Authors for correspondence:
Pavel Georgiev
e-mail: georgiev_p@mail.ru
Oksana Maksimenko
e-mail: maksog@mail.ru

THE ROYAL SOCIETY PUBLISHING

Interaction of MLE with CLAMP zinc finger is involved in proper MSL proteins binding to chromosomes in *Drosophila*

Evgeniya Tikhonova[1,2], Anastasia Revel-Muroz[1], Pavel Georgiev[2] and Oksana Maksimenko[1]

[1]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilov Street, Moscow 119334, Russia
[2]Department of the Control of Genetic Processes, Institute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilov Street, Moscow 119334, Russia

PG, 0000-0002-6509-3194; OM, 0000-0003-3502-0303

The *Drosophila* male-specific lethal (MSL) complex binds to the male X chromosome to activate transcription. It comprises five proteins (MSL1, MSL2, MSL3, male absent on the first (MOF), and maleless (MLE)) and two long noncoding RNAs (lncRNAs; roX1 and roX2). The MLE helicase remodels the roX lncRNAs, enabling the lncRNA-mediated assembly of the *Drosophila* dosage compensation complex. MSL2 is expressed only in males and interacts with the N-terminal zinc finger of the transcription factor chromatin-linked adapter for MSL proteins (CLAMP), which is important for the specific recruitment of the MSL complex to the male X chromosome. Here, we found that MLE's unstructured C-terminal region interacts with the sixth zinc-finger domain of CLAMP. *In vitro*, 4–5 zinc fingers are critical for the specific DNA-binding of CLAMP with GA repeats, which constitute the core motif at the high affinity binding sites for MSL proteins. Deleting the CLAMP binding region in MLE decreases the association of MSL proteins with the male X chromosome and increases male lethality. These results suggest that interactions of unstructured regions in MSL2 and MLE with CLAMP zinc finger domains are important for the specific recruitment of the MSL complex to the male X chromosome.

1. Introduction

Dosage compensation is a mechanism for compensating gene expression in organisms with an unbalanced number of genes between sexes. In *Drosophila*, sex is determined by the number of X chromosomes: females have two X chromosomes, while males only have one. Dosage compensation is realized via an increase in the expression of male genes on the X chromosome to which the RNA–protein complex that activates transcription binds specifically [1–6]. *Drosophila*'s *d*osage *c*ompensation *c*omplex (DCC) comprises five proteins whose gene mutations are lethal in males. These proteins are called *m*ale-*s*pecific *l*ethal 1 (MSL1), 2 (MSL2), and 3 (MSL3); maleless (MLE); and *m*ale absent *o*n the *f*irst (MOF). The complex also includes two long noncoding RNAs (lncRNAs; roX1 or roX2). MSL1, MSL3, MLE, and MOF are present in both sexes, with only MSL2 and both roX RNAs exclusively expressed in males.

MSL1's C-terminal PEHE (with characteristic amino acidic residues proline (P), glutamic acid (E), histidine (H)) domain is responsible for interacting with MSL3 and MOF [7–10], which together form the functional acetylase module. MSL1's N-terminal coiled-coil domain forms a homodimer that interacts with two MSL2 proteins [7]. It was found that the N-terminal helixes and RING finger (zinc binding domain of C3HC4 involved in ubiquitination) of MSL2 are required for interaction with MSL1 [7]. The MSL1/MSL2 core can
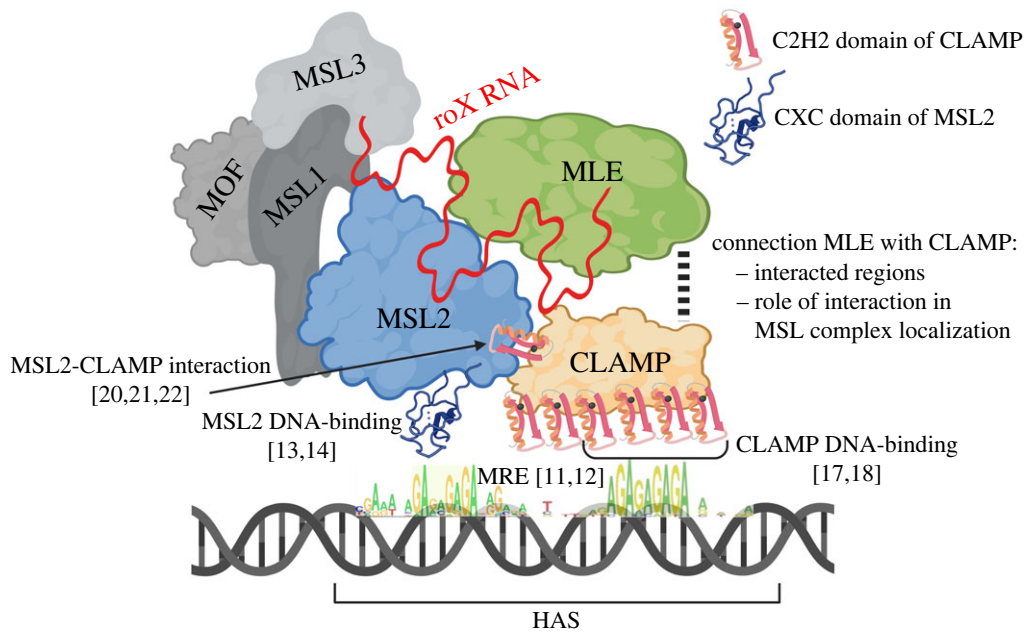
Figure 1. Schematic representation of MSL complex comprising MSL1, MSL2, MSL3, MOF, MLE, CLAMP proteins, and roX lncRNA, recruited to the HAS/CES on X chromosome. MSL2 (via CXC domain) and CLAMP (via 4–7 C2H2 domains) proteins are able to bind with (GA)-rich sequences, which are represented as part of the HAS in MRE. C2H2 domain of CLAMP directly interacts with unstructured region of MSL2. References to relevant works are presented in square brackets.

specifically bind to a reproducible set of about 200 sites on the X chromosome, which were called the primary chromatin entry (CES) [11] or high-affinity (HAS) sites [12]. A GA-rich sequence motif within the HAS/CES, the MSL recognition element (MRE), is important for DCC targeting [11,12].

MSL2 contains the conserved cysteine-rich CXC domain formed by nine cysteines coordinated by three Zn ions (Zn3Cys9) that is the only DNA-binding domain found in the *Drosophila* MSL complex [13,14]. The CXC domain binds with high specificity to MREs *in vitro* [14–16]. Ubiquitous transcription factor *c*hromatin-*l*inked *a*daptor for *M*SL *p*roteins (CLAMP) binds also to GA-rich sequences (MREs) in most of the HAS/CES [17–19] and directly interacts with MSL2's *C*LAMP *b*inding *d*omain (CBD), located distally of its CXC domain [20–22]. It is proposed that the interaction between CLAMP and MSL2 is important for the specific binding of the MSL complex to the male X chromosome [20,21,23]. The CXC domain and CBD of MSL2 are jointly required for the recruitment of MSL to the male X chromosome (figure 1) [13,20,21,23]. However, the question remains how CLAMP specifically interacts with the MSL complex associated with HAS, since there are more than ten thousand CLAMP binding sites on all chromosomes [18].

MLE, an ATP-dependent RNA/DNA helicase of the DEAD subfamily, interacts with the two noncoding roX lncRNAs with high specificity and induces their unwinding [24–30], enabling them to bind to MSL2 [31,32] and potentially other MSL proteins [24,25,32,33]. MSL2's unstructured C-terminal portion is required to integrate the roX lncRNAs into the MSL complex [31,34,35]. Without MSL2, the roX lncRNAs are not included in the MSL1-MSL3-MOF subcomplex [7]. By interacting with MSL proteins, the roX lncRNAs stimulate the assembly of complete active complexes [32–35]. MLE is mainly included in the MSL complex through its interaction with the roX lncRNAs [25,36]. A model has been proposed in which the roX lncRNAs form condensates around the male X chromosome *in vivo* that concentrate MSL proteins, leading to the subsequent specific binding of the MSL complexes to the X chromosome [34,35]. MLE has also been implicated in the spread of the DCC along the X chromosome from the HAS [29,30,37,38]. ChIP-seq binding profiles for all MSL subunits in S2 cells suggest that MSL2 and MLE determine the binding of the MSL complex to HAS [39]. In accordance with the assumption that the protein is involved in the recruitment of the MSL complex to the X chromosome, it was shown that MLE interacts with CLAMP in coimmunoprecipitation from extracts isolated from the CME W1 Cl.8+ cell line [25]. Also the deletion of either the CBD or CXC domain in MSL2 does not significantly affect specific recruiting of MSL on the X chromosome [21] suggesting that other MSL subunits can be responsible for association of the complex with the X chromosome.

The main goal of this study is to elucidate the potential role of MLE in the recruitment of the MSL complex to the X chromosome of males (figure 1). Using the yeast two-hybrid (Y2H) and coimmunoprecipitation assays, we show that MLE's unstructured C-terminal region interacts with CLAMP's six zinc-finger domain. ChIP-seq analysis showed that the deletion of the CBD in MLE affects the recruitment of the MSL complex to the male X chromosome in adult flies.

