Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter know, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Jessica Emilia Wahi April 13, 2012

Genome-wide Analysis of L(3)mbt and its Relationship to *Drosophila melanogaster* Insulators

By

Jessica Emilia Wahi

Victor G. Corces, Ph.D
Advisor

Department of Biology

Victor G. Corces, Ph.D
Advisor

John Lucchesi, Ph.D
Committee Member

Matthew Weinschenk, Ph.D
Committee Member

2012

Genome-wide Analysis of L(3)mbt and its Relationship to *Drosophila melanogaster* Insulators

By

Jessica Emilia Wahi

Victor G. Corces, Ph.D
Advisor

Department of Biology

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2012

Abstract

Genome-wide Analysis of L(3)mbt and its Relationship to *Drosophila melanogaster* Insulators

By

Jessica Emilia Wahi

Lethal (3) malignant brain tumor [L(3)mbt] is a conserved transcriptional regulator. Chromatin immunoprecipitation (ChIP) followed by high-throughput sequencing (ChIP-seq) analysis was done to determine the genome-wide distribution of L(3)mbt in *Drosophila melanogaster*. Results show that L(3)mbt is present at insulator sites in embryonic Kc cells; inverse levels of enrichment of L(3)mbt and CP190 were also observed where binding sites overlap. Furthermore, immunofluorescence microscopy of imaginal disc cells shows that L(3)mbt localizes to nuclear bodies adjacent to CP190 insulator bodies. Comparison of L(3)mbt ChIP-seq data between two cell lines revealed cell type-specific L(3)mbt enrichment sites, suggesting its role in differential regulation of gene expression patterns.

Genome-wide Analysis of L(3)mbt and its Relationship to *Drosophila melanogaster* Insulators

By

Jessica Emilia Wahi

Victor G. Corces, Ph.D
Advisor

Department of Biology

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2012

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

INTRODUCTION

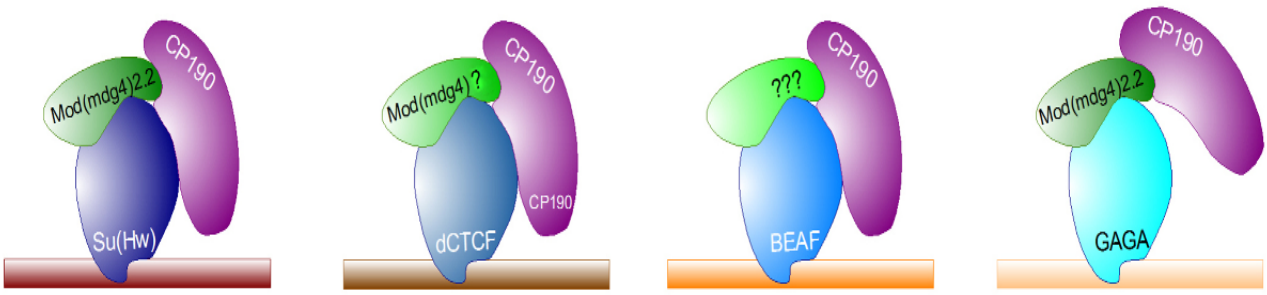
In the eukaryotic cell, transcription factors establish patterns of gene expression that are epigenetically inherited between cell generations. These patterns are maintained by alterations in the chromatin structure that include remodeling by ATP-dependent complexes, covalent histone modifications and DNA methylation. Additionally, recent research calls attention to the importance of understanding the three-dimensional organization of the eukaryotic genome in establishing proper spatial and temporal patterns of gene expression. Although the establishment and maintenance of this nuclear organization is not well characterized, insulators are arising as likely candidates in this process (Bushey et al., 2008; Phillips and Corces, 2009).

Insulators were originally discovered in *Drosophila melanogaster* and are DNA-protein complexes that mediate inter- and intra-chromosomal interactions. These interactions determine the ranges over which other regulatory elements take effect and the specific connection determines the outcome of these interactions (Phillips and Corces, 2009). For example, two insulator sites may interact to form a loop that separates an enhancer from the promoter of the gene; this would eliminate enhancer-promoter interactions and prevent gene transcription (Majumder et al., 2006). An alternative is that two insulator sites bring an enhancer in close proximity to the promoter, facilitating the transcription of the gene. In addition to the enhancer-blocking function of insulators, they have also been shown to act as boundary elements, protecting genes from position effects and preventing the spread of heterochromatin (Sun and Elgin, 1999).

CTCF is a highly conserved zinc-finger protein and has been recognized as the main insulator in vertebrates for its role in transcription regulation (Gaszner et. al., 2006). In *Drosophila,* there are five known insulators, characterized by their specific DNA-binding

proteins; additional factors that interact with the DNA-binding proteins have been characterized for some of these insulators. Such insulators include the *Drosophila* homolog of CTCF (dCTCF), Suppressor of Hairy-Wing [Su(Hw)], Boundary Element Associated Factor (BEAF), GAGA factor (GAF), and Zeste-white (Zw5); shared protein components include centrosomal protein 190 (CP190) and Mod(mdg4) (Maeda and Karch, 2007). The *mod(mdg4)* gene encodes around 30 different isoforms that arise by alternative cis- and trans-splicing (Buchner et. al., 2000). Of these different isoforms, Mod(mdg4)2.2 is the one variant that has been shown to physically interact with the Su(Hw) insulator. CP190 contains a BTB domain as well as three zinc fingers and it physically associates with both Su(Hw) and Mod(mdg4)2.2 (Pai et. al., 2004). CP190 and Mod(mdg4)2.2 interact with a variety of DNA binding proteins and it is thought that these two proteins mediate inter-insulator interactions by serving as a bridge to bring together distant insulator sites.

CP190 is necessary for insulator activity and is thought to be recruited by Su(Hw), dCTCF, and BEAF (Bushey et. al, 2009; Gurudatta and Corces, 2009). Consequently, Su(Hw), dCTCF, and BEAF may use similar mechanisms to regulate their insulator function. On the contrary, it does not seem that GAF directly interacts with CP190, although it has been shown to interact with Mod(mdg4)2.2 (Melnikova et. al., 2004). Since Mod(mdg4)2.2 can then interact with CP190, this suggests that GAF insulators may have a distinct mechanism from the other three (Figure 1). Furthermore, these observations suggest that insulator activity can be controlled through regulating the binding of Su(Hw), dCTCF or BEAF to DNA, or through regulating the interactions between these proteins and CP190.



Su(Hw) insulator dCTCF insulator BEAF insulator GAGA insulator

Figure 1: Diagram showing the structure of different *Drosophila* and vertebrate insulators. Each *Drosophila* insulator subclass contains a different binding protein that may define the specific function of the corresponding subclass. All insulators share the common protein CP190, although the role of this protein in the function of the GAGA insulator has not been demonstrated experimentally. In addition, all subclasses may also have one Mod(mdg4) isoform. The Su(Hw) insulator contains Mod(mdg4)2.2. The dCTCF and BEAF insulators lack this isoform but contain a different variant of Mod(mdg4). GAGA has been shown to interact with Mod(mdg4)2.2.

