N-terminus of Drosophila MSL1 interacts with DNA-binding proteins and is critical for the recruitment of the dosage compensation complex to the X chromosome

Valentin Babosha[1], Natalia Klimenko[2], Evgeniya Tikhonova[1], Alexander Shilovich[1], Pavel Georgiev[1,*], Oksana Maksimenko[2]*

[1]Department of the Control of Genetic Processes, Institute of Gene Biology Russian Academy of Sciences, 34/5 Vavilova St., Moscow 119334, Russia;

[2]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, 34/5 Vavilova St., Moscow 119334, Russia

*To whom correspondence should be addressed.
Oksana Maksimenko: maksog@mail.ru
Pavel Georgiev: georgiev_p@mail.ru

Running title: The MSL1 N-terminus recruits MSL complex to the X chromosome

Bullets:

1. The N-terminal region of MSL1 interacts with 14 C2H2-type zinc-finger DNA-binding proteins
2. The N-terminus of MSL1 is critical for the specific recruitment of the MSL complex to the X chromosome
3. The N-terminus of MSL1 is important for the binding of MSL2 to a small fraction of autosomal promoters

Key words: architectural proteins, zinc finger C2H2 proteins, sex determination, transcription factor, MSL, MSL2

ABSTRACT

The male-specific lethal dosage compensation complex (MSL complex or DCC), which consists of five proteins and two non-coding roX RNAs, is necessary for the transcriptional enhancement of X-linked genes to compensate for the sex chromosome monosomy in *Drosophila* XY males, compared with XX females. MSL2 is a single protein component of the DCC that is expressed only in males and is essential for the specific recruitment of the DCC to the high-affinity "entry" sites (HASs) on the X chromosome. MSL2, together with MSL1, forms the heterotetrameric DCC core. Here, we demonstrated that the N-terminal unstructured region of MSL1 interacts with many different DNA-binding proteins that contain clusters of the C2H2 zinc-finger domains. Amino acid deletions in the N-terminal region of MSL1 strongly affect the binding of the DCC to the HASs on the male X chromosome. However, the binding of MSL2 to autosomal promoters was unaffected by amino acid deletions in MSL1. Males expressing mutant variants of MSL1 died during the larvae stage, demonstrating the critical role played by the N-terminal region in DCC activity. Our results suggest that MSL1 interacts with a variety of DNA-binding proteins to increase the specificity of DCC recruitment to the male X chromosome.

INTRODUCTION

At present, how transcription complexes bind only to certain chromatin regions that do not have pronounced sequence specificity relative to other genome sites remains unknown. Dosage compensation in *Drosophila* is a striking example of the specific recruitment of transcription complexes (1). Dosage compensation is performed by increasing the level of gene expression from the X chromosome of males (X/Y) compared with the gene expression from the X chromosome of females (X/X), despite having two X chromosomes. The multisubunit dosage compensation complex (DCC), which binds only to the male X chromosome (1-4), is responsible

for the observed gene activation. The DCC consists of five proteins, male-specific lethal (MSL)1, MSL2, MSL3, males-absent on the first protein (MOF), and maleless (MLE), and includes two non-coding RNAs, roX1 (3.7 kb) and roX2 (0.6 kb), which perform partially redundant functions (2,3,5,6). The proteins MSL1, MSL3, MOF, and MLE are also expressed in females and are involved in the regulation of gene expression processes that are unrelated to dose compensation (2). The MSL2 protein is male-specific, and the expression of the *msl2* gene is controlled by the sex lethal (Sxl) protein, which binds to *msl2* mRNA in females and blocks protein translation (7). Therefore, MSL2 protein is thought to play a key role in the specific binding of the DCC to the X chromosome in males.

The N- terminus of MSL2 contains a really interesting new gene (RING) domain, which functions as a ubiquitin E3 ligase (8,9). MSL2 has been shown to induce the ubiquitinylation of itself, as well as other DCC proteins. The MSL1 protein functions as a flexible framework for the assembly of the DCC (10-13), as the N-terminal coiled-coil domain (14) of MSL1 facilitates MSL1 homodimerization and interacts with the N-terminal RING domain of MSL2. The PEHE, domain located in the C-terminal region of MSL1, is responsible for interactions with MSL3 and MOF (10,11). MSL1 and MSL2 form the structural core of the DCC, whereas MLE, an ATP-dependent RNA/DNA helicase belonging to the DEAD subfamily, interacts with the roX1 and roX2 RNAs to induce their unwinding (15-17), allowing both RNAs are to bind MSL2 and possibly MSL1 (18), joining MLE and MSL2 into a single complex.

An incomplete MSL1-MSL2 subcomplex can bind to a reproducible set of several hundreds of sites on the X chromosome, referred to as the primary chromatin entry sites (CESs) (19) or high-affinity sites (HASs) (20). A GA-rich motif within the HAS/CES, which serves as an MSL recognition element (MRE), is important for DCC targeting (19,20). Structural analysis and *in vitro* genome-wide DNA binding assay showed that MSL2 specifically recognizes and binds to a subclass of HASs named pioneering sites on the X (PionX) (21-23).

