by \( \text{FRT} \) sites (light green rectangles) so it can be excised by FLP recombinase to test for CPE (middle of figure). The insulators and mini-\( w \) are flanked by \( \text{attR} \) and \( \text{attL} \) sites. The phiC31 Dint enzyme can excise these sequences, leaving an \( \text{attP} \) site (bottom of figure). Downstream is a transcription unit made of enhancers from the \( y \) gene upstream of a \( y^+ \) cDNA to serve as a marker after mini-\( w \) is removed. There are \( \text{loxp} \) sites (green rectangles) flanking the \( y^+ \) transgene so it can be removed by CRE recombinase.

Using the M2 insulator, 23 viable lines were obtained and tested for CPE. Of these, we used 10 to generate new \( \text{attP} \) landing platforms (Table 2.1). The \( \text{attP} \) sites were confirmed by
genomic sequencing, and their locations were mapped by inverse or TAIL or splinkerette PCR
(Liu and Chen 2007; Huang et al. 2009; Potter and Luo 2010). Four of these shows clear
activating CPE, one shows silencing CPE, and the other five show weak or minimal CPE (Figure
2.3). In most cases, data on FlyBase indicates that the level of CPE correlates with the expression
of nearby genes (Gramates et al. 2017). For example, three of the four sites showing strong CPE
are associated with genes that show high expression in adult eyes (CG32638, Tsp42Ej,
l(1)G0289). One of these also shows high expression in third instar larval imaginal discs
(l(1)G0289). The fourth is associated with a gene with moderate expression in adult eyes and
third instar larval imaginal discs (Actn).

Table 2.1. Fly lines with attP sites and their CPE potential.

<table>
<thead>
<tr>
<th>Fly line</th>
<th>Chr arm</th>
<th>Genomic site</th>
<th>Orientation</th>
<th>CPE Potential</th>
<th>Integrate gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>attP-B.3</td>
<td>3R</td>
<td>27,281,983</td>
<td>+</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>attP-D.2</td>
<td>2R</td>
<td>11,912,809</td>
<td>+</td>
<td>+</td>
<td>Yes</td>
</tr>
<tr>
<td>attP-G.3</td>
<td>3R</td>
<td>4,433,992</td>
<td>-</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>attP-H.X</td>
<td>X</td>
<td>13,128,844</td>
<td>-</td>
<td>+++</td>
<td>Yes</td>
</tr>
<tr>
<td>attP-I.2</td>
<td>2R</td>
<td>7,040,089</td>
<td>+</td>
<td>+++</td>
<td>Yes</td>
</tr>
<tr>
<td>attP-M.X</td>
<td>X</td>
<td>10,366,253</td>
<td>+</td>
<td>+++</td>
<td>Yes</td>
</tr>
<tr>
<td>attP-P.3</td>
<td>3R</td>
<td>8,738,659</td>
<td>+</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>attP-R.2</td>
<td>2R</td>
<td>8,250,415</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>attP-T.X</td>
<td>X</td>
<td>2,037,411</td>
<td>+</td>
<td>+++</td>
<td>Yes</td>
</tr>
<tr>
<td>attP-W.3</td>
<td>3L</td>
<td>17,032,287</td>
<td>+</td>
<td>++ (sil)</td>
<td></td>
</tr>
</tbody>
</table>

The fly line name indicates whether the attP site is on chromosome X, 2 or 3. The chromosome
arm genomic site is in R6 coordinates. The attP-W.3 line has a silencing CPE (sil). Lines that
have been combined with a phiC31 Integrate transgene are indicated.
Figure 2.3. Testing for CPE in flies. Eyes of 2 to 3 day old heterozygous females were photographed before (M2) and after (ΔM2) flipping out the M2 insulator upstream of the mini-w gene to test for CPE. Heterozygotes were used to avoid any confounding pairing effects, although CPE is equally apparent in homozygotes. The fly lines shown were used to make attP sites. Note that the 3W.3 line shows reverse CPE, suggesting the M2 insulator is blocking a silencer rather than an enhancer.

To confirm that the attP sites are functional, and that CPE could be recapitulated, chromosomes with phiC31 Integrate transgenes were introduced into several of the lines. Three of these were used to integrate attB-containing plasmids into the attP sites. Adjacent to the attB site, the plasmids had a mini-w gene with a downstream scs insulator. One plasmid had the M2 insulator between attB and mini-w (Figure 2.4A), one had a monomer of the M sequence, and one lacked an upstream insulator. These plasmids were derived from pCaSpeR4, so P-element ends were also present. The M fragment is around 215 bp long and has a high affinity BEAF binding site. It had only been tested as a dimer, which functioned as well as a 500 bp scs’ insulator containing a low affinity BEAF binding site in addition to the high affinity site (Cuvier
et al. 1998). It was of interest to determine if the monomer with a single BEAF binding site would be as effective as the dimer in blocking CPE.

**Figure 2.4.** Testing attP sites and CPE. (A) Integration of the attB plasmid into the attP docking site results in the test insulator and downstream mini-w gene being near genomic DNA. The scs insulator and rest of the plasmid (not shown, indicated by the broken line) is between the mini-w and y⁺ transgenes. One plasmid is shown, although three plasmids were used for integration: one lacking a test insulator, one with a monomer of the M insulator fragment from scs’, and one with the M2 dimer. (B) CPE is evident after integration into the attP-H.X line without an insulator upstream of mini-w. This CPE is partially blocked by the M insulator and is more effectively blocked when M is dimerized into the M2 insulator. (C, D) CPE that can be blocked by M2 is also evident after integration into the attP-I.2 and attP-M.X lines.

All three plasmids were integrated into one attP line (Figure 2.4B). Note that additional sequences are present between the scs insulator and the y⁺ gene compared to the RLY transposon (the P 5’ and 3’ ends and pUC8 sequences), raising the question of whether CPE would be affected. The results clearly show that CPE is observed, and that a monomer of the M sequence is a weaker insulator than a dimer. Therefore, only the plasmids lacking an insulator or with M2 were integrated into the other two attP lines (Figure 2.4C, D). The results demonstrate that
functional \textit{attP} sites were generated in the flies, and CPE that could be blocked by M2 was observed in all cases.

\textbf{2.4 Discussion}

Here we describe the design and use of Dint, a phiC31 disintegrase enzyme for generating \textit{attP} landing platforms in \textit{Drosophila}. The design is based off of a report of reversible phiC31 integration (Rowley \textit{et al.} 2008). It is similar to a previously reported tool (Knapp \textit{et al.} 2015), except ours is based on a nuclear-targeted phiC31 \textit{integrase} transgene with 172 nucleotide changes to better match \textit{Drosophila} codon usage (Bischof \textit{et al.} 2007). Use of the \textit{nanos} promoter and 5’ and 3’ UTRs should also improve germline localization in embryos. The design of a \textit{P}-element based transposon with a mini-\textit{w} reporter gene for detecting CPE that can be blocked by an insulator is also described. We used this transposon with Dint to make \textit{attP} sites in 10 fly lines that show varying degrees of CPE, including one line that had silencing rather than activating CPE. Three lines that showed strong CPE were used to demonstrate that the generated \textit{attP} sites are functional. In all three cases, it was found that the CPE was recapitulated and could be blocked by the M2 insulator. M2 is a dimer of a 215 bp sequence from the 500 bp \textit{scs’} insulator (Kellum and Schedl 1991; Zhao \textit{et al.} 1995a; Cuvier \textit{et al.} 1998). While \textit{scs’} has a high affinity and a low affinity BEAF binding site, the M monomer only has the high affinity site. The BEAF binding site is important for the insulator function of M2, since mutating it to eliminate binding by BEAF impairs insulator function (Cuvier \textit{et al.} 1998). Likewise, M2 insulator function is impaired by the presence of a dominant negative form of BEAF or by a lack of BEAF caused by a null \textit{BEAF} mutation (Gilbert \textit{et al.} 2006; Roy \textit{et al.} 2007a). We found that the M monomer was not as effective as the M2 dimer at insulating against CPE, indicating that two
BEAF binding sites make a stronger insulator than a single site. Future work with these lines could further explore the relationship between scs’ sequences and insulator activity.

An advantage of this method is that it allows potential landing locations to be prescreened for expression properties. We did this by random P-element-mediated integration to look for CPE on mini-\(w\) expression in adult eyes. Different setups could be used, for instance to screen for CPE in embryos or other tissues or to find useful enhancer traps. Once locations of interest are identified, the test transgene can be excised and an \(attP\) site can be generated so other transgenes can be integrated there.

Before we started, we did not know with certainty where in the genome we would find suitable CPE. In the end we found a strong correlation between CPE and expression levels of nearby genes in adult eyes. High expression of a nearby gene correlated with strong CPE, while low expression of all nearby genes usually correlated with weak or no CPE. With hindsight, this makes sense. Sequence-specific CRISPR/Cas9-directed genome modification can readily be done in \textit{Drosophila} (Bier \textit{et al.} 2018). In principle this could be used to integrate transgenes or \(attP\) sites at locations chosen purely on high throughput expression data available through FlyBase or based on some other source of information. This could be a good strategy depending on the quality of the information used to select the integration site. However, the strategy we used provides an unbiased sampling of the genome for finding appropriate integration locations for one’s experimental needs combined with using the highly efficient phiC31 integration system. The set of 10 new \(attP\) landing platforms that we generated represent a resource that could be useful to members of the fly community.
2.5 References


CHAPTER 3. INSULATOR AND PROMOTER ACTIVITY ANALYSIS OF SCS’ INSULATOR

3.1 Introduction

Chromatin domain insulators have been defined based on their ability to block chromosomal position effects and to disrupt the communication between an enhancer and a promoter when inserted in between (Hart et al. 1997; Raab and Kamakaka 2010; Kyrchanova and Georgiev 2014). Evidence has also shown that some insulator proteins also play roles as positive or negative regulators of gene expression in addition to providing an architectural function in mediating inter- and intra-chromosomal interactions (Soeller et al. 1988; Farkas et al. 1994; Ali et al. 2016). Insulator proteins are also called architectural proteins because they are thought to influence 3D organization of chromosomes in nuclei (Gomez-Diaz and Corces 2014; Bouwman and de Laat 2015; Cubeñas-Potts et al. 2017). In vertebrates, CTCF is the only architectural protein identified to date, and plays a prominent role in maintaining chromatin loops (Rao et al. 2014; Ali et al. 2016). It is highly conserved, with 11 Zinc finger domains (C2H2-ZF) that target CTCF to thousands of genomic sites (Lupiáñez et al. 2015; Dekker and Mirny 2016; Szalaj and Plewczynski 2018). In Drosophila, several architectural proteins are known in addition to a homolog of CTCF. All of these proteins (Su(Hw), Pita, ZIPIC, Zw5, GAF, BEAF-32, Ibf1, Ibf2 and dCTCF) contain one or more zinc finger domains for specific DNA-binding. Zw5, Pita and ZIPIC, also contain a characteristic N-terminal ZAD (zinc finger associated domain) domain that is responsible for protein–protein inter- actions (Gaszner et al. 1999; Chung et al. 2002; Merkenschlager and Odom 2013; Zolotarev et al. 2016; Tsai et al. 2016).