2. Results

2.1. MLE interacts with CLAMP's sixth C2H2 domain

The 1293 amino acid (aa) ATP-dependent DEXH box RNA/DNA helicase MLE (figure 2a) contains two N-terminal RNA-binding domains (*d*ouble-*s*tranded *R*NA (dsRNA) *b*inding *d*omain 1 (dsRBD1) and 2 (dsRBD2)) [36], two highly conserved RecA
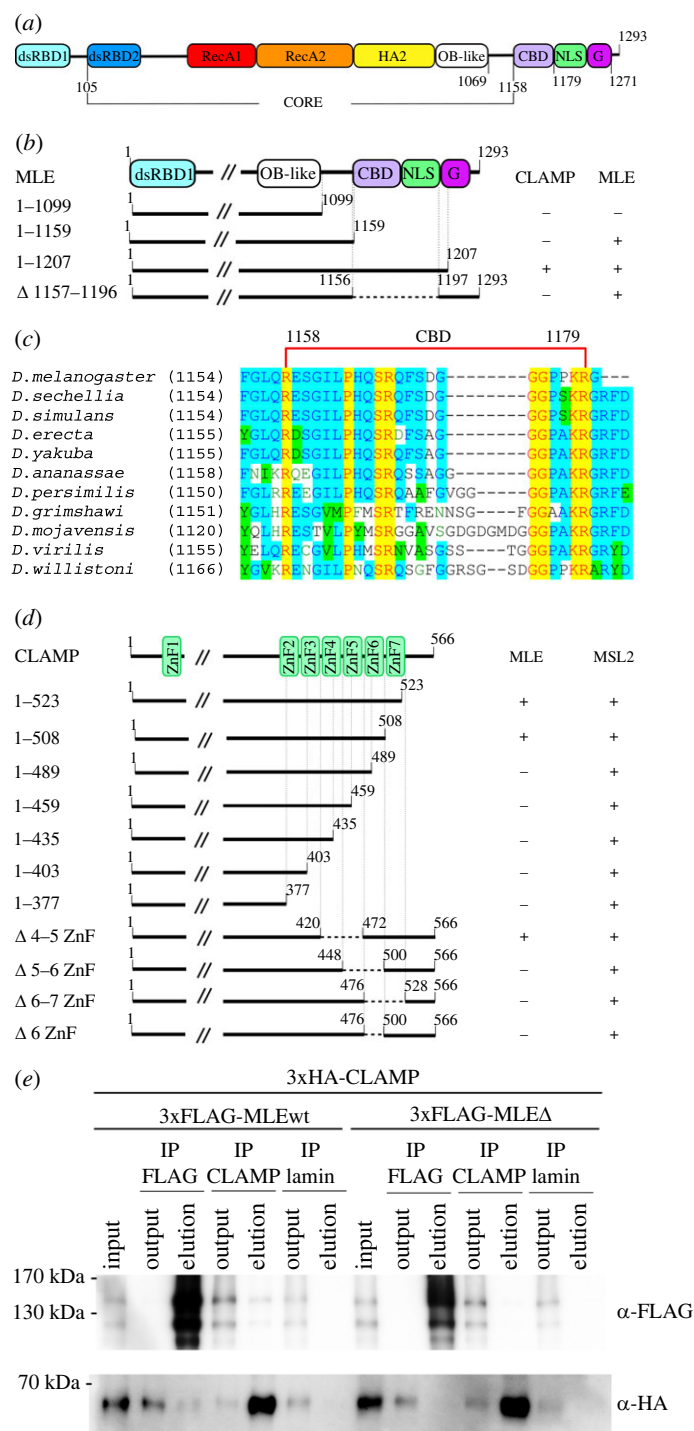
Figure 2. The mapping of interacting regions in CLAMP and MLE. (a) Schematic representation of the MLE protein. The dsRBD2, RecA1, RecA2, HA2, and OB-fold domains form the globule core of MLE. CBD, CLAMP binding domain; NLS, nuclear localization signal; G, GGGYGNN heptad repeats. (b) The mapping of the MLE domain interacting with CLAMP in the Y2H assay. Different MLE fragments were fused to the GAL4 activating domain (AD) and tested for interactions with CLAMP or MLE fused to the GAL4 DNA-binding domain (DBD). All MLE fragments were tested for the absence of interaction with the GAL4 DBD alone. (c) Alignment of the CBD of MLE among Drosophilidae. (d) The localization of the CLAMP domain interacting with MLE in the Y2H assay. ZnF, C2H2-type zinc finger domains. Different CLAMP fragments were fused to the GAL4 AD and tested for interactions with MLE or MSL2 fused to the GAL4 DBD. All CLAMP fragments were tested for the absence of interaction with the GAL4 DBD alone. (e) Total extracts from *Drosophila* S2 cells co-transfected with 3xHA-CLAMP and 3xFLAG-MLEwt (MLE^WT) or 3xFLAG-MLEΔ (MLE^ΔCBD) were immunoprecipitated with antibodies against FLAG, CLAMP or lamin (used as a negative control), and the immunoprecipitates were analysed by immunoblotting for the presence of FLAG- or HA-tagged proteins. Immunoprecipitates after elution are shown concentrated relative to the input by a factor of 5.

domains [40], a *h*elicase *a*ssociated 2 (HA2) domain, and an *o*ligonucleotide-*b*inding (OB)-fold domain. The region between 105 and 1158 aa forms the core of the MLE protein (figure 2a). At MLE's C-terminus is an unstructured region followed by imperfect GGGYGNN heptad repeats that may be involved in nonspecific RNA binding [41]. MLE's C-terminal region has been shown to contain a nuclear localization signal (NLS) and have dimerization activity [41,42].

A potential interaction between MLE and CLAMP was demonstrated only via co-immunoprecipitation and might be mediated by MSL2, which directly interacts with CLAMP. Therefore, we first confirmed that MLE and CLAMP interact directly in the Y2H

assay (figure 2b; electronic supplementary material, figure S1). We also repeatedly demonstrated that MLE can form dimers. Next, we mapped regions in MLE that interact with CLAMP. Several MLE variants with deletions in the C-terminal region were tested for interaction with CLAMP or MLE. We found that the 1159 aa MLE can interact with wild-type (WT) MLE but not CLAMP; the 1099 aa MLE lost both interaction abilities. Available data [40] showed that the region between 1099 and 1159 aa is included in the overall globular structure of the protein along with all other domains; therefore, deletion of the 1099–1158 aa region is highly likely to result in an incorrectly folded MLE protein. The 1207 aa MLE could interact both with CLAMP and MLE. Deleting the 1158–1195 aa region in MLE results in the loss of the interaction with CLAMP but preserves homodimerization activity. According the MLE structure [40], the 1158–1195 aa region is located outside of the globular core. The 1158–1195 aa region contains moderately conserved sequences only among MLEs of *Drosophila* species (figure 2c; electronic supplementary material, figure S2) and directly flanks the potential NLS. The MLE region between 1158 and 1195 aa was called the CBD.

Next, we attempted to map the region in CLAMP responsible for its interaction with MLE (figure 2d). CLAMP is a 566 aa transcription factor that contains a single C2H2-type N-terminal zinc finger (C2H2 domain) and a C-terminal cluster of six C2H2 domains (figure 2d). We previously found that the N-terminal C2H2 domain is required for its interaction with MSL2 [21], while most of the N-terminal intrinsically disordered region (46–86 aa) is responsible for homodimerization [43]. Several C-terminally truncated CLAMP variants were tested for interaction with MLE in the Y2H assay (figure 2d; electronic supplementary material, figure S1). The interaction of CLAMP variants with MSL2 was used as a control. The MLE interaction region was mapped to the region of the 6–7 C2H2 domains. Next, we tested interactions for several CLAMP variants with deletion of the 4–5, 5–6, 6–7, and 6 C2H2 domains. Only deleting the sixth C2H2 domain affected the interaction between CLAMP and MLE (figure 2d). These results suggest that MLE's unstructured region interacts with the sixth C2H2 domain of CLAMP.

Finally, we confirmed the interaction between the proteins by co-immunoprecipitation of 3×HA-tagged CLAMP and 3xFLAG-tagged MLE^{WT} in co-transfected S2 cells. At the same time 3xFLAG-tagged MLE$^{\Delta CBD}$ (deletion of 1158–1179 aa) protein demonstrated the absence of co-precipitation with 3xHA-CLAMP (figure 2e). Taken together the results suggest that CLAMP interacts with MLE's unstructured region.

2.2. CLAMP's 4–5 C2H2 domains are critical for its specific binding to GA-rich motifs

According to a previous study on CLAMP–DNA interaction [18], the sixth domain of C2H2, which interacts with the MLE protein, may also be involved in the DNA binding activity of CLAMP. Therefore, the next task was to map the C2H2 domains that determine the specific CLAMP–DNA interaction. The CLAMP motif was previously discovered via protein binding microarrays assaying CLAMP's binding to all possible 10-bp sequences [18]. CLAMP's 2–7 and 4–7 C2H2 domains bind efficiently to the same (GA)$_4$ motif *in vitro* (figure 3a). In conventional C2H2 zinc finger proteins, each finger interacts mainly with three adjacent DNA base pairs [44]. The C2H2 domains can be linked in tandem to occupy DNA of varying lengths. The predicted DNA-binding specificity of CLAMP's 2–7 C2H2 domains [45] showed that the 4 and 5 C2H2 domains can jointly bind to the GAGA motif (figure 3a), while the 6 C2H2 domain potentially recognizes GA.

To further map the C2H2 domains responsible for binding to the CLAMP motif, we tested different combinations of CLAMP's C2H2 domains for their ability to interact specifically with DNA (figure 3b). In this study, we used the 199 bp CES5C2 corresponding to CLAMP's strong binding region, which is essential for recruiting the MSL complex [11,17]. CES5C2 contains three MRE motifs (figure 3b). In addition, the (GA)$_{24}$ (64 bp) sequence was used as an alternative CLAMP binding region. Different combinations of C2H2 domains (2–7, 4–7, 2–5, 2–7Δ3, 2–7Δ4, and 2–7Δ4–5) attached to maltose-binding protein (MBP) were expressed in bacteria (figure 3c). The electrophoretic mobility shift assay confirmed that both the 2–7 and 4–7 C2H2 domains bind to the tested sequences (figure 3d). However, the 2–7 C2H2 domains bind to both tested DNAs with greater affinity. The 2–5 C2H2 domains bound to the DNA fragments more strongly than the 4–7 C2H2 domains. Deleting the single C2H2 domains in the 2–7 C2H2 array showed that the 4 C2H2 domain is essential for specific DNA binding. Deleting the 4 and 5 C2H2 domains eliminated the binding to the DNA fragments. These results are consistent with the prediction [45] (http://zf.princeton.edu), suggesting that the 4–5 C2H2 domains are required for CLAMP's specific binding to chromatin sites. Therefore, CLAMP's 6–7 C2H2 domains are not essential for specific DNA binding and may be involved in its interactions with proteins.

2.3. MLE's CLAMP binding domain is required for dosage compensation

To evaluate the functional effects of deleting the CBD in MLE *in vivo*, we created transgenic flies expressing the MLE variants tagged with the FLAGx3 epitope: MLE^{WT}-FLAG (WT) and MLE$^{\Delta CBD}$-FLAG (1158–1179 aa deletion). The MLE variants were expressed under the control of a strong ubiquitin-p63E (*Ubi*) promoter (figure 4a). In order to avoid the influence of position effects on the expression of MLE^{WT} and MLE$^{\Delta CBD}$, both transgenes were inserted into the same genomic location (86Fb) on chromosome 3 using a φC31-based integration system [46].

To test the role of the CBD in the specific recruitment of the MSL complex to the X chromosome in males, we crossed both transgenes into the null *mle* background. In this study, we used the previously characterized loss-of-function mutation in the *mle* gene, *mle^9* (*mle^{y38}*) [47,48]. Next, we compared the expression of MLE^{WT}-FLAG and MLE$^{\Delta CBD}$-FLAG in 2- to 3-day-old *mle^9* males (figure 4b).

