M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity

Indira Bag[1, 2], Shue Chen[1, 2, +], Leah F. Rosin[1, 2, +], Yang Chen[1, 2], Chen-Yu Liu[2], Guo-Yun Yu[2], and Elissa P. Lei[1, 2, *]

[1]Nuclear Organization and Gene Expression Section, [2]Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA

[+]Equal contribution

[*]Corresponding author:
E-mail: leielissa@niddk.nih.gov.

ABSTRACT

Genome organization is driven by forces affecting transcriptional state, but the relationship between transcription and genome architecture remains unclear. Here, we identified the *Drosophila* transcription factor Motif 1 Binding Protein (M1BP) in physical association with the *gypsy* chromatin insulator core complex, including the universal insulator protein CP190. M1BP is required for enhancer-blocking and barrier activities of the *gypsy* insulator as well as its proper nuclear localization. Genome-wide, M1BP specifically colocalizes with CP190 at Motif 1-containing promoters, which are enriched at topologically associating domain (TAD) borders. M1BP is required for CP190 chromatin binding at many shared sites, and CP190 also affects M1BP chromatin association. Both factors are required for Motif 1-dependent gene expression and transcription near TAD borders genome-wide. Finally, loss of M1BP alters local genome compaction. Our results reveal physical and functional interaction between CP190 and M1BP to activate transcription at TAD borders and mediate chromatin insulator-dependent genome organization.

KEYWORDS

Drosophila, Chromatin, nuclear organization, Insulator, *gypsy*, M1BP, CP190, Transcription, Promoter, TAD border, Oligopaint.

INTRODUCTION

In eukaryotic cells, the three-dimensional organization of the genome plays a critical role in achieving proper spatial and temporal patterns of gene expression during development. Chromatin insulators are DNA-protein complexes involved in the establishment, maintenance, and regulation of nuclear organization to modulate gene expression [1, 2]. Insulators regulate interactions between cis-regulatory elements such as enhancers and promoters and demarcate silent and active chromatin regions to ensure their proper regulation. They can inhibit the interaction between an enhancer and a promoter when positioned between the two elements and can act as a barrier to stop repressive chromatin from spreading over active genes. Furthermore, chromatin insulators can promote intra- and inter-chromosomal looping to control topology of the genome. Certain insulator proteins are highly enriched at the self-interacting boundaries of topologically associating domains (TADs) throughout the genome. In mammals, only a single insulator protein, CCCTC-binding Factor (CTCF), has thus far been identified, and CTCF indeed is enriched at TAD borders and is required for TAD formation. In contrast, *D. melanogaster* CTCF is not particularly enriched at TAD borders, and a recent study suggests that CTCF may not play any role in TAD formation in flies [3]. In fact, *Drosophila* harbors a variety of insulator protein complexes, all of which contain the protein Centrosomal protein 190 (CP190). Finally, CP190 is highly enriched at TAD borders, suggesting a central role in TAD formation.

The *Drosophila gypsy* insulator, also known as the Suppressor of Hairy wing [Su(Hw)] insulator, was the first characterized CP190-containing insulator complex. The zinc-finger DNA-binding protein Su(Hw) provides binding specificity of the complex, and both CP190 and the Modifier of mdg4 [Mod(mdg4)] 67.2 kDa isoform [Mod(mdg4)67.2] contain an N-terminal Broad-Complex, Tramtrack and Bric a brac (BTB) domain that can homodimerize or heterodimerize to facilitate insulator–insulator interactions and promote formation of long range insulator-mediated loops [4-7]. Initially, the *gypsy*

insulator complex was characterized as binding the 5′-untranslated region of the *gypsy* retroelement. However, the core complex also binds thousands of endogenous sites throughout the genome and can function similarly at least at a subset of those sites [8-12]. Moreover, the three *gypsy* insulator core components do not co-localize absolutely at all binding sites throughout the genome, and each protein can interact with other insulator proteins [10, 13-16]. In diploid interphase nuclei, *gypsy* insulator proteins coalesce into large foci termed insulator bodies. The function of these bodies is still not clear, but there is a high correlation between proper insulator function and insulator body localization [6, 13, 17-24]. In summary, the *gypsy* insulator complex contributes to higher order nuclear organization on several levels.

CP190 also associates with a variety of additional DNA-binding proteins that likely impart specificity of the respective complex. The BED finger-containing proteins BEAF-32, Ibf1, and Ibf2 interact with CP190 and promote insulator function [14, 25]. Three additional zinc finger proteins Pita, ZIPIC, and CTCF also interact with CP190 and contribute to insulator activity [15, 26]. Recently, the zinc-finger protein CLAMP was demonstrated to positively affect *gypsy* insulator activity and to colocalize particularly with CP190 at promoters throughout the genome [13]. Additionally, previous work showed that CP190 preferentially binds Motif 1-containing promoters [27], but the functional significance of this observation is currently unknown. The precise functions of CP190, its associated factors, as well as their relationship with transcription regulation have not yet been elucidated.

Motif 1 binding protein (M1BP) is a ubiquitously expressed transcriptional activator that is required for the expression of predominantly constitutive genes. A zinc finger DNA-binding protein, M1BP specifically binds to the core promoter element Motif 1 consensus sequence that is distinct from the canonical TATA box and mainly controls the expression of constitutively active genes that are transiently paused [28]. For example, M1BP interacts with the TATA-binding protein-related factor 2 (TRF2) to activate transcription of ribosomal protein genes in a Motif 1-dependent manner [29]. Finally, recent studies found that Motif 1 and M1BP are highly enriched at TAD boundaries along with CP190 and BEAF-32 [30-33]. Depletion of M1BP led to increased inter-chromosomal Hi-C contacts; however, concomitant cell cycle disruption precluded

interpretation of these results [31]. The possible role of M1BP-dependent transcriptional regulation in genome organization has not yet been interrogated in detail.

In this study, we identify M1BP as a physical interactor and positive regulator of the *gypsy* insulator complex. Depletion of M1BP decreases *gypsy*-dependent enhancer blocking and barrier activities and reduces the association of the core insulator complex with the *gypsy* insulator sequence. ChIP-seq analysis reveals extensive genome-wide overlap of M1BP particularly with promoter-bound CP190, and depletion of M1BP results in extensive loss of CP190 chromatin association genome-wide. Depletion of CP190 also disrupts M1BP binding at many of its binding sites. Finally, nascent euRNA-seq (neuRNA-seq) analysis of M1BP- or CP190-depleted cells indicates that both factors co-regulate a similar set of genes genome-wide. In particular, loss of gene activation correlates with disrupted M1BP and CP190 binding, and these events are frequently observed at TAD borders. Finally, depletion of M1BP disrupts *gypsy* insulator body localization within the nucleus and alters local genome compaction. Taken together, our findings identify a novel mechanistic relationship between M1BP and CP190 to activate Motif 1-dependent transcription as well as to promote chromatin insulator activity and nuclear organization.

RESULTS

M1BP associates physically with the core *gypsy* insulator proteins

In order to identify novel *gypsy* insulator interactors, we performed immunoaffinity purification of *Drosophila* embryonic nuclear extracts using antibodies specific for Su(Hw) or CP190 and analyzed the eluates using quantitative mass spectrometry. In addition to the third *gypsy* core component Mod(mdg4)67.2 and previously reported interactors such as Ibf1, Ibf2, BEAF-32, and CTCF, we identified M1BP in both purifications (Supplementary Table S1). CP190 was also previously identified as a physical interactor by coimmunoprecipitation of M1BP by mass spec analysis using *Drosophila* embryonic nuclear extract and subsequently confirmed by western blotting [34]. We verified that all three *gypsy* core components interact with M1BP by performing anti-Su(Hw), anti-CP190, and anti-Mod(mdg4)67.2 purifications and western blotting (Fig. 1A, 1B and 1C). Finally, we verified that Su(Hw), CP190, and Mod(mdg4)67.2

could be immunopurified by anti-M1BP but not normal serum (Fig. 1D). In contrast, an unrelated factor, Polycomb, was not co-immunoprecipitated with any of these antibodies. These results indicate that M1BP is associated physically either directly or indirectly with the core *gypsy* insulator proteins.

M1BP promotes enhancer blocking and barrier activities of the *gypsy* insulator

Given their physical interaction, we sought to test whether M1BP may play a role in *gypsy* insulator function. To deplete M1BP levels *in vivo*, we utilized an *M1BP*^{RNAi} fly line expressing a hairpin under upstream activating sequence (UAS) control that generates siRNAs against *M1BP* when combined with a Gal4 driver. To validate the knockdown efficiency of this line, we performed western blot analysis of larval extracts from flies expressing this RNAi construct using the ubiquitously expressed *Act5C-Gal4* driver compared to a control line expressing driver alone (Fig. 2A). Importantly, depletion of M1BP does not have any effect on overall Su(Hw), CP190 or Mod(mdg4)67.2 protein levels. Using *Act5C-Gal4*, *M1BP*^{RNAi} resulted in 100% late larval lethality, and complete pupal lethality is observed using the muscle-specific *Mef2-Gal4* driver (Supplementary Table S2). With CNS enriched *l(3)31-1-Gal4* or wing-expressed *Ser-Gal4*, *M1BP*^{RNAi} expression, flies remain completely viable.

In order to investigate whether M1BP affects *gypsy*-dependent enhancer-blocking activity, we examined the effect of M1BP depletion on the well-characterized allele *ct*^{6}. This loss-of-function allele results from *gypsy* retrotransposon insertion between the promoter and distal wing margin enhancer of *cut* (Fig. 2B). The *gypsy* insulator blocks communication between the two elements, reducing *cut* expression and causing disruption of the wing margin visible in the adult fly [35]. We used a scoring scale from 0–4 with increasing severity of wing margin notching corresponding to higher insulator activity. We found that knockdown of *M1BP* driven by *Ser-Gal4* compared to driver alone restored wing margin tissue, consistent with a decrease in enhancer-blocking activity (Fig. 2C). We also determined that knockdown of *M1BP* does not result in changes in wing margin induced by the *gypsy*-independent *ct*^{n} loss-of-function allele, which is caused by insertion of the *roo* transposable element (Supplementary Fig. S1).

These results indicate that M1BP is required for *gypsy*-dependent enhancer blocking activity.

We next investigated the effect of M1BP depletion on *gypsy*-dependent barrier activity in a variety of tissues. In this assay, we used a *UAS-luciferase* reporter that is either insulated by flanking Su(Hw)-binding sites or not insulated, with either reporter inserted into the same genomic site (Fig. 2D) [22]. We performed this quantitative luciferase-based assay using different tissue-specific Gal4 drivers to control both the luciferase reporter and $M1BP^{RNAi}$. Ubiquitously expressed *Act5C-Gal4* promotes high luciferase expression in insulated compared to non-insulated control larvae (Fig. 2E, Supplementary Table S3). As a positive control, knockdown of *su(Hw)* resulted in dramatic decrease in luciferase expression only in the insulated line, suggesting loss of barrier function. Likewise, knockdown of *M1BP* driven by *Act5C-Gal4* resulted in statistically significant reduction in luciferase activity compared to the insulated control line. Furthermore, knockdown using either the muscle-specific *Mef2-Gal4* driver or the CNS-enriched *l(3)31-1-Gal4* driver also resulted in significant decreases in luciferase activity (Fig. 2F, 2G, Supplementary Table S3). We conclude that M1BP promotes *gypsy*-dependent barrier activity in all tissues tested.

M1BP extensively colocalizes with CP190 throughout the genome

In order to obtain high-resolution information about the genome-wide chromatin association of M1BP and its relationship with *gypsy* core components, we performed ChIP-seq analysis for M1BP, CP190, Su(Hw) and Mod(mdg4)67.2 in the embryonic Kc167 (Kc) hemocyte cell line. Using previously validated antibodies, we identified 3121 M1BP, 8022 CP190, 4638 Su(Hw), and 3536 Mod(mdg4)67.2 peaks. By western blot analysis, we validated efficient protein depletion of M1BP and no effect on insulator protein levels five days after *M1BP* double stranded RNA (dsRNA) transfection (Fig. 3A, Fig. S2A). We also verified that this level of M1BP depletion did not greatly alter cell viability or lead to accumulation in M-phase (Supplementary Fig. S2B, S2D-E), in contrast to a previous study that used more dsRNA for a longer time period [31]. Our ChIP-seq analyses confirm that depletion of M1BP by RNAi dramatically reduced M1BP binding to chromatin throughout the genome, also confirming the specificity of the

antibody (Fig. 3B). Interestingly, we found that 79% (2461) of total M1BP peaks overlap with CP190 genome-wide (31% of total CP190 peaks) (Fig. 3B, 3C). In contrast, we found very low overlap of M1BP with either Su(Hw) (5.2%) or Mod(mdg4)67.2 (7.0%) (Fig. 3C). Finally, M1BP also overlaps considerably with BEAF-32 (41%) but less substantially with CLAMP, ZIPIC, Pita, Ibf1, Ibf2, and CTCF (Fig. S3).

We further examined the distribution of M1BP, CP190, and shared M1BP-CP190 sites with respect to genomic features. We noted that M1BP binding is not observed at well-characterized individual endogenous insulator sequences such as 1A-2, *Fab-7*, or *Fab-8* [36-38] (data not shown). Consistent with earlier reports, we verified that both M1BP and CP190 are most frequently observed at M1BP and CP190 binding, but not Su(Hw) or Mod(mdg4)67.2 binding, is enriched at transcription start sites (TSS) across the genome (Fig. 3E). As expected, we verified that these promoters frequently harbor Motif 1 consensus sequence [28, 39] (Fig. 3F, Supplementary Table S4, 54% of M1BP, 35% of CP190, and 54% co-occupied promoter peaks). Therefore, our ChIP-seq results indicate that M1BP colocalizes primarily with CP190 throughout the genome, particularly at Motif 1-containing promoters.