Architectural proteins are critical in regulating enhancer–promoter interaction specificity and those interactions between enhancers and promoters significantly contribute to the
generation of 3D chromatin architecture. In *Drosophila*, this idea is supported by the observation that most architectural protein sites in the genome correspond to enhancers and promoters (Cubeñas-Potts et al. 2017). Study done by Cubeñas-Potts and colleagues also suggest that hkCP enhancers bring multiple TSSs together to increase the local concentration of RNA Polymerase II and general transcription factors, while dCP enhancers often are associated with single TSS. In either case the enhancer and promoter interactions are associated with architectural proteins.

Genomic studies indicate that most (>85%) of BEAF binding regions were centered within 300bp of transcription start sites. Around 85% of the genes that BEAF binds are housekeeping genes, suggesting a general role in promoter activity and enhancer–promoter communication (Jiang et al. 2009). BEAF was also shown to have genetic interactions with some transcription factors, which suggests a role in gene regulation (Roy et al. 2007b, 2011). Most BEAF associated genes are transcriptionally active and are marked by the presence of RNA polymerase II, H3K4me2 and histone variant H3.3. All these results suggests that BEAF plays a role in maintaining the environment of promoter regions favorable for active transcription (Jiang et al. 2009). Similarly, in budding and fission yeast, certain Pol III promoters as well as the budding yeast pol II promoter of the *CHA1* gene have been found to act as barriers to heterochromatin spreading (Donze and Kamakaka 2001; Simms et al. 2004; Scott et al. 2006). This raises the question of whether BEAF-associated promoters play a role in insulator activity.

Among all BEAF binding sites, around half are between divergently transcribed genes (Jiang et al. 2009). One good example is the scs’ insulator element. It is one of the first insulators described in *Drosophila*. Located at one end of the 87A7 hsp70 heat shock locus, scs’ has both a high and low affinity BEAF binding site, both with clusters of 3 CGATA motifs (Zhao et al. 1995b). The CGATA motifs are important for scs’ insulator function, and clusters are found in
other sequences that BEAF binds to that have been shown to have insulator activity (Cuvier et al. 1998). A model termed Dual-core was proposed in which BEAF binds 3 or more CGATA motifs clustered in a 100 bp region (Emberly et al. 2008). A dual core has two such binding elements separated by less than 800 bp of generally AT-rich DNA. Genome-wide mapping found 1800 to 3000 BEAF peaks and confirmed that CGATA clusters are frequent in BEAF binding regions (Jiang et al. 2009; Nègre et al. 2010b). However, there is high variability in the relative orientations and spacing of CGATA motifs in clusters and many peak regions have only one or no CGATA motifs. Electrophoretic mobility shift assays (EMSA) found that CGATA clusters are not sufficient to guarantee binding by BEAF, and BEAF can bind sequences with a single CGATA (Jiang et al. 2009).

To gain insight into BEAF function, we did an analysis of the scs’ insulator element. scs’ is a highly studied BEAF-binding insulator of the Drosophila genome (Udvardy et al. 1985; Kellum and Schedl 1991, 1992; Zhao et al. 1995; Cuvier et al. 1998; Blanton et al. 2003). The half of the scs’ sequence with the high affinity BEAF binding site was tested as a dimer, M2 fragment. It has equal insulating potential as scs’ (Cuvier et al. 1998), proving half of scs’ is dispensable. In this study, we used the M2 fragment to determine the minimal sequence required for scs’ insulator function. M2 has been designed such that it contains two high affinity BEAF binding sites at same distance apart as the two BEAF binding sites in scs’. A linker scanning (LS) assay was done to determine if there were any sequences necessary for insulator function in addition to BEAF binding sites. A 20bp long sequence (LS4) appeared to be important for proper insulator function of M2 fragment. Based on this preliminary result we further divided the M fragment into several derivative sequences to check additional important sequences for insulator function. We used the PhiC31 integrase (Thorpe and Smith 1998) system to make a standardized
site-specific integration and test all the M derivatives at the same genomic location. Previously created attP fly lines, showing good CPE (Chromosomal Position Effect) (Maharjan et al. 2018), were used to integrate attB plasmids having different deletions or mutations of the M sequence of scs’. Additionally, it has long been known that scs’ contains two divergent promoters (Glover et al. 1995), but the significance of this has never been investigated. Along with insulator function, we also tested for promoter function, to see if these activities correlate or can be separated. Evidence so far suggests that BEAF could play a role in keeping promoters accessible to the transcription machinery, or perhaps play a more direct role in Pol II recruitment. Insight into mechanisms will be gained by identifying scs’ sequences that work with the BEAF binding site and further identifying the proteins that bind these sequences. In the future, expanding this analysis to additional BEAF binding regions will help us understand the role of BEAF and perhaps also other architectural proteins in genome architecture and gene regulation.

3.2 Materials and Method

Plasmids and DNA

All plasmids used for testing insulator activity were made from pC4scs as described previously (Pirrotta 1988; Maharjan et al. 2018). Three fly lines (attP-H-X; Int.4, attP-I-2; Int.X and attP-M-X; Int.4) which showed strong CPE in adult eyes were used. We tested derivatives of the M fragment (Cuvier et al. 1998) for its minimal sequence for insulator function. M was divided into five sub fragments: 5’ end, LS4 region, spacer region, H site, and 3’ end. Δ5, Δ3 and Δ53 fragments were made from the M fragment by deleting the 5’ end, the 3’ end, or both the 5’ and 3’ ends, respectively. ΔSp and Sp* fragment were made by deleting or mutating the spacer region (the sequence between LS4 and the BEAF binding site changed to lambda DNA sequences to maintain the spacing) respectively. The M2 dimer is reversed in direction in M2rev.
Additionally, in H* and LS4*, the M fragment is mutated at the BEAF binding H site and at the LS4 site.

**Growing S2 cells and media**

S2 cells were grown in Shields and Sang M3 Insect media (Sigma #S8398). S2 complete media was prepared by mixing M3 media with 10% FBS (Corning 35-010-CV) and 1x antibiotics, anti-mycotic (Gibco™ 15240062). Cells were grown in 25cm² T flasks and were maintained with cell splitting every 4 days with 1:4 dilutions into new media and flask. The cells were not allowed to grow more than $10^7$/ml before it was split.

**Transfections and luciferase assay**

Transfections were done in 24 well plates. 7.5 x $10^5$ cell were placed in each well. The total volume was brought to 1ml by adding S2 complete media and incubated for 24 hours at 25°C. Next day the cells should have 70% confluency. The DNA mix, total 1ug, was prepared by adding 30% and 70% of Firefly M-fragment and Renilla experimental Plasmid DNAs, respectively. 5ul of lipofectamine reagent (Invitrogen 11668-019) and 500ul of serum free S2 media was added to the DNA mix and incubated at room temperature for 15 minutes. Meanwhile, cells in the plate were carefully washed with serum free media and DNA mix solution was added to cells and incubated for at least 4 hours. After incubation, the transfection mix was replaced by S2 complete media and incubated for 48 hours.

For luciferase assay a dual luciferase kit was used from Promega (#E1910). The assay was done as described in the manufacturer’s instructions. The readings were taken after 48 hours of incubation of the transfected cell in S2 complete media.
3.3 Results

LS4 site is also necessary for insulator function of SCS’

Previously, a 225 bp sequence from scs’ containing a high affinity BEAF binding site (H site), termed the M fragment was dimerized and found to insulate against CPE as well as scs’ does (Cuvier et al. 1998) (Figure 3.1). As with scs’, only 10% of fly lines with M2 showed activation of a mini-white reporter gene by CPE. The rationale was to make the distance between the duplicated high affinity BEAF binding sites the same as the distance between the high and low affinity binding sites in scs’ (Figure 3.1). Mutations in the H site (M*2) that eliminated BEAF binding eliminated insulator activity in this assay (Figure 3.1). The smaller dimer fragments S2 and X2 (dimers of 110 bp and 48 bp sequences), were also tested for insulator activity. They both showed reduced activity. This suggested that regions in M2 missing in S2 or X2 increase the insulator activity of the BEAF binding site (Figure 3.1). This was confirmed using a Kolmogorov-Smirnov statistical analysis. A linker-scanning analysis was performed on the 115 bp present in M2 but not in S2, using six 20 bp steps. Analysis of over 200 transgenic fly lines found that the LS4 linker-scanning mutation clearly impaired insulator activity (Figure 3.1). This supports the hypothesis that there is an accessory protein or proteins that work with BEAF for full insulator activity and overall, the H site and LS4 region were found to be essential for full scs’ insulator function. No obvious protein binding or promoter motif is apparent in the LS4 sequence. Additionally, motif-finding programs did not find an LS4-related sequence associated with other BEAF binding regions. Still, possibilities include binding by a sequence-specific binding protein or a complex such as TFIID or RNA polymerase II. To investigate further, we wanted to test if only the H site and LS4 region are sufficient for full insulator activity or the spacing between the H site and the LS4 region are also important for insulator function.
 Regions except the 5’ end is important for insulator function