Due to the fact that there are several insulator subclasses, each with distinct DNA-binding proteins, but with shared components, calls into question whether each subclass exerts a similar role in regulating gene expression. Throughout various stages of development, many insulator complexes are common at certain sites in different cell types. However, different insulator subclasses may have specialized functions (Gurudatta and Corces, 2009). Some insulators display distinct localization towards genes involved in different cellular processes, genomic landmarks and levels of gene expression. These findings foster the idea that different insulators have different roles in chromatin organization and gene regulation (Bushey et. al., 2009). Interestingly, the conserved transcriptional regulator Lethal (3) malignant brain tumor [L(3)mbt] seems to play a role at chromatin insulator sites (Richter et. al., 2011).

Homozygous mutations in *l(3)mbt* have been shown to cause malignant growth in *Drosophila* larval brain, identifying *l(3)mbt* as a tumor suppressor gene. *l(3)mbt* encodes a protein of 1477 amino acids with three repeats of the 100 amino acid long Malignant Brain

Tumor (MBT) domain (Wismar, 1995). The MBT domain is a protein module that recognizes mono- and di-methylated lysines at different positions on histone H3 and H4 tails (Bonasio, 2010). The human homolog of L(3)mbt is a member of the Polycomb group (PcG) proteins. PcG proteins are transcriptional regulators required for the repression of developmental control genes by modifying chromatin of target genes (Schwartz et. al., 2007). Previous studies have suggested a role for L(3)mbt in chromatin compaction, (Trojer et. al, 2007) binding to insulator elements in larval brain, (Richter et. al, 2011) and colocalizing in the genome with CP190.

Together these results suggest that L(3)mbt may be involved as a regulator of insulator function. Therefore, this study investigates the *in vivo* binding profile of L(3)mbt in *Drosophila* by chromatin immunoprecipitation (ChIP) and sequencing analysis (ChIP-seq) in embryonic Kc-167 cells to better understand the relationship between chromatin insulators and L(3)mbt. This analysis demonstrates that the binding pattern of L(3)mbt correlates with known insulator-associated proteins dCTCF, BEAF, Su(Hw) and CP190. Furthermore, comparison of the genome-wide localization of L(3)mbt in two different cell lines reveals cell line-specific localization of L(3)mbt. These findings suggest that L(3)mbt may play different roles in regulating gene expression levels in particular cell types.

MATERIALS AND METHODS

Antibodies

For L(3)mbt antibody production, an N-terminal MBP-tagged protein consisting of amino acids 1243-1477 of the *l(3)mbt* gene was purified from *Escherichia coli* and was used to immunize guinea pigs and rabbits using standard procedures. Rabbit α- Centrosomal Protein 190 (CP190) was prepared as described (Pai et al., 2004).

Chromatin immunoprecipitation

ChIP was carried out using chromatin extracts from 4 x 10^7 *Drosophila* Kc cells. The chromatin was cross-linked with 1% formaldehyde for 10 min at room temperature. Nuclear lysates were sonicated 16 times for 10 s intervals, generating DNA fragments between 200-1000 bp. All samples were pre-cleared with GE Healthcare Protein A Sepharose Fast Flow beads. ChIP was then performed with 25μL of guinea pig α-L(3)mbt antibody overnight at 4°C. Next, the pull-down used 50 μL beads/sample of Protein A Sepharose beads after incubation for 2 h at 4°C. The sample was washed with low salt, high salt and LiCl buffers. Subsequently, the DNA was eluted using IP elution buffer (1% SDS, 0.1M NaHCO3), and crosslinks reversed at 65°C overnight with proteinase K. ChIP DNA was then isolated by phenol:chloroform extraction and DNA precipitation.

The sequencing library was generated by preparing ChIP DNA for adaptor ligation by end repair with End-it DNA End Repair Kit and addition of "A" base to 3' ends. Ilumina adaptors were titrated according to the prepared DNA ChIP sample concentration and ligated with T4 ligase. The ligated ChIP samples were PCR-amplified using Illumina primers and Physion DNA polymerase and size selected for 200-300 bp by gel extraction. ChIP libraries

were sequenced at the HudsonAlpha Institute for Biotechnology using an Illumina HiSeq2000 sequencer.

ChIP-seq mapping and data analysis

Sequences were mapped to the *Drosophila melanogaster* genome (FlyBase 5.27) using Bowtie 0.12.5 using default settings (Langmead et al., 2009). MACS software was then used as a peakfinder for the 36 million reads in the data sets using equal numbers of unique reads for input and ChIP samples and a p value cutoff of 1×10^{-10} (Zhang et al., 2008). ChIP-seq data for other insulator-associated proteins were obtained from previously published data (Wood et al., 2011). Genome distribution analysis was performed with Galaxy (http://main.g2.bx.psu.edu). To determine the position of a peak relative to genomic elements, the summit was used as its location in the genome, and gene annotations were obtained from the UCSC database. For Venn diagram counts, two or more regions that overlap with at least one base were merged and defined as a common region.

Motif sequence identification

The 2,919 identified L(3)mbt peaks were expanded to 301 bp centered on the peak summit, and were analyzed for enriched motifs. *De novo* motif analysis was conducted using the motif discovery program MEME (multiple em for motif elicitation) (Bailey et. al., 2009) with parameters: distribution of zero or one occurrence per sequence model, allowing sites on + and – strands, 6 as minimum and 30 as maximum motif width.

Gene expression analysis

Gene expression analysis of both Kc cells and larval brain tissues were based on previously published data (Bushey et. al., 2009 and Janic, et. al, 2010). The genes were ranked

by expression signal and then divided equally into three distinct groups based upon the intensity of the expression signal.

Gene ontology analysis

Gene ontology analysis for genes associated with L(3)mbt stable sites, Kc-specific sites, and brain-specific sites was performed with Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov). Flybase IDs were used to determine statistically enriched biological process categories on the basis of a background list of all annotated genes in the *Drosophila* genome.

Immunofluorescence

Immunofluorescence analysis of imaginal disc tissue was done as described previously (Ivaldi et al., 2007). Briefly, third-instar larvae were dissected in PBS and fixed for 15–20 min in 4% paraformaldehyde in PBS with 0.1% Triton X-100. Tissues were stained with primary antibodies in antibody dilution buffer (1 × PBS, 0.1% Tween20, 1% BSA) overnight at 4°C (1:1000 rabbit α-CP190, 1:250 guinea pig α-L(3)mbt], mounted in Vectashield containing DAPI (Vector Laboratories).

RESULTS

L(3)mbt is present at insulator sites

In order to elucidate sites of L(3)mbt protein occupancy on a genome-wide scale, a ChIP-seq assay was performed with an antibody against L(3)mbt. This analysis was carried out in Kc-167 cells, a *Drosophila* cell line originating from a heterogeneous embryonic culture. Sequences were mapped to the *Drosophila* genome using Bowtie (Langmead et al., 2009) and peaks were called with MACS (Zhang et al., 2008). This analysis found 2,919 regions enriched for L(3)mbt binding. Representative ChIP-seq data over a 15.3 kb region of Chromosome 2R is illustrated in Figure 2. To understand how L(3)mbt relates to various insulator proteins, the genome-wide distribution of L(3)mbt was compared to that of insulator proteins dCTCF, BEAF and Su(Hw), from previously published data (Wood et. al., 2011).