The zinc-finger protein chromatin-linked adaptor for MSL proteins (CLAMP) was found to be important for the binding between the DCC complex and the X chromosome (24,25). CLAMP is an essential protein in *Drosophila*, expressed in both sexes, which binds thousands of GA-rich sequences genome-wide (25-27). CLAMP interacts directly with MSL2 and opens chromatin at the HAS, which promotes MSL complex recruitment (27-29). The CLAMP-interaction and CXC domains have partially redundant functions for the recruitment of the MSL complex to the X chromosome (29). The inactivation of both domains in MSL2 strongly reduces but does not eliminate MSL association with the X chromosome, although this reduction correlates with the low viability of mutant males (29). Thus, additional components of the MSL complex are likely primarily responsible for the specific recruitment of the DCC to the X chromosome.

A previous study (30) showed that the N-terminal 265 aa of MSL1 could bind to approximately 30 sites on polytene chromosomes in females that express MSL2 exogenously, and the first 26 aa are essential for such binding. The authors suggested that the N-terminal sequences of MSL proteins can bind to DNA and facilitate the specific recruitment of MSL to the male X chromosome. However, further studies examining the mechanisms of DCC recruitment to the male X chromosome have concentrated on the MSL2 protein because this is the only DCC component that is exclusively expressed in males [see review (1-3)].

Although MSL1 represents a core protein of the DCC (11-13), MSL1 also interacts with gene promoters independently of the DCC (31,32). Many transcription factors containing an array of C2H2 type zinc-finger domains (C2H2 proteins) can form a promoter architecture (33-40); therefore, we asked whether MSL1 can directly interact with this class of proteins. We have identified 14 C2H2 proteins that interact directly with MSL1. Among these, ten proteins interact with the N-terminal 15-aa region of MSL1. The deletions of the 1–15-aa and the 41–85-aa regions in the N-terminal domain of MSL1 strongly affected the recruitment of the DCC to the X chromosome and were associated with male lethality. These results suggested that the interaction

between the N-terminal region of MSL1 and C2H2 proteins is essential for the recruitment of the DCC to the X chromosome.

MATERIALS AND METHODS

Yeast two-hybrid assay

Yeast two-hybrid assay was carried out using yeast strain pJ69-4A, with plasmid vectors from Clontech and as described in (28). 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. Details of the cloning procedures and plasmids used for plasmid construction are available upon request. The sequences of PCR primers used to produce MSL1 variants are listed in Supplementary Table S1.

Fly crosses and transgenic lines

Drosophila strains were grown at 25°C and 75% humidity under standard culture conditions. The transgenic constructs were injected into preblastoderm embryos using the φC31-mediated site-specific integration system at locus 86Fb (41). The emerging adults were crossed with the *y ac* w^{1118} flies, and the progeny carrying the transgene in the 86Fb region were identified by y^{+} pigmented cuticle.

The *msl1*$^{-}$ background corresponds to *msl1*$^{\gamma 269}$/*msl1*^{L60} heterozygote. The *msl1*^{L60}/*CyO* (42)(Chang and Kuroda 1998) and *msl1*$^{\gamma 269}$ *cn*^{1}*bw*^{1}/*CyO* (12) stocks with the null *msl1* mutations were kindly donated by M. Kuroda and J. Lucchesi. The *w[1118]; sna[Sco]/CyO, P[ActGFP.w[-]]CC2* stock no. 9325 was obtained from the Bloomington stock center. To access the viability of males carrying different MSL1 variants (MSL1*), 2-3 days old *msl1*^{L60}/*CyO*-GFP; MSL1*/MSL1* females were crossed with 2-3 days old *msl1*$^{\gamma 269}$/*CyO*-

GFP; MSL1*/MSL1* males. MSL1* is any transgene expressing one of the tested MSL1 variants. *msl1*^{L60}/*msl1*$^{\gamma 269}$; MSL1*/MSL1* males and females were counted after hatching. The percentage of male viability with different MSL* variants was estimated by taking the viability of females as 100%. To assess viability of females expressed MSL2 and one of MSL1 variants (MSL*), 2-3 days old MSL1*/MSL1* females were crossed with 2-3 days old MSL2-3xFLAG/TM6,Tb males. MSL2-3xFLAG is transgene (86Fb) expressing the MSL2 cDNA under the control of the Ubi promoter at the 86Fb site (29). The percentage of MSL2/MSL* female viability was estimated by taking the viability of MSL2/TM6,Tb females as 100%.

Details of the crosses used for genetic analysis are available upon request.

Antibodies

Antibodies against MSL1[423-1030], MSL2[421-540], MSL3, GAF were raised in rabbits and purified from the sera by ammonium sulfate fractionation followed by affinity purification on CNBr-activated Sepharose (GE Healthcare, United States) or Aminolink Resin (ThermoFisher Scientific) according to standard protocols. Mouse monoclonal anti-FLAG M2 antibodies were from Sigma (Unites States).

Fly extract preparation

20 adult flies were cooled in liquid nitrogen, homogenized during 30 s with a pestle in 200 μL of extraction buffer (20 mM HEPES, pH 7.5, 100 mM KCl, 5% glycerol, 10 mM EDTA, 0.1% Triton, 1 mM DTT, 5 mM PMSF, 1:100 Calbiochem Complete Protease Inhibitor Cocktails VII and V) and incubated in ice for 10 min. Suspension was sonicated 3 times for 5 s at 5 W. Then, 50 μL of 5xSDS-PAGE sample buffer was added to the homogenate. Extracts was incubated for 10 min at 100°C, microcentrifuged at 16,000 *g* for 10 min, and loaded on an SDS-PAGE gel.