The M fragment was further divided into five sub-fragments: 5’ end, LS4 region, spacer region, H site, and 3’ end (Figure 3.2). We recently found that the M fragment monomer containing a single H site does not work as well as M2 containing two H sites (Maharjan et al. 2018). Since the dimer construct worked well, all the derivatives were tested as dimers and constructs were cloned into the pC4 attB plasmid for fly injection to detect insulator activity in position effect assays. Here, we used the PhiC31 integrase mediated site-specific insertion system which mediates specific integration of an attB site into a transgenic attP landing site already in
the fly genome (Bischof et al. 2007; Maharjan et al. 2018). We found that 5’ end sequence and the sequence between LS4 and H site is not important for insulator function. Deleting the spacer region, 3’ end or mutating LS4 or the H site all reduced insulator function (Figure 3.2). It was a reconfirmation that the LS4 sequence is important. At two genomic locations tested, the result of the LS4 site mutation matched with the result when insulator was absence. It is not clear why the LS4 mutant insulator worked well in the attP-H-X line. Perhaps the requirement for LS4 and presumably LS4 binding proteins, depends on the chromosomal context. Because both the 5’ deletion and 3’ deletion changed the spacing between BEAF binding sites, it is likely that 3’ sequences rather than spacing are important for insulator function. However, there is a 3bp difference in spacing between H site for the 5’ deletion (163 bp) compared to the 3’ deletion (166 bp), so DNA helical phasing cannot be ruled out.
Figure 3.2. M fragment derivative sequences and their activity in CPE. 5’ end and spacer sequence are dispensable for insulator function. A) Derivative sequences of scs’ inserted in plasmid with attB site. Colored boxes are not drawn to scale. B) Eye color of transgenic flies that are injected with the derivative sequences. attB: no insulator; M2: full insulator; M2rev: M2 reversed; H*, LS4* and SP*: BEAF, LS4 and SP site mutated; Δ5, Δ3 and ΔSp: 5,3 and spacer sequence deleted.
**BEAF binding site relates to promoter function**

Over 85% of BEAF peaks are centered within 300 bp of a TSS (Figure 3.3). Many of these are between head-to-head divergent gene pairs, representing at least one-third of the genes organized in this fashion. These could fit the dual-core model. An example of this is scs’, with BEAF binding sites next to both TSSs (Emberly *et al.* 2008; Jiang *et al.* 2009). To see the relationship between the promoter and insulator activity of scs’, M fragment derivatives were tested for promoter activity by transient transfection of S2 cells (Figure 3.4 B). The constructs were placed upstream of a Luciferase reporter gene (Figure 3.4 A). Promoter activity of the M fragment was detected (Figure 3.4C). Only the 3’ end of the M fragment was found to be unnecessary for promoter activity. BEAF binding was found to be essential for promoter function of M fragment. Interestingly, our result also showed the M fragment has bidirectional transcription which is consistent with other studies (Meers *et al.* 2018; Henriques *et al.* 2018).

![Figure 3.3](image)

**Figure 3.3.** BEAF binds near TSSs and is important for its associated promoters. A) Positions of the centers of 1820 BEAF peaks relative to the nearest annotated TSS. (source of picture Jiang *et al.* 2009).
Figure 3.4. Promoter assay with M fragment. A) Construct with luciferase gene at downstream of M fragment. B) M fragment derivatives used for promoter assay. All the descriptions and symbols are same as Figure 3.2. C) Promoter activity of M insulator is showed by all of the derivatives except 3’ end. Mean value of luciferase activity was taken from six individual biological replicates. Luciferase value is normalized with M fragment luciferase activity value. Deletion is denoted by ‘Del’ and mutation in the fragments are denoted by ‘*’ mark. All abbreviations of the different derivative M fragments are as described in text.
3.4 Discussion

Previous studies have indicated that BEAF binding sites can contribute to insulator function (Cuvier et al. 1998; Sultana et al. 2011). It was shown that a dimer made of two copies of half of scs’ (M2) containing the high affinity BEAF binding site insulated against CPE as well as scs’ does. We recently reported that a single copy is not as effective (Maharjan et al. 2018). We dissected the M fragment for insulator and promoter activity. Our result reconfirms the importance of BEAF binding for insulator function. Additionally, for the first time, our result also indicates the BEAF binding site is needed for promoter function of a BEAF associated promoter. While our finding shows an overlap of sequences needed for insulator and promoter function, it also indicates the insulator and promoter function can be separated.

Initially, the M fragment was further tested by making shorter (S) and extra-short (X) dimer fragments both containing the H site. Neither of the short fragments maintained full insulator activity. Hence the 5’ sequence that were present in M but deleted in S and X fragments were found to be important. This sequence was further dissected in linker scanning (LS) assay. The sequence was divided into six 20bp sequences with 3-4 bps overlap that were mutated and tested. Only the fourth LS sequence (termed as LS4) was found to be important for insulator function. After we found that the LS4 sequence is important, we went on to test additional variations of the M fragment. In further dissection of M fragment, we found that only the 52bp 5’ sequence could be deleted. However, the spacing between LS4 and BEAF binding site turned out to be important but the sequence itself was not consistent with the LS5, LS6 result. So, a large sequence of 173bp plays role in insulator function (but LS5 and LS6 can be mutated). This indicates for full insulator function there must be more proteins binding to the insulator than just BEAF. However, analysis with motif finding program such as MEME and JASPAR did not find
other obvious protein binding sites. A similar scenario is seen for the *Fab-7* boundary region of the Bithorax complex, where the LBC (Late Binding Complex) responsible for the boundary function recognize minimum sequence of >65bp (Wolle *et al.* 2015; Cleard *et al.* 2017). Known proteins of LBC are GAF, Mod(mdg4) and E(y)2, with the entire complex being >700 kDa. In case of scs’, the proteins that bind to facilitate insulator function might be structure dependent rather than sequence dependent, accounting for the lack of obvious motifs, perhaps binding together with BEAF stabilizes their binding. The insulator was also tested in the reverse orientation and insulator function was not affected, which suggest that the mechanism is different than CTCF (Guo *et al.* 2015) and might not involve directional chromatin looping.

Most of the BEAF sites are found near TSSs (Jiang *et al.* 2009; Schwartz *et al.* 2012; Fujioka *et al.* 2013). In addition, scs’ has 2 promoters at its two ends with BEAF binding sites by each. The M fragment contains one of these promoters. Therefore, we tested the M fragment and its BEAF binding site for promoter activity. A related question was whether insulator and promoter activity rely on the same sequences. Even though our initial prediction was that deleting the 3’end will affect the promoter activity as it might contain the aurA promoter, this was not observed. Surprisingly, only the 3’ end was found to be not important for the promoter activity, indicating sequence after the BEAF binding site is not important in the M fragment. BEAF binding seems to be very important for promoter activity as well, the possible explanation being that BEAF might help in making the promoter (or TSS) site more accessible for the factors needed for promoter functioning. That is also consistent with the finding that BEAF is mostly found near housekeeping gene TSSs (Shrestha *et al.* 2018), which are constitutively expressed, meaning they have to be made accessible for the transcription machinery and BEAF might play a role in making promoters active.
It is interesting we found that insulator and promoter activities overlap by 110 bps but can be separated. Only the middle portion of M fragment i.e. LS4, spacer and H site seems to be important for both insulator and promoter function. BEAF is necessary for both, but insulator function requires 50 bp additional downstream DNA sequences while promoter activity requires 50 bp additional upstream sequences. However, the spacer sequence cannot be mutated and retain promoter function but is not important for insulator function. Furthermore, for promoter analysis, linker scanning might show some important sequence within 5’ region. In conclusion, this means insulator and promoter activity can be separated and probably a common set of proteins, including BEAF, bind for these functions and some unique protein set also works for these two activities (Figure 3.5).

Figure 3.5. M fragment showing the different regions that proteins bind for insulator function and promoter function. LS4, spacer and H site seems to be important for both insulator and promoter function. 5’ region plays role in promoter function and 3’ region in insulator function.

3.5 References


Henriques T., Scruggs B. S., Inouye M. O., Muse G. W., Williams L. H., et al., 2018 Widespread
transcriptional pausing and elongation control at enhancers. Genes Dev. 32: 26–41.


Wolle D., Cleard F., Aoki T., Deshpande G., Schedl P., et al., 2015 Functional Requirements for Fab-7 Boundary Activity in the Bithorax Complex.

CHAPTER 4. IDENTIFICATION OF BEAF-ASSOCIATED PROTEINS

4.1 Introduction

Two of the first identified insulator elements are scs and scs’ sequences which bracket two Hsp70 genes at the 87A locus of Drosophila (Udvardy et al. 1985). BEAF was identified as a Boundary Element-Associated Factor that binds to the scs’ insulator and was subsequently shown to immunolocalize to hundreds of inter-bands and band/inter-band boundaries of polytene chromosomes (Zhao et al. 1995a). A single gene of BEAF gives two isoforms, BEAF-32A and BEAF-32B, both being 32 kD. These two isoforms presumably originate from alternative promoters and differ only in around 80 amino acids at the long N-terminus. Isoform 32A is not essential for Drosophila survival but 32B is. In this chapter we focus on BEAF isoform 32B. BEAF-32B is 282 amino acids long (Zhao et al. 1995a; Avva and Hart 2016). The C-terminal region of 80 amino acids has a putative leucine zipper and a BESS domain and mediates BEAF-BEAF interactions (Hart et al. 1997; Avva and Hart 2016). It has an 80 amino acid N terminal region with a DNA binding domain, while the structure and function of the middle region of 120 amino acids is unknown. The scope of the BEAF interacting protein network is not known and it is not possible to know the exact role of BEAF in gene regulation or in nuclear architecture without identifying its interacting proteins and their general functions.

Although many proteins have been found to colocalize with BEAF in genome-wide ChIP-chip and ChIP-seq data, very few proteins that have been found to physically interact with BEAF (Ali et al. 2016). An abundant chromosomal protein D1, resembling a large version of mammalian HMGA (formerly HMG-I) proteins, predominantly binds to AT-rich satellite DNA sequences. It has been reported to cooperatively bind to certain DNA sequences with BEAF to perhaps act as a local boundary between hetero- and euchromatin (Cuvier et al. 2002). Another
protein reported to interact with BEAF is Zw5 (Blanton et al. 2003), a protein that binds to the scs insulator (Gaszner et al. 1999). Centrosomal Protein 190 (CP190), was also reported to bind with BEAF and perhaps makes chromatin loops for regulation of gene expression and maybe Polymerase-II pausing (Ahanger et al. 2013; Liang et al. 2014). BEAF together with CP-190 also might interact with the dREAM complex to maintain and organize transcriptional domains and cell cycle dependent gene regulation (Sadasivam and DeCaprio 2013; Korenjak et al. 2014; Vogelmann et al. 2014). Some insulator proteins like Chromator and mod(mdg4) might interact with BEAF to somehow affect chromatin dynamics (Ahanger et al. 2013; Vogelmann et al. 2014; Ong and Corces 2014; Beagan et al. 2017). With the exception of Chromator, evidence for physical interactions with BEAF are not conclusive and the functional consequences is not clear. There are reports for other interactions with BEAF like: Sry-delta, pzg, Dref and fs(1)h (Gan et al. 2011; Kellner et al. 2013; Lhoumaud et al. 2014; Rhee et al. 2014).