From the results, it is evident that L(3)mbt binds at narrow sites in the genome and that its distribution is not random. Furthermore, results demonstrate an overlap in the binding patterns of insulator proteins with L(3)mbt; 65% of the L(3)mbt peaks found in the genome overlap with the binding of dCTCF, BEAF and Su(Hw) (Figure 3). A more stringent data set, where only the peaks of dCTCF, BEAF and Su(Hw) containing their respective consensus sequence were considered, shows 49% of L(3)mbt sites overlap with the binding sites of these insulator proteins.

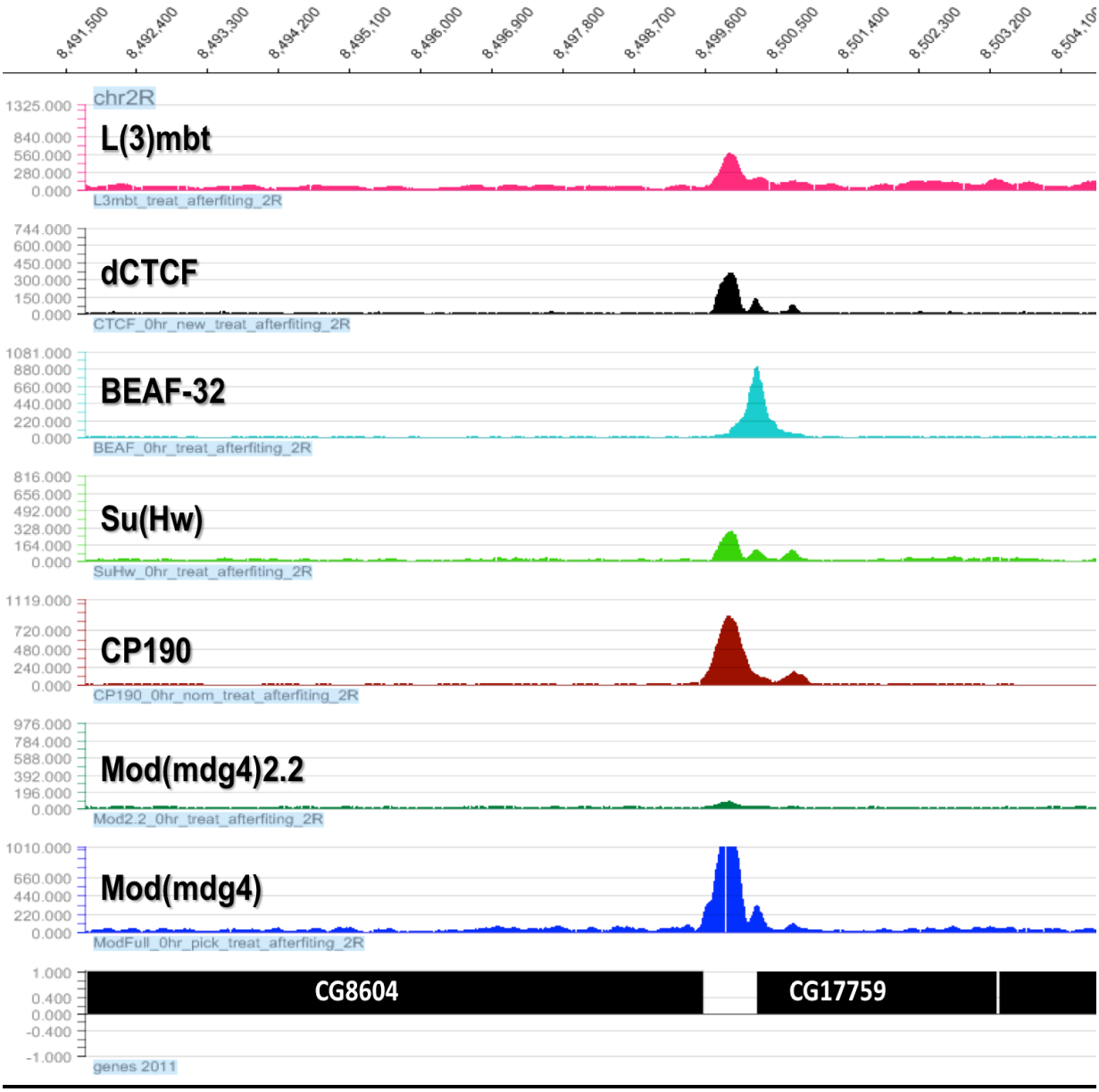


Figure 2. Genome-wide localization of L(3)mbt. Representative Chip-seq tracks for L(3)mbt, dCTCF, BEAF, Su(Hw), CP190, Mod(mdg4) and Mod(mdg4)2.2 over a 15.3 kb region of Chromosome 2R. The X-axis represents the linear genome and the Y-axis measures the fold enrichment. The data show a strong overlap of L(3)mbt binding sites with other chromatin insulator proteins.