M1BP and CP190 regulate transcription of a similar set of genes across the genome

Given that M1BP and CP190 colocalize extensively at promoters enriched for the presence of Motif 1, we compared how each factor affects transcription genome-wide. We performed neuRNA-seq after a 1h pulse labeling of Kc cells to examine newly synthesized transcripts in mock transfected control cells versus knockdown of *M1BP*, *Cp190*, or *mod(mdg4)* treatment specific for the 67.2 KDa isoform. We found that depletion of M1BP results in up-regulation of nascent transcription of 1315 genes and down-regulation of 607 genes (Fig. 4A). Importantly, we found that only 22% of promoters of up-regulated genes (FET, P = 1.6e-02, odds ratio = 0.8) but 46% of promoters of down-regulated genes harbor M1BP chromatin binding based on ChIP-seq analysis in the control condition (FET, P < 2.2e-16, odds ratio = 2.8). These results are consistent with an earlier report that M1BP is mainly directly involved in transcriptional activation [28]. Likewise, CP190 binds promoters of only 40% of genes that are up-

regulated after M1BP knockdown (FET, *P* = 7.3e-02, odds ratio = 0.9), but CP190 binding is highly enriched at gene promoters that are down-regulated (76%, FET, *P* < 2.2e-16, odds ratio = 4.4) resulting from M1BP depletion. Taken together, these results suggest that CP190 may cooperate with M1BP to activate transcription (Fig. 4B, Supplementary Fig. S4A).

Intriguingly, we found that depletion of CP190 results in similar changes in nascent transcript levels compared to depletion of M1BP. After *Cp190* knockdown, 1894 genes are up-regulated, and 1382 genes are down-regulated (Fig. 4C). Overall, the extent of nascent transcription changes is correlated with that of *M1BP* knockdown (R = 0.69), with 901 up-regulated and 319 down-regulated genes in common (Figure 4E). In contrast, knockdown of *mod(mdg4)* resulted in little overlap of nascent transcription changes compared to *M1BP* knockdown (R = 0.24, Fig. 4F) or to *Cp190* knockdown (R = 0.29, Fig. 4G). Importantly, CP190 binding at the promoter in the control condition is also enriched at down-regulated genes (73%, FET, *P* < 2.2e-16, odds ratio = 4.1) but not up-regulated genes (41%, FET, *P* = 9.5e-02, odds ratio = 0.9) after CP190 depletion (Fig. 4C). M1BP is also associated with promoters of down-regulated genes (49%, FET, *P* < 2.2e-16, odds ratio = 3.4) but not up-regulated genes (21%, FET, *P* = 4.1e-04, odds ratio = 0.8) after *Cp190* knockdown (Fig. 4D). Finally, we verified that both M1BP plus CP190 binding together is also similarly statistically enriched predominantly at down-regulated genes in either *M1BP* (43%, FET, *P* < 2.2e-16, odds ratio = 2.9) or *Cp190* knockdown condition (44%, FET, *P* < 2.2e-16, odds ratio = 3.3) (Supplementary Fig. S4), suggesting a direct and specific relationship between CP190 and M1BP in transcriptional activation in particular. Although M1BP and CP190 appear to negatively regulate a similar set of genes, neither M1BP nor CP190 binding is enriched at the promoters of upregulated genes; therefore, these effects are likely to be indirect. Finally, we verified the association of M1BP and CP190 at promoters of several common down-regulated genes dependent on M1BP and CP190 using directed ChIP-qPCR (Fig. 4H).

M1BP and CP190 are both required for expression of Motif 1-dependent genes

To investigate whether CP190 is required for expression of genes dependent on Motif 1, we performed plasmid-based luciferase reporter assays in transfected Kc cells.

We monitored luciferase expression driven by Motif 1-containing *RpLP1* or *RpL30* promoters relative to *Renilla*, driven by the Motif-1 independent *RpIII128* promoter and expressed on a co-transfected control plasmid. (Fig. 5A, 5B). Cells were mock transfected or knocked down for *M1BP*, *Cp190*, *su(Hw)*, or *mod(mdg4)*. Both M1BP or CP190 depletion decreased Motif 1-dependent *RpLP1* and *RpL30*-driven luciferase expression, while *su(Hw)* and *mod(mdg4)67.2* knockdown had no effect (Fig. 5C, Supplementary Fig. S5). Importantly, *Cp190* knockdown did not affect luciferase expression from constructs driven by *RpLP1* or *RpL30* promoters that harbor point mutations in Motif 1 that abolish M1BP binding [28, 29]. These results suggest that CP190 is required for Motif 1-dependent expression and raise the possibility that CP190 may also affect M1BP association with Motif 1-containing promoters.

CP190 is required for M1BP chromatin association at a subset of endogenous binding sites

In order to test whether CP190 does indeed play a role in recruitment of M1BP to chromatin, we performed ChIP-seq of M1BP after depletion of CP190 in Kc cells. Western blot analysis illustrated that knockdown of CP190 successfully reduced CP190 levels but did not have any effect on M1BP protein levels (Fig. 6A). Additionally, we observed no major effect on cell proliferation or indication of mitotic arrest (Supplementary Fig. S2C-E). Using ChIP-seq, we verified substantial reduction of CP190 chromatin association (Fig. 6B). In order to identify statistically significant signal loss of called peaks by ChIP-seq, we applied the DiffBind algorithm [40] at *P*-value <0.05 and identified 806 M1BP sites (22% of total) with significantly decreased M1BP binding after CP190 depletion. This corresponds to loss of M1BP binding at 18% of co-occupied sites. The majority (81%) of reduced M1BP peaks colocalize with CP190 binding in the control condition, suggesting that many effects are direct and that M1BP binding at these sites is CP190-dependent. We validated six CP190-dependent M1BP sites by directed ChIP-qPCR (Fig. 6C, sites 1-5) and another six sites that are bound by both M1BP and CP190 at which M1BP remains unaffected after CP190 depletion (Fig. 6C, sites 6-11, Supplementary Fig. S4D). These data suggest that CP190 affects recruitment of M1BP to a subset of its binding sites within the genome.

We next examined the genome-wide relationship between both gene expression and M1BP chromatin association that is dependent on CP190. We found that CP190-dependent M1BP binding sites are enriched at genes that require CP190 for expression (Fig. 6D, 9% of genes, FET, P = 1.7e-10, odds ratio = 2.0) but not for genes that are upregulated when CP190 is depleted (5% of genes, FET, P = 7.0e-01, odds ratio = 0.9). Taken together, CP190-dependent loss of M1BP at these sites may culminate in a reduction of gene expression.

Chromatin association of CP190 is highly dependent on M1BP and correlates with M1BP-dependent activation of gene expression

We next tested the opposite scenario: the non-exclusive possibility that M1BP affects CP190 recruitment throughout the genome. We thus performed ChIP-seq analysis of M1BP or CP190 in mock-treated versus M1BP-depleted Kc cells. Using the DiffBind algorithm, we identified 2018 CP190 peaks (25% of total) that were reduced after M1BP depletion (Fig. 3B), and 34% of co-occupied sites show a decrease of CP190 binding. Again, we found that decreased CP190 peaks overlapped with M1BP binding in the control condition in a large proportion of cases (42%, FET, P < 2.2e-16, odds ratio = 2.0), suggesting that direct effects are observed. We validated a decrease of both CP190 and M1BP binding at co-occupied sites (Sites 1-5) in M1BP knockdown using directed ChIP-qPCR (Fig. 6C, Supplementary Table S6). We also verified one CP190 binding site lacking M1BP that remained unchanged after M1BP depletion (Fig. 6C, site 12, Supplementary Fig. S4C, Supplementary Table S6).

We then compared the genome-wide relationship between M1BP-dependent CP190 chromatin association and M1BP-dependent changes in nascent transcription. M1BP-dependent CP190 binding sites are enriched at promoters of genes that require M1BP for expression (19% of genes, FET, P = 4.3e-09, odds ratio = 1.9) but not genes that are up-regulated in *M1BP* knockdown (10% of genes, FET, P = 4.6e-01, odds ratio = 0.9). These results suggest that M1BP is directly required for CP190 recruitment throughout the genome, and loss of these factors may manifest in reduced gene expression.

Reduced gene expression and loss of CP190 binding after M1BP depletion occurs close to TAD borders

Because both M1BP and CP190 are enriched at promoters of genes located near TAD borders throughout the genome, we examined whether M1BP and CP190 activate transcription in proximity of TAD borders. We examined the distance from the promoter to the nearest TAD border for either upregulated or downregulated genes in either *M1BP* or *Cp190* knockdown cells and found that genes that require M1BP (FDR 9.1e-35) or CP190 (FDR 1.3e-80) for their activation are positioned significantly closer to Kc TAD borders compared to unchanged genes (Fig. 6E). In contrast, genes that are upregulated after *M1BP* (FDR 0.43) or *Cp190* (FDR 0.33) knockdown are located a similar distance from TAD borders compared to unaffected genes. Similar trends are observed for genes located in all four classes of TADs (active, inactive, PcG, and HP1, Supplementary Fig. S6) as previously defined [31]. Overall our observations suggest that M1BP and CP190 promote transcription near TAD borders genome-wide independently of TAD chromatin state.

M1BP and CP190 binding to the *gypsy* insulator are interdependent and required for core insulator complex recruitment

Since M1BP rarely overlaps with Su(Hw) and Mod(mdg4)67.2 across the genome, we wanted to determine the mechanism by which M1BP promotes *gypsy* insulator activity in the context of the retrotransposon. To this end, we examined whether M1BP is also responsible for CP190 recruitment to *gypsy* retrotransposon sites. Interestingly, we did detect substantial M1BP chromatin association along with the three core insulator components at the 12 Su(Hw)-binding sites of the *gypsy* retrotransposon in Kc cells (Fig. 6F). No binding of M1BP or core insulator proteins is observed at *TART,* another element found at multiple locations throughout the genome. After depletion of either M1BP or CP190, chromatin association of all four factors is dramatically reduced at the *gypsy* insulator binding site. These results are consistent with a direct effect of M1BP on *gypsy* chromatin insulator activity and on the recruitment of the core insulator components.

Depletion of M1BP alters nuclear localization of *gypsy* insulator bodies

Since M1BP is required for core complex recruitment to *gypsy* insulator sites as well as enhancer blocking and barrier activities, we examined the effect of M1BP on the nuclear localization of *gypsy* insulator bodies, of which proper localization correlate with insulator activities. We performed whole-mount immunostaining of dissected brains and imaginal discs of third instar larvae using antibodies against CP190 to detect insulator body localization. In the control line, approximately one insulator body per focal plane was observed in brain optic lobe, eye, leg and wing discs (Fig. 7A). In contrast, multiple insulator bodies were observed in $M1BP^{RNAi}$ driven by *Act5C-Gal4*, and the differences in number are statistically significant (Fig. 7B). These results suggest that M1BP ubiquitously affects nuclear localization of insulator bodies.

M1BP regulates local genome compaction but not large scale chromosome territory formation or intermingling

Changes observed in insulator body organization prompted us to examine whether M1BP or CP190 may affect global nuclear organization. Therefore, we examined the 3D localization of chromosome territories (CTs) for chromosome arms 2L and 2R using Oligopaint FISH in Kc control versus *M1BP* or *Cp190* knockdown cells. No significant differences in territory volume relative to nuclear volume, CT contact, or CT intermingling were observed (Supplementary Fig. S7).

To investigate if local compaction is altered after M1BP or Cp190 depletion, we used Oligopaint FISH to measure the nuclear spatial positioning of two closely spaced regions (~44 kb apart) in distinct TADs separated by M1BP and CP190 binding sites. These regions are also of interest because M1BP- and CP190-dependent nascent transcription is observed within this vicinity (Fig. 8A). Examination of these two 30 kb probes in control versus M1BP- or CP190-depleted cells showed that depletion of M1BP but not CP190 resulted in a statistically significant decrease of inter-focal distance (Fig. 8C, E, Supplementary Fig. S8). Moreover, we examined a second pair of probes located ~13 kb apart spanning the previously characterized Nhomie and Homie insulator sites respectively [41], which flank a Polycomb Group repressed TAD and also feature M1BP and CP190 binding as well as M1BP- and CP190-dependent nascent

transcription (Fig. 8B). Distance between these probes was also significantly decreased in M1BP- but not CP190-depleted cells (Fig 8D, E, Supplementary Fig. S8). Finally, we analyzed the distance between two 50 kb regions located ~3.1 Mb apart and found no differences in M1BP- or CP190-depleted cells (Supplementary Fig. S9). In summary, we conclude that loss of M1BP can alter local but not large scale genome compaction.

DISCUSSION

Here we show that M1BP is required for proper *gypsy* insulator function and insulator body localization and that M1BP and CP190 together activate transcription at TAD borders. We found that M1BP physically associates with CP190 as well as core *gypsy* components and promotes enhancer blocking and barrier activities. Genome-wide, M1BP colocalizes mainly with CP190 at Motif 1-containing promoters, which are enriched at TAD borders. Depletion of M1BP extensively reduces chromatin association of CP190, and loss of either factor reduces gene expression at TAD borders. Furthermore, recruitment of M1BP is dependent on CP190 at a subset of genomic sites. Finally, M1BP is required for proper nuclear localization of insulator bodies and local genome compaction. Taken together, our findings suggest that M1BP may play a role in 3D genome organization through a CP190- and transcription-dependent mechanism.