Despite the identification of several BEAF partners, the scope of the BEAF interacting protein network is unknown. Hence, the exact functions of the BEAF protein remain poorly understood. One way to identify protein partners of BEAF is to immunoprecipitate (IP) it from Drosophila nuclear extracts and identify proteins that co-immunoprecipitate. In this study we have used a combination of co-immunoprecipitation and proteomic mass spectrometry to identify proteins associated with BEAF in embryo nuclear extracts. We generated flies containing transgenes encoding an epitope-tagged version of BEAF-32B driven by its native promoter: FLAG-32B-EGFP. Population cages were used to collect grams of embryos, and nuclear protein extracts (NE) were prepared from these embryos. After IP using antibodies against FLAG, covalently coupled to magnetic bead, eluted proteins were sent to a proteomics facility to identify co-immunoprecipitated proteins.
Our data suggests that BEAF probably plays a role in recruiting chromatin remodelers and help in the maintenance of proper chromatin state and also in transcriptional regulation by interacting with several transcription factors. Significant proteins identified include ATP-dependent chromatin remodeler subunits from PBAP and NURF complexes, both histone chaperon, FACT subunits, transcription factors, architectural proteins and histone demethylase and acetyltransferase proteins. While physical interactions between specific proteins and BEAF needs to be confirmed using additional methods, this study establishes a very strong foundation for future. Following up on these proteins will provide insight into molecular mechanisms of BEAF function that could apply to architectural proteins in general.

4.2 Materials and method

![Population cage and grape juice agar plates with yeast paste.](http://haplotypewriter.com/2013/05/29/how-to-grow-massive-amounts-of-fruit-flies/)

**Figure 4.1.** Population cage and grape juice agar plates with yeast paste. A) Population cage B) Yeast paste applied over the grape juice agar prepared in Styrofoam plates. Fly population will feed upon the yeast and lay their eggs over the agar plates. source of pictures: [http://haplotypewriter.com/2013/05/29/how-to-grow-massive-amounts-of-fruit-flies/](http://haplotypewriter.com/2013/05/29/how-to-grow-massive-amounts-of-fruit-flies/)
**Fly line generation and growing a large *Drosophila* population**

A transgenic fly line was generated to express BEAF-32B in vivo. The 32B-GFP plasmid (Avva and Hart 2016) was modified by adding N-terminus FLAG epitope tag. The transgene was incorporated into fly genome by P-element transformation. Injections of pre-blastoderm embryos to generate transgenic flies were done by GenetiVision (Houston, TX).

First the fly line was grown in plastic food storage boxes (9’’x 6’’) with a nitex opening (4’’x 2’’) cut in the top. A thin layer of absorbent cotton was placed on the bottom and the fly food solution (water 200ml, propionic acid 0.13m, phosphoric acid 0.75ml, live yeast 54g and sucrose 27g) was poured evenly over it (http://www.personal.psu.edu/zcl1/lab/protocols/Rubin%20lab%20manual%20'90.pdf). Once the boxes were full of newly eclosed *Drosophila* they were transferred to medium sized population cages (Genesee Scientific) (Figure 4.1A). The population of flies were fed with yeast paste applied over a plate of grape juice agar (Genesee #47-102) (Figure 4.1B).

**Embryo collection**

*Drosophila* population were grown at 18°C. Embryos, 0-24 hours old, were collected once a day. With the help of a brush, the agar plates were gently washed with distilled water to dislodge the embryos from the yeast paste and were collected in a sieved tube, prepared in lab with 50ml Falcon tubes and nylon sieve (125 micron). The collected embryos were then thoroughly washed and dechorionated with a 50% bleach solution for 3 – 5 minutes, to dechorionate them. Embryos were thoroughly but gently washed with 0.7% NaCl for 4 to 5 times to remove bleach. The embryos were placed into 1.7ml microcentrifuge tubes, weighed, flash frozen in liquid nitrogen and stored at -80°C until used for nuclear extract (NE) preparation.
**Nuclear extract preparation**

NE extract was prepared from 10 grams of embryos as described in (Zhao et al. 1995a). The NE was then dialyzed with dialysis membrane (Sigma D0530) of 12.4KD cutoff in nuclear extract buffer (25mM HEPES pH 7.6, 150mM NaCl and 0.1mM PMSF), flash frozen in liquid nitrogen and stored at -80°C.

**Antibody and Co-Immunoprecipitation**

Anti-FLAG M2 covalently linked to magnetic beads (Sigma M8823) were used to immunoprecipitate the tagged 32B. The co-immunoprecipitation (co-IP) was done following manufacturer’s protocol using a modified elution buffer (100mM Glycine [pH 2.5], 10% Glycerol). The supernatant was neutralized with 1M Tris [pH 8.0]. The co-immunoprecipitated supernatant, washes and eluates were flash frozen in liquid nitrogen, stored at -80°C. Mock co-IPs were done with NE lacking FLAG-32B.

**Mass spectrometry and Statistical analysis**

After confirming the quality of the co-IP samples by SDS-PAGE followed by detection of BEAF by western blotting and total proteins by silver staining, the eluates were directly sent for the proteomic analysis at Thermo Fisher Scientific Center for Multiplexed Proteomics, Harvard Medical School. Three biological replicates were performed for experimental and mock Co-IP and proteomic differences were evaluated for statistical significance (P< 0.05) by student t-tests and corrected for multiple testing using the Benjamini–Hochberg (B&H) correction (Benjamini and Hochberg 1995). Student t-tests and B&H corrections were performed using Excel 2016 (Microsoft). The experimental to mock ratio cut-off value of 1.35 was used for experimental vs mock proteomic values. The false discovery rate (FDR) value used for B&H correction was ≤0.10.
4.3 Result and discussion

Because the FLAG epitope works well, we generated flies expressing FLAG-32B-EGFP (Riising et al. 2008; Baillat and Shiekhattar 2009). We performed four experimental co-IPs: three using FLAG antibodies (with nuclear extracts from 5 g, 10 g and 10 g of embryos, respectively) and one using GFP antibodies (with a nuclear extract from 5 g embryos). For each experimental co-IP, we performed an equivalent co-IP using nuclear extract prepared from embryos lacking epitope-tagged BEAF, as control (or mock) co-IP. We visualized proteins by silver staining after polyacrylamide gel electrophoresis to confirm co-precipitation and sent the samples for protein identification by mass spectrometry analysis (Thermo Fisher Center for Multiplexed Proteomics at Harvard Medical School). Inspection of the MS data suggested that the two co-IPs using the FLAG antibody and NE from 10 g of embryos had the highest quality data, so we focused on those samples and their matched wild-type controls. A total of 680 proteins were identified by MS analysis (Table 4.1). Omitting proteins with only one quantified peptide reduced this to 472 proteins. Using a cut-off value of 1.35 for the ratio of experimental to wild-type control Normalized % Relative Abundance to minimize identification of background proteins left 151 proteins. We performed t-tests on these 151 proteins to determine significant differences in abundance between the experimental and control (P < 0.05). This left 116 proteins. To correct for multiple testing, we applied the B&H method with an FDR of 0.1 (Benjamini and Hochberg 1995), which reduced the number of proteins to 95 (Figure 4.2) (Full table is in the appendix).
Table 4.1. Number of proteins identified by MS analysis of Co-IPs.

<table>
<thead>
<tr>
<th>Selection criteria</th>
<th>Total no of Proteins Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>680</td>
</tr>
<tr>
<td>Proteins having &gt;1 quantified peptide</td>
<td>472</td>
</tr>
<tr>
<td>Experimental/Control (Ratio ≥ 1.35)</td>
<td>151</td>
</tr>
<tr>
<td>P &lt; 0.05</td>
<td>116</td>
</tr>
<tr>
<td>B&amp;H P-value correction (FDR&lt;0.1)</td>
<td>95</td>
</tr>
</tbody>
</table>

**BEAF is associated with ATP dependent chromatin remodelers**

We were able to identify ATP dependent chromatin remodelers from the IP samples. Most of the protein subunits were statistically significant (P-values < 0.05) (Table 4.2).

Components from SWI/SNF, ISWI, and the histone chaperone FACT complexes were detected. In *Drosophila* the SWI/SNF related complex is called the Brahma (BRM) complex. Two distinct subclasses of Brahma exist, BAP and PBAP. The PBAP complex has the largest number of subunits identified at significant level, including the signature subunit Polybromo. In contrast, the signature subunit of the BAP complex, OSA, appeared in the list but was not significant.

Two of the most significant proteins found were the two subunits of the FACT complex, SSRP and SPT16. The FACT complex can assist chromatin remodelers especially during transcription. Subunits of the remodelers NURF (ISWI related), dNURD (CHD related) and INO80 were also identified at a significant level. All these subunits provide a compelling argument that BEAF-32B interacts with ATP dependent chromatin remodelers.
**Figure 4.2.** Volcano plot illustrates significantly differentially abundant proteins. The $-\log_{10}$ (Benjamini–Hochberg corrected P value) is plotted against the log2 (fold change: FLAG/Mock). The non-axial vertical line denotes 1.35-fold change while the non-axial horizontal line denotes $P= 0.05$, which is our significance threshold (prior to logarithmic transformation). The proteins in upper right (shaded) quadrant are significant.
Table 4.2. BEAF associated ATP-dependent chromatin remodelers.