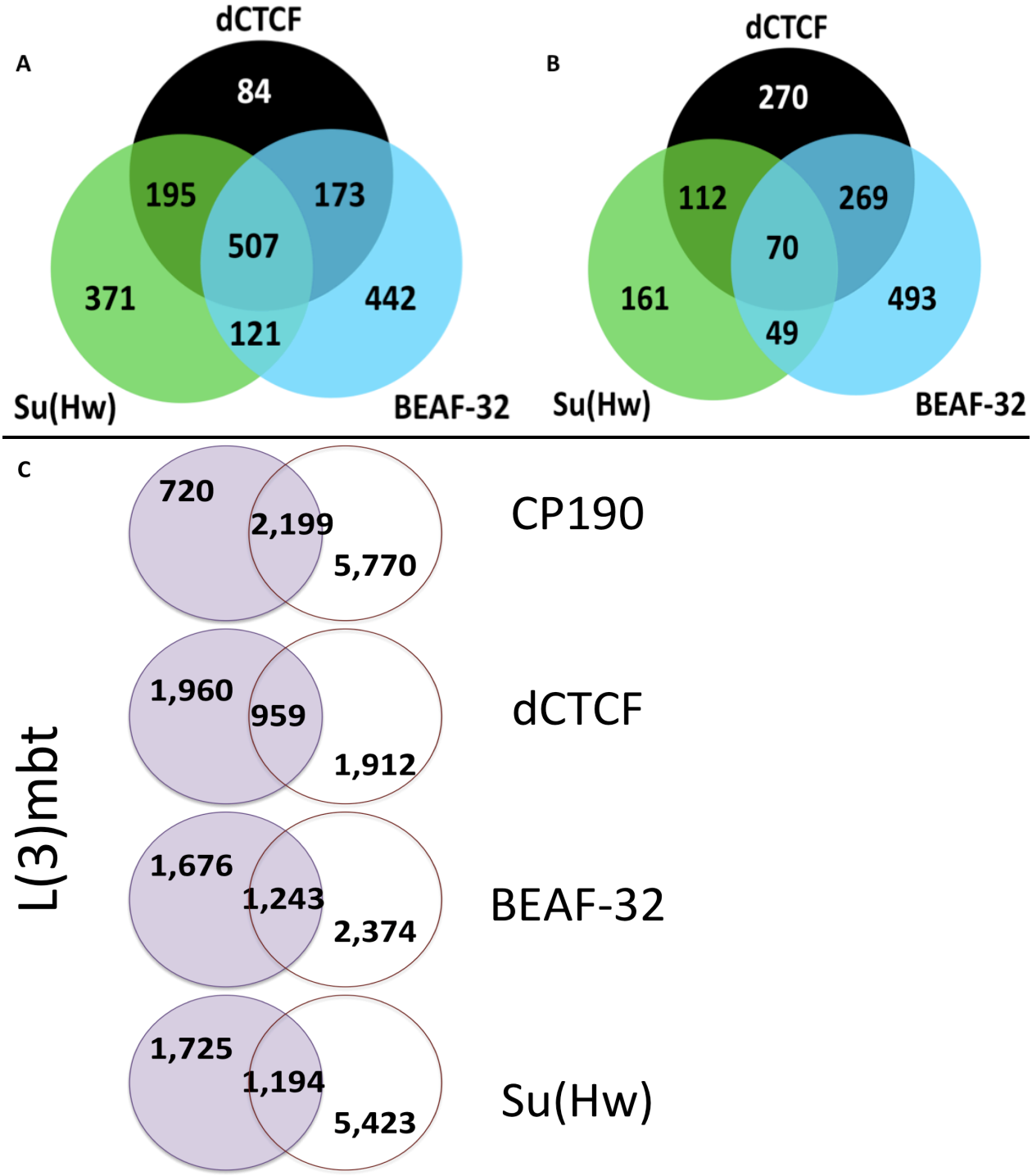


Figure 3. Distribution of L(3)mbt sites with insulators. (A) Venn diagram shows that 1,893 of the total 2,919 L(3)mbt sites overlap with the binding sites of chromatin insulators dCTCF, Su(Hw) and BEAF-32. (B) Venn diagram of L(3)mbt distribution with insulator proteins selected for consensus sequence shows that 1,424 of the total 2,919 L(3)mbt sites overlap. (C) Diagram indicates the overlap of regions bound by L(3)mbt and regions bound by insulator-associated proteins.

L(3)mbt sequence motif corresponds to CP190 enriched motif

The binding specificity of L(3)mbt was analyzed by searching for DNA motifs enriched among L(3)mbt binding sites. *De novo* identified sequence motifs significantly enriched for L(3)mbt correspond to the consensus sequences for the DNA binding proteins BEAF-32, dCTCF and Su(Hw), as well as CP190 enriched sites. CP190 itself cannot directly bind DNA in a sequence-specific manner and is thought to be recruited by BEAF-32, dCTCF and Su(Hw) in order to establish sequence-specific localization on chromatin (Pai et. al., 2004; Bushey et. al., 2009). The DNA motif most enriched for L(3)mbt binding is illustrated in Figure 4A; Figure 4B illustrates a DNA motif enriched for CP190 binding, where CP190 is present in 92% of these sites. There is a striking similarity between the L(3)mbt enriched motif and the CP190 enriched motif. This similarity suggests some possible relationship between the recruitment of CP190 and the recruitment of L(3)mbt to an insulator site.

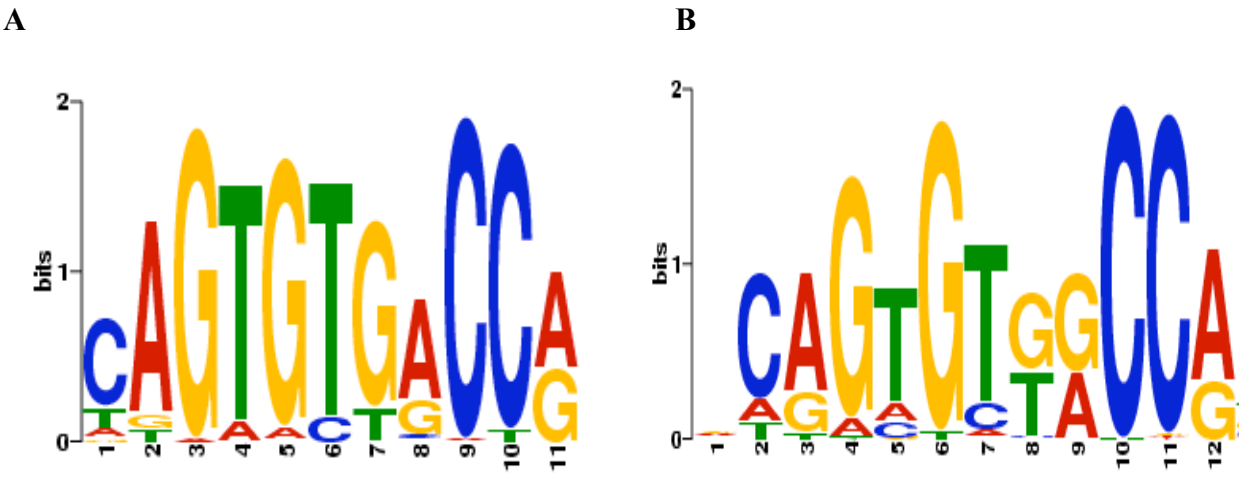


Figure 4. L(3)mbt motif. (A) Sequence analysis indicated *de novo* motif enriched in L(3)mbt-bound regions, depicted as a sequence logo. (B) CP190 enriched DNA motif.

L(3)mbt localizes to nuclear bodies near CP190 bodies

It has been shown previously that insulator function requires the formation of chromatin loops by the interaction among multiple insulators (Byrd and Corces, 2003). These interactions can be visualized in the form of insulator bodies, sites in the nucleus composed of large aggregates of insulators. Specifically, CP190 has been shown to be essential for the formation of insulator bodies (Pai et. al., 2004). The distribution of L(3)mbt in diploid nuclei of imaginal disc cells was analyzed to better ascertain its role and its relation to insulator proteins.

Immunofluorescence microscopy of CP190 and L(3)mbt suggests that L(3)mbt regions localize to nuclear bodies near CP190 insulator bodies. However, L(3)mbt regions do not overlap or colocalize to the same body where CP190 is present (Figure 5). These results suggest that L(3)mbt bodies are distinct from CP190 bodies, possibly suggesting different biological roles for L(3)mbt and CP190 and/or the differential regulation of these two proteins.