M1BP promotes *gypsy* insulator activity by aiding core complex recruitment to insulator sites

As M1BP is ubiquitously expressed throughout development, we observe effects on insulator activity and complex localization after M1BP depletion in all tissues and stages of development tested. M1BP associates physically with chromatin at the Su(Hw)-binding sites of the *gypsy* insulator in conjunction with core insulator proteins, and M1BP is required for the binding of all three factors. These findings suggest that M1BP directly affects insulator activity by aiding the recruitment of *gypsy* core components to *gypsy* insulator sites, all three of which are required for proper insulator activity [5-7]. Interestingly, Motif 1 is not present at this sequence, and M1BP binding at this site is also dependent on the presence of CP190. One scenario is that binding of the two factors could be cooperative, and M1BP recruitment may additionally help

stabilize the multimerization and/or higher order organization of insulator complexes. Consistent with this possibility, depletion of M1BP results in increased numbers of insulator bodies, similar to the effect of complete loss of the BTB-containing core insulator protein Mod(mdg4)67.2 [17, 24]. Since CP190 is a universal insulator protein in *Drosophila*, mislocalization of CP190 may result in, or at least serve as an indicator of, disrupted genome organization when M1BP is depleted.

Dependency of CP190 binding on M1BP throughout the genome

Although M1BP physically interacts with each of the core *gypsy* insulator components, we observed that M1BP co-localizes mainly with just CP190 throughout the genome, particularly at Motif 1-containing promoters. Importantly, recruitment of CP190 is dependent on M1BP and vice versa at many co-occupied sites genome-wide. Why M1BP is only partially dependent on CP190 for binding is unclear, but these results are consistent with the known ability of M1BP to bind DNA directly [28] whereas CP190 is believed to require interaction with a specific DNA-binding protein in order to associate with chromatin [4, 14, 15, 25, 42, 43]. Aside from the insulator protein BEAF-32, which binds an AT-rich dual core sequence [44], we did not generally observe a large extent of overlap between M1BP and other DNA-binding insulator proteins that have been shown to be involved in recruiting CP190 to DNA. Future studies may reveal a possible functional relationship between M1BP and BEAF-32 in insulator activity or regulation of gene expression.

Distinct relationships between CP190 and M1BP compared to CP190 with CLAMP

Our results suggest that M1BP promotes *gypsy* insulator function through interaction with CP190 in a manner distinct from CP190 interaction with the zinc finger DNA-binding protein CLAMP. We did not observe a large extent of genome-wide overlap between M1BP and CLAMP, a recently identified positive regulator of *gypsy* insulator activity [13]. The sequence binding specificity of CLAMP is similar to that of GAF [45], while M1BP and GAF bind to and regulate distinct sets of promoters [28]. Although either CLAMP or M1BP depletion reduces *gypsy* enhancer blocking and barrier activities as well as alter insulator body localization, unlike M1BP, CLAMP depletion

does not affect CP190 chromatin association throughout the genome [13]. In fact, CP190 depletion had a substantial effect on CLAMP chromatin association, again suggesting that CP190 may affect the ability of certain DNA-binding proteins, including M1BP, to associate with chromatin through cooperative or higher order physical interactions.

CP190 promotes transcription through interaction with M1BP

Our results suggest that CP190 may play a more direct role in transcriptional regulation than previously appreciated, likely through interaction with M1BP. Genome-wide profiling studies have shown that CP190 preferentially associates with promoters genome-wide [10, 13, 46]. CP190 was found to be enriched particularly at active promoters and was shown to affect steady state gene expression when depleted; however, direct and indirect effects as well as transcriptional and posttranscriptional effects could not be separated [46]. In order to avoid the complication of interpreting steady state gene expression profiles, we performed neuRNA-seq after either CP190 or M1BP depletion in order to measure newly synthesized transcripts. Intriguingly, nascent RNA expression profiles of CP190 or M1BP-depleted cells showed a remarkably high level of correlation. Since both M1BP and CP190 are particularly associated with promoters of genes that require either factor for adequate expression, it is likely that both factors mainly function in transcriptional activation rather than repression. Furthermore, we found that CP190 is specifically required for Motif 1-dependent expression of two particular previously characterized ribosomal protein genes. Whether TRF2 recruitment to Motif 1-containing promoters is affected by CP190 as well as the precise mechanism by which CP190 contributes to M1BP-dependent transcription will be important topics of further study.

Relationship between M1BP and CP190-dependent transcription regulation and higher order chromatin organization

We found that genes that require M1BP and CP190 for adequate expression are frequently located at TAD borders, as are both proteins. It has previously been proposed that active transcription, particularly at/near TAD borders (also referred to as “compartmental domains”), may be a defining or at least key feature of overall genome organization in cells [30, 33, 47] and throughout development [48]. CP190 was previously

shown to be associated particularly with Motif 1-containing promoters [27], and CP190 was observed to be specifically enriched at TAD borders [32, 33]. Recently, Motif 1 was also found to be an enriched sequence at TAD borders present in Kc cells [31] and is apparent when TAD borders first appear in embryonic development [48], findings consistent with these previous studies. A previous study of M1BP involvement in genome organization provided limited evidence using Hi-C to suggest that chromosome intermingling may be increased after M1BP depletion [31]. However, the extent and duration of M1BP depletion in their study caused a major disruption of cell cycle and cellular growth, thus obscuring interpretation of those results. We did not observe changes in CT intermingling in our Oligopaint FISH experiments in M1BP-depleted G1 cells, nor did we observe any difference in distance between two distant regions on the same chromosome. We did observe increased local genome compaction after M1BP depletion coincident with altered local transcription, suggesting that M1BP-dependent transcriptional changes might affect local chromatin structure. However, CP190 depletion did not result in a similar effect, perhaps reflecting the multifunctional nature of this universal insulator protein contributing to opposing forces. Our work shows the requirement of M1BP for accurate CP190 binding throughout the genome as well as *gypsy*-dependent chromatin insulator activity and nuclear localization through interaction with CP190 and other core insulator proteins. Overall, our results provide evidence that the presence of M1BP and CP190 at TAD borders and their ability to activate transcription of genes located at TAD borders may play a role in genome organization.

M1BP has been shown to activate transcription of genes at which RNA Pol II is transiently paused in the promoter proximal region, and these promoter regions may themselves possess insulator activity. Intriguingly, a previous study showed that stalled Hox promoters, including Motif 1-containing *Abd-B*, possess intrinsic enhancer blocking insulator activity [49]. However, the second paused promoter identified in this study, *Ubx*, does not harbor Motif 1; thus, M1BP may not necessarily be involved in the enhancer blocking activity of all stalled promoters. Recently, it was shown that M1BP promoter binding can prime the recruitment of the Hox protein Abd-A to the promoter in order to release paused Pol II and activate transcription [50]. We find that M1BP is similarly

required for CP190 recruitment at a large number of sites throughout the genome and thus propose that M1BP and CP190 are together required to maintain active gene expression near TAD borders. Active transcription at these sites may be needed for higher order chromatin organization such as TAD insulation, formation of active compartmental domains, and/or proper local genome structure. Future studies will elucidate the precise mechanisms by which M1BP and CP190 contribute to higher order chromatin organization.

MATERIALS AND METHODS

Drosophila strains

Fly lines were maintained on standard cornmeal medium at 25°C. We used lines expressing dsRNA against *su(Hw)* (10724 GD) [22] and *M1BP* (110498 KK) from the Vienna Drosophila RNAi Center. *y,w[1118]; P{attP,y[+],w[3']* (60100 KK) was used as a control for *M1BP* KK RNAi lines. *Act5C-Gal4*, *Mef2-Gal4* and *l(3)31-1-Gal4* driver lines were obtained from Bloomington Drosophila Stock Center. We scored the ct^6 phenotype on the first day after eclosion [13, 22]. *UAS-luciferase* constructs were inserted into the *attP3* landing site using phiC31 site-specific integration [51]. Protein extracts from anterior thirds of larvae were used for western blotting. Embryos aged 0–24 h were collected from a population cage reared at RT as described previously [52] to produce nuclear extracts.

Luciferase insulator barrier activity assay

Insulator barrier activity by luciferase assay was carried out as described previously using Bright-Glo™ Luciferase Assay System (Promega) [13, 22]. Luciferase signal was quantified using a Spectramax II Gemini EM plate reader (Molecular Devices). Luciferase levels were measured for twelve individual whole third instar male for all genotypes indicated in a single panel simultaneously. Luciferase value was normalized to total protein of each larva determined by BCA reagent (Thermo Scientific). The relative luciferase activity of a population of a single genotype was aggregated into a box and whisker plot. Populations were compared with one-way

ANOVA followed by a Tukey HSD post-hoc test to obtain *P*-values for each pairwise comparison. The *P*-values for pairwise comparisons between the control and RNAi lines within both non-insulated and insulated groups are listed in Supplementary Table S3.

Immunostaining of imaginal discs

Imaginal discs and brains were dissected from at least six larvae of each genotype and immunostained with rabbit serum against CP190 following whole-mount staining methods as previously described [26] mounted with ProLong Diamond mounting media (Life Technologies). Number of insulator foci per nucleus were counted manually.

Co-immunoprecipitation

Embryonic nuclear extract was prepared from 20 g of mixed stage (0–24 h) *Drosophila* embryos as described previously [20]. Nuclei were lysed with 5 mL nuclear lysis buffer (60 mM HEPES pH 7.5, 10 mM MgCl2, 100 mM KCl, 0.1% Triton X-100, 10% glycerol, Roche cOmplete protease inhibitor) and sonicated for 8 cycles with 10 s on and 50 s off [13]. The soluble fraction of extracts was collected by centrifugation. First, two sets of 20 µL of Protein A or Protein G Sepharose beads (GE Healthcare) were washed three times with nuclear lysis buffer for immunoprecipitation with antibody raised in guinea pig or in rabbit respectively. Then 3 µL of guinea pig normal serum (Covance Research Products), anti-serum against CP190 (guinea pig), anti-serum against Su(Hw) (guinea pig), rabbit normal serum (Covance Research Products), anti-serum against Mod(mdg4)67.2 (rabbit), or anti-serum against M1BP (rabbit) (3 µl) were incubated with sepharose beads for 1 h at 4°C, and unbound antibodies were removed by centrifugation at 1500 g for 1 min. Beads were washed three times with 0.2 M of sodium borate, pH 9, and then crosslinked with 20 mM DMP in sodium borate for 30 min at RT [53]. Next, beads were collected by centrifugation and washed once with ethanolamine and three times with lysis buffer. After crosslinking, 500 µg of nuclear extract was used for each immunoprecipitation and incubated with antibody-bound beads overnight at 4°C. The next day, beads were collected by centrifugation and washed three times with nuclear lysis buffer. Samples were eluted with SDS sample buffer by boiling, separated using SDS-PAGE, transferred to nitrocellulose membrane in

10 mM CAPS, pH 11, and detected using western blotting. The co-immunoprecipitation efficiency was calculated for each immunoprecipitated protein based on the percentage of total input protein.

IP and mass spectrometry

Nuclear extracts from 18 g of mixed stage (0–24 h) *Drosophila* embryos [20] were lysed in 5 mL HBSMT-0.3% + 1 M KCl (50 mM HEPES, 150 mM NaCl, 1 M KCl, 3 mM MgCl2, 0.3% Triton X-100 [v/v] at pH 7) including 1 mM PMSF, and Complete protease inhibitor cocktail (Roche). Immunoprecipitation was performed using previously described methods, and beads were washed twice with HBSMT-0.3% + 1 M KCl, once with HBSM (50 mM HEPES, 150 mM NaCl, 5 mM KCl, 3 mM $MgCl_2$), and eluted with 1% Sodium Dodecanoate for mass spectrometry.

Proteins were analyzed using tandem HPLC-mass spectrometry at the NIDDK Mass Spectrometry Facility. Mass from eluted peptides was queried in the UniProt database, and results were analyzed by MaxQuant.

Cell lines

Kc167 cells were grown in CCM3 media (Thermo Scientific HyClone, Logan, UT). Cells were maintained in monolayer at 25°C.

DsRNA knockdowns

1×10^7 Kc167 cells were transfected with either 20 µg of dsRNA against M1BP [28], 12 µg of dsRNA against *Cp190* [13, 26], 5 µg of dsRNA against *su(Hw)*, 5 µg of dsRNA against *mod(mdg4)* (specific to deplete the Mod(mdg4)67.2 isoform), or no RNA (mock) using Amaxa cell line Nucleofector kit V (Lonza) and electroporated using G-30 program. Cells were incubated for 5 d at 25°C to obtain efficient depletion of each protein but minimal effects on cell viability and growth. The primers for generating the templates for *in vitro* synthesis of RNAi are listed in Supplementary Table S5.

Quantification of cell viability

Cell viability was quantified by Trypan Blue exclusion test. 10 μL of cells mixed with equal amount of trypan blue were examined to determine the percentage of cells that have clear cytoplasm (viable cells) versus cells which have blue cytoplasm (nonviable cells). Number of viable cells of mock, *M1BP* dsRNA, or *Cp190* dsRNA treated for each day of knockdown were recorded. Viability graphs were generated from the values of viable cell numbers and S.D was calculated from two replicates.

Quantification of cell proliferation and mitotic index

Cells were collected each day of knockdown for mock, *M1BP* dsRNA, or *Cp190* dsRNA treatments. Cells were immunostained with rabbit antibody against Histone H3 (phospho S10) (1: 5000) to identify mitotic cells and mouse anti-Tubulin (Sigma; 1:500) to assay cell viability and mitotic stage as previously described [54], stained with DAPI and mounted with ProLong Diamond mounting media (Life Technologies). Numbers of mitotic cells were counted manually. Statistical tests were performed using Prism 8 software by GraphPad.

FISH with Oligopaints

Oligopaints were designed to have 80 bp of homology using a modified version of the OligoMiner pipeline [55]**.** Oligo pools were synthesized by Twist Biosciences, and probes were synthesized as previously described [56]**.** Coordinates for all probes can be found in Supplementary Table S8. Chromosome 2L and 2R Oligopaints were a gift from E. Joyce and label the same domains as previously described [57].