Immunostaining of polytene chromosomes

Drosophila 3rd instar larvae were cultured at 18°C under standard conditions. Polytene chromosome staining was performed as described (43). The following primary antibodies were

used: rabbit anti-MSl1 at 1:100 dilution, mouse anti-FLAG at 1:100 dilution. The secondary antibodies were Alexa Fluor 488 goat anti-rabbit 1:2,000 and Alexa Fluor 555 goat anti-mouse 1:2,000 (Invitrogen). The polytene chromosomes were co-stained with DAPI (AppliChem). Images were acquired on the Nikon Elclipse T*i* fluorescent microscope using Nikon DS-Qi2 digital camera, processed with ImageJ 1.50c4 and Fiji bundle 2.0.0-rc-46. 3-4 independent staining and 4-5 samples of polytene chromosomes were performed with each MSL1-expressing transgenic line.

ChIP-seq

Chromatin preparation was performed as described (44,45) with some modifications. Samples of 500 mg each of adult flies were ground in a mortar in liquid nitrogen and resuspended in 10 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, 0.5 mM DTT) supplemented with 0.5 mM PMSF and Calbiochem Complete Protease Inhibitor Cocktail V. The suspension was then homogenized subsequently in a Potter (10 strokes) and a Dounce (tight pestle, 15 strokes) homogenizers and filtered through 70 µm MACS SmartStrainers (Miltenyi Biotec, United States). The homogenate was 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, 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, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine, 0.1% SDS, 0.5 mM PMSF, Calbiochem Complete Protease Inhibitor Cocktail V). The suspension was incubated for 30 min at 4°C and then sonicated with Covaris ME220 (50x15 s with 30 s intervals at 6°C, peak power 75.0, duty % factor 25, cycles/burst 1000), and 50 µL aliquots were used to test the extent of sonication and to measure DNA concentration. Chromatin was then transferred to 1.5 ml-eppendorf tubes and

centrifuged at 14000 rpm for 10 min at 4 °C. Finally, supernatants were pooled, aliquoted and frozen in liquid nitrogen.

For immunoprecipitation chromatin was pre-cleared with Protein A Dynabeads (Thermo Fisher), blocked with 0.5% BSA. Samples containing 20 µg of DNA equivalent in 0,5 mL of nuclear lysis buffer were incubated overnight, at 4°C on a rotary shaker, with rabbit antibodies against MSL1 (1:300), MSL2 (1:200), MSL3 (1:200), or with nonspecific IgG purified from rabbit preimmune sera (control). Chromatin–antibody complexes were collected using blocked Protein A Dynabeads (Thermo Fisher) at 4°C over 5 h. After three rounds of washing with nuclei 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), treated with RNase A for 30 min and proteinase K overnight, incubated at 65 °C for 6 hours and extracted with ChIP DNA Clean&Concentrator (Zymo Research). At least two independent biological replicates were made for each chromatin sample.

The ChIP-seq libraries were prepared with NEBNext® Ultra™ II DNA Library Prep kit, as described in the manufacturer's instructions. In short, eluted DNA was end-repaired and terminal adenosine residues were added using the NEBNext reagents. Indexed adapters were ligated, after which the material was size selected at ~200-600 bp with Ampure XP beads (Beckman Coulter). PCR amplification was performed using NEB primers for 15-16 cycles using the Q5 Hot Start HiFi PCR Master Mix (NEB). The PCR-amplified library was purified using Ampure XP beads and its quality was assessed on a Bioanalyzer 2100 system (Agilent). Diluted libraries were clustered on a pair-read flowcell and sequenced using a NovaSeq 6000 system (Illumina, United States).

ChIP-seq data processing and sequence analysis

All ChIP-seq raw data were presented as two biological replicates with paired-end reads (except MSL1_M1[wt]♂ sample, for which only one biological replicate was obtained). Trimming and

mapping were performed using cutadapt software (46)https://paperpile.com/c/p9fu25/vT8KE and Bowtie version 2 (47), as described previously (48). The dm6 version of the *Drosophila melanogaster* genome was used as a reference genome. After merging replicates, coverage tracks (BedGraph) were obtained using the deepTools (49) bamCoverage function, with bin widths of 100 bp and normalized by reads per kilobase of transcript, per million mapped reads (RPKM). Raw and processed data were deposited in the NCBI Gene Expression Omnibus (GEO) under accession number .

Peak calling was performed only for the MSL2♂ and MSL2_M1[wt]♂ samples, as described previously (48). Briefly:

1) duplicates were removed with Picard (http://broadinstitute.github.io/picard/);
2) blacklist filtration was performed;
3) peak calling was performed using MACS version 2 against nonspecific IgG (50), with a soft p-value threshold of 1×10^{-2} for further irreproducible discovery rate (IDR) analysis (https://sites.google.com/site/anshulkundaje/projects/idr).

Reproducibility was assessed using the IDR pipeline, with p-value thresholds of 0.05 for true replicates and 0.01 for pseudoreplicates. All samples showed ideal reproducibility [both the rescue ratio (RR) and self-consistency ratio (SR) were less than 2, Supplementary Table S2]; thus, an "optimal" set of highly-reproduced peaks was chosen for each sample.