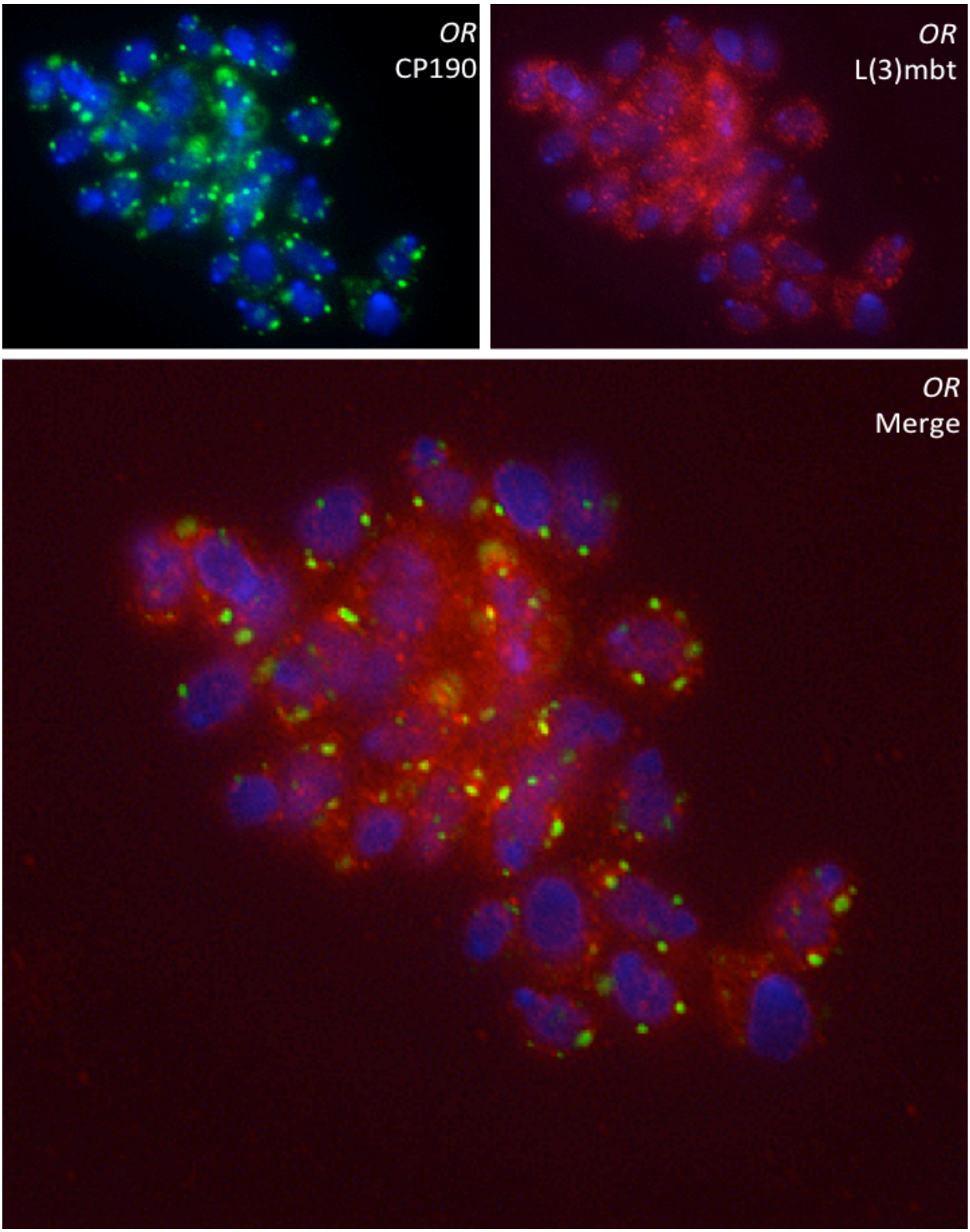


Figure 5. L(3)mbt localizes to nuclear bodies near CP190 insulator bodies. Immunolocalization of CP190 (green) and L(3)mbt (red) in diploid nuclei of wild type cells; DNA is stained with DAPI (blue) in all images.

Inverse enrichment levels of L(3)mbt and CP190 where binding sites overlap

Further analysis of the relationship between L(3)mbt and CP190 led to investigating the sites in the genome where their binding patterns overlap. Careful analysis demonstrates an inverse relationship in the enrichment levels of these two proteins: a high enrichment for L(3)mbt correlates with low enrichment of CP190. Computational analysis involved anchoring all the L(3)mbt sites and extracting the enrichment values for CP190 over a 4 Kb window (Figure 6). Two separate clusters were established: cluster 1 represents stronger signal intensity for CP190, whereas cluster 2 represents stronger signal intensity for L(3)mbt. Interestingly, in approximately 80% of the sites where these two proteins are both present, the average enrichment of L(3)mbt is very high, whereas there is a weak enrichment for CP190. Conversely, in the remainder of the sites where the two proteins are both located, there is an average higher CP190 enrichment level and less enrichment for L(3)mbt. These results suggest the possibility that there is a functional difference between the recruitment of CP190 and the recruitment of L(3)mbt to chromatin.