For FISH, slides were prepared by dispensing 1.5 x 10^5 cells onto 0.01% poly-L-lysine coated slides and fixed with 4% paraformaldehyde for 15 min at RT, followed by washes two times, each for 5 min in PBS-T. Then slides were washed once in 2x SSCT for 5 min at RT, once in 2x SSCT/ 50% formamide for 5 min at RT, once in 2x SSCT/ 50% formamide at 92˚C for 2.5 min, and once in 2xSSCT/50% formamide at 60˚C for 20 min. For hybridization, 50-150 pmol of each probe was used per slide in a final volume of 25 μL. After the applying of primary Oligopaint probes, slides were covered with a coverslip and sealed. Denaturation was performed at 92˚C for 2.5 min. Slides were transferred to a 37˚C humidified chamber and incubated for 16–18 h. Then slides were

washed in 2x SSCT at 60˚C for 15 min, 2x SSCT at RT for 15 min, and 0.2x SSC at RT for 5 min. Fluorophore conjugated secondary probes (0.4 pmol/μL) were added to slides, covered and sealed. Slides were incubated at 37˚C for 2 h in a humidified chamber followed by washes as for the primary probes. All slides were washed with DAPI DNA stain (1:10,000 in PBS) for 5 min, followed by 2 x 5 min washes in PBS before mounting in Prolong Diamond (Life Technologies).

Imaging, quantification, and data analysis

Images were captured at RT on a Leica DMi 6000B widefield fluorescence microscope using a 1.4 NA 63x or 100x oil-immersion objective and Leica DFC9000 sCMOS Monochrome Camera. DAPI, CY3, CY5, and FITC filter cubes were used for image acquisition. Images were acquired using LasX Premium software and deconvolved using Huygens Professional software (Scientific Volume Imaging). After deconvolution, images were segmented and measured by a TANGO 3D-segmentation plugin for ImageJ as described [58]. All nuclei were segmented using the 'Hysteresis' algorithm. For CT measurements, CTs were segmented using the 'Spot Detector 3D' algorithm. Foci were segmented using the 'Hysteresis' algorithm. To avoid confounding effects of possible differences based on cell cycle stage, imaged nuclei were sorted into G1 and G2 cell cycle stages based on nuclear volume as measured in TANGO (where G1 nuclei were less than 150 μm^3 and G2 nuclei were between 150-300 μm^3) [57]. Only G1 nuclei, which would have recently divided during the time of the experiment, were included for all FISH analyses. Statistical tests were performed using Prism 8 software by GraphPad. Signal plot profiles were created in ImageJ using single Z slices.

ChIP and ChIP-seq library preparation

2–3×10^7 cells were fixed by adding 1% formaldehyde directly to cells in culture medium for 10 min at RT with gentle agitation. Formaldehyde was quenched with 0.125 M glycine with gentle agitation for 5 min. Then cells were pelleted by centrifugation at 2000 x *g* and washed twice with ice-cold PBS. Next, chromatin was prepared and chromatin immunoprecipitation (ChIP) was carried out following previously described methods [13]. Details of antibodies can be found in Table S7. Libraries were constructed

by pooling two immunoprecipitation (IP) samples using TruSeq adapters (Illumina) according to the TruSeq Illumina ChIP-seq sample preparation protocol with minor modification as described previously [13]. All samples were sequenced with HiSeq2500 (Illumina) using 50 bp single-end sequencing.

ChIP-quantitative PCR

We performed quantitative PCR using ChIP DNA samples of normal serum IP (negative control), M1BP IP, CP190 IP, Su(Hw) IP, and Mod(mdg4)67.2 IP from both mock-treated Kc cells and Kc cells transfected with dsRNA against *M1BP* or *Cp190*. ChIP DNA samples were amplified on an Applied Biosystems real-time PCR machine using site-specific primer sets (Supplementary Table S6) and quantified using SYBR Green (Applied Biosystems) incorporation. Experiments were performed in two independent biological replicates, and each sample was quantified using four technical replicates. The *P*-values were calculated by Student's t-test.

Nascent EU-RNA labelling and library preparation

Nascent EU RNA seq (neuRNA-seq) labeling and capture were done by using Click-iT Nascent RNA Capture Kit (Thermo Fisher Scientific) according to manufacturer's protocol. Kc cells were incubated with 0.2 mM EU for 1 h and RNA was extracted with Trizol (Thermo Fisher Scientific). Next, RNA was chemically fragmented for 5 min at 70˚C with RNA Fragmentation Reagents (Thermo Fisher Scientific), followed by DNase I treatment (Roche). Then RNA was ethanol precipitated after Phenol:Chloroform (Thermo Fisher Scientific) purification. The Click-iT reaction was performed with 0.5 mM biotin azide using 5 µg of EU-RNA, and biotinylated RNA was captured with 12 µL T1 beads. The nascent EU-RNA was used to generate RNA-seq libraries with Ovation RNA-seq Systems 1–16 for Model Organisms (Nugen). Samples were sequenced on HiSeq2500 (Illumina) using 50 bp single-end sequencing at the NIDDK Genomics Core Facility.

Luciferase reporter assays for ribosomal protein gene promoters

5 ×10^6 Kc cells were transfected with dsRNA against *M1BP*, *Cp190*, *su(Hw)* and *mod(mdg4)* as described above. After 2 d of incubation with dsRNA at 25°C, luciferase reporter construct expressing either wild type or Motif 1 mutant of pGL3-(RpLP1 [-500 to + 50]) or pGL3-(RpLP30 [-500 to + 50]) was co-transfected with RpL-polIII-Renilla into cells. Cells were incubated for an additional 3 d and then assayed in tandem for firefly and Renilla luciferase activity. Cells were lysed and assayed using the dual-luciferase reporter assay system (Promega) according to the manufacturer's protocol.

ChIP-seq data analysis

FASTQ files of sequenced single-end 50 bp reads were trimmed using cutadapt v1.8.1 [59] with arguments '--quality-cutoff 20', '-a AGATCGGAAGAGC', '--minimum-length 25' and '--overlap 10'. Then trimmed reads were mapped to the Flybase r6-24 dm6 genome assembly with Bowtie2 v2.3.5 [60] with default arguments. Multimapping reads were removed mapped reads using samtools v1.9 [61] view command with the argument -q 20. Duplicates were removed from mapped, uniquely mapping reads with picard MarkDuplicates v2.20.2 (http://broadinstitute.github.io/picard/index.html). MACS2 v2.2.5 [62] (https://github.com/taoliu/MACS) was used to call peaks by providing replicate IPs and inputs as multiple BAMs, effectively calling peaks on pooled/merged samples and using additional arguments '-f BAM', '--gsize=dm', '--mfold 3 100' (the latter to include a larger set of preliminary peaks for fragment size estimation).

Binary heatmaps were generated using pybedtools [63, 64]. Since peaks for one protein can potentially overlap multiple peaks for other proteins, the output represents unique genomic regions as determined by bedtools multiinter with the -cluster argument. Therefore, as a result, when considering the multi-way overlap with other proteins, the sum of unique genomic regions for a protein is not guaranteed to sum to the total number of called peaks for that protein.

Pairwise comparison for co-localization of different factors was performed using the BEDTools 'jaccard' command [63, 64]. The heatmap was clustered with the scipy.cluster.hierarchy module, using 'euclidean' as the distance metric and 'Ward' as the clustering method. As the Jaccard statistic is independent of the order of comparison and is symmetric across the diagonal, only the upper triangle is shown.

FlyBase release 6.24 annotations were used to annotate peaks following into genomic regions in Fig. 3D as follows. Exons were defined as any exon from any transcript of any gene. Introns were defined as the space between exons derived in a per-transcript manner by using the gffutils (https://github.com/daler/gffutils) method FeatureDB.create_introns(). Promoters were defined as the TSS of each transcript plus 1500 bp upstream. Intergenic regions were defined as all regions between gene bodies. Shared co-bound peaks were determined by using pybedtools BedTool.intersect with the v=True or u=True argument, respectively. Each set of peaks was intersected with the annotations using this hierarchy 'promoter>exon>intron>intergenic' where a peak was classified according to the highest priority feature. Here, a peak simultaneously intersecting a promoter of one isoform and an intron of a different isoform would be classified as 'promoter'. To compare percentages across annotated peaks in different types of peaks (All M1BP, all CP190, shared M1BP and CP190 peaks and all *gypsy* sites) the number of peaks in each class was divided by the total number of peaks of that type.

Heat maps were generated by using default deepTools packages in Galaxy. ChIP reads were normalized to reads of input samples and mapped in a 3 kb window centered on the TSS containing Motif 1 consensus sequence.

Differential ChIP-seq

To detect differential ChIP-seq binding, we used the Diffbind v2.14.0 /R package [40] using the config object 'data.frame(RunParallel=TRUE, DataType=DBA_DATA_FRAME, AnalysisMethod=DBA_EDGER, bCorPlot=FALSE, bUsePval=FALSE, fragmentSize=300)' and otherwise used defaults. Input files consisted of the final peak calls mentioned above and the IP and input BAM files for each replicate as described above with multimappers and duplicates removed. Then the final results were exported with the dba.report function with parameters 'th=1, bCalled=TRUE, bNormalized=TRUE, bCounts=TRUE' and final differentially gained or lost peaks were those that had a log2 fold change of >0 or <0, respectively, and *P*-value <0.05.

neuRNA-seq (mapping and read counting)

Low quality bases and adaptors were trimmed from sequence reads using cutadapt v2.7 [59] with parameters "–q 20 -minimum-length 25 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA". Resulting reads were mapped to the FlyBase r6-24 reference genome using HISAT2 v2.1.0 [65] with default parameters. Aligned reads were counted using subread featureCounts v1.6.4 [66] with default parameters, except that "-t gene" option was used to quantify the reads on gene feature for EuSeq, and "-s1" option specifically for this sense-stranded library. Differential expression analysis was done with DESeq2 v1.22.1 [67] in v3.5.1 R version.

neuRNA-seq (differential expression)

Counts tables were loaded into DESeq2 v1.10.1 for [67] neuRNA-seq analysis. Counts tables from independent neuRNA-seq experiments were independently imported and normalized using simple design "~treatment" and using otherwise default parameters. Differentially expressed genes were those with $p_{adj} < 0.05$.

Fisher's exact tests

These tests were performed using number of gene promoters occupied/not occupied by M1BP or CP190 or both from the group of up-regulated (log2fold >0, $p_{adj} < 0.05$), down-regulated (log2fold < 0, $p_{adj} < 0.05$) and not affected gene promoters, respectively analyzed from the DESeq2. Prior to determining the number of peaks intersecting with gene promoter regions from the selected group, the promoter regions were merged if there were any overlapping promoter regions. For intersections with ChIP peaks, a gene promoter region was defined as 1 kb up and down stream of TSS. The Flybase r6.24 gene annotation gtf file was used to define gene start sites.

Fisher's exact tests (FET) were used to test whether the up/down-regulated genes show preference on the overlapped peaks. FET is implemented by Python scipy package (scipy.stats.fisher_exact). FET and odds ratios are based on the 2×2 contingency table. All genes are used. FET test uses two-tailed test.

Genes	Up/Down	Others	Row total
Overlap with ChIP-peaks	a	b	a+b

No overlap	c	d	c+d
Column total	a+c	b+d	a+b+c+d=n

The *P* value is computed as $p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$.

The odds ratio = $\frac{a/b}{c/d}$

Promoter distance analysis from closest to TAD border

Four types of TADs (Active, inactive, HP1 and PcG TADs) were downloaded from (http://chorogenome.ie-freiburg.mpg.de/) [31] and lifted over to the dm6 assembly. The distance of the gene promoters to the TAD border was computed for each type of TAD, and then pybedtools packages including sort and closest operations were used to search the nearest TAD for each gene promoter. Next, gene promoters inside the TAD were selected and computed for the minimum distance to the TAD border. Seaborn package (seaborn.kdeplot) was used to calculate the kernel density of distance for up, down, and unchanged genes, and cumulative distribution plots were generated based on the kernel density estimation. The matplotlib package was used to plot the density of gene distance to TAD border as a histogram (matplotlib.pyplot.hist, bin=100).

Mann-Whitney U test

The Mann-Whitney U test ("two-sided") was used for each set of changed genes against unchanged genes to test whether the difference of both mean TAD border distances is significant. False Discovery Rate (FDR) correction was performed using multiple test correction (Benjamini-Hochberg method).

Motif 1 derivation

Motif 1 analysis was derived from the MEME analysis of promoters [39]. Motif scanning was carried out using FIMO [68] for all gene promoter regions. The Motif 1_input was a position probability matrix of Motif 1. The position probability matrix of Motif 1 G[A/C]TACGGTCACACTG was obtained by transformation of a position weight matrix from Supplementary Table S4 from a previous study [39]. The transformation converts the weight matrix to the probability position matrix by using the definition -$\ln\left(\frac{N_{ij}+p_i}{N+1}\right) -$

$\ln(p_i) = M_{ij}$. Here, *Mij* is the entry of weight matrix and *pi* the background probability of nucleotide *i, and N* is the total number of sequences used in motif prediction. Assume p_i is equal to 0.25. Since N is much larger than pi, the equation can be rewritten into $\ln\left(\frac{N_{ij}}{N+1}\right) - \ln(p_i) = M_{ij}$ or $\ln(P_{ij}) - \ln(p_i) = M_{ij}$.

P_{ij} is an entry of position probability matrix [69]. After rounding all entries to the 19th decimal, only the probability of the second position was assigned as 0.5 for nucleotide A and C, respectively. The remaining positions in the matrix were assigned to 1 for each corresponding nucleotide. Other than that, all entries were assigned to zero. The motif search cutoff used was $p \leq 7.12e\text{-}5$.

Data Availability

The accession numbers for the raw data FASTQ files, processed files, and BigWig files for all sequencing data deposited in NCBI GEO are GEO: GSE142533 (https://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE142533).

Acknowledgments

We thank D. Gilmour for sharing previously unpublished results, rabbit anti-M1BP sera, and for comments on the manuscript. We also thank R. Dale for initial software development, A. Beyer for anti-Pep, P. O'Farrell for anti-Pc antibodies, S. Nguyen and E. Joyce for whole chromosome Oligopaints, and J. Kassis, S. Nguyen, and members of the Lei laboratory for critical reading of the manuscript.