Further analysis was performed in R version 3.6.3 (51). Colocalization analysis was performed using the ChIPpeakAnno package version 3.20.1 (52). Average signal calculation and heatmaps were constructed using the ChIPseeker package version 1.22.1 (53)https://paperpile.com/c/p9fu25/WAMtI. Gene ontology enrichment analysis was performed with the clusterProfiler package (54)https://paperpile.com/c/p9fu25/6BWi.

HAS regions for which ChIP-seq signal enrichment was calculated were obtained by combining HAS variants from three different studies (19,55,56). Only HAS regions located on the X chromosome were considered in this analysis, resulting in 301 HAS regions

(Supplementary Table S3). Autosomal regions were obtained by overlapping the highly-reproducible peaks from the MSL2♂ analysis identified near autosomal promoters (± 200 bp) with promoter regions of genes for which MSL2 enrichment has previously been demonstrated at the 3rd-instar larval stage (57), resulting in 348 autosomal regions (Supplementary Table S4).

RESULTS

The N-terminal domain of MSL1 interacts with many DNA binding zinc-finger proteins

The MSL1 protein (Figure 1A) contains a coiled-coil domain (aa 101-155), which is involved in MSL1 homodimerization and the interaction with MSL2, and the PEHE domain (aa 846-973), which binds to MSL3 and MOF (10,11,14,58). To test whether MSL1 is able to interact with transcription factors containing arrays of C2H2 domains, we prepared a library consisting of 128 zinc-finger proteins fused to the galactose-responsive transcription factor 4 (GAL4) activation domain (Supplementary Table S5). In a yeast two-hybrid assay (Y2H), we screened these zinc-finger proteins against an MSL1 protein fused to the DNA-binding domain of GAL4. As a result (Figure 1B, Supplementary Table S5), we identified interactions between MSL1 and 14 C2H2 proteins, including activators (Ci and Odd) and repressors (Kr and Snail) of early *Drosophila* development, and general transcription factors (Z4 and Row). As a control, we tested MSL2 using the same Y2H assay and only identified an interaction between MSL2 and CLAMP (Supplementary Table S1). This interaction was previously characterized (28,29), which supports the effectiveness of the Y2H assay. We found that 10 of the identified C2H2 proteins interacted with the 178-aa N-terminal region of MSL1 (Figure 1C). The other four proteins (Z4, CG30020, Jim, Snail) interacted with other parts of the MSL1 protein.

To accurately map the protein-protein interacting regions of MSL1, we generated several mutants containing truncations and internal deletions within the 178-aa region and tested them for the ability to interact with the 10 identified C2H2 proteins (Figure 1C). In all cases, the

interaction was lost when the N-terminal 15-aa region was deleted. These results suggested that the N-terminal 15-aa region is able to interact with many different C2H2 proteins. With the exception of the coiled-coil region, the first 15-aa region represents the most conserved region within the N-terminal MSL1 region among Drosophilidae. However, among different families of Diptera (Figures 1D, S1), a high level of homology was only observed for the coiled-coil region, which is required for interactions with MSL2 (14). The N-terminal amino acids are conserved among representatives within each of the Diptera families (Figure S1), but not across families. For example, the MSL1 of *Anopheles* mosquitoes, which are closely related to *Drosophila melanogaster*, features a completely different N-terminal domain from that in *Drosophila*.

Viability of transgenic lines expressing different Msl1 mutants, featuring deletions in the N-terminal domain

To evaluate the functional effects of N-terminal domain deletions in the MSL1 protein *in vivo*, we created transgenic flies expressing wild-type ($MSL1^{WT}$) and mutant variants of Msl1 ($MSL1^{\Delta 1\text{-}15}$, $MSL1^{\Delta 1\text{-}85}$, $MSL1^{\Delta 41\text{-}85}$, and $MSL1^{\Delta 66\text{-}85}$, Figure 2A). The MSL1 proteins were expressed under the control of a strong ubiquitin (Ubi) promoter. To avoid the influence of position effects on the expression of the MSL1 variants, all transgenes were inserted into the same genomic location (86Fb) on the 3rd chromosome, using a φC31-based integration system (41).

To test the roles played by these N-terminal sequences during the specific recruitment of the DCC in males, we crossed these transgenes into the null *msl1*$^{-}$ background (the *msl1*$^{-}$ background corresponds to $msl1^{\gamma 269}/msl1^{L60}$ heterozygote) (12). To compare the expression of the MSL1 variants in transgenic lines, we examined the levels of MSL1 expression in 2-3-day-old *msl1*$^{-}$ females, which are homozygous for the transgenes expressing MSL1 variants. Immunoblot analysis showed that all MSL1 variants were expressed in transgenic females at nearly equivalent levels (Figure 2B).