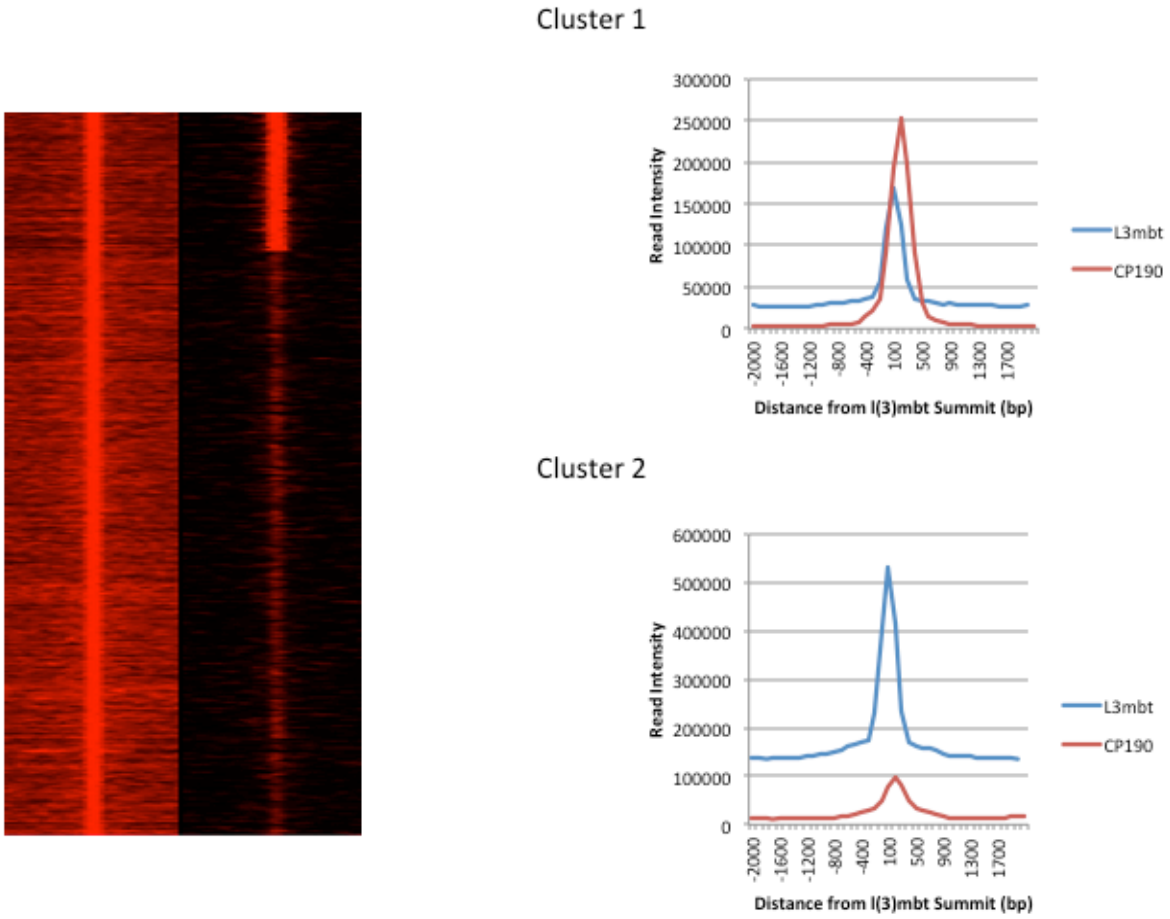


Figure 6. L(3)mbt enrichment correlates with lower CP190 levels. (A) Heat map where L(3)mbt signal intensity is shown in the left column and CP190 signal intensity is shown in the right column. (B) Graph where the X-axis is a 4 Kb window of the linear distance centered on L(3)mbt peaks. The Y-axis expresses the read intensity of protein enrichment.

Functional analysis of L(3)mbt

Previous studies have demonstrated cell line-specific localization sites for CP190, dCTCF, Su(Hw), and BEAF, indicating multiple levels at which insulators can be regulated to affect gene expression (Bushey et. al., 2009). To test this possibility for L(3)mbt, the L(3)mbt ChIP-seq data using embryonic Kc cells was compared with published L(3)mbt ChIP-seq data from larval brain-disc tissue (Richter et. al, 2011).

Comparison between the two cell lines revealed that a fraction of the sites are constant; 1,255 L(3)mbt sites were present in both Kc cells and in larval brain. At a 1% FDR, 58% of L(3)mbt sites in Kc cells and 63% of l(3)mbt sites in larval brian were found to be cell type-specific. Since a subset of the localization sites was found to be cell type-specific for L(3)mbt, this may generate changes in chromatin organization necessary to establish different gene expression profiles during cell differentiation.

To further investigate the significance of the L(3)mbt localization patterns, a gene ontology analysis based on biological process was done. Using the Database for Annotation, Visualization and Integrated Discovery (DAVID), genes within 5 Kb either upstream or downstream of L(3)mbt sites were identified. The data presented in Tables 1-3 show much similarity in the biological processes of genes near L(3)mbt tissue-specific sites, although this does not supply any additional information regarding the function exerted by L(3)mbt on these processes.

Table 1. Gene ontology analysis for biological processes of genes near L(3)mbt stable sites

	P-value:
Multicellular organismal development	*2.0E-17*
Anatomical structure development	*2.3E-14*
Cellular developmental process	*1.1E-10*
Pattern specification process	*1.2E-10*
Cell cycle	*2.9E-10*
Anatomical structure morphogenesis	*5.1E-10*
Regulation of biological process	*1.6E-9*
Regulation of cellular process	*3.8E-9*

Table 2. Gene ontology analysis for biological processes of genes near L(3)mbt Kc-specific sites

	P-value:
Regulation of biological process	*1.0E-13*
Regulation of metabolic process	*2.7E-12*
Regulation of cellular process	*3.2E-12*
Organelle organization	*7.5E-11*
Cellular metabolic process	*3.6E-9*
Cellular component assembly	*1.9E-6*
Nitrogen compound metabolic process	*3.4E-6*
Regulation of biological quality	*2.1E-5*

Table 3. Gene ontology analysis for biological processes of genes near L(3)mbt brain-specific sites

	P-value:
Regulation of biological process	*1.3E-17*
Regulation of cellular process	*1.6E-16*
Cellular developmental process	*7.3E-12*
Anatomical structure development	*3.3E-11*
Multicellular organismal development	*2.6E-9*
Regulation of developmental process	*4.7E-9*
Anatomical structure morphogenesis	*1.4E-8*
Post-embryonic development	*8.1E-7*

The differential distribution of L(3)mbt in different cell lineages led to explore the relationship between L(3)mbt sites and levels of gene expression. Interactions among insulators in the genome may give rise to a specific chromatin organization pattern that can be representative of the transcriptional status of the cell. Since gene expression patterns are cell lineage-specific, comparison of the genome-wide distribution of L(3)mbt in Kc and larval brain may give insight into the regulation of L(3)mbt function and possible mechanisms employed during cell differentiation to establish different patterns of gene expression.

Published gene expression levels from both Kc cells and wild-type larval brain tissue (Bushey et. al., 2009; Janic et. al., 2010) were used; all the genes were divided into three groups based upon their expression signal: high, medium and low. The first site analyzed was an L(3)mbt site specific to Kc cells and its relationship to the closest gene (Figure 7). In this case, *larval serum protein 1 β* is a gene encoding a protein reported to act as a nutrient reservoir. This function is to be expected, as table 2 illustrates that Kc-specific sites have a significant association with genes regulating metabolic processes. Interestingly, the transcript of this gene is only moderately expressed in Kc cells, whereas it is highly expressed in larval brain where this L(3)mbt site is absent.