<table>
<thead>
<tr>
<th>Gene names</th>
<th>Annotation symbol</th>
<th>Flybase ID</th>
<th>No. of quantified peptides</th>
<th>Ratio (FL/M)</th>
<th>P value</th>
<th>B&amp;H correction significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FACT Complex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dre4 sp16</td>
<td>CG1828</td>
<td>FBgn0002183</td>
<td>276</td>
<td>5.07</td>
<td>0.003</td>
<td>significant</td>
</tr>
<tr>
<td>Ssrp CF5 SSRP1</td>
<td>CG4817</td>
<td>FBgn0010278</td>
<td>209</td>
<td>4.09</td>
<td>0.002</td>
<td>significant</td>
</tr>
<tr>
<td><strong>SWI/SNF sub family</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PBAP/BAP Complex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mor</td>
<td>CG18740</td>
<td>FBgn0002783</td>
<td>15</td>
<td>2.34</td>
<td>0.002</td>
<td>significant</td>
</tr>
<tr>
<td>Bap60</td>
<td>CG4303</td>
<td>FBgn0025463</td>
<td>7</td>
<td>2.33</td>
<td>0.000</td>
<td>significant</td>
</tr>
<tr>
<td>brm</td>
<td>CG5942</td>
<td>FBgn0000212</td>
<td>2</td>
<td>1.60</td>
<td>0.011</td>
<td>significant</td>
</tr>
<tr>
<td>Bap55</td>
<td>CG6546</td>
<td>FBgn0025716</td>
<td>5</td>
<td>1.38</td>
<td>0.009</td>
<td>significant</td>
</tr>
<tr>
<td>Bap111/dalao</td>
<td>CG7055</td>
<td>FBgn0030093</td>
<td>4</td>
<td>1.61</td>
<td>0.011</td>
<td>significant</td>
</tr>
<tr>
<td>Act5C</td>
<td>CG4027</td>
<td>FBgn0000042</td>
<td>29</td>
<td>0.86</td>
<td>0.300</td>
<td>Not- significant</td>
</tr>
<tr>
<td>Snr1</td>
<td>CG1064</td>
<td>FBgn0011715</td>
<td>2</td>
<td>0.85</td>
<td>0.270</td>
<td>Not- significant</td>
</tr>
<tr>
<td><strong>PBAP specific subunits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>polybromo</td>
<td>CG11375</td>
<td>FBgn0039227</td>
<td>3</td>
<td>1.35</td>
<td>0.002</td>
<td>significant</td>
</tr>
<tr>
<td>Bap170</td>
<td>CG3274</td>
<td>FBgn0042085</td>
<td>1</td>
<td>0.92</td>
<td>0.418</td>
<td>Not- significant</td>
</tr>
<tr>
<td>SAYP/e(y)3</td>
<td>CG12238</td>
<td>FBgn0087008</td>
<td>2</td>
<td>0.54</td>
<td>0.062</td>
<td>Not- significant</td>
</tr>
<tr>
<td><strong>BAP specific subunits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>osa eld CG7467</td>
<td></td>
<td>FBgn0261885</td>
<td>4</td>
<td>0.93</td>
<td>0.384</td>
<td>Not- significant</td>
</tr>
<tr>
<td><strong>ISWI subfamily</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>NURF complex</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caf1</td>
<td>CG4236</td>
<td>FBgn00263979</td>
<td>7</td>
<td>1.42</td>
<td>0.016</td>
<td>significant</td>
</tr>
<tr>
<td>Iswi</td>
<td>CG8625</td>
<td>FBgn0011604</td>
<td>16</td>
<td>1.21</td>
<td>0.045</td>
<td>Not- significant</td>
</tr>
<tr>
<td>E(bx) / Nurf301</td>
<td>CG7022</td>
<td>FBgn00000541</td>
<td>4</td>
<td>0.94</td>
<td>0.367</td>
<td>Not- significant</td>
</tr>
<tr>
<td>Nurf-38</td>
<td>CG4634</td>
<td>FBgn0016687</td>
<td>8</td>
<td>0.94</td>
<td>0.436</td>
<td>Not- significant</td>
</tr>
<tr>
<td>associated protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pzg Z4</td>
<td>CG7752</td>
<td>FBgn0259785</td>
<td>38</td>
<td>3.11</td>
<td>0.003</td>
<td>significant</td>
</tr>
<tr>
<td><strong>ACF complex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acf1</td>
<td>CG1966</td>
<td>FBgn0027620</td>
<td>2</td>
<td>1.23</td>
<td>0.044</td>
<td>Not- significant</td>
</tr>
<tr>
<td>Iswi</td>
<td>CG8625</td>
<td>FBgn0011604</td>
<td>16</td>
<td>1.21</td>
<td>0.045</td>
<td>Not- significant</td>
</tr>
<tr>
<td><strong>CHD subfamily</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dNURD (Mi-2) complex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caf1</td>
<td>CG4236</td>
<td>FBgn00263979</td>
<td>7</td>
<td>1.42</td>
<td>0.016</td>
<td>significant</td>
</tr>
<tr>
<td>ttk FTZ-F2</td>
<td>CG1856</td>
<td>FBgn0003870</td>
<td>1</td>
<td>5.65</td>
<td>0.007</td>
<td>significant</td>
</tr>
<tr>
<td>Rpd3 HDAC1</td>
<td>CG7471</td>
<td>FBgn0015805</td>
<td>6</td>
<td>2.69</td>
<td>0.026</td>
<td>Not- significant</td>
</tr>
<tr>
<td>Mi-2</td>
<td>CG8103</td>
<td>FBgn0262519</td>
<td>5</td>
<td>0.99</td>
<td>0.483</td>
<td>Not- significant</td>
</tr>
<tr>
<td>simj</td>
<td>CG32067</td>
<td>FBgn0010762</td>
<td>2</td>
<td>0.46</td>
<td>0.038</td>
<td>Not- significant</td>
</tr>
<tr>
<td>MEP-1</td>
<td>CG1244</td>
<td>FBgn0035357</td>
<td>1</td>
<td>0.90</td>
<td>0.370</td>
<td>Not- significant</td>
</tr>
<tr>
<td><strong>INO80</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pont</td>
<td>CG4003</td>
<td>FBgn00040078</td>
<td>3</td>
<td>2.17</td>
<td>0.006</td>
<td>significant</td>
</tr>
<tr>
<td>Act5C</td>
<td>CG4027</td>
<td>FBgn0000042</td>
<td>29</td>
<td>0.86</td>
<td>0.300</td>
<td>Not-significant</td>
</tr>
</tbody>
</table>
ATP dependent chromatin remodeler subunits associated with BEAF32B are listed. The ratio of normalized abundance value of each peptide with FLAG vs mock (FL/M) is shown. Statistical significance was analyzed using the Student’s t-test and corrected for multiple testing using the Benjamini–Hochberg correction. Only significant proteins are highlighted, but other subunits from the same complex are also shown even though they were not significant.

**BEAF associated with other architectural proteins**

Among architectural proteins we found ones that have been reported to interact with BEAF. We found Chromator (Chriz), mod(mdg4), putzig (Z4), MAGE and DREF in our list at a significant level (Table 4.3). The Map60 (also known as CP60) is a microtubule associated protein that interacts with the important architectural protein CP190 (Kellogg et al. 1995) was also found. However, CP190, pita and subunits from cohesin and condensin I appeared in our list but were not significant. This suggests that BEAF might be doing its function in cooperation with other architectural or boundary function proteins.

**Table 4.3. BEAF associated chromosome architecture factors.**

<table>
<thead>
<tr>
<th>Gene names</th>
<th>Annotation symbol</th>
<th>Flybase ID</th>
<th>No. of quantified peptides</th>
<th>Ratio (FL/M)</th>
<th>P value</th>
<th>B&amp;H correction significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>mod(mdg4)</td>
<td>CG32491</td>
<td>FBgn0002781</td>
<td>15</td>
<td>3.39</td>
<td>0.003</td>
<td>significant</td>
</tr>
<tr>
<td>Map60</td>
<td>CG1825</td>
<td>FBgn0010342</td>
<td>2</td>
<td>7.57</td>
<td>0.003</td>
<td>significant</td>
</tr>
<tr>
<td>pzg Z4</td>
<td>CG7752</td>
<td>FBgn0259785</td>
<td>38</td>
<td>3.11</td>
<td>0.003</td>
<td>significant</td>
</tr>
<tr>
<td>MAGE</td>
<td>CG10059</td>
<td>FBgn0037481</td>
<td>2</td>
<td>10.70</td>
<td>0.000</td>
<td>significant</td>
</tr>
<tr>
<td>Dref</td>
<td>CG5838</td>
<td>FBgn0015664</td>
<td>3</td>
<td>2.61</td>
<td>0.018</td>
<td>significant</td>
</tr>
<tr>
<td>Chro</td>
<td>CG10712</td>
<td>FBgn0044324</td>
<td>33</td>
<td>2.81</td>
<td>0.018</td>
<td>significant</td>
</tr>
<tr>
<td>pita</td>
<td>CG3941</td>
<td>FBgn0034878</td>
<td>1</td>
<td>1.27</td>
<td>0.015</td>
<td>Not-significant</td>
</tr>
</tbody>
</table>

Cohesin Complex

<table>
<thead>
<tr>
<th>Gene names</th>
<th>Annotation symbol</th>
<th>Flybase ID</th>
<th>No. of quantified peptides</th>
<th>Ratio (FL/M)</th>
<th>P value</th>
<th>B&amp;H correction significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMC3 Cap</td>
<td>CG9802</td>
<td>FBgn0015615</td>
<td>5</td>
<td>0.95</td>
<td>0.380</td>
<td>Not- significant</td>
</tr>
<tr>
<td>SMC1</td>
<td>CG6057</td>
<td>FBgn0040283</td>
<td>3</td>
<td>1.31</td>
<td>0.069</td>
<td>Not- significant</td>
</tr>
<tr>
<td>pds5</td>
<td>CG17509</td>
<td>FBgn0260012</td>
<td>2</td>
<td>0.71</td>
<td>0.064</td>
<td>Not- significant</td>
</tr>
</tbody>
</table>

Condensin I

<table>
<thead>
<tr>
<th>Gene names</th>
<th>Annotation symbol</th>
<th>Flybase ID</th>
<th>No. of quantified peptides</th>
<th>Ratio (FL/M)</th>
<th>P value</th>
<th>B&amp;H correction significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>glu</td>
<td>CG11397</td>
<td>FBgn0015391</td>
<td>2</td>
<td>0.26</td>
<td>0.007</td>
<td>Not- significant</td>
</tr>
</tbody>
</table>

Proteins associated with chromosome architecture are listed and the significant ones are highlighted. Other designations are as in Table 4.2.
BEAF is associated with transcription

In addition to BEAF’s association with chromatin remodeling and architectural proteins another interesting group of proteins present in the co-IP are proteins associated with transcription (Table 4.4). BEAF function is thought to be associated with promoters since BEAF sites are mostly found within 300bp of TSSs. This supports the idea that BEAF might associate with transcription factors to regulate associated promoters. Previous studies have shown association of BEAF with Sry-delta and Dref (Hart et al. 1999; Matsukage et al. 2007; Rhee et al. 2014; Lhoumaud et al. 2014). Domains of interest in identified transcription factors include BTB/POZ domains (rib, psq and lola) and WD40 domains (ebi). Both domains can mediate protein-protein interactions. Interestingly, some proteins had histone methylase (Lid) and acetyltransferase activity (wds, pont). Pont is also a subunit of chromatin remodeler INO80 where it catalyzes acetylation of histone residues to remodel chromatin.
Proteins associated with transcription are listed and the significant ones are highlighted. Other designations are as in Table 4.2.