Homozygous *msl1*$^{-}$ females displayed normal viability, whereas males died during the late larva and early pupa stages (Figure 2C). The MSL1^{WT} transgene, expressed in a heterozygous or homozygous state, fully compensated for the viability of *msl1*$^{-}$ males, which suggested that MSL1^{WT} was functional. A similar result was observed for the expression of MSL1$^{\Delta 66\text{-}85}$, which also fully compensated for the absence of the endogenous MSL1 in *msl1*$^{-}$ male flies. In contrast, all other tested mutant MSL1 proteins (MSL1$^{\Delta 1\text{-}15}$, MSL1$^{\Delta 1\text{-}85}$, and MSL1$^{\Delta 41\text{-}85}$) failed to compensate for the *msl1*$^{-}$ mutation in males (Figure 2C). Taken together, these results suggested that the N-terminal 1–60-aa region of MSL1 is critical for the functional activity of the MSL complex.

N-terminal sequences of MSL1 are critical for the specific recruitment of the dosage compensation complex to the X chromosome

We next examined the extent to which deletions in the N-terminal region of MSL1 would affect 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 of proteins on interphase chromatin and has been used extensively to study dosage compensation (5,30,59-62). Because males expressing the mutant MSL1 variants survived to the late larvae stage, we initially attempted to compare the specific recruitment of MSL1 variants and MSL2 to the polytene chromosomes in male larvae. To visualize the proteins, we used antibodies against MSL2 (29). As expected, the MSL1^{WT} and MSL1$^{\Delta 66\text{-}85}$ proteins were able to recruit the DCC to the X chromosome in males (data not shown). Unfortunately, in the other mutant males, the polytene chromosomes had very poor morphology, and we were unable to obtain unambiguous results regarding the ability of the MSL complex to bind the X chromosome.

Due to the inability to study the binding of the mutant MSL1 variants in males, we used a previously described sensitive model system to study the factors required for the recruitment of

the DCC to the X chromosome in females (42,59,63-65). This model was based on the finding that the ectopic expression of the MSL2 protein in females induced the formation of the DCC on the X chromosome. In a previous study (29), we showed that MSL1 and MSL2 were specifically recruited to the X chromosome in transgenic females expressing MSL2^{WT}-3xFLAG under the control of the Ubi promoter (86Fb). Even heterozygous MSL2^{wt}/+ females displayed reduced viability relative to males. The simultaneous overexpression of MSL1 and MSL2 led to a further decrease in female viability (Figure 2D). High levels of MSL1 and MSL2 were likely to facilitate the formation of a large number of productive MSL complexes, resulting in a significant increase in gene expression from the X chromosome and, consequently, to a significant imbalance in the expression of X-chromosome-encoded genes relative to genes located on autosomes. Females that expressed MSL1$^{\Delta 1\text{-}15}$ or MSL1$^{\Delta 1\text{-}85}$, in combination with MSL2, had nearly normal viability. Mutant MSL1 variants likely form unproductive complexes with MSL2, decreasing the number of functional complexes that can be formed between endogenous MSL1 with MSL2. Females that expressed MSL1$^{\Delta 66\text{-}85}$ and MSL2 displayed low viability, confirming that MSL1$^{\Delta 66\text{-}85}$ functions as a wild-type protein. Interestingly, the co-expression between MSL2 and MSL1$^{\Delta 41\text{-}85}$ also affects female viability, suggesting that MSL1$^{\Delta 41\text{-}85}$ is partially functional.

When MSL2^{WT}-3xFLAG was expressed in the *msl1*$^{-}$ background, we did not observe the binding of the DCC to the female X chromosome (data not shown). However, the binding of MSL1 and MSL2 to the X chromosome was restored in *msl1*$^{-}$ females when MSL2^{WT}-3xFLAG was expressed simultaneously with either MSL1^{WT} or MSL1$^{\Delta 66\text{-}85}$ (Figure 3). These results further supported that the 66–85 aa region is not essential for MSL1 function *in vivo*. In contrast, the expression of all MSL1 variants with deletions of the first 15-aa region (MSL1$^{\Delta 1\text{-}15}$ and MSL1$^{\Delta 1\text{-}85}$) resulted in the near-complete loss of DCC binding to the X chromosome in females (Figure 3). Interestingly, additional MSL1/MSL2 bands were identified on the polytene autosomes suggesting that the MSL proteins began to bind to lower affinity sites on autosomes.

Previous studies (5,59-62) showed that the inactivation of MSL3, MLE, or roX RNAs resulted in the recruitment of the MSL complex to a small proportion of bands that correspond to the main HASs, including regions of the *Rox1* (3F) and *Rox2* (10C) genes. In polytene chromosomes in the MSL1$^{\Delta 1\text{-}15}$ and MSL1$^{\Delta 1\text{-}84}$ lines, we found only two intensive bands that corresponded to the strongest HAS, which is associated with the roX genes (3F and 10C). Several faint MSL1/MSL2 bands could be mapped to other HASs (Figure S2, Supplementary Table S6).

The deletion of the 41–85-aa region also significantly reduced the binding of MSL proteins. MSL1$^{\Delta 41\text{-}85}$ and MSL2 bind to the HASs in *roX* genes and to several accessory HASs on the X chromosome, many of which can also be identified as weak bands in the MSL1$^{\Delta 1\text{-}15}$ and MSL1$^{\Delta 1\text{-}85}$ lines (Figure S2, Supplementary Table S6). In comparison with MSL1$^{\Delta 1\text{-}15}$ and MSL1$^{\Delta 1\text{-}85}$, MSL proteins bind more efficiently with the MSL1$^{\Delta 41\text{-}85}$ polytene chromosomes. However, the binding of the MSL complex to the X chromosome in MSL1 mutants was much weaker than previously suggested for a minimal MSL complex lacking either MSL3 or MLE (62). Taken together, these results suggested that the 1–15-aa and 41–65-aa regions in MSL1 are critical for the recruitment of the DCC to the X chromosome.