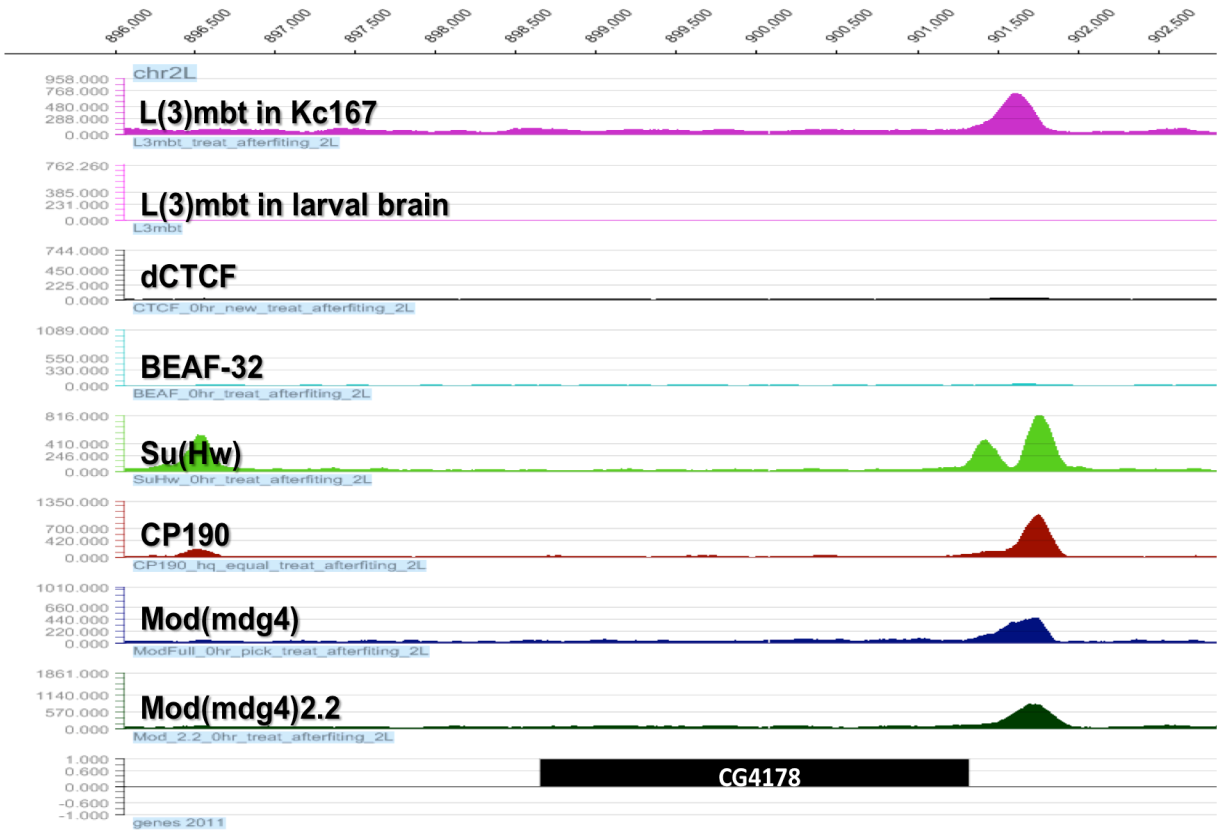


Figure 7. Analysis of L(3)mbt Kc-specific site. Representative ChIP-seq data of L(3)mbt and other insulators. At this locus, L(3)mbt enrichment is found only in Kc cells and absent from the larval brain. The closest gene is labeled with its Flybase ID CG4178, just downstream of the insulator sites.

The next site analyzed was a locus where an L(3)mbt site was found in the larval brain, but was absent from the Kc data. Two brain-specific L(3)mbt sites are illustrated in Figure 8. Here, the gene *organic anion transporting polypeptide 30B* (Flybase ID CG3811) is found flanked between two insulator sites. The molecular function of its product is involved in the biological process of transporting anions. Consistent with the previous observations, this gene is expressed at low levels in the larval brain, but at moderate levels in Kc cells. Upstream of the second insulator site is the gene referred to in Flybase by CG17855. Although the molecular function of the product of this protein coding gene is unknown, its expression is minimal in larval brain and moderately expressed in Kc cells. It is an interesting finding that particular L(3)mbt tissue-specific peaks seem to correlate with genes with lower expression levels in those tissues. Although a more extensive analysis of L(3)mbt and gene expression patterns is needed to prove these findings significant, these results suggest that L(3)mbt may contribute to the differential gene expression between Kc and brain disc.

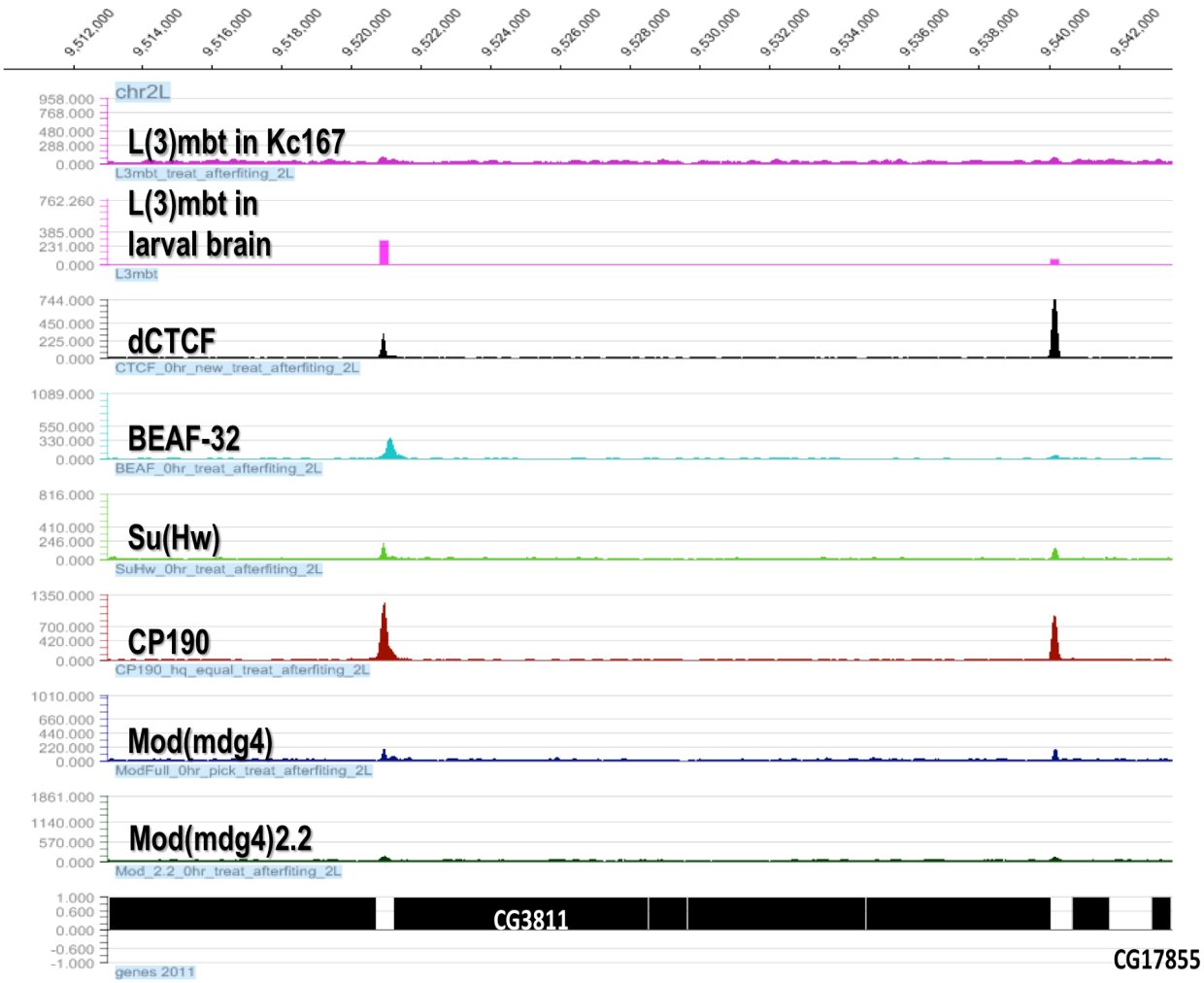


Figure 8. Analysis of L(3)mbt brain-specific site. Representative ChIP-seq data of L(3)mbt and other insulators. At this locus, L(3)mbt enrichment is found only in larval brain and absent from the Kc sample. The closest genes to each L(3)mbt site are labeled with their Flybase ID.