Author Contributions

Conceptualization: I.B., E.P.L.; Methodology: I.B.; Mass-spec: S.C.; Oligopaints: L.F.R.; Software: Y.C., C.L., G.Y.; Bioinformatics analysis: I.B., S.C.; Validation: I.B., E.P.L.; Formal analysis: I.B., E.P.L.; Investigation: I.B.; Data curation: C.L.; Writing - original draft: I.B., E.P.L.; Supervision: E.P.L.; Funding acquisition: E.P.L.

Funding

This work was funded by the Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (DK015602

to E.P.L.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of Interests

The authors declare no competing interests.

Figure Legends

Fig. 1. Co-immunoprecipitation of core *gypsy* components with M1BP.

(A) Nuclear extracts (NE) from embryos aged 0–24 h were immunoprecipitated with either normal guinea pig serum or guinea pig anti-Su(Hw) antibody. Unbound supernatant (Sup) and bound (IP) fractions are shown. Polycomb (Pc) is shown as a negative control. 10% of NE is loaded in the gel. 5.2% of total CP190, 7.2% of Su(Hw), 1.9% of Mod(mdg4)67.2, and 5.3% of M1BP were immunoprecipitated with anti-Su(Hw) antibody.

(B) Immunoprecipitation of M1BP with normal guinea pig serum or guinea pig anti-CP190 is shown. 12% of total CP190, 7.0% of Su(Hw), 8.2% of Mod(mdg4)67.2, and 3.8% of M1BP were immunoprecipitated with anti-CP190 antibody.

(C) Immunoprecipitation of M1BP with normal rabbit serum or rabbit anti-Mod(mdg4)67.2 is shown. 11% of total CP190, 8.5% of Su(Hw), 10% of Mod(mdg4)67.2, and 3.0% of M1BP were immunoprecipitated with anti-Mod(mdg4)67.2 antibody.

(D) Immunoprecipitation of M1BP with normal rabbit serum or rabbit anti-M1BP is shown. 3.8% of total CP190, 2.6% of Su(Hw), 1.4% of Mod(mdg4)67.2, and 3.7% of M1BP were immunoprecipitated with anti-M1BP antibody.

Fig. 2. M1BP promotes *gypsy*-dependent enhancer-blocking and barrier activities.

(A) Western blotting of male third instar larval extracts for M1BP, insulator proteins, and Pep loading control in control and $M1BP^{RNAi}$ knockdown flies using *Act5C-Gal4* driver.

(B) M1BP promotes enhancer-blocking activity at ct^6. Top: schematic diagram of reporter system. The *gypsy* retrotransposon is inserted in between the promoter of *cut* and wing margin (En_{Wing}) enhancer. Bottom: insulator activity for ct^6 was scored in male flies on a scale of 0-4. 0, no notching; 1, slight notching in anterior tip of wing; 2, mild notching throughout posterior wing; 3, extensive notching both in anterior and posterior wing; 4, severe notching throughout the anterior and posterior wing.

(C) Graph represents quantification of ct^6 wing phenotype of male wild-type (+/+) and $M1BP^{RNAi}$ using *Ser-Gal4* driver. n, total number of flies scored.

(D) Depletion of M1BP shows reduced *gypsy*-dependent barrier activity in all tissues tested. Schematic diagram of non-insulated *UAS-luciferase* system shows spreading of repressive chromatin can reduce luciferase expression, but presence of the *gypsy* insulator acts as a barrier and allows for luciferase activity.

(E) Relative luciferase activity of insulated or non-insulated male larvae of control and *M1BP*^{RNAi} driven by *Act5C-Gal4* driver,

(F) *Mef2-Gal4* driver, and

(G) *l(3)31-1-Gal4* driver. Luciferase values of all genotypes are plotted as box and whisker plots using one-way ANOVA followed by Tukey HSD post hoc tests to calculate *P*-values for pairwise comparisons. The box represents the 25–75th percentiles, and the median is indicated. The whiskers show the minimum and maximum values. For each genotype, n=12 individual larvae. Bracket indicates *P*-values of two-way comparisons (**P* <0.5) and all comparisons are shown in Supplementary Table S3.

Fig. 3. M1BP extensively co-localizes specifically with CP190 throughout the genome.

(A) Western blotting of total lysates from Kc control and *M1BP* knockdown cells showing knockdown efficiency of *M1BP* and no effects on protein levels of insulator proteins, with Pep as loading control.

(B) Screenshot example of ChIP-seq profiles showing in Kc cells M1BP co-localizing with CP190 but not Su(Hw) or Mod(mdg4)67.2. Statistically significant decreased CP190 ChIP-seq peaks after depletion of M1BP are shown in black. Asterisk indicates a particular peak of interest measured in Fig. 6C, site 7.

(C) Binary heatmap of M1BP, CP190, Mod(mdg4)67.2 and Su(Hw) binding sites in control cells ordered by supervised hierarchical clustering. Each row represents a single independent genomic location, and a black mark in a column represents the presence of a particular factor: 2461 (79%, purple bar) peaks of 3121 total M1BP peaks overlap with CP190, 163 (5.2%) M1BP peaks overlap with Su(Hw) peaks and 217 (7%) M1BP peaks overlap with Mod(mdg4)67.2.

(D) Bar plot shows distribution of M1BP, CP190 and *gypsy* (Su(Hw)/Mod(mdg4)67.2/CP190 all present) binding sites with respect to genomic features.

(E) Heatmaps of M1BP, CP190, Su(Hw), and Mod(mdg4)67.2 peaks in control cells sorted by decreasing average ChIP-seq signals normalized to inputs. Reads are centered on TSS containing Motif 1 site. The horizontal axis corresponds to distance from TSS with Motif 1 site.

(F) Motif 1 consensus sequence.

Fig. 4. M1BP and CP190 transcriptionally regulate a common set of genes.

(A, B) MA plots showing changes in neuRNA levels upon depletion of M1BP. Statistically significant changes include 1315 up-regulated genes (red) and 607 down-regulated genes (blue) using $p_{adj} < 0.05$. Unchanged genes are indicated in grey. Gene promoters containing M1BP peaks (A) or CP190 peaks (B) are additionally colored yellow.

(C, D) MA plots showing neuRNA-seq affected genes after depletion of CP190. Significantly up-regulated genes (1894, red) and down-regulated genes (1382, blue) are shown. Genes containing CP190 (C) or M1BP (D) at their promoters are shown in yellow.

(E, F, G) Scatter plots comparing neuRNA-seq profiles of *M1BP* and *Cp190* knockdowns (E), *M1BP* and *mod(mdg4)* knockdowns (F), or *Cp190* and *mod(mdg4)* knockdowns (G). Pearson's R corresponds to correlation coefficient between two profiles. Common up regulated genes are indicated in red, and common down regulated genes are indicated in blue.

(H) Differentially decreased CP190 peaks in $M1BP^{RNAi}$ associated with the promoter of downregulated genes affected in both *M1BP* and *Cp190* knockdowns are verified by ChIP-qPCR. Percent input DNA precipitated is shown for each primer set. Error bars indicate s.d. from two replicates. ns, not significant; $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$ by t-test. Detailed description of each site is summarized in Supplementary Table S6.

Fig. 5. M1BP and CP190 both activate Motif 1-containing gene promoters.

(A) Schematic diagram of plasmids containing wild type and Motif 1 mutants of ribosomal protein (RP) gene promoter driving luciferase reporters.
(B) Schematic diagram showing RP gene luciferase reporter assay. *RpIII128* promoter lacks Motif 1 and serves as a transfection control.
(C) Firefly/Renilla luciferase ratio of relative light unit measurements. Experiments are performed in indicated knockdown conditions. Error bar indicates s.d. from two replicates. ns, not significant; *P ≤0.05; **P ≤0.01; ***P ≤0.001 by t-test.

Fig. 6. M1BP association to chromatin is CP190-dependent at a subset of sites.
(A) Western blotting of total lysates from Kc control and *Cp190* knockdown cells to show knockdown efficiency of *Cp190* and no effect on M1BP level, with Pep used as loading control.
(B) Example screenshot of ChIP-seq profiles for lost CP190 peak in $M1BP^{RNAi}$ and lost M1BP peak in $Cp190^{RNAi}$ are shown. Asterisk indicates the particular interdependent peak measured in (C), site 1. Called ChIP-seq peaks in either knockdown condition shown in black are significantly decreased in the knockdown of the opposite factor.
(C) Differentially bound M1BP and CP190 ChIP-seq peaks in *M1BP* knockdown and *Cp190* knockdown validated by ChIP-qPCR. Validation of selected differentially decreased peaks of M1BP and CP190 (Sites 1-5) and negative control sites (Sites 10-11). Validation of decreased peaks of CP190 (Sites 6-9) and negative control site 12. Percentage input chromatin DNA precipitated is shown for each primer set, and error bars indicate s.d. of two biological replicates. Each measurement corresponds to four technical replicates. ns, not significant; *P ≤0.05; **P ≤0.01; ***P ≤0.001 by t-test. Detailed description of each site labeled is summarized in Supplementary Table S6.
(D) Example screenshot showing that promoter association of M1BP and CP190 is interdependent for down regulated genes in either knockdown of *M1BP* or *Cp190*. In most cases, both proteins localize at TAD borders. For simplicity, only replicate 1 of mock, $M1BP^{RNAi}$ and $Cp190^{RNAi}$ for are shown for neuRNA-seq tracks. Called ChIP-seq peak in either knockdown condition shown in black is significantly decreased in the knockdown of the opposite factor.

(E) Cumulative histograms of promoter distance from closest TAD border classified by change in nascent expression in *M1BP* (left) or *Cp190* (right) knockdown cells. Downregulated (blue), upregulated (red), or unchanged (black) genes are indicated. Table indicates the median distance from closest TAD border for down and upregulated genes compared to unchanged genes in either knockdown condition. Mann-Whitney U test for each set of changed genes against unchanged genes are shown. Analysis of TADs classified by chromatin state is shown in Supplementary Fig. S6.

(F) Percentage of precipitated input chromatin DNA from ChIP-qPCR of M1BP, Su(Hw), Mod(mdg4)67.2 and CP190 at Su(Hw)-binding sites of *gypsy* or *TART* transposon sites as a negative control in Kc cells either mock-treated or subjected to $M1BP^{RNAi}$ or $Cp190^{RNAi}$.

Fig. 7. Knockdown of *M1BP* alters nuclear localization of insulator bodies.

(A) Epifluorescence imaging of insulator body localization using anti-CP190 in whole-mount brain, eye, leg or wing imaginal disc tissues. *M1BP* knockdown is driven by *Act5C-Gal4* driver. Insets show zoom of single nucleus outlined with dashed line in larger panel. Scale bars: 5 μm.

(B) Histograms showing the number of insulator bodies per nucleus in the experiment exemplified in (A). In all tissues, the number of insulator bodies is statistically significantly increased in *M1BP* knockdown (Kruskal–Wallis test; all Benjamini–Hochberg corrected $P < 5\times10^{-17}$).

Fig. 8. Knockdown of *M1BP* increases local genome compaction.

(A) Screenshot of regions detected by A and B probes spaced ~44 kb apart. TADs with state classification, longest isoform of genes, ChIP-seq signals of M1BP and CP190, and called peaks are shown. Upregulated genes in either M1BP or CP190 knockdown are shown in purple or green, respectively. Downregulated genes in either M1BP or CP190 knockdowns are outlined with a box and shown in purple or green, respectively. Called ChIP-seq peaks shown in black are significantly decreased in the knockdown of the opposite factor.

(B) Screenshot of regions detected by C and D probes spaced ~13 kb apart.

(C-D) Left: Representative G1 nuclei labeled with probes shown in A and B, respectively. Single Z slices are shown. Dashed line represents nuclear edge. Center: Zoom of FISH signals. Right: 3D mesh rendering from TANGO.

(E) Average center-to-center distance between probes A-B (left) and C-D (right). Single cell distances were normalized to nuclear radius before population averages were calculated. All averages were normalized to their respective mock controls. *A-B, $p=0.05$, *C-D, $p=0.03$, Paired T-test of means before normalization to controls. Data shown are from G1 cells from four biological replicates. $n>100$ cells were measured per replicate. Error bars show standard error of the mean.