The N-terminal domain of MSL1 is essential for the recruitment of the MSL complex to HASs in females expressing MSL2

For a more complete understanding of the role played by the N-terminal region of MSL1 in the recruitment of the MSL complex to the X chromosome, we tested the binding of the MSL1, MSL2, and MSL3 proteins with the HAS regions by chromatin immunoprecipitation-sequencing (ChIP-seq) analysis. We obtained MSL protein profiles from two biological replicates of 3-day-old adult males and females. We used a standard protocol for chromatin fragmentation based on sonication. Previously, this protocol has been shown to partially disrupt MSL complexes, reducing the observed binding of MSL1 and MSL3 proteins to HASs (31). Therefore, the MSL2

profile provides the most reliable information regarding the binding efficiency of the MSL complex to HASs.

Because the binding profiles of the MSL complex has been well-studied in cell lines, we compared the MSL1, MSL2, and MSL3 binding efficiency at the HASs (Figure S3) with previously published data from S2 cells (31). The ChIP-seq data from adult y^1w^1 males showed MSL binding to approximately 90% of the previously identified HASs from three different studies (19,55,56) (Supplementary Table S3). As a control, the MSL proteins do not bind to HASs in females. These results suggested that most HASs are common between S2 cells and adult males.

Next, we compared the binding profiles of MSL1, MSL2, and MSL3 proteins obtained in *msl1*$^-$; MSL2^{WT}-3xFLAG/MSL1^{WT} females and *msl1*$^-$; MSL1^{WT}/TM6,Tb males. We found a good correlation between the obtained profiles, demonstrating that MSL2 expressed in females induces the binding of the MSL complex to the same HAS regions (Figure 4A). Interestingly, we observed stronger binding of the MSL1 and MSL3 proteins in *msl1*$^-$; MSL2^{WT}-3xFLAG /MSL1^{WT} males compared with that in control y^1w^1 males (Figures 4A and S3), which may be a consequence of the higher concentrations of MSL proteins in the transgenic line due to being expressed under the control of the Ubi promoter. In *msl1*$^-$; MSL2^{WT}-3xFLAG /MSL1^{WT} females, the MSL proteins also bind to most HASs, but with weaker efficiency (Figure 4A and B). These results are consistent with previous studies, which demonstrated that the ectopic expression of MSL2 in females was sufficient to trigger the correct targeting of MSL proteins, but with lower efficacy compared with males (57).

In addition to X-linked binding, MSL2 has been shown to associate with a number of autosomal gene promoters (31,57,61,66). MSL2 binding to autosomal promoters does not require roX RNAs and MLE (57), suggesting that a different mechanism is responsible for the MSL2 recruitment to these sites. We identified at least 1,500 autosomal MSL2 peaks in y^1w^1 and *msl1*$^-$;

MSL1^{WT}-3xFLAG/TM6,Tb males (Figure S4). Most of these promoters are associated with genes necessary for various morphogenetic processes (Figure S4 C-E). In addition, 348 peaks identified in adult males (Figure S4) overlapped with previously characterized MSL2 peaks, obtained from female larvae expressing MSL2 (57). A portion of these 348 MSL2 peaks were also identified in *msl1*$^{-}$; MSL2^{WT}-3xFLAG/MSL1^{WT} females (Figure 4A and C). These results suggested that MSL2 stably binds to autosomal promoters in different tissues and stages of *Drosophila* development. These peaks demonstrated only very weak enrichment of MSL1 and MSL3 in *msl1*$^{-}$; MSL2^{WT}/MSL1^{WT} females (Figure 4A and C), which is consistent with the hypothesis of the existence of an alternative mechanism for the recruitment of the MSL complex.

To test the role played by the N-terminal region of MSL1, we compared the binding of MSL proteins in females expressing MSL2 combined with MSL1^{WT}, MSL1$^{\Delta 1\text{-}15}$, or MSL1$^{\Delta 41\text{-}85}$ (Figure 5). As demonstrated above, MSL2 and MSL1^{WT} bound effectively with most HASs in females. MSL3 was associated with extended regions surrounding the HAS, consistent with previously reported data (67,68). Our results agree with previous observations (Figure 5A and B). MSL1$^{\Delta 41\text{-}85}$ strongly reduced, and MSL1$^{\Delta 1\text{-}15}$ almost completely eliminated the recruitment of MSL proteins to HASs (Figure 5A and B). These results confirmed the critical role played by the N-terminal sequence in the specific recruitment of the MSL complex to the X chromosome. In particular, we found identified a crucial decrease in the level of MSL2 binding to the HASs (Figure 5C), for example, with *roX1* and *roX2* (Figure 5E), in the MSL1$^{\Delta 1\text{-}15}$ and MSL1$^{\Delta 41\text{-}85}$ lines.

Interestingly, N-terminal deletions in MSL1 did not significantly affect the binding of the MSL2 protein to autosomal sites (Figure 5D and F). These results supported the assumption (57) that MSL2 is recruited to autosomal promoters through a different mechanism than that involved in the recruitment to the X chromosome.