Immunoblot analysis showed that *Ubi:mle^{WT}* transgene in a heterozygous state was expressed at a similar level to the *Ubi:mle$^{\Delta CBD}$* transgene in a homozygous state. Next we compared expression of the MLE variants in transgenic lines and MLE in the *y^{1}w^{1118}* line (figure 4b). In *mle^9/mle^9;Ubi:mle$^{\Delta CBD}$/Ubi:mle$^{\Delta CBD}$* and *mle^9/mle^9;Ubi:mle^{WT}/TM6* males MLE variants are expressed at least twice as strongly as MLE in *y^{1}w^{1118}* males. As expected in *mle^9/mle^9; Ubi:mle^{WT}/Ubi:mle^{WT}* males MLE are expressed much stronger. We also compared MSL1 and MSL2 expression (electronic supplementary material, figure S3) in *y^{1}w^{1118}*, *mle^9/mle^9;Ubi:*
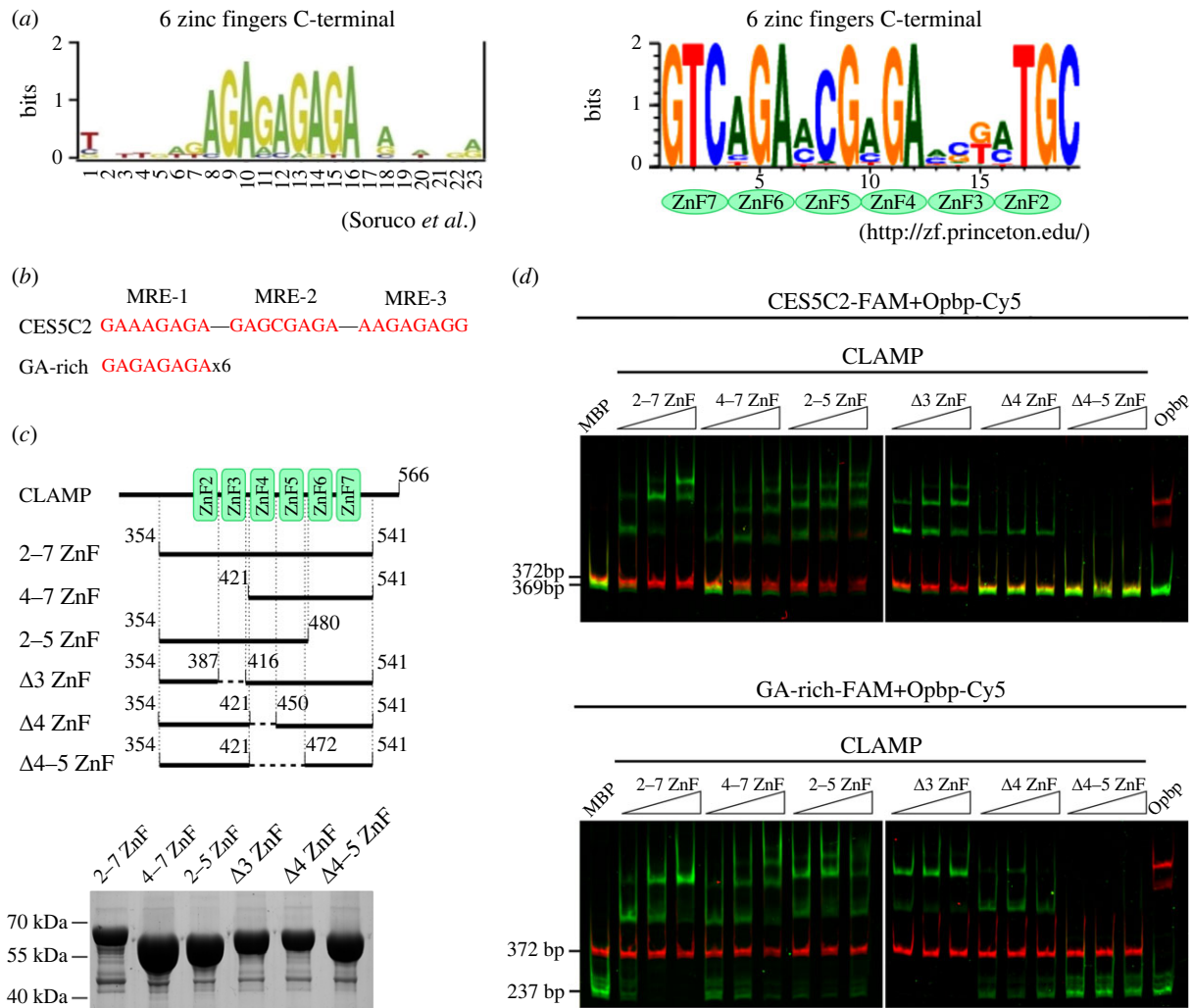
Figure 3. Mapping the C2H2 domains required for CLAMP's specific binding to DNA. (*a*) On the left is a CLAMP motif discovered via protein binding microarrays assessing CLAMP binding to all possible 10-bp sequences [18]. On the right is the predicted binding site for the CLAMP 2–7 C2H2 array (http://zf.princeton.edu; polynomial SVG) [45]. (*b*) The DNA fragments tested for CLAMP binding: 199 bp CES5C2 (with three MRE motifs [11]) and 64 bp (GA)$_{24}$. (*c*) Schematic representation of the CLAMP 2–7 C2H2 zinc finger (ZnF) cluster and its various truncated derivatives that were fused to MBP and expressed in *Escherichia coli*. (*d*) Electrophoretic mobility shift assay of different variants of recombinant zinc finger domains of CLAMP with two DNA fragments. Different zinc finger domains of CLAMP fused with MBP or five zinc finger domains of Opbp fused with MBP or MBP alone were incubated with fluorescently labelled DNA fragments; CLAMP binding sites labelled with FAM (CES5C2-FAM and GA-rich-FAM) and Opbp binding sites (used as a negative control) labelled with Cy5 (Opbp-Cy5). Signals were detected for the FAM-labelled fragment (green bands) at an Ex/Em of 500/535 nm and the Cy5-labeled fragment (red bands) at an Ex/Em of 630/700 nm. Interaction specificity was demonstrated by incubation with DNA fragments (100 ng each) with different amounts of truncated CLAMP variants presented as a series of twofold dilutions (5, 10, and 20 pmol). Note that no binding was observed on the control DNA region (Cy5 signal).

mle[WT]/*TM6* and *mle*[9]/*mle*[9];*Ubi:mle*[ΔCBD]/*Ubi:mle*[ΔCBD] lines. Immunoblot analysis revealed no significant difference of the expression of MSL1 and MSL2 in these lines.

Like other null *mle* mutants, homozygous *mle*[9] females were viable, while males died during development. We compared the viability of *mle*[9] females expressing either MLE[WT]-FLAG or MLE[ΔCBD]-FLAG (figure 4c; electronic supplementary material, table S1). Unexpectedly, strong overexpression of MLE[WT] (*Ubi:mle*[WT]/*Ubi:mle*[WT]) affected viability of males and females in the *mle*[9]/*mle*[9] but not in the *mle*[9]/*CyO* background. It is likely that the *mle*[9] chromosome contains additional recessive mutations/allelic variants that interfere with MLE overexpression. Lower expression of MLE[WT] (*Ubi:mle*[WT]/*TM6*) resulted in almost complete restoration of *mle*[9] male survival. Moreover, *mle*[9]/*mle*[9]; *Ubi:mle*[WT]/*TM6* line is stable, and males and females of this line have approximately the same survival rate (electronic supplementary material, table S2). These results suggest that *Ubi:mle*[WT]/*TM6* almost completely compensates the *mle*[9] mutation and restores dosage compensation.

In contrast to MLE[WT], the viability of *mle*[9]/*mle*[9] flies increased with the expression of the mutant protein MLE[ΔCBD] in the *Ubi:mle*[ΔCBD] homozygotes (figure 4c; electronic supplementary material, table S1). In the *Ubi:mle*[ΔCBD]/*TM6* background, only *mle*[9]/*mle*[9] females are viable. Increasing of the MLE[ΔCBD] expression in *Ubi:mle*[ΔCBD]/*Ubi:mle*[ΔCBD] homozygotes led to better survival of males, which appear in a ratio of one to five females. *mle*[9]/*mle*[9];*Ubi:mle*[ΔCBD]/*Ubi:mle*[ΔCBD] males were weak and died within one week after hatching. In addition, we were unable to support the *mle*[9]/*mle*[9];*Ubi:mle*[ΔCBD]/*Ubi:mle*[ΔCBD] line, which was lost within a few generations (electronic supplementary material, table S2). Altogether, these results suggest that deleting the CBD in MLE affects dosage compensation in males.