REFERENCES

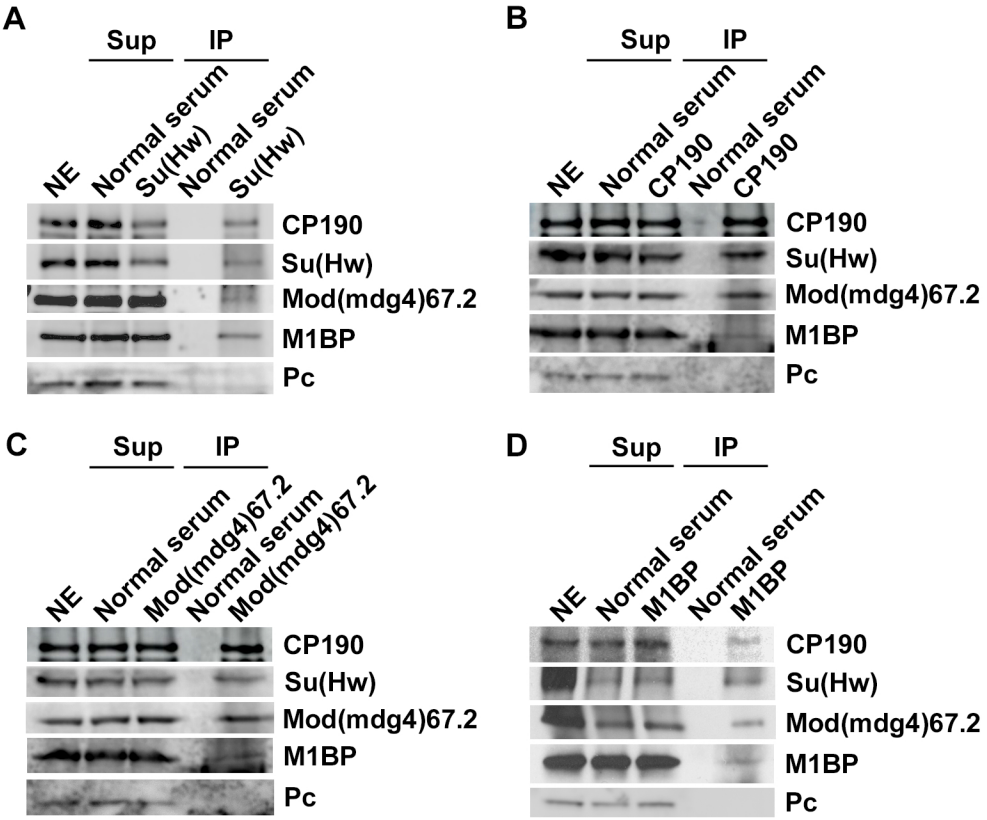
1. Chen, D. & Lei, E.P. Function and regulation of chromatin insulators in dynamic genome organization. *Curr Opin Cell Biol* **58**, 61-68 (2019).
2. Rowley, M.J. & Corces, V.G. Organizational principles of 3D genome architecture. *Nature reviews. Genetics* **19**, 789-800 (2018).
3. Gambetta, M.C. & Furlong, E.E.M. The Insulator Protein CTCF Is Required for Correct Hox Gene Expression, but Not for Embryonic Development in Drosophila. *Genetics* **210**, 129-136 (2018).
4. Bonchuk, A., Denisov, S., Georgiev, P. & Maksimenko, O. Drosophila BTB/POZ domains of "ttk group" can form multimers and selectively interact with each other. *J Mol Biol* **412**, 423-436 (2011).
5. Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O. & Corces, V.G. A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell* **82**, 587-597 (1995).
6. Pai, C.Y., Lei, E.P., Ghosh, D. & Corces, V.G. The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Mol Cell* **16**, 737-748 (2004).
7. Parkhurst, D.F., Wong, S.C., Farquhar, G.D. & Cowan, I.R. Gradients of Intercellular CO(2) Levels Across the Leaf Mesophyll. *Plant Physiol* **86**, 1032-1037 (1988).
8. Brasset, E. & Vaury, C. Insulators are fundamental components of the eukaryotic genomes. *Heredity (Edinb)* **94**, 571-576 (2005).
9. Bushey, A.M., Ramos, E. & Corces, V.G. Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. *Genes Dev* **23**, 1338-1350 (2009).
10. Negre, N. *et al.* A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet* **6**, e1000814 (2010).
11. Roseman, R.R., Pirrotta, V. & Geyer, P.K. The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. *EMBO J* **12**, 435-442 (1993).
12. Spana, C., Harrison, D.A. & Corces, V.G. The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. *Genes Dev* **2**, 1414-1423 (1988).
13. Bag, I., Dale, R.K., Palmer, C. & Lei, E.P. The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in Drosophila. *J Cell Sci* **132** (2019).
14. Cuartero, S., Fresan, U., Reina, O., Planet, E. & Espinas, M.L. Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *EMBO J* **33**, 637-647 (2014).
15. Maksimenko, O. *et al.* Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Res* **25**, 89-99 (2015).
16. Van Bortle, K. *et al.* Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol* **15**, R82 (2014).
17. Gerasimova, T.I. & Corces, V.G. Polycomb and trithorax group proteins mediate the function of a chromatin insulator. *Cell* **92**, 511-521 (1998).
18. Capelson, M. & Corces, V.G. The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol Cell* **20**, 105-116 (2005).

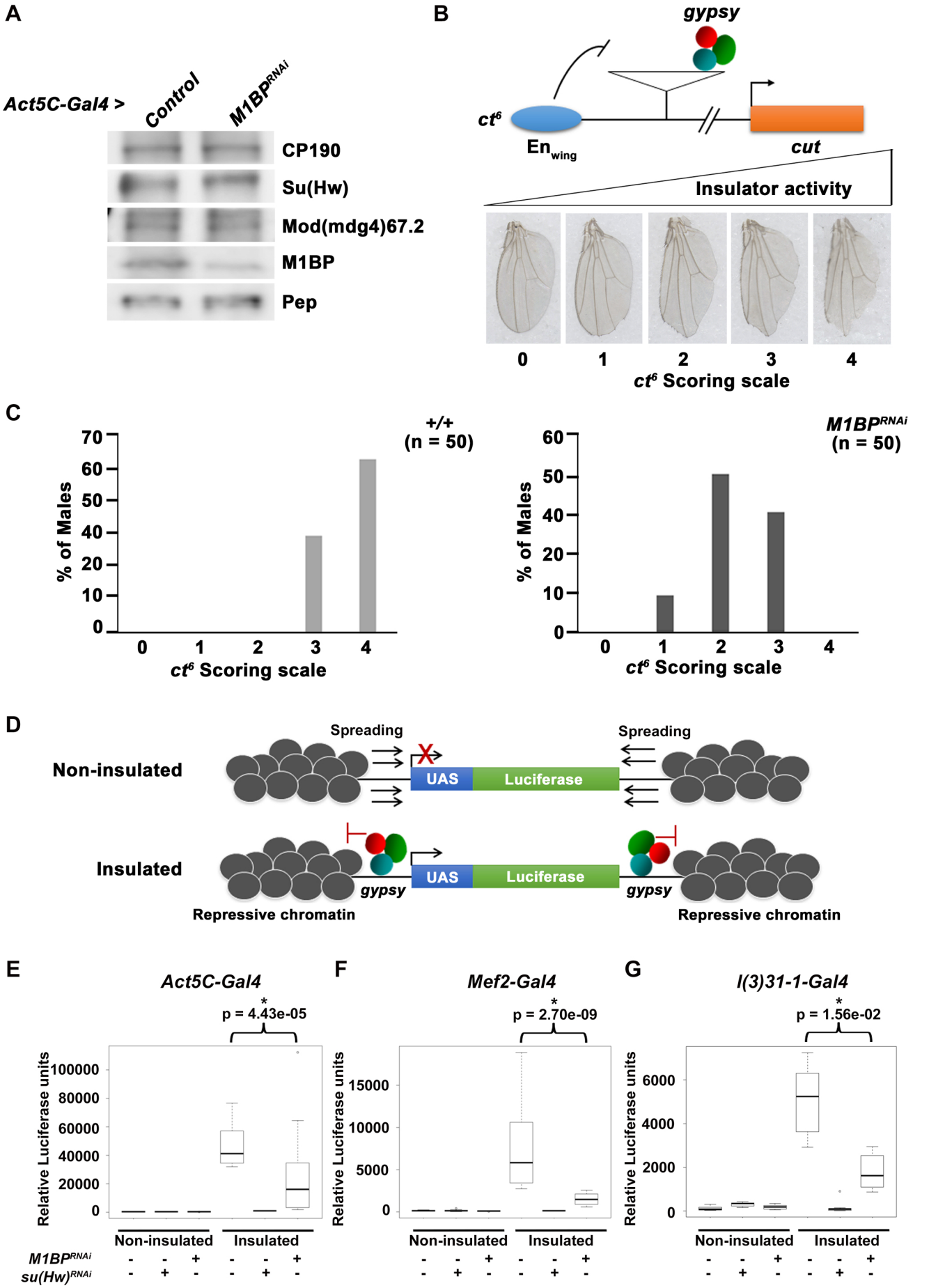
19. Capelson, M. & Corces, V.G. SUMO conjugation attenuates the activity of the gypsy chromatin insulator. *EMBO J* **25**, 1906-1914 (2006).
20. Lei, E.P. & Corces, V.G. RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat Genet* **38**, 936-941 (2006).
21. Golovnin, A., Volkov, I. & Georgiev, P. SUMO conjugation is required for the assembly of Drosophila Su(Hw) and Mod(mdg4) into insulator bodies that facilitate insulator complex formation. *J Cell Sci* **125**, 2064-2074 (2012).
22. Matzat, L.H., Dale, R.K., Moshkovich, N. & Lei, E.P. Tissue-specific regulation of chromatin insulator function. *PLoS Genet* **8**, e1003069 (2012).
23. Matzat, L.H., Dale, R.K. & Lei, E.P. Messenger RNA is a functional component of a chromatin insulator complex. *EMBO Rep* **14**, 916-922 (2013).
24. King, M.R., Matzat, L.H., Dale, R.K., Lim, S.J. & Lei, E.P. The RNA-binding protein Rumpelstiltskin antagonizes gypsy chromatin insulator function in a tissue-specific manner. *J Cell Sci* **127**, 2956-2966 (2014).
25. Lim, S.J., Boyle, P.J., Chinen, M., Dale, R.K. & Lei, E.P. Genome-wide localization of exosome components to active promoters and chromatin insulators in Drosophila. *Nucleic Acids Res* **41**, 2963-2980 (2013).
26. Moshkovich, N. *et al.* RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Genes Dev* **25**, 1686-1701 (2011).
27. Rach, E.A. *et al.* Transcription Initiation Patterns Indicate Divergent Strategies for Gene Regulation at the Chromatin Level. *PLoS Genet* **7**, e1001274 (2011).
28. Li, J. & Gilmour, D.S. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *EMBO J* **32**, 1829-1841 (2013).
29. Baumann, D.G. & Gilmour, D.S. A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. *Nucleic Acids Res* **45**, 10481-10491 (2017).
30. Cubeñas-Potts, C. *et al.* Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic acids research* **45**, 1714-1730 (2017).
31. Ramirez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**, 189 (2018).
32. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472 (2012).
33. Hou, C., Li, L., Qin, Z.S. & Corces, V.G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Molecular cell* **48**, 471-484 (2012).
34. Baumann, D.G. Core promoter function in transcriptional regulation of constitutively expressed genes in Drosophila. *The Pennsylvania State University*, 164 (2018).
35. Gdula, D.A., Gerasimova, T.I. & Corces, V.G. Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. *Proc Natl Acad Sci U S A* **93**, 9378-9383 (1996).
36. Zhou, J., Ashe, H., Burks, C. & Levine, M. Characterization of the transvection mediating region of the abdominal-B locus in Drosophila. *Development* **126**, 3057-3065 (1999).

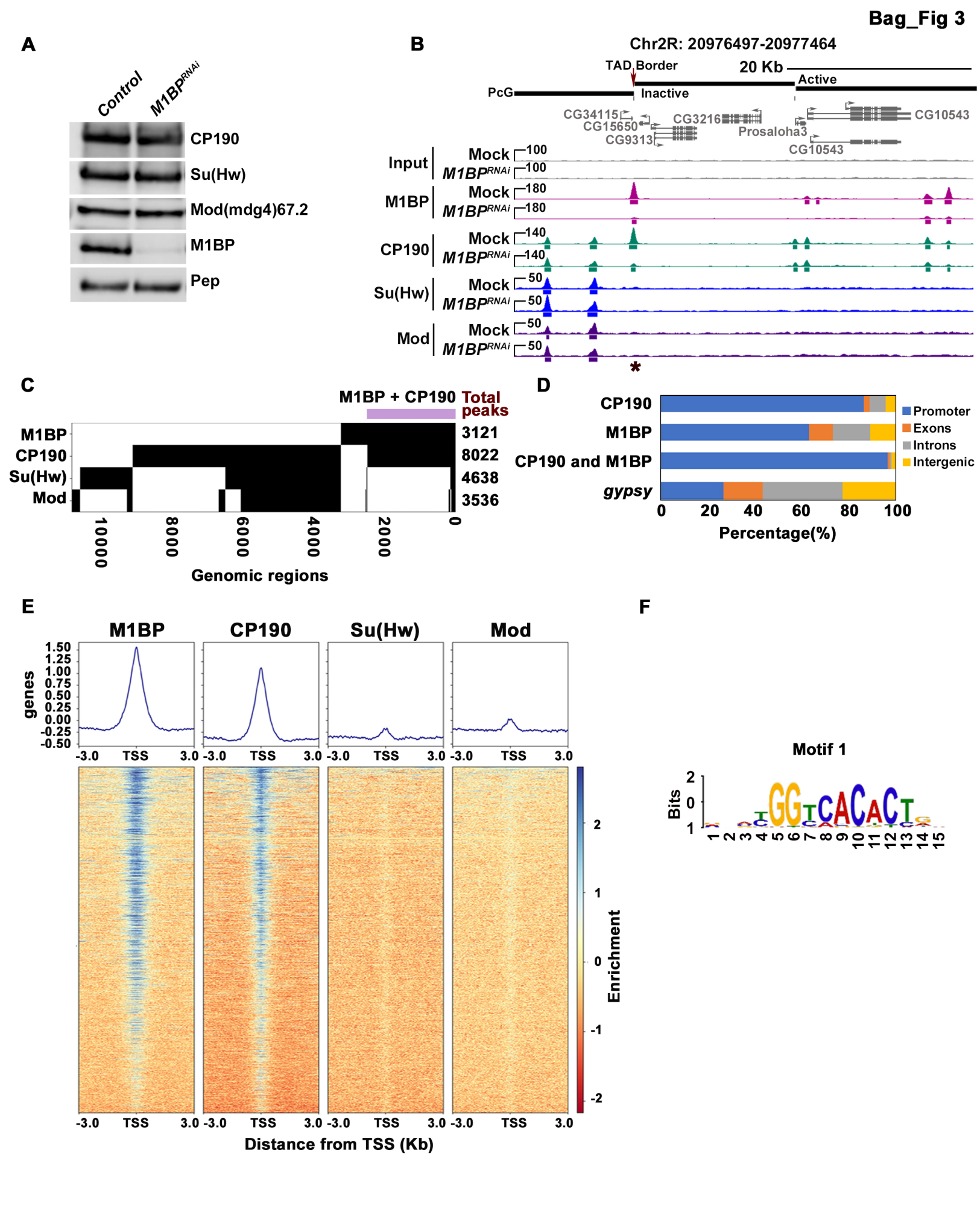
37. Soshnev, A.A., Li, X., Wehling, M.D. & Geyer, P.K. Context differences reveal insulator and activator functions of a Su(Hw) binding region. *PLoS Genet* **4**, e1000159 (2008).
38. Gyurkovics, H., Gausz, J., Kummer, J. & Karch, F. A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. *Embo j* **9**, 2579-2585 (1990).
39. Ohler, U., Liao, G.-c., Niemann, H. & Rubin, G.M. Computational analysis of core promoters in the Drosophila genome. *Genome biology* **3**, RESEARCH0087-RESEARCH0087 (2002).
40. Ross-Innes, C.S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).
41. Fujioka, M., Mistry, H., Schedl, P. & Jaynes, J.B. Determinants of Chromosome Architecture: Insulator Pairing in cis and in trans. *PLOS Genetics* **12**, e1005889 (2016).
42. Mohan, M. *et al.* The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *The EMBO journal* **26**, 4203-4214 (2007).
43. Schwartz, Y.B. *et al.* Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Res* **22**, 2188-2198 (2012).
44. Shrestha, S., Oh, D.H., McKowen, J.K., Dassanayake, M. & Hart, C.M. 4C-seq characterization of Drosophila BEAF binding regions provides evidence for highly variable long-distance interactions between active chromatin. *PLoS One* **13**, e0203843 (2018).
45. Kaye, E.G. *et al.* Differential Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage Compensation Complex to the Male X Chromosome. *Cell Rep* **22**, 3227-3239 (2018).
46. Bartkuhn, M. *et al.* Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J* **28**, 877-888 (2009).
47. Rowley, M.J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Molecular cell* **67**, 837-852.e837 (2017).
48. Hug, C.B., Grimaldi, A.G., Kruse, K. & Vaquerizas, J.M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216-228.e219 (2017).
49. Chopra, V.S., Cande, J., Hong, J.-W. & Levine, M. Stalled Hox promoters as chromosomal boundaries. *Genes & development* **23**, 1505-1509 (2009).
50. Zouaz, A. *et al.* The Hox proteins Ubx and AbdA collaborate with the transcription pausing factor M1BP to regulate gene transcription. *The EMBO journal* **36**, 2887-2906 (2017).
51. Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E. & Perrimon, N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* **40**, 476-483 (2008).
52. Caravaca, J.M. & Lei, E.P. Maintenance of a Drosophila melanogaster Population Cage. *J Vis Exp* (2016).
53. Harlow, E. Using antibodies, A laboratory manual. *Cold spring harbor laboratory press, New York* (1998).