DISCUSSION

This study focused on examining the role played by the MSL1 protein in the recruitment of the DCC in *Drosophila* to HASs. We found that MSL1 interacts with many C2H2 proteins and is critical for the recruitment of the MSL complex to the male X chromosome. Several transcription factors that interact with MSL1 are well-known transcriptional regulators. Snail and Kruppel are *Drosophila* transcriptional repressors (69) that occasionally function as transcriptional activators (70,71). The Odd and Ci proteins are well-known, highly conserved DNA-binding transcription factors that regulate genes involved in various *Drosophila* developmental processes (72-74). Z4 was previously identified as a partner of the MSL complex (75) that binds to housekeeping gene promoters (76,77) and is a component of the telomere repeat binding factor 2 (TRF2)-dependent promoter complex (78). Other transcription factors that were identified as MSL1-interactors have not yet been characterized.

In this study, we did not analyze the mechanisms of interaction between MSL1 and C2H2 proteins. Our results suggested that the N-terminal 15-aa region of MSL1 is involved in the interaction with many different C2H2 proteins. The unstructured region of MSL1 appears to be able to specifically recognize sequences within the C2H2 proteins. Further study remains necessary to map and characterize the interacting regions in C2H2 proteins. In accordance with the important role played in the recruitment of the DCC to the male X chromosome, this N-terminal region is highly conserved among Drosophilidae. However, MSL1 orthologs in other families of Diptera have completely different N-terminal sequences, despite the presence of high homology in the dimeric coiled-coil region. The lack of conservation between N-terminal regions may be a consequence of the rapid evolution of C2H2 proteins, including several orthologs that are specific to Drosophilidae (79,80). Even between *Anopheles gambiae* and *Drosophila melanogaster*, the X chromosome is thought to have formed independently from the same

ancestral chromosome (81-83). The co-evolution of the MSL1 and C2H2 proteins appears likely to have occurred during the formation of the X chromosome due to regulation by the DCC. The generally accepted model has suggested that, initially, MSL complexes bind with numerous HASs on the male X chromosome (17,84-87). Then, the MSL complexes spread from the HASs to active X chromosomal genes through distance chromatin contacts, formed by preexisting chromosome architecture (88,89), roX RNAs, which modulate the activity and conformation of the MSL complex (16,18,90,91), and MSL3, which recruits the complex to the H3K36me3 nucleosomes that label transcriptionally active genes (67).

Most HASs contain low-complexity GAGA motifs, referred to as the MRE (19,56). The CLAMP protein binds to MREs and is responsible for chromatin opening (28,92). Some C2H2 proteins that interact with MSL1 can functionally cooperate with CLAMP and bind to HASs. Because these proteins can bind to different sites, the HAS sequences can display high levels of variation that can be difficult to identify using bioinformatic approaches. MSL2 contributes to the specific recognition of HASs through the direct binding of GAGA motifs (23) and interactions with CLAMP (28,29). However, our results suggested that the protein-protein interactions mediated by the N-terminal domain of MSL1 are the most critical interactions that define the specific binding of MSL to the X chromosome. This finding is in agreement with previous findings that the N-terminal 265 aa of MSL1, in cooperation with MSL2 and possibly with MLE, but not with MSL3 and MOF, can specifically bind to several sites on female polytene chromosomes (30). Rieder et al. (92) showed that at the initiation of embryogenesis, the DCC first binds to all chromosomes before becoming concentrated on the X chromosome. Weak interactions with chromatin sites on autosomes likely lead to the rapid dissociation of the DCC, resulting in the gradual concentration of the DCC on HASs. In contrast with the recruitment of the MSL complex to HASs on the X chromosome, the binding of MSL2 to autosomal promoters does not require the N-terminal region of MSL1. Although MSL1 and MSL3 co-localize with most of the autosomal binding regions of MSL2, MSL2 is likely able to bind to autosomal sites

independently. MSL1, in cooperation with MSL3 and MOF, is also associated with many autosomal promoters, independently of MSL2 (31,32). Autosomal sites likely create a pool of less specific sites for the full MSL complex that are more effectively bound by the MSL subcomplexes.

Complete dose compensation is established during embryogenesis, and the effectiveness of this process is directly correlated with the distance of the genes from the HASs (89). The chromosome architecture supports local interactions between HASs and neighboring active genes, which facilitates the spread of DCC complexes to the regions containing genes whose transcription should be increased in males (88,93). A comparison of the X chromosome architecture in distant *Drosophila* species showed that spatial contacts with potential HAS regions are conserved, despite the existence of a large number of genomic rearrangements between the studied species (94). The C2H2 proteins that interact with MSL1 and CLAMP are likely to contribute both to distance interactions between HASs and genes and to the formation of secondary (weak) MSL recruitment sites at target genes. Further study remains necessary to confirm the current model and to understand interaction mechanisms between MSL1 and C2H2 proteins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

We thank M Kuroda (Harvard Medical School), and E. Larschan (Brown University) for kindly providing fly strains. We thank Farhod Hasanov and Aleksander Parshikov for fly injection. This study was performed using the equipment of the Center for Precision Genome Editing and

Genetic Technologies for Biomedicine of IGB RAS supported by the Ministry of Science and Education of the Russian Federation.