DISCUSSION

Chromatin insulators are thought to play a role in chromatin organization and in the regulation of gene expression by mediating inter- and intra-chromosomal interactions. In *Drosophila*, insulators Su(Hw), dCTCF, and BEAF have been shown to have different distribution patterns with respect to gene locations and expression levels, suggesting diverse roles for these three subclasses of insulators. This study uses chromatin immunoprecipitation and sequencing analysis (ChIP-seq) to analyze the genome-wide localization of the transcriptional regulator L(3)mbt in Kc cells. Results from this analysis agree with previous work demonstrating that L(3)mbt binds at insulator sites.

Analysis of the subcellular localization of L(3)mbt demonstrates that this protein aggregates in nuclear bodies that are distinct, but adjacent to CP190 insulator bodies. Chromatin is organized in structures that localize to specific nuclear domains and insulator bodies are just one example of these domains. Another example is PcG bodies, where PcG proteins are concentrated in nuclear foci and colocalize with repressed *Hox* genes in *Drosophila* (Pirrota and Li, 2011). If L(3)mbt functions as a PcG, like its human homolog, then a model can be proposed where insulators establish a particular organization in the genome and the localization of PcG proteins is dictated by insulator proteins. An interesting experiment in the future would be to investigate the effect of knocking down CP190 and other insulator proteins on the localization of L(3)mbt to nuclear bodies.

Computational analysis shows inverse enrichment levels at sites where L(3)mbt and CP190 are both present. This pattern observed between CP190 and L(3)mbt suggests a functional difference between the recruitment of CP190 and the recruitment of L(3)mbt to chromatin. Comparison of the genome-wide localization of L(3)mbt in different cell types

demonstrates that approximately 60% of L(3)mbt sites are cell type-specific. Furthermore, the transcriptional analysis in this study suggests that L(3)mbt cell type-specific sites associate with genes that are repressed or transcribed at low levels. These results support a cell-specific model in which L(3)mbt is recruited to certain loci, establishing gene expression patterns. However, more extensive functional analyses need to be done to investigate the role of L(3)mbt in establishing gene expression patterns.

REFERENCES

Bantignies, et. al. Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell. 2011 Jan 21;144(2):214-26.

Bonasio, R., Lecona, E. & Reinberg, D. MBT domain proteins in development and disease. Semin Cell Dev Biol. 2010 Apr;21(2):221-30.

Büchner K, Roth P, Schotta G, Krauss V, Saumweber H, Reuter G, Dorn R. Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in Drosophila. Genetics. 2000 May;155(1):141-57.

Boccuni et. al. The human L(3)MBT polycomb group protein is a transcriptional repressor and interacts physically and functionally with Tel (ETV6). J Biol Chem. 2003 Apr 25; 278(17):15412-20.

Bushey, A.M., Dorman, E.R., and Corces, V.G.. Chromatin insulators: regulatory mechanisms and epigenetic inheritance. Mol Cell. 2008 Oct 10;32(1):1-9.

Bushey, A. M., Ramos, E. and Corces, V. G.. Three subclasses of a Drosophila insulator show distinct and cell type- genomic distributions. Genes Dev. 2009 Jun 1;23(11):1338-50.

Capelson, M., and Corces, V.G..The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. Mol Cell. 2005 Oct 7;20(1):105-16.

Gaszner, M., and Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet. 2006 Sep;7(9):703-13.

Gerasimova, T.I., Lei, E.P., Bushey, A.M., and Corces, V.G.. Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. Mol Cell. 2007 Dec 14;28(5):761-72.

M.S. Ivaldi, C.S. Karam, V.G. Corces. Phosphorylation of histone H3 at Ser10 facilitates RNA polymerase II release from promoter-proximal pausing in Drosophila. Genes Dev. 2007 Nov 1;21(21):2818-31.

Ana Janic, Leire Mendizabal, Salud Llamazares, David Rossell, and Cayetano Gonzalez. Ectopic Expression of Germline Genes Drives Malignant Brain Tumor Growth in Drosophila. Science. 2010 Dec 24;330(6012):1824-7.

Langmead, B., Trapnell, Pop, Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

Lei, E.P., and Corces, V.G.. RNA interference machinery influences the nuclear organization of a chromatin insulator. Nat Genet. 2006 Aug;38(8):936-41.

Maeda, R.K., and Karch, F.. Making connections: boundaries and insulators in Drosophila. Curr Opin Genet Dev. 2007 Oct;17(5): 394-399.

Majumder, P., Gomez, J.A., and Boss, J.M.. The human major histocompatibility complex class II HLA-DRB1 and HLA-DQA1 genes are separated by a CTCF-binding enhancer-blocking element. J Biol Chem. 2006 Jul 7;281(27): 18435-18443.

Melnikova, L. et. al. Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. Proc Natl Acad Sci USA. 2004 Oct 12;101(41):14806-11.

Pai, C.Y., Lei, E.P., Ghosh, D., and Corces, V.G.. The centrosomal protein CP190 is a component of the gypsy chromatin insulator. Mol Cell. 2004 Dec 3; 16(5): 737-748.

Phillips, J.E., and Corces, V.G.. CTCF: master weaver of the genome. Cell. 2009 Jun 26; *137*(7): 1194-1211.

Pirrotta, Vincenzo and Li, Hua-Bing. A view of nuclear Polycomb bodies. Curr Opin Genet Dev. 2011 Dec 16.

Richter, C., Oktaba, K., Steinmann, J. Müller, J. and Knoblich, J.A.. The tumor suppressor L(3)mbt inhibits neuroepithelial proliferation and acts on insulator elements. Nat Cell Biol. 2011 August 21;13(9): 1029-1039.

Schwartz, Yuri et. Al. Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet. 2007 Jan;8(1): 9-22.

Sun, FL and Elgin, SC. Putting boundaries on silence. Cell. 1999 Nov 24;99(5):459-62.

Timothy L. Bailey, et. al. "MEME SUITE: tools for motif discovery and searching", *Nucleic Acids Research*, 37:W202-W208, 2009.

Trojer, P. et al. L3MBTL1, a histone-methylation-dependent chromatin lock. Cell 2007 Jun 1;129(5): 915–928.

Morgane Thomas-Chollier, Carl Herrmann, Matthieu Defrance, Olivier Sand, Denis Thieffry, and Jacques van Helden RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets.
Nucleic Acids Res. 2012 Feb 1;40(4):e31.

Wallace JA, Felsenfeld G. We gather together: Insulators and genome organization. Curr Opin Genet Dev 2007 17: 400–407.

Wendt, K.S., and Peters, J.M.. How cohesin and CTCF cooperate in regulating gene expression. Chromosome Res 2009; *17*(2): 201-214.

Wismar, T. Löffler, N. Habtemichael, O. Vef, M. Geissen, R. Zirwes *et al.* The *Drosophila melanogaster* tumor suppressor gene lethal(3)malignant brain tumor encodes a proline-rich protein with a novel zinc finger. Mech Dev. 1995 Sep;53(1):141–154

Wood AM, Van Bortle K, Ramos E, Takenaka N, Rohrbaugh M, Jones BC, Jones KC, Corces VG. Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. Mol Cell. 2011 Oct 7;44(1):29-38.

Yang and Corces. Chromatin insulators: a role in nuclear organization and gene expression. Adv Cancer Res. 2011; 110:43-76.

Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, X.S. Liu. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9(9):R137