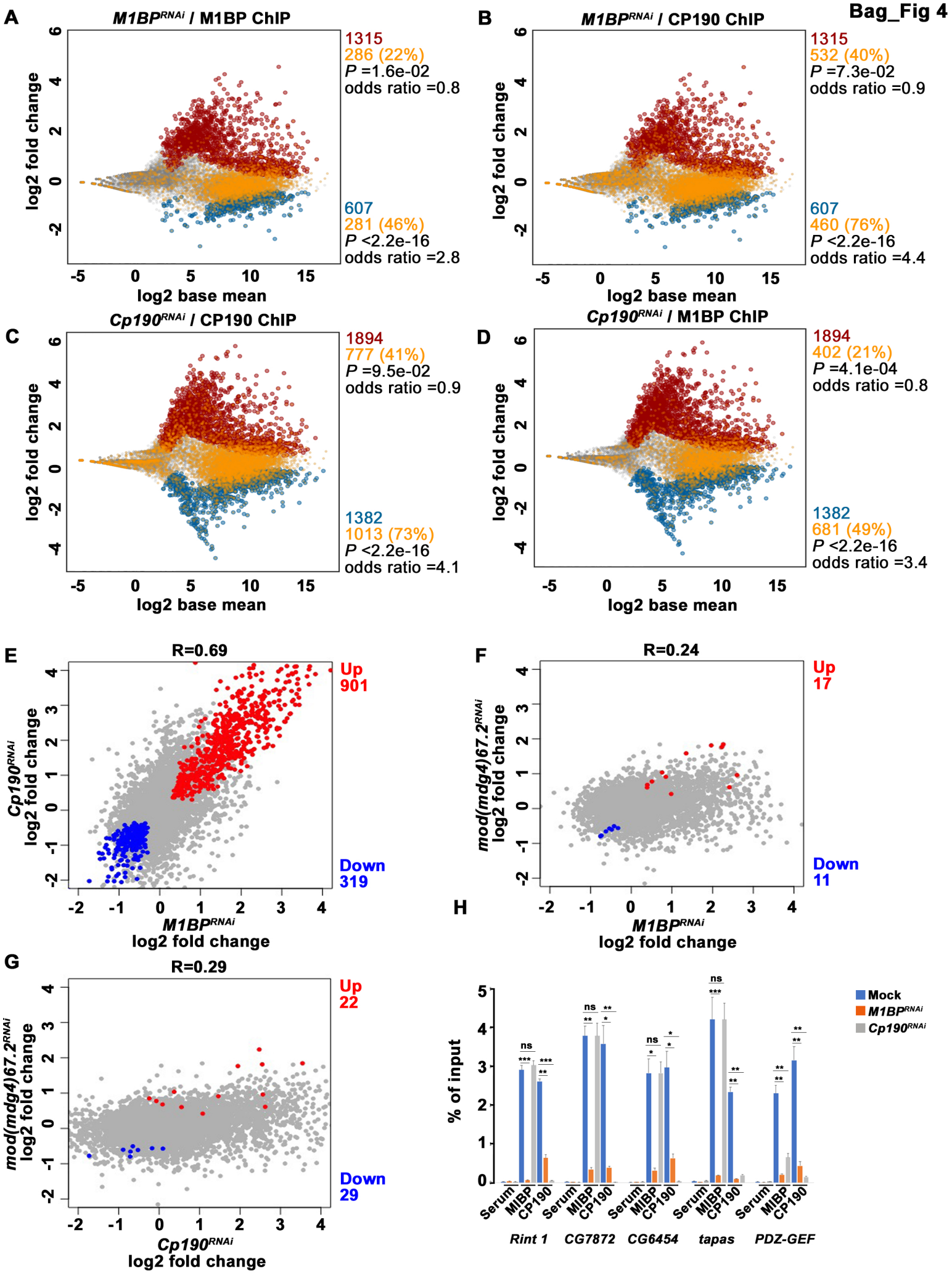
54. Nazer, E., Dale, R.K., Chinen, M., Radmanesh, B. & Lei, E.P. Argonaute2 and LaminB modulate gene expression by controlling chromatin topology. *PLoS Genet* **14**, e1007276-e1007276 (2018).
55. Beliveau, B.J. *et al.* OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proceedings of the National Academy of Sciences* **115**, E2183 (2018).
56. Nguyen, S.C. & Joyce, E.F. Programmable Chromosome Painting with Oligopaints. *Methods Mol Biol* **2038**, 167-180 (2019).
57. Rosin, L.F., Nguyen, S.C. & Joyce, E.F. Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in Drosophila nuclei. *PLoS Genet* **14**, e1007393-e1007393 (2018).
58. Ollion, J., Cochennec, J., Loll, F., Escudé, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840-1841 (2013).
59. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17** (2011).
60. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
61. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
62. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
63. Dale, R.K., Pedersen, B.S. & Quinlan, A.R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423-3424 (2011).
64. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
65. Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907-915 (2019).
66. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2013).
67. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
68. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018 (2011).
69. Hertz, G.Z. & Stormo, G.D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**, 563-577 (1999).

Bag_Fig 1

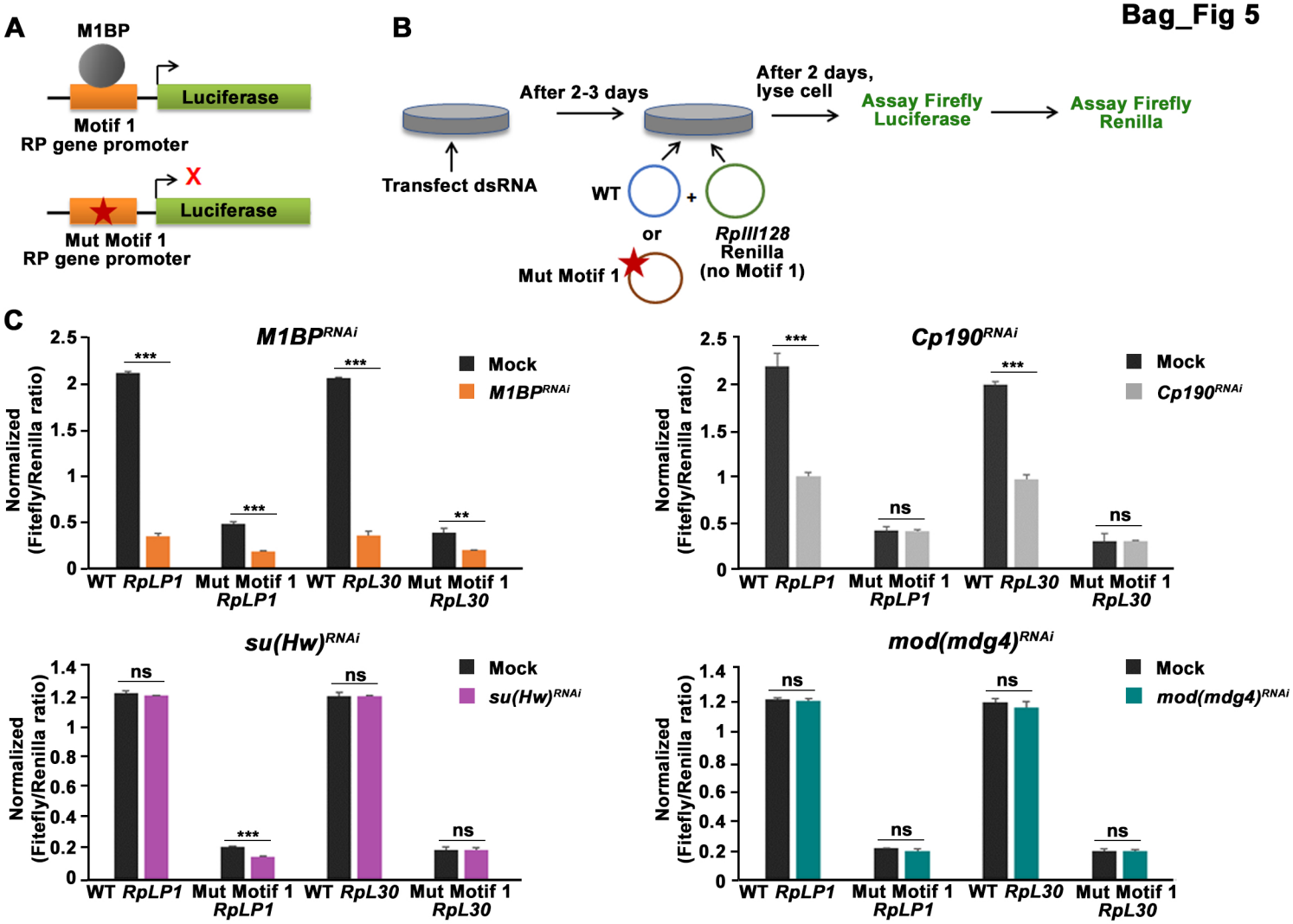

A
Sup
IP
NE
Normal serum
Su(Hw)
Normal serum
Su(Hw)
CP190
Su(Hw)
Mod(mdg4)67.2
M1BP
Pc
B
Sup
IP
NE
Normal serum
CP190
Normal serum
CP190
CP190
Su(Hw)
Mod(mdg4)67.2
M1BP
Pc
C
Sup
IP
NE
Normal serum
Mod(mdg4)67.2
Normal serum
Mod(mdg4)67.2
CP190
Su(Hw)
Mod(mdg4)67.2
M1BP
Pc
D
Sup
IP
NE
Normal serum
M1BP
Normal serum
M1BP
CP190
Su(Hw)
Mod(mdg4)67.2
M1BP
Pc