### 4.4 Conclusion

Our results provide a glimpse into how BEAF might mediate chromatin domain insulator activity by implicating it in processes like maintaining chromatin architecture, remodeling chromatin and transcription regulation. The occurrence of SWI/SNF and possibly ISWI, CHD and INO80 family of chromatin remodelers strongly suggest BEAF is associated with chromatin remodeling. More than 85% of BEAF sites are found at promoters, totaling over 3000 genes. Of these genes, 85% are housekeeping and highly active genes. Hence, BEAF might act as more than just an insulator. BEAF might play a role in keeping promoters active, and one of the ways might be by creating nucleosome depleted regions (NDRs) at promoters. This possibility is supported by the presence of both subunits of the FACT histone chaperone complex in the co-IP. Moreover, physical interactions with transcription factors and BEAF suggest that BEAF might activate associated promoters by looping interactions with enhancers.
Overall, we found very interesting results that will help in finding how BEAF works inside the nucleus. In contrast to previous studies, we did not find the association of BEAF with CP190, zw5 or dMes4 (Lhoumaud et al. 2014; Liang et al. 2014; Blanton et al. 2003). Assuming these reports are correct, the reason might be our co-IP conditions, our nuclear extract preparation conditions or the stage of embryos (0-24hrs). Follow up experiments like yeast 2 hybrid and pull-down assays are necessary to test co-IP results. Then other experimental approaches will be needed to explore the relevance of the interactions. Further, the BEAF associated proteins we have identified are probably not the complete list. Indirectly or weakly associating proteins might not be seen in our co-IP, and developmental stage-specific or low abundance proteins also might not appear in the list. Regardless, the proteins we found will provide the basis for future experiments aimed at understanding molecular mechanisms of BEAF function.

4.5 References


CHAPTER 5. CONCLUSION AND FUTURE DIRECTIONS

5.1 Conclusion

Our lab is mainly focused on understanding the role of insulator proteins, particularly BEAF, and how they impact gene expression, nuclear organization and higher order chromatin organization through loop formation and TAD formation. Previous studies from our lab have found that BEAF functions to maintain polytene chromatin structure, affects chromatin dynamics (tested by position effect variegation), and has preferential binding in promoter regions of housekeeping genes (Gilbert et al. 2006; Roy et al. 2007b; Jiang et al. 2009a). Genomic data showed that 85% of 1820 BEAF peaks had their centers within 300 bp of a transcription start site (TSS). This led to the hypothesis that BEAF helps establish or maintain an architecture favorable for transcription either as a consequence of or in addition to its role in insulator function (Jiang et al. 2009a). This led to my project, which focuses on a detailed analysis of scs’ element as an insulator and a promoter (Chapters 2 and 3).

Furthermore, other BEAF binding sites have been tested to have insulator activity (Cuvier et al. 1998, 2002; Sultana et al. 2011). Genetic interactions tested by a rough eye phenotype enhancement showed BEAF interaction with various transcription factors (mostly in antennapedia complex), other insulator binding proteins and chromatin proteins (Roy et al. 2007b). Yeast two hybrid studies and pulldown assays done in our lab also found that BEAF interacts with proteins like Scr (sex comb reduced), bin1 (bicoid interaction protein 1), spnE (spindle E), polybromo, bcd (bicoid), srydelta (serendipity-delta) and Chromator (Avva et al., unpublished). Despite the identification of several BEAF partners, the scope of the BEAF interacting protein network is unknown. This motivated a search to detect interacting protein
partners of BEAF through Co-IP (co-immunoprecipitation) and proteomic MS (mass-spectrometry).

Interestingly, scs’ analysis found, for the first time, that BEAF is directly involved in promoter function. Mutating the BEAF binding site near the aurA promoter of scs’ inactivated the promoter. Our finding showed an overlap of sequence needed for insulator and promoter function. However, it also indicated that insulator and promoter function can be separated.

Additionally, a large sequence of 173bp for full insulator function indicates that there could be more proteins binding to the insulator than just BEAF. Our motif analysis for protein binding sites in the region did not find any obvious binding sites. This suggests that it is possible that proteins binding with BEAF to facilitate insulator function at scs’ might be structure dependent rather than sequence dependent, making the binding more flexible than sequence specific bindings. The scs’ case might be similar to the Fab-7 boundary region of the Bithorax complex, where the LBC (Late Binding Complex) is responsible for the boundary function. LBC recognizes a minimum sequence of >65bp (Wolle et al. 2015; Cleard et al. 2017). The only obvious protein binding motif is a GAF binding motif, but the complex consists of GAF, Mod(mdg4) and E(y)2 making a huge protein complex of >700 kDa. Additionally, the orientation independence of the M fragment suggest that the mechanism is different than CTCF (Guo et al. 2015) and might not involve directional chromatin looping.
Although studies have suggested BEAF functions largely as a chromatin domain boundary protein. Promoter assay and co-IP/MS data support the idea that BEAF might be involved in other processes related to promoter activation. Most significantly, both of my projects suggest that BEAF might play a role in remodeling chromatin and in transcriptional regulation. The occurrence of ISWI, SWI/SNF and CHD family of chromatin remodelers strongly suggest BEAF is associated with chromatin remodeling. The presence of both subunits of the FACT complex enhances the idea that BEAF might function in remodeling or removing nucleosomes and hence chromatin dynamics (Gilbert et al. 2006; Roy et al. 2007a; Maharjan et al. 2018). Interactions with the identified transcription factors combined with promoter binding by BEAF suggests the Chromatin remodeling could be associated with gene activation (Figure 5.1).

**Figure 5.1.** Proposed model of BEAF function. BEAF might interact with Transcription Factors (TFs) to bring them into promoter proximity by binding near TSS and interacting with TF. The promoter site might be made nucleosome free by the help of recruitment of histone chaperone and chromatin remodelers and nucleosomes are either slid away or evicted to make it more accessible for RNA Polymerase II binding. For simplicity the nucleosomes downstream of TF binding site in the DNA loop are not shown.

Although studies have suggested BEAF functions largely as a chromatin domain boundary protein. Promoter assay and co-IP/MS data support the idea that BEAF might be involved in other processes related to promoter activation. Most significantly, both of my projects suggest that BEAF might play a role in remodeling chromatin and in transcriptional regulation. The occurrence of ISWI, SWI/SNF and CHD family of chromatin remodelers strongly suggest BEAF is associated with chromatin remodeling. The presence of both subunits of the FACT complex enhances the idea that BEAF might function in remodeling or removing nucleosomes and hence chromatin dynamics (Gilbert et al. 2006; Roy et al. 2007a; Maharjan et al. 2018). Interactions with the identified transcription factors combined with promoter binding by BEAF suggests the Chromatin remodeling could be associated with gene activation (Figure 5.1).
5.1). This study establishes a very strong background for the further elucidation of the function of BEAF together with partner proteins.

5.2 Potential future direction

In our scs’ analysis, we tested derivatives of scs’ as dimers in adult flies. Deletion of 5’ or 3’ sequences therefore alters the spacing between the two BEAF binding sites, complicating interpretation of results. In the future, the assay can be done with these sequences mutated rather than deleted. If the sequences are important, they can be further analyzed with linker-scanning mutations. Ultimately, it is of interest to use required sequences to purify and identify proteins that work with BEAF for insulator or promoter activity.

Overall, our co-IP results will help elucidate the function of BEAF inside the nucleus. Follow up experiments like yeast 2 hybrid, bimolecular fluorescence complementation assays, rough eye genetic interaction assay and pull-down assays can give us further details about physical interactions to add confidence to our co-IP result. However, the BEAF associated proteins we have identified here might not be the complete list. In contrast to previous studies, we did not find the association of BEAF with CP190, zw5 or dMes4 (Lhoumaud et al. 2014; Liang et al. 2014). The reason might be our binding conditions or our nuclear protein source (0-24hrs). Indirectly or weakly associating proteins might not be seen in our co-IP. But pursuing the proteins leads we found should provide insight into how BEAF provides insulator function and contributes to promoter function.

5.3 References


Wolle D., Cleard F., Aoki T., Deshpande G., Schedl P., et al., 2015 Functional Requirements for Fab-7 Boundary Activity in the Bithorax Complex.
### APPENDIX. COMPLETE LIST OF SIGNIFICANT PROTEINS FROM CO-IP WITH BEAF