We next examined whether deleting the CBD in MLE affects the recruitment efficiency of the MSL complex to the male X chromosome. Immunostaining of polytene chromosomes from the salivary glands of *Drosophila* larvae allows for the visualization
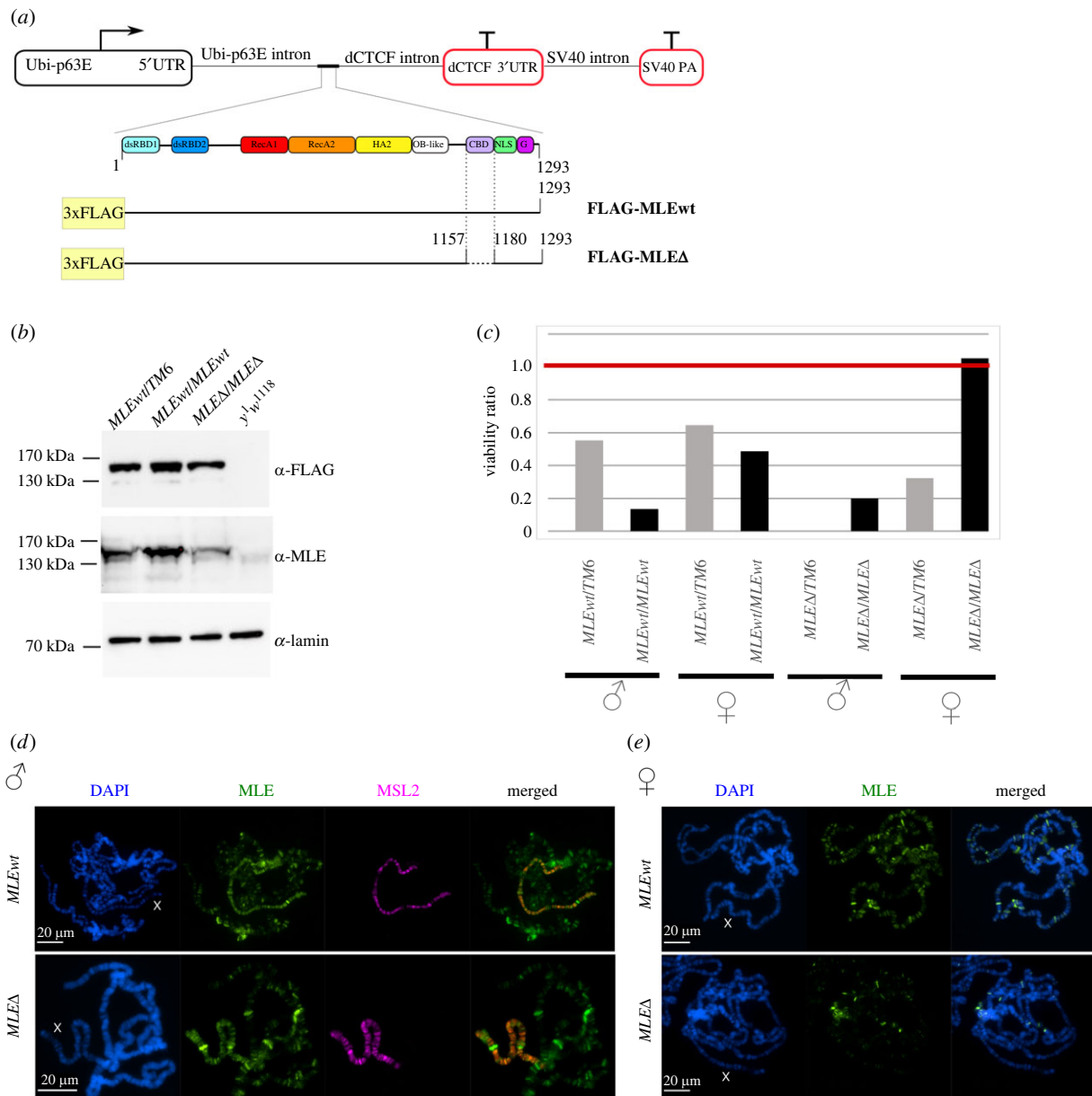
Figure 4. Comparison of the functional activity of MLE^{WT} and MLE$^{\Delta CBD}$ *in vivo*. (a) Schematic representation of the MLE^{WT}-FLAG (MLEwt) and MLE$^{\Delta CBD}$-FLAG (MLEΔ) transgenes and proteins expressed in flies. (b) Immunoblot analysis of protein extracts prepared from adult flies carrying MLEwt/TM6 (y^{1}w^{1118}; mle^9/mle^9; Ubi:mle^{WT}-FLAG/TM6); MLEwt/MLEwt (y^{1}w^{1118}; mle^9/mle^9; Ubi:mle^{WT}-FLAG/Ubi:mle^{WT}-FLAG); MLEΔ/MLEΔ (y^{1}w^{1118}; mle^9/mle^9; Ubi:mle$^{\Delta CBD}$-FLAG/Ubi:mle$^{\Delta CBD}$-FLAG); y^{1}w^{1118} (control with endogenous MLE). Immunoblot analysis using mouse anti-FLAG, rabbit anti-MLE and mouse anti-lamin Dm0 antibodies (internal control). (c) The viability of males and females expressing MLE^{WT}-FLAG (MLEwt) and MLE$^{\Delta CBD}$-FLAG (MLEΔ) in the mle^9-null background was scored in the progeny of crosses of males and females with genotypes y^{1}w^{1118}, mle^9/CyO; Ubi:mle*/TM6, where mle* is MLEwt or MLEΔ. The viability was tested in males and females homozygous (mle*/mle*) or heterozygous (mle*/TM6) for the transgene. The expected viability ratio should be close to 1.0. At least 500 flies were scored in two independent experiments (see electronic supplementary material, table S1). The percentage of obtained females and males with a certain genotype (y-axis) was calculated as the ratio of the existing number of flies to the expected number of flies. The expected number of females and males with a particular genotype was calculated based on the best surviving genotype of females (mle^9/CyO; Ubi:mle*/Ubi:mle*), the viability of which was taken as 100%. (d) Distribution of the MSL complex on the polytene chromosomes of male third-day larvae expressing MLEwt (y^{1}w^{1118}; mle^9/mle^9; Ubi:mle^{WT}/TM6) and MLEΔ (y^{1}w^{1118}; mle^9/mle^9; Ubi:mle$^{\Delta CBD}$/Ubi:mle$^{\Delta CBD}$). Panels show protein immunostaining using mouse anti-FLAG (green) and anti-MSL2 (magenta) antibodies. DNA was stained with DAPI (blue). (e) Binding of the MLE variants to the polytene chromosomes of female third-day larvae expressing MLEwt and MLEΔ. Panels show protein immunostaining using mouse anti-FLAG antibody (green). DNA was stained with DAPI (blue).

of proteins on interphase chromatin and has been used extensively to study dosage compensation [9,49–53]. Polytene chromosomes had the same MSL1 and MSL2 distribution pattern in *Ubi:mle^{WT}* and *Ubi:mle$^{\Delta CBD}$* larvae: only the X chromosome was covered by these proteins (figure 4d). MLE (detectable by anti-FLAG antibodies) colocalized with MSL2 (figure 4d) and MSL1 (electronic supplementary material, figure S4) on the X chromosome but was also found in bands and puffs on all chromosomes. The same MLE binding pattern was observed in females lacking the MSL complex (figure 4e). The C-terminal 353 aa of MLE (lacking the helicase domain) has been previously shown to bind to all chromosomes [54]. This association is RNA sensitive, implicating MLE's C-terminal domain in nonspecific interactions with RNA. Taken together, the immunostaining of polytene chromosomes showed that MLE^{WT}-FLAG and MLE$^{\Delta CBD}$-FLAG bind with the same efficiency to the MSL complex on the X chromosome, confirming that deletion of CBD in the MLE protein does not affect its presence in the MSL complex.

2.4. Deletion of CLAMP binding region of MLE affects the recruitment of the MSL complex to the X chromosome in adult males

The results presented in the previous section show that deletion of CBD in the MLE protein significantly reduces the survival of males compared to females, indicating a decrease the efficiency of dosage compensation. On the other hand, on the polytene chromosomes of larvae we do not observe any effect on the binding of the MSL proteins upon expression of the MLE$^{\Delta CBD}$ mutant in comparing with MLE^{WT}. It was previously shown that the deletion of CLAMP binding domain in the MSL2 protein also does not affect the binding of the complex to the X chromosome on the polytene chromosomes of male larvae [21]. Therefore ChIP-seq (chromatin immunoprecipitation-sequencing) as a more sensitive approach was used to investigate the potential role of CBD in MLE to recruit the MSL complex to the adult male X chromosome.

We compared the binding of MSL1, MSL2, and MSL3 in males expressing either MLE^{WT}-FLAG (mle^9/mle^9; $Ubi:mle^{WT}/TM6$) or MLE$^{\Delta CBD}$-FLAG (mle^9/mle^9; $Ubi:mle^{\Delta CBD}/Ubi:mle^{\Delta CBD}$) (figure 5). We performed ChIP-seq analysis of the chromatin collected from three-day-old flies. For ChIP we used antibodies against the MSL1, MSL2, and MSL3 for both MLE^{WT}-FLAG and MLE$^{\Delta CBD}$-FLAG transgenic lines, followed by massively parallel sequencing using an Illumina NovaSeq platform. Peaks were identified for MSL1 (294 and 568, respectively), MSL2 (223 and 82), and MSL3 (2948 and 2271) in MLE^{WT}-FLAG and MLE$^{\Delta CBD}$-FLAG males (electronic supplementary material, table S3). All peaks were divided into three groups: coincident with HAS on the X chromosome, other peaks on the X chromosome (outside the HAS) (X), and autosomal peaks (AUT; figure 5a). It is worth noting that the peaks and their signals found for the MLE lines are mainly consistent with the control y^1w^{1118} line (electronic supplementary material, figure S5). CLAMP binds to most HAS but is not enriched at autosomal peaks associated with MSL proteins [15,18,20,55].

Notably, MSL2 binding was significantly (Wilcoxon signed-rank test, p-value = 3.28×10^{-23}) weakened in the HAS regions of the MLE$^{\Delta CBD}$-FLAG line, with more than half of the peaks disappearing. In addition, there was a detectable 2- to 3-fold reduction in MSL3 binding levels across the X chromosome, with approximately 70% of the original peaks retained. However, MSL1 binding to the X chromosome remained almost unchanged (figure 5a–c). Interestingly, MSL1 binding to the autosomes was noticeably enhanced in the MLE$^{\Delta CBD}$-FLAG line (figure 5a,d).

The ChIP-seq data from adult MLE^{WT}-FLAG males showed that the MSL proteins bound to approximately 90% of the previously identified HAS from three studies [11,12,56] (figure 6; electronic supplementary material, figure S5). MSL1 and MSL2 bound directly to HAS, while MSL3 was associated with extended regions surrounding the HAS (figure 5b).

MSL2 bound predominantly to all HAS on the X chromosome in MLE^{WT}-FLAG males (figure 6a). MSL2 was partially redistributed from HAS (only 25% remained significantly enriched for MSL2) to a group of autosomal peaks in MLE$^{\Delta CBD}$-FLAG males. The average signal for MSL2 was also reduced for the remaining HAS in MLE$^{\Delta CBD}$-FLAG males.

Unlike MSL2, the number of peaks on the X chromosome for MSL1 did not differ much between MLE$^{\Delta CBD}$-FLAG and MLE^{WT}-FLAG males (figure 6b). However, approximately 330 new peaks for MSL1 appeared on the autosomes in MLE$^{\Delta CBD}$-FLAG males, which predominantly showed promoter localization (electronic supplementary material, figure S6a,b). The average signal for MSL1 decreased slightly for HAS and remained almost unchanged for other parts of the X chromosome. For the new peaks on the autosomes, we detected significant signal increases (Wilcoxon signed-rank test, p-value = 1.77×10^{-39}) in MLE$^{\Delta CBD}$-FLAG flies (figure 6b; electronic supplementary material, figure S6a,c). Therefore, deleting the CBD in MLE causes a redistribution of MSL1 to autosome sites with simultaneous preservation of the general binding profile along the X chromosome.

MSL3 showed peaks at HAS and at an additional 800 sites on the X chromosome and about 1900 sites on autosomes (figure 6c). The average signal for MSL3 peaks on the X chromosome was several times larger than for peaks on the autosomes. The number of peaks just slightly decreased in MLE$^{\Delta CBD}$-FLAG males. However, the average signal for the X-chromosome peaks was significantly reduced (Wilcoxon signed-rank test, p-value = 2.26×10^{-106}). A comparison of MSL3 profiles on the X chromosome outside of HAS and the autosomes demonstrated the predominant preservation of its binding profile in MLE$^{\Delta CBD}$-FLAG flies (figure 6c; electronic supplementary material, figure S6). In a large proportion of cases, protein binding regions extend into the body of genes (electronic supplementary material, figure S6b). We found a more significant increasing of MSL1 binding in MLE$^{\Delta CBD}$-FLAG flies at the peaks with MSL3 colocalization (Wilcoxon signed-rank test, p-value = 1.15×10^{-6}) (electronic supplementary material, figure S6c).