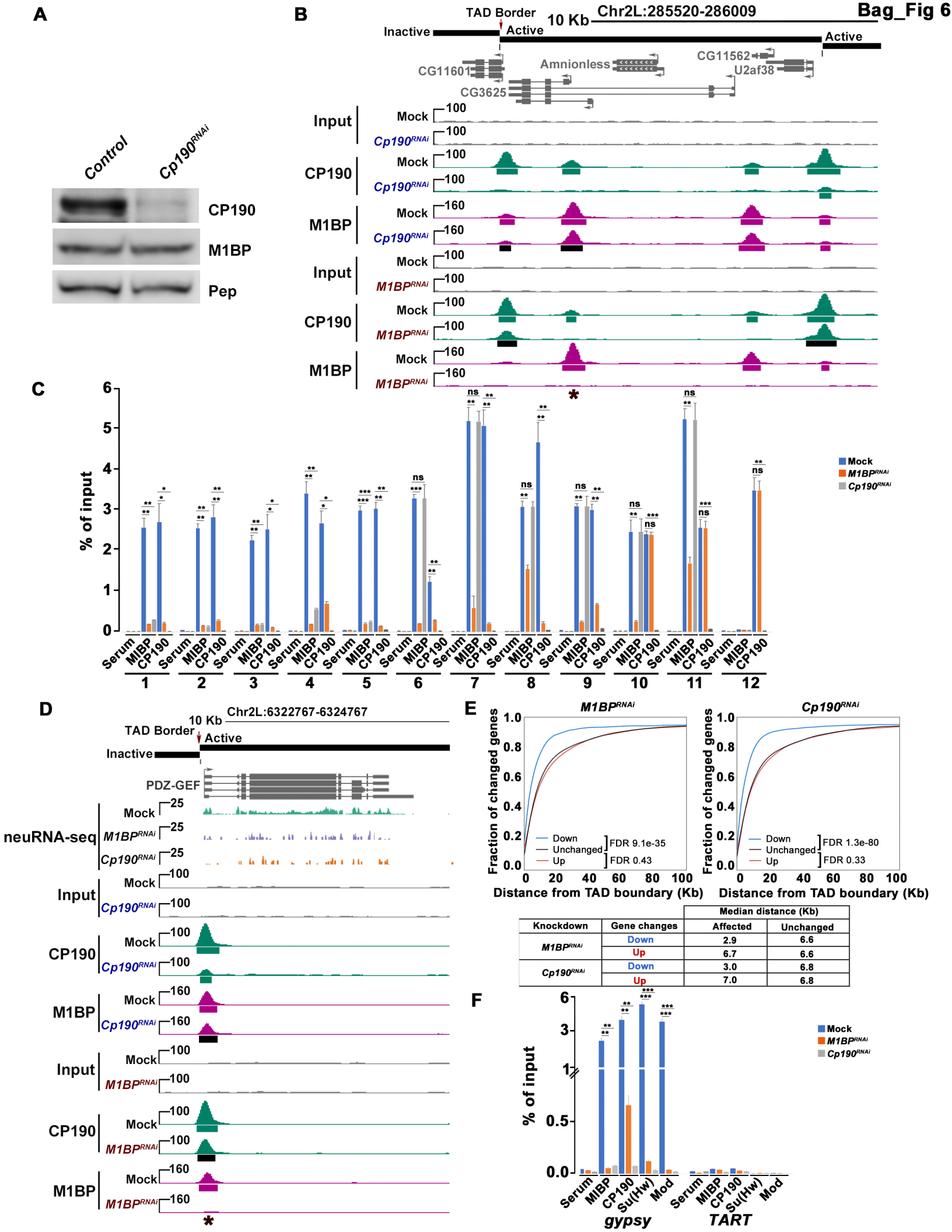

A
Act5C-Gal4 >
Control
M1BP^RNAi
CP190
Su(Hw)
Mod(mdg4)67.2
M1BP
Pep
B
gypsy
ct6
En wing
cut
Insulator activity
0
1
2
3
4
ct6 Scoring scale
C
+/+
(n = 50)
% of Males
ct6 Scoring scale
M1BP^RNAi
(n = 50)
% of Males
ct6 Scoring scale
D
Non-insulated
Spreading
UAS
Luciferase
Spreading
Insulated
gypsy
UAS
Luciferase
gypsy
Repressive chromatin
Repressive chromatin
E
Act5C-Gal4
p = 4.43e-05
Relative Luciferase units
Non-insulated
Insulated
M1BP^RNAi
su(Hw)^RNAi
F
Mef2-Gal4
p = 2.70e-09
Relative Luciferase units
Non-insulated
Insulated
G
l(3)31-1-Gal4
p = 1.56e-02
Relative Luciferase units
Non-insulated
Insulated


A
Control
M1BP RNAi
CP190
Su(Hw)
Mod(mdg4)67.2
M1BP
Pep
B
Chr2R: 20976497-20977464
TAD Border
20 Kb
PcG
Inactive
Active
CG34115
CG15650
CG9313
CG3216
Prosaloha3
CG10543
CG10543
Input
M1BP
CP190
Su(Hw)
Mod
Mock
M1BP RNAi
C
M1BP + CP190
Total peaks
M1BP
CP190
Su(Hw)
Mod
3121
8022
4638
3536
10000
8000
6000
4000
2000
0
Genomic regions
D
CP190
M1BP
CP190 and M1BP
gypsy
Promoter
Exons
Introns
Intergenic
0
20
40
60
80
100
Percentage(%)
E
M1BP
CP190
Su(Hw)
Mod
genes
-3.0
TSS
3.0
Enrichment
Distance from TSS (Kb)
F
Motif 1
Bits


Bag_Fig 4
A
M1BP RNAi / M1BP ChIP
1315
286 (22%)
P =1.6e-02
odds ratio =0.8
607
281 (46%)
P <2.2e-16
odds ratio =2.8
log2 fold change
log2 base mean
B
M1BP RNAi / CP190 ChIP
1315
532 (40%)
P =7.3e-02
odds ratio =0.9
607
460 (76%)
P <2.2e-16
odds ratio =4.4
C
Cp190 RNAi / CP190 ChIP
1894
777 (41%)
P =9.5e-02
odds ratio =0.9
1382
1013 (73%)
P <2.2e-16
odds ratio =4.1
D
Cp190 RNAi / M1BP ChIP
1894
402 (21%)
P =4.1e-04
odds ratio =0.8
1382
681 (49%)
P <2.2e-16
odds ratio =3.4
E
R=0.69
Up
901
Down
319
Cp190 RNAi log2 fold change
M1BP RNAi log2 fold change
F
R=0.24
Up
17
Down
11
mod(mdg4)67.2 RNAi log2 fold change
M1BP RNAi log2 fold change
G
R=0.29
Up
22
Down
29
mod(mdg4)67.2 RNAi log2 fold change
Cp190 RNAi log2 fold change
H
% of input
Mock
M1BP RNAi
Cp190 RNAi
Serum
MIBP
CP190
Rint 1
CG7872
CG6454
tapas
PDZ-GEF
ns

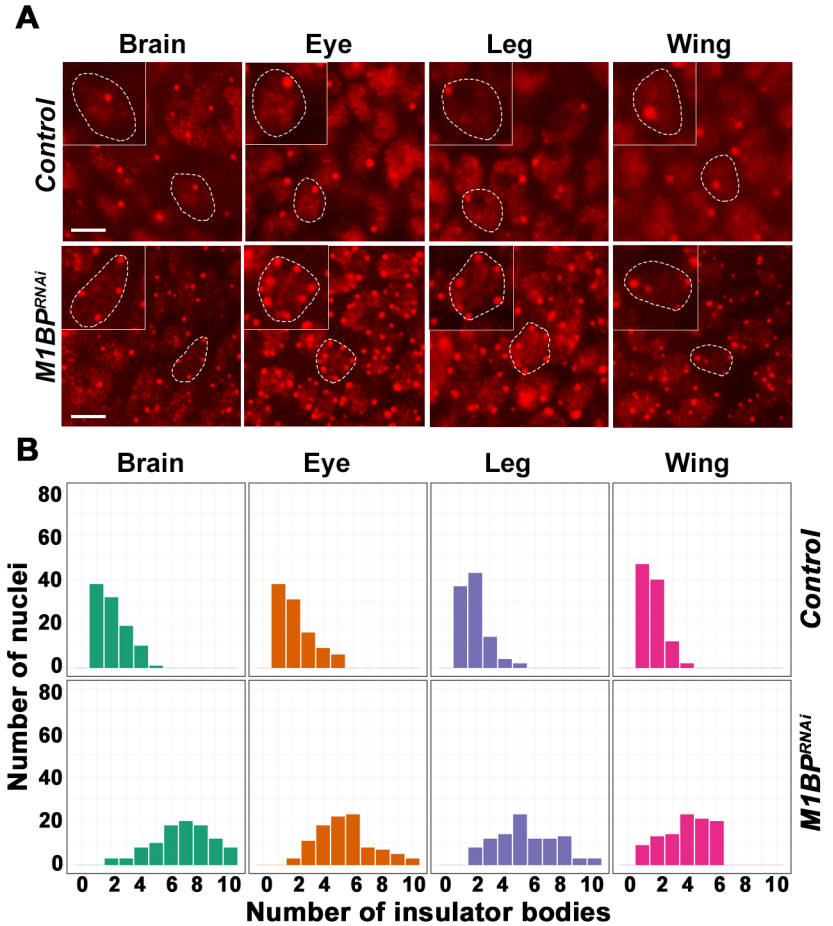
Bag_Fig 5

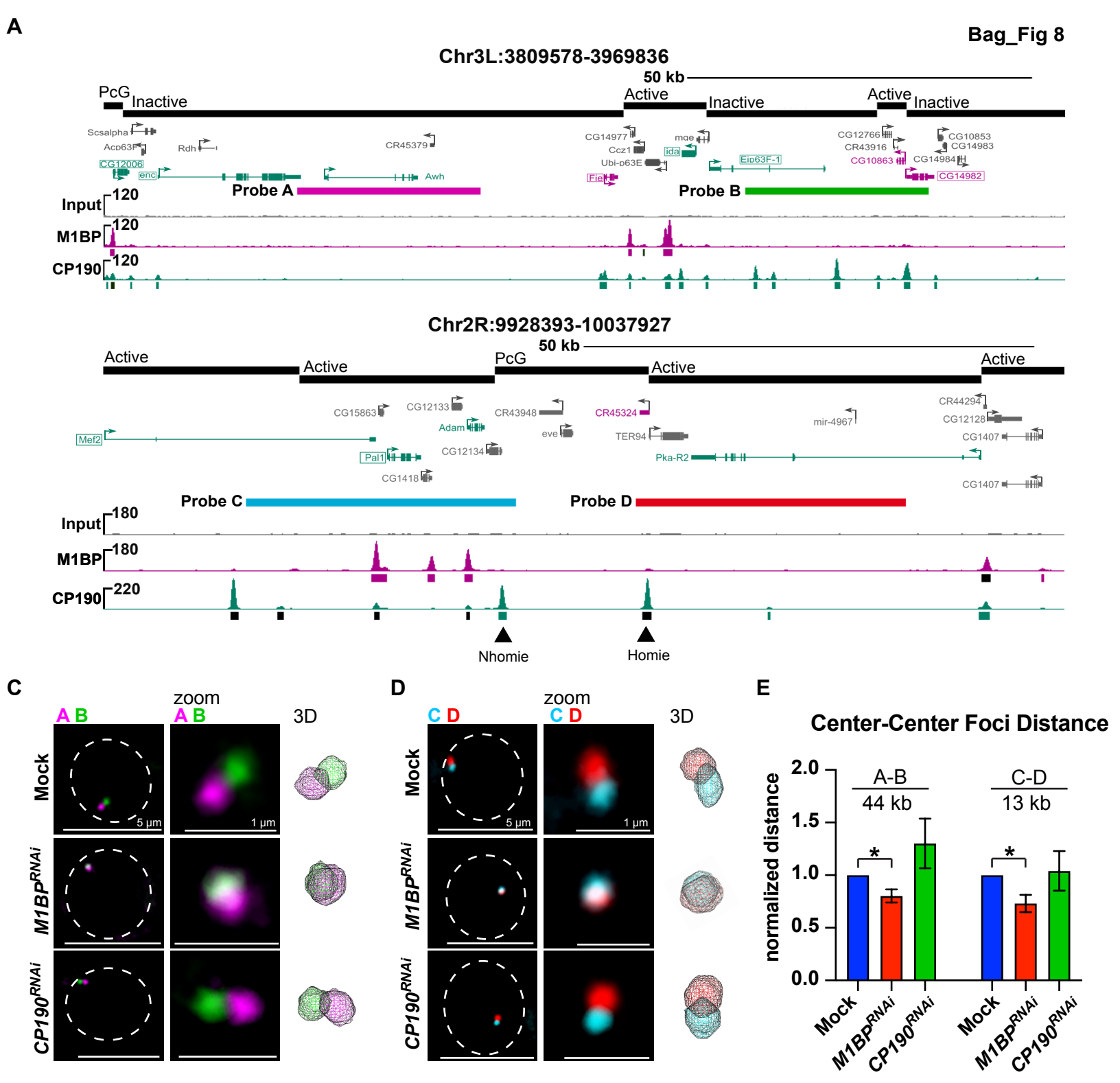

A
M1BP
Luciferase
Motif 1
RP gene promoter
X
Luciferase
Mut Motif 1
RP gene promoter
B
After 2-3 days
After 2 days, lyse cell
Assay Firefly Luciferase
Assay Firefly Renilla
Transfect dsRNA
WT
+
or
Mut Motif 1
RpIII128
Renilla
(no Motif 1)
C
M1BP^RNAi
Mock
M1BP^RNAi
Normalized (Fitefly/Renilla ratio)
WT RpLP1
Mut Motif 1 RpLP1
WT RpL30
Mut Motif 1 RpL30
Cp190^RNAi
Mock
Cp190^RNAi
su(Hw)^RNAi
Mock
su(Hw)^RNAi
mod(mdg4)^RNAi
Mock
mod(mdg4)^RNAi
ns
***
**



Knockdown	Gene changes	Median distance (Kb)	
		Affected	Unchanged
M1BPRNAi	Down	2.9	6.6
	Up	6.7	6.6
Cp190RNAi	Down	3.0	6.8
	Up	7.0	6.8

Bag_Fig 7


A
Brain
Eye
Leg
Wing
Control
M1BP^RNAi
B
Brain
Eye
Leg
Wing
Control
M1BP^RNAi
Number of nuclei
0
20
40
60
80
0 2 4 6 8 10
Number of insulator bodies


A
Bag_Fig 8
Chr3L:3809578-3969836
50 kb
PcG
Inactive
Active
Inactive
Active
Inactive
Probe A
Probe B
Input
M1BP
CP190
Chr2R:9928393-10037927
50 kb
Active
Active
PcG
Active
Active
Probe C
Probe D
Input
M1BP
CP190
Nhomie
Homie
C
zoom
A B
3D
Mock
M1BP^RNAi
CP190^RNAi
5 µm
1 µm
D
zoom
C D
3D
Mock
M1BP^RNAi
CP190^RNAi
5 µm
1 µm
E
Center-Center Foci Distance
normalized distance
A-B
44 kb
C-D
13 kb
Mock
M1BP^RNAi
CP190^RNAi
Mock
M1BP^RNAi
CP190^RNAi