<table>
<thead>
<tr>
<th>Gene names</th>
<th>Uniprot Protein ID</th>
<th>Flybase ID</th>
<th>No. of quantified peptides</th>
<th>Fold Enrichment</th>
<th>P value</th>
<th>B&amp;H correction significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>His3.3A</td>
<td>sp</td>
<td>P84249</td>
<td>H33_DRO</td>
<td>FBgn0004828</td>
<td>6</td>
<td>2.13</td>
</tr>
<tr>
<td>His4 H4</td>
<td>sp</td>
<td>P84040</td>
<td>H4_DRO</td>
<td>FBgn0001200</td>
<td>22</td>
<td>2.40</td>
</tr>
<tr>
<td>CG12032</td>
<td>tr</td>
<td>Q9VZU0</td>
<td>Q9VZU0</td>
<td>FBgn0052278</td>
<td>2</td>
<td>3.08</td>
</tr>
<tr>
<td>Bap60</td>
<td>sp</td>
<td>Q9VYG2</td>
<td>BAP60</td>
<td>FBgn0025463</td>
<td>7</td>
<td>2.33</td>
</tr>
<tr>
<td>BEAF-32</td>
<td>tr</td>
<td>Q7JN06</td>
<td>Q7JN06_</td>
<td>FBgn0015602</td>
<td>10</td>
<td>6.21</td>
</tr>
<tr>
<td>BEAF-32</td>
<td>tr</td>
<td>Q95153</td>
<td>Q95153_D</td>
<td>FBgn0015602</td>
<td>49</td>
<td>5.80</td>
</tr>
<tr>
<td>MAGE</td>
<td>tr</td>
<td>Q97L4</td>
<td>Q97L4_D</td>
<td>FBgn0037481</td>
<td>2</td>
<td>10.70</td>
</tr>
<tr>
<td>bel CG9748</td>
<td>sp</td>
<td>Q9VHP0</td>
<td>DDX3_</td>
<td>FBgn0263231</td>
<td>19</td>
<td>2.48</td>
</tr>
<tr>
<td>apt apt-RC</td>
<td>tr</td>
<td>O61602</td>
<td>O61602_D</td>
<td>FBgn0015903</td>
<td>39</td>
<td>4.23</td>
</tr>
<tr>
<td>CG2982</td>
<td>tr</td>
<td>Q7K4H4</td>
<td>NO66_D</td>
<td>FBgn0266570</td>
<td>6</td>
<td>2.41</td>
</tr>
<tr>
<td>CtbP</td>
<td>tr</td>
<td>A0A0B4</td>
<td>KG99_A</td>
<td>FBgn0020496</td>
<td>9</td>
<td>2.13</td>
</tr>
<tr>
<td>NHP2</td>
<td>sp</td>
<td>Q9VU2</td>
<td>NHP2_D</td>
<td>FBgn0029148</td>
<td>3</td>
<td>3.55</td>
</tr>
<tr>
<td>CG3756</td>
<td>tr</td>
<td>Q9VMX3</td>
<td>Q9VMX</td>
<td>FBgn0031657</td>
<td>2</td>
<td>1.51</td>
</tr>
<tr>
<td>sle</td>
<td>sp</td>
<td>Q81NM3</td>
<td>SLE_DR</td>
<td>FBgn0037810</td>
<td>7</td>
<td>3.30</td>
</tr>
<tr>
<td>Sry-delta</td>
<td>sp</td>
<td>P07664</td>
<td>SRYD_D</td>
<td>FBgn003512</td>
<td>6</td>
<td>2.46</td>
</tr>
<tr>
<td>SdhA Sce-fp</td>
<td>sp</td>
<td>Q94523</td>
<td>SDHA_D</td>
<td>FBgn0261439</td>
<td>3</td>
<td>1.88</td>
</tr>
<tr>
<td>CG1316</td>
<td>tr</td>
<td>Q9VZE4</td>
<td>Q9VZE4</td>
<td>FBgn0035526</td>
<td>3</td>
<td>2.43</td>
</tr>
<tr>
<td>CG4038</td>
<td>tr</td>
<td>Q7KVQ0</td>
<td>GAR1_</td>
<td>FBgn0011824</td>
<td>9</td>
<td>1.98</td>
</tr>
<tr>
<td>mor</td>
<td>tr</td>
<td>Q9VF03</td>
<td>Q9VF03_</td>
<td>FBgn0002783</td>
<td>15</td>
<td>2.34</td>
</tr>
<tr>
<td>polybromo</td>
<td>tr</td>
<td>Q9VC36</td>
<td>Q9VC36_</td>
<td>FBgn0039227</td>
<td>3</td>
<td>1.35</td>
</tr>
<tr>
<td>rad50</td>
<td>sp</td>
<td>Q9W252</td>
<td>RAD50_</td>
<td>FBgn0034728</td>
<td>3</td>
<td>1.70</td>
</tr>
<tr>
<td>Ssrp CF5</td>
<td>sp</td>
<td>Q05344</td>
<td>SSRP1_D</td>
<td>FBgn0010278</td>
<td>209</td>
<td>4.09</td>
</tr>
<tr>
<td>Nup214</td>
<td>sp</td>
<td>Q9W1X4</td>
<td>NU214_</td>
<td>FBgn0010660</td>
<td>3</td>
<td>2.85</td>
</tr>
<tr>
<td>bl</td>
<td>tr</td>
<td>A1ZBW1</td>
<td>A1ZBWW</td>
<td>FBgn0267791</td>
<td>9</td>
<td>1.51</td>
</tr>
<tr>
<td>SmB</td>
<td>sp</td>
<td>Q05856</td>
<td>RSMB_D</td>
<td>FBgn0262601</td>
<td>5</td>
<td>2.67</td>
</tr>
<tr>
<td>UQCR-C1</td>
<td>tr</td>
<td>Q9VFF0</td>
<td>Q9VFF0_</td>
<td>FBgn0038271</td>
<td>17</td>
<td>2.80</td>
</tr>
<tr>
<td>rib CG7230</td>
<td>tr</td>
<td>Q7KF43</td>
<td>Q7KF43_</td>
<td>FBgn0003254</td>
<td>3</td>
<td>1.84</td>
</tr>
<tr>
<td>gkt Tdp1</td>
<td>tr</td>
<td>Q9VMQ4</td>
<td>TYPD1_</td>
<td>FBgn0260817</td>
<td>2</td>
<td>2.78</td>
</tr>
<tr>
<td>dre4 sp16</td>
<td>sp</td>
<td>Q8IRG6</td>
<td>SPT16_</td>
<td>FBgn0002183</td>
<td>276</td>
<td>5.07</td>
</tr>
<tr>
<td>mod(mdg4)</td>
<td>sp</td>
<td>Q86B87</td>
<td>MMD4_</td>
<td>FBgn0002781</td>
<td>15</td>
<td>3.39</td>
</tr>
<tr>
<td>ebi CG4063</td>
<td>sp</td>
<td>Q9SRJ9</td>
<td>EBI_DRO</td>
<td>FBgn0263933</td>
<td>4</td>
<td>3.62</td>
</tr>
<tr>
<td>pzg Z4</td>
<td>tr</td>
<td>Q9VP57</td>
<td>Q9VP57_</td>
<td>FBgn0259785</td>
<td>38</td>
<td>3.11</td>
</tr>
<tr>
<td>elF5B</td>
<td>tr</td>
<td>Q9VZP5</td>
<td>Q9VZP5_</td>
<td>FBgn0026259</td>
<td>6</td>
<td>1.60</td>
</tr>
<tr>
<td>Hel25E</td>
<td>sp</td>
<td>Q27268</td>
<td>DX39B_</td>
<td>FBgn0014189</td>
<td>6</td>
<td>1.66</td>
</tr>
<tr>
<td>Map60</td>
<td>tr</td>
<td>Q7K180</td>
<td>Q7K180_</td>
<td>FBgn0010342</td>
<td>2</td>
<td>7.57</td>
</tr>
<tr>
<td>HDAC6</td>
<td>tr</td>
<td>Q8IR38</td>
<td>Q8IR38_</td>
<td>FBgn0026428</td>
<td>2</td>
<td>2.42</td>
</tr>
<tr>
<td>CG12262</td>
<td>sp</td>
<td>Q9VSA3</td>
<td>ACADM</td>
<td>FBgn0035811</td>
<td>8</td>
<td>2.37</td>
</tr>
<tr>
<td>psq</td>
<td>tr</td>
<td>A4UZC9</td>
<td>A4UZC9_</td>
<td>FBgn0263102</td>
<td>2</td>
<td>2.65</td>
</tr>
<tr>
<td>SmD2</td>
<td>sp</td>
<td>Q9V10</td>
<td>SMMD2_D</td>
<td>FBgn0261789</td>
<td>3</td>
<td>1.36</td>
</tr>
<tr>
<td>Psi CG8912</td>
<td>tr</td>
<td>A1ZAK7</td>
<td>A1ZAK7_</td>
<td>FBgn0014870</td>
<td>12</td>
<td>1.63</td>
</tr>
<tr>
<td>mRNA-cap</td>
<td>tr</td>
<td>Q9YV44</td>
<td>Q9YV44_</td>
<td>FBgn0030556</td>
<td>3</td>
<td>2.24</td>
</tr>
<tr>
<td>CG1218</td>
<td>sp</td>
<td>Q9VNI3</td>
<td>U609_D</td>
<td>FBgn0037377</td>
<td>4</td>
<td>1.79</td>
</tr>
<tr>
<td>cdc2 cdk1</td>
<td>sp</td>
<td>P23572</td>
<td>CDK1_D</td>
<td>FBgn0004106</td>
<td>2</td>
<td>6.11</td>
</tr>
<tr>
<td>me31B</td>
<td>sp</td>
<td>P23128</td>
<td>DDX6_D</td>
<td>FBgn0004419</td>
<td>2</td>
<td>2.06</td>
</tr>
<tr>
<td>gfzf</td>
<td>tr</td>
<td>Q6NP69</td>
<td>Q6NP69_</td>
<td>FBgn0250732</td>
<td>3</td>
<td>2.58</td>
</tr>
<tr>
<td>Protein</td>
<td>Accession</td>
<td>FBgn</td>
<td>Ratio</td>
<td>p-value</td>
<td>Significant</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>------</td>
<td>-------</td>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>CG32409</td>
<td>tr</td>
<td>Q8I937</td>
<td>FBgn00052409</td>
<td>3</td>
<td>2.19</td>
<td>0.0041</td>
</tr>
<tr>
<td>RpS19a</td>
<td>sp</td>
<td>P39018</td>
<td>R519A_D</td>
<td>FBgn0010412</td>
<td>13</td>
<td>2.10</td>
</tr>
<tr>
<td>l(2)k14505</td>
<td>tr</td>
<td>Q9VID7</td>
<td>FBgn0021856</td>
<td>3</td>
<td>1.60</td>
<td>0.0046</td>
</tr>
<tr>
<td>Nup205</td>
<td>tr</td>
<td>Q8IQV9</td>
<td>FBgn0031078</td>
<td>2</td>
<td>1.62</td>
<td>0.0050</td>