Altogether, these results indicate that deleting the CBD in MLE decreases MSL2 and MSL3 binding to HAS on the X chromosome. It also significantly increases MSL1 binding and redistributes MSL3 binding to autosomes in adult males. Taken together these results suggest that the CBD in the MLE protein participates in the process of the recruitment of the MSL complex to the male X chromosome.

3. Discussion

Our results provide evidence that MLE's small unstructured region between its main core region and C-terminal NLS [41] interacts with the sixth zinc-finger domain of CLAMP. MLE's primary function is the assembly of the MSL complex since only MSL1/MSL2 are found at the entry sites in its absence [49,57]. MLE specifically interacts with two roX lncRNAs and induces their unwinding [24–30,58], enabling both to interact with MSL proteins [24,25,31–33]. Interactions between the MSL proteins and roX lncRNAs increase the efficiency of the complex formation. Our Y2H results showed that MLE can dimerize and support a previous finding that the glycine-rich heptad repeats are involved in this process [41]. MLE dimerization may promote the formation of centres for assembling MSL complexes.

Here, we have described a new function of MLE in dosage compensation associated with a specific interaction with CLAMP. Our genetic and ChIP-seq experiments showed that the CBD in MLE contributes to the specific recruitment of MSL complexes to the male
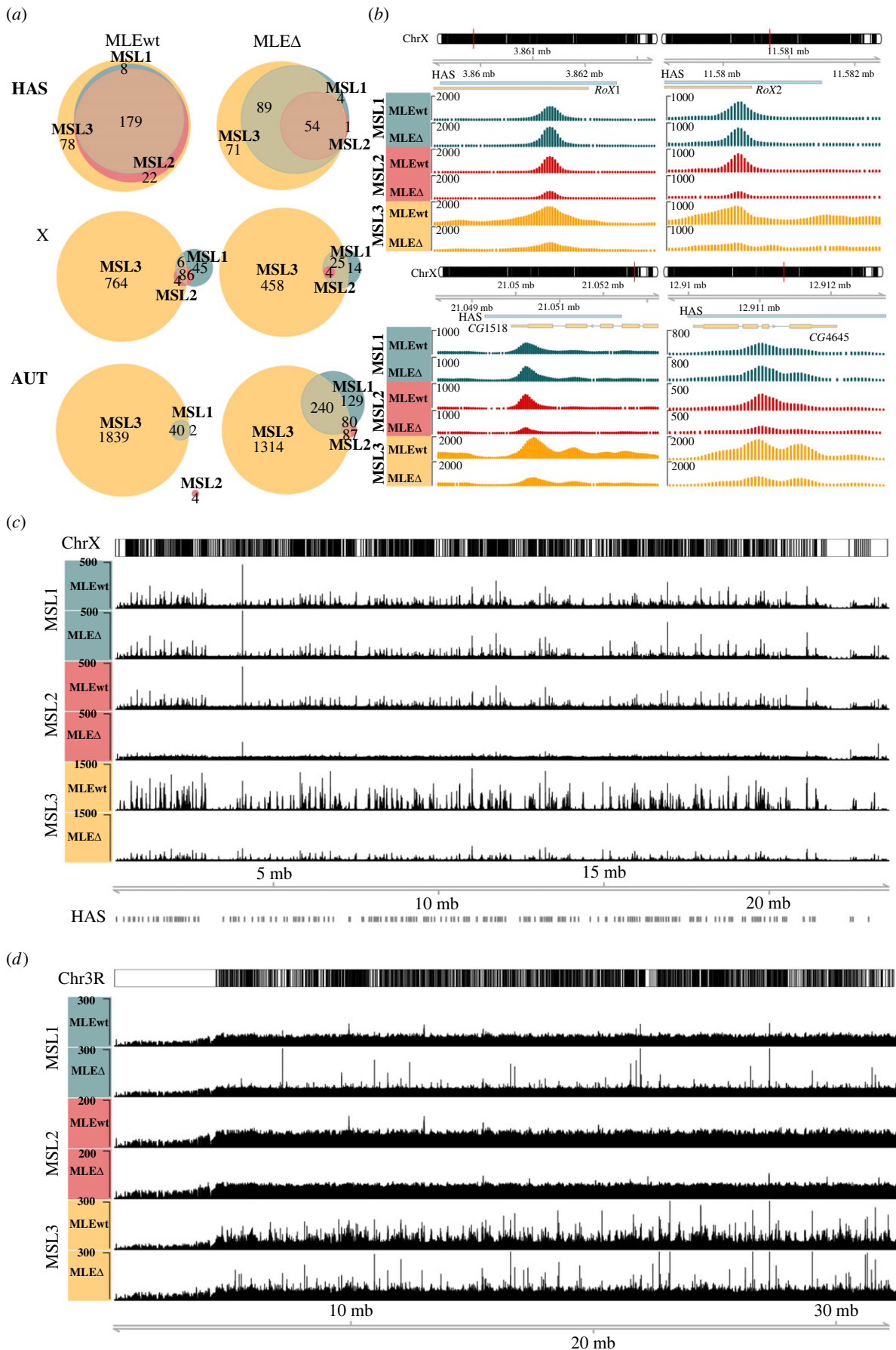
Figure 5. Deletion of the MLE CBD causes chromosome-scale MSL protein redistribution. To study the functional role of MLE's CBD in recruiting the DCC to chromatin sites, we compared the ChIP-Seq profiles of MSL1, MSL2, and MSL3 in three-day males expressing MLEwt (y^1w^{1118}; mle^9/mle^9; $Ubi:mle^{WT}/TM6$) and MLEΔ (y^1w^{1118}; mle^9/mle^9; $Ubi:mle^{ΔCBD}/Ubi:mle^{ΔCBD}$). (a) A Venn diagram of MSL1, MSL2, and MSL3 peak intersections for the MLEwt and MLEΔ lines. All peaks were divided into three groups: coincident with HAS on the X chromosome (HAS), other peaks on the X chromosome (X), and autosomal peaks (AUT). (b) Examples of MSL1, MSL2, and MSL3 signal distributions on HAS for *RoX1*, *RoX2*, *CG1518*, and *CG4645* areas. Blue and yellow tracks represent HAS and gene positions, respectively. (c) Normalized (reads per kilobase per million mapped reads (RPKM)) signals for the MSL1, MSL2, and MSL3 proteins on chromosome X for the MLEwt and MLEΔ lines. (d) Normalized (RPKM) signals for the MSL proteins on chromosome 3R for the MLEwt and MLEΔ lines.
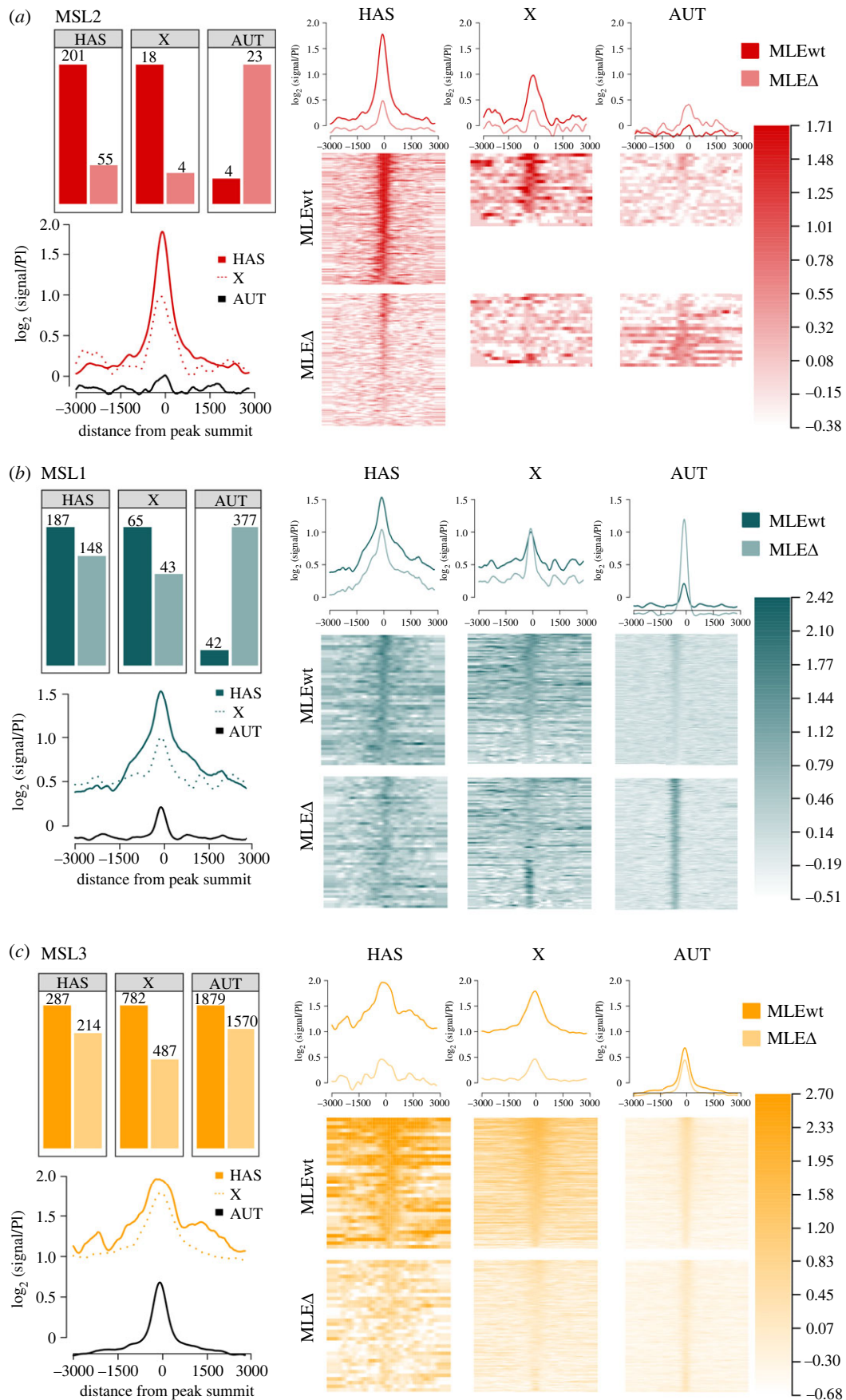
Figure 6. Deleting MLE's CBD alters MSL protein binding. (a–c) A histogram of peak counts (top left), and average log fold-change (FC) between normalized (RPKM) signals and nonspecific IgG signals in different genomic regions (HAS, X chromosome, AUTosomes) of the MLEwt line (bottom left). Histograms show average log FC between normalized (RPKM) test signals and nonspecific IgG signals of the MLEwt and the MLEΔ lines in different genomic regions (HAS, X chromosome, AUTosomes) (top right), and heatmaps (bottom right) for MSL2, MSL1, and MSL3, respectively. Peak lists and scales, unified between fly lines, were used to build heatmaps and profiles separately for each MSL protein. Peaks were arranged in descending order of the average signal calculated at vicinity of ±250 bp from the middle of the peak for the MLEwt. LogFC was calculated after smoothing signals using the Daniell kernel with kernel size 100 and adding a pseudo count.