</tr>
<tr>
<td>Sym</td>
<td>sp</td>
<td>Q8MSU4</td>
<td>SYMPK</td>
<td>FBgn0037371</td>
<td>3</td>
<td>1.63</td>
</tr>
<tr>
<td>sgg gsk3</td>
<td>sp</td>
<td>P18431</td>
<td>SGG_DR</td>
<td>FBgn0037371</td>
<td>6</td>
<td>1.65</td>
</tr>
<tr>
<td>tid CG9088</td>
<td>sp</td>
<td>Q9VMJ7</td>
<td>KDM5</td>
<td>FBgn0031759</td>
<td>20</td>
<td>1.82</td>
</tr>
<tr>
<td>elf-4B</td>
<td>tr</td>
<td>Q7PLL3</td>
<td>FBgn0020660</td>
<td>59</td>
<td>3.21</td>
<td>0.0060</td>
</tr>
<tr>
<td>pont</td>
<td>sp</td>
<td>Q9VH07</td>
<td>RUVB1</td>
<td>FBgn0040078</td>
<td>3</td>
<td>2.17</td>
</tr>
<tr>
<td>mub</td>
<td>tr</td>
<td>A4V2A4</td>
<td>A4V2A4</td>
<td>FBgn0262737</td>
<td>2</td>
<td>1.98</td>
</tr>
<tr>
<td>Adfl</td>
<td>sp</td>
<td>P05552</td>
<td>ADFL1_DR</td>
<td>FBgn0000054</td>
<td>2</td>
<td>1.85</td>
</tr>
<tr>
<td>CG1737-RA</td>
<td>tr</td>
<td>Q9VZ00</td>
<td>Q9VZ00</td>
<td>FBgn0030293</td>
<td>10</td>
<td>3.97</td>
</tr>
<tr>
<td>Nup62</td>
<td>tr</td>
<td>Q7JXF5</td>
<td>FBgn0034118</td>
<td>11</td>
<td>1.62</td>
<td>0.0078</td>
</tr>
<tr>
<td>Nup75</td>
<td>tr</td>
<td>A1YK02</td>
<td>A1YK02</td>
<td>FBgn0034310</td>
<td>2</td>
<td>2.69</td>
</tr>
<tr>
<td>CG8142</td>
<td>tr</td>
<td>Q9VX15</td>
<td>Q9VX15</td>
<td>FBgn0030871</td>
<td>4</td>
<td>1.36</td>
</tr>
<tr>
<td>elf-4a</td>
<td>sp</td>
<td>P02748</td>
<td>IF4A_DR</td>
<td>FBgn0001942</td>
<td>15</td>
<td>2.76</td>
</tr>
<tr>
<td>Ski6</td>
<td>tr</td>
<td>Q9VK19</td>
<td>Q9VK19</td>
<td>FBgn0032487</td>
<td>3</td>
<td>3.16</td>
</tr>
<tr>
<td>ClkIbeta</td>
<td>sp</td>
<td>P08182</td>
<td>CSK2B_D</td>
<td>FBgn0000259</td>
<td>5</td>
<td>2.41</td>
</tr>
<tr>
<td>lolA</td>
<td>sp</td>
<td>P42284</td>
<td>LOLA2_D</td>
<td>FBgn005630</td>
<td>2</td>
<td>2.00</td>
</tr>
<tr>
<td>Rrp40</td>
<td>tr</td>
<td>Q8IPX7</td>
<td>Q8IPX7</td>
<td>FBgn0260648</td>
<td>2</td>
<td>3.25</td>
</tr>
<tr>
<td>Bap55</td>
<td>tr</td>
<td>Q7K012</td>
<td>Q7K012</td>
<td>FBgn0025716</td>
<td>5</td>
<td>1.38</td>
</tr>
<tr>
<td>SelD PTF1</td>
<td>sp</td>
<td>Q18373</td>
<td>SPS1_DR</td>
<td>FBgn0261270</td>
<td>10</td>
<td>2.13</td>
</tr>
<tr>
<td>Kap-alpha3</td>
<td>tr</td>
<td>Q9V455</td>
<td>Q9V455</td>
<td>FBgn0027338</td>
<td>2</td>
<td>1.79</td>
</tr>
<tr>
<td>Hrb27C</td>
<td>tr</td>
<td>Q9VZ12</td>
<td>Q9VZ12</td>
<td>FBgn0028484</td>
<td>24</td>
<td>3.10</td>
</tr>
<tr>
<td>Ack</td>
<td>tr</td>
<td>Q9VZ12</td>
<td>Q9VZ12</td>
<td>FBgn0000212</td>
<td>2</td>
<td>1.60</td>
</tr>
<tr>
<td>brm</td>
<td>sp</td>
<td>P25439</td>
<td>BRM_DR</td>
<td>FBgn0000093</td>
<td>4</td>
<td>1.61</td>
</tr>
<tr>
<td>dala</td>
<td>tr</td>
<td>Q9W384</td>
<td>Q9W384</td>
<td>FBgn030993</td>
<td>9</td>
<td>1.61</td>
</tr>
<tr>
<td>CG8036</td>
<td>tr</td>
<td>Q9VHN7</td>
<td>Q9VHN7</td>
<td>FBgn0037607</td>
<td>10</td>
<td>1.91</td>
</tr>
<tr>
<td>mbo Nup88</td>
<td>sp</td>
<td>Q9GYU8</td>
<td>NUP88</td>
<td>FBgn0026207</td>
<td>17</td>
<td>2.21</td>
</tr>
<tr>
<td>row CG8092</td>
<td>tr</td>
<td>Q5U156</td>
<td>Q5U156</td>
<td>FBgn0033998</td>
<td>8</td>
<td>2.00</td>
</tr>
<tr>
<td>Klpl0A</td>
<td>sp</td>
<td>Q960Z0</td>
<td>K10A_D</td>
<td>FBgn0030268</td>
<td>4</td>
<td>1.70</td>
</tr>
<tr>
<td>Ppr19</td>
<td>tr</td>
<td>Q7KWL9</td>
<td>Q7KWL9</td>
<td>FBgn0261119</td>
<td>7</td>
<td>1.94</td>
</tr>
<tr>
<td>wds l(1)3Ad</td>
<td>sp</td>
<td>Q9V3J8</td>
<td>WDS_D</td>
<td>FBgn0040066</td>
<td>2</td>
<td>2.55</td>
</tr>
<tr>
<td>RpS28b</td>
<td>tr</td>
<td>Q9W334</td>
<td>RS28_D</td>
<td>FBgn0030136</td>
<td>4</td>
<td>2.45</td>
</tr>
<tr>
<td>Cal1</td>
<td>sp</td>
<td>Q24572</td>
<td>CAL1_D</td>
<td>FBgn0263979</td>
<td>7</td>
<td>1.42</td>
</tr>
<tr>
<td>CG7920</td>
<td>tr</td>
<td>Q9VAC1</td>
<td>Q9VAC1</td>
<td>FBgn0039737</td>
<td>17</td>
<td>1.59</td>
</tr>
<tr>
<td>Dref</td>
<td>tr</td>
<td>Q94883</td>
<td>Q94883_D</td>
<td>FBgn0015664</td>
<td>3</td>
<td>2.61</td>
</tr>
<tr>
<td>Chro Chr-</td>
<td>tr</td>
<td>Q86BS3</td>
<td>Q86BS3</td>
<td>FBgn0044324</td>
<td>33</td>
<td>2.81</td>
</tr>
<tr>
<td>Map205</td>
<td>tr</td>
<td>P23226</td>
<td>MA205_D</td>
<td>FBgn00002645</td>
<td>13</td>
<td>1.49</td>
</tr>
<tr>
<td>isoQC</td>
<td>tr</td>
<td>Q7KTY3</td>
<td>Q7KTY3</td>
<td>FBgn0036999</td>
<td>5</td>
<td>1.67</td>
</tr>
<tr>
<td>CG30122</td>
<td>tr</td>
<td>A8E6M1</td>
<td>A8E6M1</td>
<td>FBgn0050122</td>
<td>9</td>
<td>1.48</td>
</tr>
<tr>
<td>CG7928</td>
<td>tr</td>
<td>Q9VAB8</td>
<td>Q9VAB8</td>
<td>FBgn0039740</td>
<td>2</td>
<td>1.69</td>
</tr>
<tr>
<td>M1BP</td>
<td>tr</td>
<td>Q9VHM3</td>
<td>Q9VHM</td>
<td>FBgn0037621</td>
<td>11</td>
<td>1.63</td>
</tr>
<tr>
<td>CG9650</td>
<td>tr</td>
<td>A8JVL1</td>
<td>A8JVL1</td>
<td>FBgn0029939</td>
<td>27</td>
<td>3.04</td>
</tr>
<tr>
<td>EF2 Er2b</td>
<td>tr</td>
<td>P13060</td>
<td>EF2_DRO</td>
<td>FBgn0000559</td>
<td>35</td>
<td>2.38</td>
</tr>
<tr>
<td>Ufd1-like</td>
<td>sp</td>
<td>Q9VTF9</td>
<td>UFD1_D</td>
<td>FBgn0036136</td>
<td>2</td>
<td>3.02</td>
</tr>
<tr>
<td>CG3800</td>
<td>sp</td>
<td>Q8T81</td>
<td>R1Y3800_D</td>
<td>FBgn0034802</td>
<td>7</td>
<td>1.66</td>
</tr>
<tr>
<td>CG5642</td>
<td>sp</td>
<td>Q9VNU4</td>
<td>EF3L</td>
<td>FBgn0036258</td>
<td>2</td>
<td>1.66</td>
</tr>
<tr>
<td>Dis3</td>
<td>tr</td>
<td>Q9VC93</td>
<td>Q9VC93</td>
<td>FBgn0039183</td>
<td>9</td>
<td>2.14</td>
</tr>
<tr>
<td>Spindly</td>
<td>tr</td>
<td>Q9VSQ4</td>
<td>Q9VSQ4</td>
<td>FBgn0031549</td>
<td>6</td>
<td>1.62</td>
</tr>
</tbody>
</table>

List of all significant proteins from co-IP with FLAG-BEAF-32B. Fold enrichment value is the ratio of normalized abundance value of each protein with FLAG vs mock. Statistical significance
was analyzed using the Student’s t-test and corrected for multiple testing using the Benjamini–Hochberg correction (FDR < 0.1).
VITA

Mukesh Maharjan was born in Kathmandu, Nepal to Mr. Ram Chandra Maharjan and Mrs. Dhana Devi Maharjan. He completed bachelor's degree in B.Sc. Biotechnology from SANN International College, affiliated to Purbanchal University, Nepal in 2010. After 2 years of work experience as a research technician in NARC (Nepal Agricultural Research Council), Kathmandu, he started his Ph.D. program at Louisiana State University, Baton Rouge, Louisiana in January 2013 and is expected to graduate in May 2019.