X chromosome. The interaction of MLE with CLAMP could increase the specificity of MSL binding to HAS. Previous Chip-seq analysis in S2 cells showed that MLE and MSL2 are located closer to the DNA and thus might be important for recruiting the MSL complex to the HAS [39]. Interestingly, deletion of only the CBD in either MLE or MSL2 has only a weak effect on the specific recruitment of the MSL complex to the X chromosome, which is especially evident from the efficient binding of the complex to the polytene X chromosome of males expressing mutant variants of MLE or MSL2 [21]. This suggests that partially redundant CLAMP–MLE and CLAMP–MSL2 interactions cooperatively improve the stability and efficiency of MSL recruitment to the X chromosome.

The interaction between MLE's unstructured region and the sixth C2H2 domain in CLAMP is a new example of the involvement of C2H2 domains in specific protein–protein interactions. We have previously shown an interaction between MSL2 and the N-terminal domain of CLAMP zinc finger [21,22]. These results demonstrate for the first time that the C2H2 domains of a single architectural protein, such as CLAMP, can simultaneously interact with different unstructured regions of two proteins that form the transcription complex, resulting in more specific and stable recruitment to regulatory elements. Interactions between C2H2 domains in DNA binding proteins and intrinsically disordered regions of subunits of transcription complexes may be widespread for the specific recruitment of transcription complexes to regulatory elements. Despite the key role of C2H2 proteins in the formation of promoter architecture [59,60], the involvement of C2H2 domains in the recruitment of transcription complexes remains poorly understood. It is likely that identifying other protein–protein interactions that determine the specificity of recruitment of the MSL complex to the X chromosome will allow us to better understand the contribution of C2H2 domains to this process.

We found that only the 4–5 C2H2 domains are critical for CLAMP's binding to DNA motifs *in vitro*, while the 3, 6, and 7 C2H2 domains improve binding effectiveness. Since MSL2's CXC domain also binds to GA repeats [13,14], CLAMP likely binds to part of the GA repeats in HAS. We hypothesize that CLAMP binds to HAS/CES in cooperation with additional DNA-binding proteins that together form a chromatin module leaving CLAMP's zinc fingers 1 and 6 free to interact with the MSL complex. At autosomal sites, CLAMP presumably binds in cooperation with alternative protein groups, which form chromatin sites in which interactions with proteins or DNA mask its 1 and 6 C2H2 domains. For example, in embryos CLAMP functions together with a key pioneer transcription factor Zelda that also contains C2H2 domains [61]. The model can explain the recent results that ectopic expression of MSL2 in females results in preferential binding of MSL2 to autosomal promoters enriched in TGTG motifs (1093) rather than GAGA motifs and CLAMP sites (363) [62]. Since CLAMP binds to more than 10 000 sites on autosomes [18], it seems likely that the N-terminal zinc finger domain of CLAMP is masked to interact with MSL2 at autosomal sites.

CLAMP's 6 and 7 C2H2 domains are highly conserved in insects. In contrast, the region of MLE corresponding to its CBD is only conserved in the *Drosophila* genus. Moreover, even in species remote from *Drosophila melanogaster*, like *Drosophila virilis*, there are many substitutions and insertions in this region of MLE. Similarly, the CBD in MSL2 is also only conserved in the *Drosophila* genus [21,34]. It seems likely that unstructured regions in MSL proteins can rapidly evolve to form new specific interactions with the zinc finger domains of DNA-binding proteins, allowing the specific interactions between DNA-bound architectural proteins and transcriptional complexes to be organized and improved.

4. Material and methods

4.1. Plasmid construction

Plasmids for the yeast two-hybrid assay were prepared using the full-sized and truncated versions with separate domains of MLE and CLAMP as C-termini fused with pGBT9 and pGAD424 vectors from Clontech. Different full-sized variants of MLE were fused with 3xFLAG at the N-terminus and cloned into an expression vector. This vector contains *attB* site for *φC31*-mediated recombination, *Ubi63E* promoter with its 5′UTR, 3′UTR with SV40 polyadenylation signal, intronless *yellow* gene as a reporter for detection of transformants. Details of the cloning procedures, primers and plasmids used for plasmid construction are available upon request.

4.2. Yeast two-hybrid assay

Yeast two-hybrid assay was carried out using yeast strain pJ69-4A, with plasmids and protocols from Clontech. For growth assays, plasmids were transformed into yeast strain pJ69-4A by the lithium acetate method, as described by the manufacturer, and plated on media without tryptophan and leucine. After 2 days of growth at 30°C, the cells were plated on selective media without tryptophan, leucine, histidine and adenine, and their growth was compared after 2–3 days. Each assay was repeated three times.

4.3. Antibodies

Antibodies against MSL1[423-1030], MSL2[421-540], CLAMP and MSL3 were raised in rabbits and purified from the sera by ammonium sulfate fractionation followed by affinity purification on CNBr-activated Sepharose (GE Healthcare, USA) or Amino-link Resin (ThermoFisher Scientific, USA) according to standard protocols. Other antibodies: mouse monoclonal anti-FLAG, clone M2 (F1804, Sigma, USA), mouse monoclonal anti-HA, clone HA-7 (H3663, Sigma, USA), mouse monoclonal anti-lamin Dm0, clone ADL84.12 (ADL84.12, DSHB, USA).

4.4. Co-immunoprecipitation assay

Drosophila S2 cells were co-transfected by 3xHA-CLAMP and 3xFLAG-MLE^WT or 3xFLAG-MLE^ΔCBD plasmids with MACSFectin (Miltenyi Biotec, USA). After transfection, the cells were incubated for 48 h and then collected by centrifugation at 700*g* for 5 min,

washed once with 1×PBS, and resuspended in 20 packed cell volumes of hypotonic lysis buffer (20 mM Tris–HCl, pH 7.4, with 10 mM KCl, 10 mM MgCl$_2$, 2 mM EDTA, 10% glycerol, 1% Triton X-100, 1 mM DTT, 1 mM PMSF and Calbiochem Complete Protease Inhibitor Cocktail V). After incubation on ice for 10 min, the cells were sonicated (Bioruptor (Diagenode, USA) for 2 min on setting L, 15 s ON/45 s OFF), NaCl was added to a final concentration of 420 mM, and incubation on ice continued for 60 min, with periodic mixing. Sonication was repeated as above to reduce viscosity, cell debris was pelleted by centrifugation at 10 000g for 30 min at 4°C, and the supernatant was collected for immunoprecipitation with anti-FLAG-, anti-lamin-conjugated Protein G Magnetic beads, and anti-CLAMP-conjugated Protein A Magnetic beads (NEB, USA) (by incubating in the PBST on a rotary shaker at room temperature for 1 h) equilibrated in incubation buffer-150 (20 mM Tris–HCl, pH 7.4, with 150 mM NaCl, 10 mM MgCl$_2$, 1 mM EDTA, 1 mM EGTA, 10% glycerol, and 0.1% NP-40). The protein extract (50 µg protein) was adjusted to a volume of 500 µl with buffer-150, mixed with antibody-conjugated beads (30 µl), and incubated on a rotary shaker overnight at 4°C. The beads were then washed with five portions of buffer-150, and resuspended in SDS-PAGE loading buffer, boiled and analysed by immunoblot analysis. Proteins were detected using the ECL Plus Western Blotting substrate (Thermo Fisher Scientific, USA).

4.5. Fly crosses and transgenic lines

Drosophila strains were grown at 25°C under standard culture conditions. The transgenic constructs were injected into preblastoderm embryos using the *φC31*-mediated site-specific integration system at locus 86Fb [46]. The emerging adults were crossed with the *y^1 ac w^{1118}* flies, and the progeny carrying the transgene in the 86Fb region were identified by *y$^+$* pigmented cuticle. Details of the crosses and primers used for genetic analysis are available upon request.

4.6. Fly extract preparation

Twenty adult flies were homogenized with a pestle in 200 µl of 1×PBS containing 1% β-mercaptoethanol, 10 mM PMSF and 1 : 100 Calbiochem Complete Protease Inhibitor Cocktail VII. Suspension was sonicated 3 times for 5 s at 5 W. Then, 200 µl of 4×SDS-PAGE sample buffer was added and the mixture was incubated for 10 min at 100°C and centrifuged at 16 000g for 10 min.

4.7. Electrophoretic mobility shift assay

Aliquots of purified recombinant proteins were incubated with fluorescently labelled DNA fragments in the presence of nonspecific binding competitor poly(dI-dC). Incubation was performed in PBS (pH 8.0) containing 5 mM MgCl$_2$, 0.1 mM ZnSO$_4$, 1 mM DTT, 0.1% NP-40 and 10% glycerol at room temperature for 30 min. The mixtures were then resolved by nondenaturing 5% PAGE (79 AA:1 BAA) in 0.5 × TBE buffer at 5 V cm^{-1}. Signals were detected for FAM-labelled fragment at the Ex 500 nm/Em 535 nm and for Cy5-labelled fragment at the Ex 630 nm/Em 700 nm.

4.8. Immunostaining of polytene chromosomes

Drosophila 3rd instar larvae were cultured at 18°C under standard conditions. Polytene chromosome staining was performed as described [21]. The following primary antibodies were used: rabbit anti-MSL1 at 1 : 500 dilution, rabbit anti-MSL2 at 1 : 500 dilution, and monoclonal mouse anti-FLAG at 1 : 100 dilution. The secondary antibodies were Alexa Fluor 488 goat anti-mouse 1 : 2000 and Alexa Fluor 555 goat anti-rabbit 1 : 2000 (Invitrogen). The polytene chromosomes were co-stained with DAPI (Appli-Chem). Images were acquired on a Nikon Eclipse Ti fluorescent microscope using a Nikon DS-Qi2 digital camera, processed with ImageJ 1.50c4 and Fiji bundle 2.0.0-rc-46. Three to four independent stainings and 4–5 samples of polytene chromosomes were performed with each MLE-expressing transgenic line.

4.9. ChIP-Seq

Chromatin was prepared from two- to three-day-old adult males. One gram of adult flies was ground in a mortar in liquid nitrogen and resuspended in 20 ml of buffer A (15 mM HEPES–KOH, pH 7.6, 60 mM KCl, 15 mM NaCl, 13 mM EDTA, 0.1 mM EGTA, 0.15 mM spermine, 0.5 mM spermidine, 0.5% NP-40, and 0.5 mM DTT) supplemented with 0.5 mM PMSF and Calbiochem Complete Protease Inhibitor Cocktail V. The suspension was then homogenized in a Potter and Dounce homogenizer with a tight pestle, filtered through a 100 µm nylon cell strainer (Miltenyi Biotec, USA), and cross-linked with 1% formaldehyde for 15 min at room temperature. Cross-linking was stopped by adding glycine to a final concentration of 125 mM. The nuclei were washed with three 10 ml portions of wash buffer (15 mM HEPES–KOH, pH 7.6, 60 mM KCl, 15 mM NaCl, 1 mM EDTA, 0.1 mM EGTA, 0.1% NP-40, and protease inhibitors) and one 5 ml portion of basic nuclear lysis basic buffer (15 mM HEPES, pH 7.6, 140 mM NaCl, 1 mM EDTA, 0.1 mM EGTA, 1% Triton X-100, 0.5 mM DTT, 0.1% sodium deoxycholate, and protease inhibitors) and resuspended in 1 ml of nuclear lysis buffer (15 mM HEPES, pH 7.6, 140 mM NaCl, 1 mM EDTA, 0.1 mM EGTA, 1% Triton X-100, 0.5 mM DTT, 0.1% sodium deoxycholate, 0.5% SLS, 0.1% SDS, and protease inhibitors). The suspension was sonicated in a Covaris ME220 focused-ultrasonicator (40 alternating 15-s ON and 45-s OFF intervals, peak power 75, duty % factor 25), and 50 µl aliquots were used to test the extent of sonication and measure the DNA concentration. Debris was removed by centrifugation at 14 000g, 4°C, for 10 min, and chromatin was pre-cleared with Protein A Dynabeads (Invitrogen, USA). Corresponding antibodies were incubated for 1 hour at room temperature with 20 µl aliquots of Protein A (anti-MSL1, 1 : 100; anti-MSL2, 1 : 100; anti-MSL3, 1 : 500) Dynabeads (Invitrogen, USA) mixed with 200 µl of PBST. Then antibody–Dynabead complexes were washed and equilibrated in nuclear lysis buffer. Chromatin samples containing 10–20 µg of DNA equivalent in

200 µl nuclear lysis buffer (2 µl aliquots of pre-cleared chromatin as input material) were incubated overnight at 4°C with anti-body–Dynabead complexes. After 3 rounds of washing with lysis buffer supplemented with 500 mM NaCl and TE buffer (10 mM Tris–HCl, pH 8; 1 mM EDTA), the DNA was eluted with elution buffer (50 mM Tris–HCl, pH 8.0; 1 mM EDTA, 1% SDS), the cross-links were reversed, and the precipitated DNA was extracted using a ChIP DNA Clean &Concentrator kit (Zymo Research, USA).

The ChIP-seq libraries were prepared with NEBNext Ultra II DNA Library Prep kit per the manufacturer's instructions. Amplified libraries were quantified using fluorometry with DS-11 (DeNovix, USA) and Bioanalyzer 2100 (Agilent, USA). Diluted libraries were clustered on a pair-read flowcell and sequenced using a NovaSeq 6000 system (Illumina, USA).

ChIP-seq analysis was performed on paired-end reads data. Each sample was presented in 2 biological replicas. Preprocessing and peak calling was performed based on the pipeline described previously [63]. The cutadapt [64], Bowtie2 [65] and MACS2 [66] were employed for trimming, mapping and peak calling respectively. dm6 version of *Drosophila melanogaster* genome was used as a reference genome. Upon merging the replicates, coverage tracks in BedGraph format were generated using the deepTools [67] bamCoverage function, with a bin width of 50 bp, and extendReads option applied. The coverage tracks were further normalized by reads per kilobase of transcript, per million mapped reads (RPKM).

We assessed reproducibility using the IDR pipeline (https://sites.google.com/site/anshulkundaje/projects/idr), with *p*-value thresholds for true replicates, pseudoreplicates, and MACS2 was selected 0.05 and 0.01 respectively achieving rescue ratio (RR) or self-consistency ratio (SR) being less than 2.

Further analysis was conducted using R version 4.2.1 (http://www.r-project.org). To track changes in the ChIP-seq signal level between different lines, a unified list of peaks was created for each protein, using the findOverlapsOfPeaks function from ChIP-peakAnno package [68]. The average signal was calculated using getTagMatrix from ChIPseeker package [69]. Part of the peaks were assigned to HAS, by intersecting them with a list of combined HAS variants from three different studies [11,12,56]. Additionally closest to peaks genomic features were annotated using the annotatePeak function from clusterprofiler [70] with tssRegion set to 200. Gviz [71] were used for genomic track visualization. To measure the significance level of the signal change between $MLE^{\Delta CBD}$ and MLE^{WT} fly lines, we conducted Wilcoxon signed-rank tests on groups of the average signals calculated at vicinity of ±1000 bp from the middle of the peaks.

References

1. Chery J, Larschan E. 2014 X-marks the spot: X-chromosome identification during dosage compensation. *Biochim. Biophys. Acta (BBA)* **1839**, 234–240. (doi:10.1016/j.bbagrm.2013.12.007)

2. Kuroda MI, Hilfiker A, Lucchesi JC. 2016 Dosage compensation in *Drosophila*: a model for the coordinate regulation of transcription. *Genetics* **204**, 435–450. (doi:10.1534/genetics.115.185108)

3. Lucchesi JC, Kuroda MI. 2015 Dosage compensation in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **7**, a019398. (doi:10.1101/cshperspect.a019398)

4. Menon DU, Meller VH. 2015 Identification of the *Drosophila* X chromosome: the long and short of it. *RNA Biol.* **12**, 1088–1093. (doi:10.1080/15476286.2015.1086864)

5. Samata M, Akhtar A. 2018 Dosage compensation of the X chromosome: a complex epigenetic assignment involving chromatin regulators and long noncoding RNAs. *Annu. Rev. Biochem.* **87**, 323–350. (doi:10.1146/annurev-biochem-062917-011816)

6. Straub T, Becker PB. 2011 Transcription modulation chromosome-wide: universal features and principles of dosage compensation in worms and flies. *Curr. Opin. Genet. Dev.* **21**, 147–153. (doi:10.1016/j.gde.2011.01.012)

7. Hallacli E, Lipp M, Georgiev P, Spielman C, Cusack S, Akhtar A, Kadlec J. 2012 Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in Drosophila. *Mol. Cell* **48**, 587–600. (doi:10.1016/j.molcel.2012.09.014)

8. Kadlec J, Hallacli E, Lipp M, Holz H, Sanchez-Weatherby J, Cusack S, Akhtar A. 2011 Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat. Struct. Mol. Biol.* **18**, 142–149. (doi:10.1038/nsmb.1960)

9. Li F, Parry DA, Scott MJ. 2005 The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. *Mol. Cell. Biol.* **25**, 8913–8924. (doi:10.1128/MCB.25.20.8913-8924.2005)

10. Scott MJ, Pan LL, Cleland SB, Knox AL, Heinrich J. 2000 MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in *Drosophila*. *EMBO J.* **19**, 144–155. (doi:10.1093/emboj/19.1.144)

11. Alekseyenko AA et al. 2008 A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* **134**, 599–609. (doi:10.1016/j.cell.2008.06.033)

12. Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB. 2008 The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS Genetics* **4**, e1000302. (doi:10.1371/journal.pgen.1000302)

13. Fauth T, Muller-Planitz F, Konig C, Straub T, Becker PB. 2010 The DNA binding CXC domain of MSL2 is required for faithful targeting the dosage compensation complex to the X chromosome. *Nucleic Acids Res.* **38**, 3209–3221. (doi:10.1093/nar/gkq026)

14. Zheng S, Villa R, Wang J, Feng Y, Becker PB, Ye K. 2014 Structural basis of X chromosome DNA recognition by the MSL2 CXC domain during *Drosophila* dosage compensation. *Genes Dev.* **28**, 2652–2662. (doi:10.1101/gad.250936.114)

15. Villa R, Schauer T, Smialowski P, Straub T, Becker PB. 2016 PionX sites mark the X chromosome for dosage compensation. *Nature* **537**, 244–248. (doi:10.1038/nature19338)

16. Zheng S, Wang J, Feng Y, Wang J, Ye K. 2012 Solution structure of MSL2 CXC domain reveals an unusual Zn3Cys9 cluster and similarity to pre-SET domains of histone lysine methyltransferases. *PLoS ONE* **7**, e45437. (doi:10.1371/journal.pone.0045437)

17. Larschan E *et al.* 2012 Identification of chromatin-associated regulators of MSL complex targeting in *Drosophila* dosage compensation. *PLoS Genet.* **8**, e1002830. (doi:10.1371/journal.pgen.1002830)

18. Soruco MM *et al.* 2013 The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes Dev.* **27**, 1551–1556. (doi:10.1101/gad.214585.113)

19. Wang CI *et al.* 2013 Chromatin proteins captured by ChIP–mass spectrometry are linked to dosage compensation in *Drosophila*. *Nat. Struct. Mol. Biol.* **20**, 202–209. (doi:10.1038/nsmb.2477)

20. Albig C, Tikhonova E, Krause S, Maksimenko O, Regnard C, Becker PB. 2019 Factor cooperation for chromosome discrimination in *Drosophila*. *Nucleic Acids Res.* **47**, 1706–1724. (doi:10.1093/nar/gky1238)

21. Tikhonova E, Fedotova A, Bonchuk A, Mogila V, Larschan EN, Georgiev P, Maksimenko O. 2019 The simultaneous interaction of MSL2 with CLAMP and DNA provides redundancy in the initiation of dosage compensation in *Drosophila* males. *Development* **146**, dev179663. (doi:10.1242/dev.179663)

22. Tikhonova E, Mariasina S, Efimov S, Polshakov V, Maksimenko O, Georgiev P, Bonchuk A. 2022 Structural basis for interaction between CLAMP and MSL2 proteins involved in the specific recruitment of the dosage compensation complex in *Drosophila*. *Nucleic Acids Res.* **50**, 6521–6531. (doi:10.1093/nar/gkac455)

23. Eggers N, Gkountromichos F, Krause S, Campos-Sparr A, Becker PB. 2023 Physical interaction between MSL2 and CLAMP assures direct cooperativity and prevents competition at composite binding sites. *Nucleic Acids Res.* **51**, gkad680. (doi:10.1093/nar/gkad680)

24. Maenner S, Muller M, Frohlich J, Langer D, Becker PB. 2013 ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol. Cell* **51**, 174–184. (doi:10.1016/j.molcel.2013.06.011)

25. Quinn JJ, Ilik IA, Qu K, Georgiev P, Chu C, Akhtar A, Chang HY. 2014 Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat. Biotechnol.* **32**, 933–940. (doi:10.1038/nbt.2943)

26. Lee CG, Hurwitz J. 1993 Human RNA helicase A is homologous to the maleless protein of *Drosophila*. *J. Biol. Chem.* **268**, 16 822–16 830. (doi:10.1016/S0021-9258(19)85490-X)

27. Kuroda MI, Kernan MJ, Kreber R, Ganetzky B, Baker BS. 1991 The maleless protein associates with the X chromosome to regulate dosage compensation in drosophila. *Cell* **66**, 935–947. (doi:10.1016/0092-8674(91)90439-6)

28. Ilik IA, Maticzka D, Georgiev P, Gutierrez NM, Backofen R, Akhtar A. 2017 A mutually exclusive stem-loop arrangement in roX2 RNA is essential for X-chromosome regulation in *Drosophila*. *Genes Dev.* **31**, 1973–1987. (doi:10.1101/gad.304600.117)

29. Ilik IA *et al.* 2013 Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell* **51**, 156–173. (doi:10.1016/j.molcel.2013.07.001)

30. Lee CG, Chang KA, Kuroda MI, Hurwitz J. 1997 The NTPase/helicase activities of *Drosophila* maleless, an essential factor in dosage compensation. *EMBO J.* **16**, 2671–2681. (doi:10.1093/emboj/16.10.2671)

31. Li F, Schiemann AH, Scott MJ. 2008 Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the *Drosophila* MSL1/MSL2 complex. *Mol. Cell. Biol.* **28**, 1252–1264. (doi:10.1128/MCB.00910-07)

32. Muller M, Schauer T, Krause S, Villa R, Thomae AW, Becker PB. 2020 Two-step mechanism for selective incorporation of lncRNA into a chromatin modifier. *Nucleic Acids Res.* **48**, 7483–7501. (doi:10.1093/nar/gkaa492)

33. Akhtar A, Zink D, Becker PB. 2000 Chromodomains are protein–RNA interaction modules. *Nature* **407**, 405–409. (doi:10.1038/35030169)

34. Villa R, Jagtap PKA, Thomae AW, Campos Sparr A, Forne I, Hennig J, Straub T, Becker PB. 2021 Divergent evolution toward sex chromosome-specific gene regulation in *Drosophila*. *Genes Dev.* **35**, 1055–1070. (doi:10.1101/gad.348411.121)

35. Valsecchi CIK *et al.* 2021 RNA nucleation by MSL2 induces selective X chromosome compartmentalization. *Nature* **589**, 137–142. (doi:10.1038/s41586-020-2935-z)

36. Ankush JPK *et al.* 2019 Structure, dynamics and roX2-lncRNA binding of tandem double-stranded RNA binding domains dsRBD1,2 of *Drosophila* helicase Maleless. *Nucleic Acids Res.* **47**, 4319–4333. (doi:10.1093/nar/gkz125)

37. Gu W, Wei X, Pannuti A, Lucchesi JC. 2000 Targeting the chromatin-remodeling MSL complex of *Drosophila* to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities. *EMBO J.* **19**, 5202–5211. (doi:10.1093/emboj/19.19.5202)

38. Franke A, Baker BS. 1999 The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol. Cell* **4**, 117–122. (doi:10.1016/s1097-2765(00)80193-8)

39. Straub T, Zabel A, Gilfillan GD, Feller C, Becker PB. 2013 Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. *Genome Res.* **23**, 473–485. (doi:10.1101/gr.146407.112)

40. Prabu JR, Müller M, Thomae AW, Schüssler S, Bonneau F, Becker PB, Conti E. 2015 Structure of the RNA helicase MLE reveals the molecular mechanisms for uridine specificity and RNA-ATP coupling. *Mol. Cell* **60**, 487–499. (doi:10.1016/j.molcel.2015.10.011)

41. Izzo A, Regnard C, Morales V, Kremmer E, Becker PB. 2008 Structure-function analysis of the RNA helicase maleless. *Nucleic Acids Res* **36**, 950–962. (doi:10.1093/nar/gkm1108)

42. Morra R, Yokoyama R, Ling H, Lucchesi JC. 2011 Role of the ATPase/helicase maleless (MLE) in the assembly, targeting, spreading and function of the male-specific lethal (MSL) complex of *Drosophila*. *Epigenetics Chromatin* **4**, 6. (doi:10.1186/1756-8935-4-6)

43. Tikhonova E, Mariasina S, Arkova O, Maksimenko O, Georgiev P, Bonchuk A. 2022 Dimerization activity of a disordered N-terminal domain from *Drosophila* CLAMP protein. *IJMS* **23**, 3862. (doi:10.3390/ijms23073862)

44. Choo Y, Klug A. 1994 Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc. Natl Acad. Sci. USA* **91**, 11 168–11 172. (doi:10.1073/pnas.91.23.11168)

45. Persikov AV, Singh M. 2014 De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* **42**, 97–108. (doi:10.1093/nar/gkt890)

46. Bischof J, Maeda RK, Hediger M, Karch F, Basler K. 2007 An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317. (doi:10.1073/pnas.0611511104)

47. Kernan MJ, Kuroda MI, Kreber R, Baker BS, Ganetzky B. 1991 napts, a mutation affecting sodium channel activity in *Drosophila*, is an allele of mle, a regulator of X chromosome transcription. *Cell* **66**, 949–959. (doi:10.1016/0092-8674(91)90440-a)

48. Hilfiker A, Yang Y, Hayes DH, Beard CA, Manning JE, Lucchesi JC. 1994 Dosage compensation in *Drosophila*: the X-chromosomal binding of MSL-1 and MLE is dependent on Sxl activity. *EMBO J.* **13**, 3542–3550. (doi:10.1002/j.1460-2075.1994.tb06661.x)

49. Palmer MJ, Richman R, Richter L, Kuroda MI. 1994 Sex-specific regulation of the male-specific lethal-1 dosage compensation gene in *Drosophila*. *Genes Dev.* **8**, 698–706. (doi:10.1101/gad.8.6.698)

50. Lyman LM, Copps K, Rastelli L, Kelley RL, Kuroda MI. 1997 *Drosophila* male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* **147**, 1743–1753. (doi:10.1093/genetics/147.4.1743)

51. Demakova OV, Kotlikova IV, Gordadze PR, Alekseyenko AA, Kuroda MI, Zhimulev IF. 2003 The MSL complex levels are critical for its correct targeting to the chromosomes in *Drosophila melanogaster*. *Chromosoma* **112**, 103–115. (doi:10.1007/s00412-003-0249-1)

52. Dahlsveen IK, Gilfillan GD, Shelest VI, Lamm R, Becker PB. 2006 Targeting determinants of dosage compensation in *Drosophila*. *PLoS Genetics* **2**, e5. (doi:10.1371/journal.pgen.0020005)

53. Meller VH, Rattner BP. 2002 The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J.* **21**, 1084–1091. (doi:10.1093/emboj/21.5.1084)

54. Richter L, Bone JR, Kuroda MI. 1996 RNA-dependent association of the *Drosophila* maleless protein with the male X chromosome. *Genes Cells* **1**, 325–336. (doi:10.1046/j.1365-2443.1996.26027.x)

55. Rieder LE, Jordan WT, Larschan EN. 2019 Targeting of the dosage-compensated male X-chromosome during early *Drosophila* development. *Cell Rep.* **29**, 4268–4275. (doi:10.1016/j.celrep.2019.11.095)

56. Ramirez F *et al.* 2015 High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X Chromosome in Drosophila. *Molecular Cell* **60**, 146–162. (doi:10.1016/j.molcel.2015.08.024)

57. Gu W, Szauter P, Lucchesi JC. 1998 Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev. Genet.* **22**, 56–64. (doi:10.1002/(SICI)1520-6408(1998)22:1<56::AID-DVG6>3.0.CO;2-6)

58. Morra R, Smith ER, Yokoyama R, Lucchesi JC. 2008 The MLE subunit of the *Drosophila* MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. *Mol. Cell. Biol.* **28**, 958–966. (doi:10.1128/MCB.00995-07)

59. Kyrchanova OV, Bylino OV, Georgiev PG. 2022 Mechanisms of enhancer-promoter communication and chromosomal architecture in mammals and *Drosophila*. *Front. Genet.* **13**, 1081088. (doi:10.3389/fgene.2022.1081088)

60. Maksimenko OG, Fursenko DV, Belova EV, Georgiev PG. 2021 CTCF As an example of DNA-binding transcription factors containing clusters of C2H2-type zinc fingers. *Acta Naturae* **13**, 31–46. (doi:10.32607/actanaturae.11206)

61. Duan J *et al.* 2021 CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. *eLife* **10**, e69937. (doi:10.7554/eLife.69937)

62. Valsecchi CIK, Basilicata MF, Semplicio G, Georgiev P, Gutierrez NM, Akhtar A. 2018 Facultative dosage compensation of developmental genes on autosomes in *Drosophila* and mouse embryonic stem cells. *Nat. Commun.* **9**, 3626. (doi:10.1038/s41467-018-05642-2)

63. Sabirov M, Kyrchanova O, Pokholkova GV, Bonchuk A, Klimenko N, Belova E, Zhimulev IF, Maksimenko O, Georgiev P. 2021 Mechanism and functional role of the interaction between CP190 and the architectural protein Pita in *Drosophila melanogaster*. *Epigenetics Chromatin* **14**, 16. (doi:10.1186/s13072-021-00391-x)

64. Martin M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12. (doi:10.14806/ej.17.1.200)

65. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. (doi:10.1038/nmeth.1923)

66. Zhang Y *et al.* 2008 Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. (doi:10.1186/gb-2008-9-9-r137)

67. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014 deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191. (doi:10.1093/nar/gku365)

68. Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, Green MR. 2010 ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinform.* **11**, 237. (doi:10.1186/1471-2105-11-237)

69. Yu G, Wang LG, He QY. 2015 ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383. (doi:10.1093/bioinformatics/btv145)

70. Yu G, Wang LG, Han Y, He QY. 2012 clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: J. Integr. Biol.* **16**, 284–287. (doi:10.1089/omi.2011.0118)

71. Helaers R, Bareke E, De Meulder B, Pierre M, Depiereux S, Habra N, Depiereux E. 2011 gViz, a novel tool for the visualization of co-expression networks. *BMC Res. Notes* **4**, 452. (doi:10.1186/1756-0500-4-452)

72. Tikhonova E, Revel-Muroz A, Georgiev P, Maksimenko O. 2024 Interaction of MLE with CLAMP zinc finger is involved in proper MSL proteins binding to chromosomes in *Drosophila*. Figshare. (doi:10.6084/m9.figshare.c.7090092)