FUNDING

This work was supported by the grant 075-15-2019-1661 from the Ministry of Science and Higher Education of the Russian Federation. The work to produce mutant variants of MSL1 in vivo was supported by RFBR 18-29-07081.

Funding for open access charge: Ministry of Science and Education of the Russian Federation.

Conflict of interest statement. None declared.

REFERENCES

1. Lucchesi, J.C. (2018) Transcriptional modulation of entire chromosomes: dosage compensation. *J Genet*, **97**, 357-364.
2. Samata, M. and 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.
3. Kuroda, M.I., Hilfiker, A. and Lucchesi, J.C. (2016) Dosage Compensation in Drosophila-a Model for the Coordinate Regulation of Transcription. *Genetics*, **204**, 435-450.
4. Ferrari, F., Alekseyenko, A.A., Park, P.J. and Kuroda, M.I. (2014) Transcriptional control of a whole chromosome: emerging models for dosage compensation. *Nature structural & molecular biology*, **21**, 118-125.
5. Meller, V.H. and Rattner, B.P. (2002) The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *The EMBO journal*, **21**, 1084-1091.
6. Kim, M., Faucillion, M.L. and Larsson, J. (2018) RNA-on-X 1 and 2 in Drosophila melanogaster fulfill separate functions in dosage compensation. *PLoS Genet*, **14**, e1007842.
7. Beckmann, K., Grskovic, M., Gebauer, F. and Hentze, M.W. (2005) A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila. *Cell*, **122**, 529-540.
8. Villa, R., Forne, I., Muller, M., Imhof, A., Straub, T. and Becker, P.B. (2012) MSL2 combines sensor and effector functions in homeostatic control of the Drosophila dosage compensation machinery. *Mol Cell*, **48**, 647-654.

9. Wu, L., Zee, B.M., Wang, Y., Garcia, B.A. and Dou, Y. (2011) The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. *Mol Cell*, **43**, 132-144.

10. Kadlec, J., Hallacli, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S. and Akhtar, A. (2011) Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nature structural & molecular biology*, **18**, 142-149.

11. Scott, M.J., Pan, L.L., Cleland, S.B., Knox, A.L. and Heinrich, J. (2000) MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila. *The EMBO journal*, **19**, 144-155.

12. Palmer, M.J., Mergner, V.A., Richman, R., Manning, J.E., Kuroda, M.I. and Lucchesi, J.C. (1993) The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes a novel protein that associates with the X chromosome in males. *Genetics*, **134**, 545-557.

13. Gorman, M., Kuroda, M.I. and Baker, B.S. (1993) Regulation of the sex-specific binding of the maleless dosage compensation protein to the male X chromosome in Drosophila. *Cell*, **72**, 39-49.

14. Hallacli, E., Lipp, M., Georgiev, P., Spielman, C., Cusack, S., Akhtar, A. and 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.

15. Maenner, S., Muller, M., Frohlich, J., Langer, D. and Becker, P.B. (2013) ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol Cell*, **51**, 174-184.

16. Ilik, I.A., Maticzka, D., Georgiev, P., Gutierrez, N.M., Backofen, R. and Akhtar, A. (2017) A mutually exclusive stem-loop arrangement in roX2 RNA is essential for X-chromosome regulation in Drosophila. *Genes & development*, **31**, 1973-1987.

17. Meller, V.H., Gordadze, P.R., Park, Y., Chu, X., Stuckenholz, C., Kelley, R.L. and Kuroda, M.I. (2000) Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila. *Current biology : CB*, **10**, 136-143.

18. Muller, M., Schauer, T., Krause, S., Villa, R., Thomae, A.W. and Becker, P.B. (2020) Two-step mechanism for selective incorporation of lncRNA into a chromatin modifier. *Nucleic Acids Res*, **48**, 7483-7501.

19. Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J. *et al.* (2008) A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell*, **134**, 599-609.

20. Straub, T. and Becker, P.B. (2008) DNA sequence and the organization of chromosomal domains. *Current opinion in genetics & development*, **18**, 175-180.

21. Zheng, S., Villa, R., Wang, J., Feng, Y., Becker, P.B. and Ye, K. (2014) Structural basis of X chromosome DNA recognition by the MSL2 CXC domain during Drosophila dosage compensation. *Genes & development*, **28**, 2652-2662.

22. Fauth, T., Muller-Planitz, F., Konig, C., Straub, T. and Becker, P.B. (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.

23. Villa, R., Schauer, T., Smialowski, P., Straub, T. and Becker, P.B. (2016) PionX sites mark the X chromosome for dosage compensation. *Nature*, **537**, 244-248.

24. Larschan, E., Soruco, M.M., Lee, O.K., Peng, S., Bishop, E., Chery, J., Goebel, K., Feng, J., Park, P.J. and Kuroda, M.I. (2012) Identification of chromatin-associated regulators of MSL complex targeting in Drosophila dosage compensation. *PLoS Genet*, **8**, e1002830.

25. Soruco, M.M., Chery, J., Bishop, E.P., Siggers, T., Tolstorukov, M.Y., Leydon, A.R., Sugden, A.U., Goebel, K., Feng, J., Xia, P. *et al.* (2013) The CLAMP protein links the

MSL complex to the X chromosome during Drosophila dosage compensation. *Genes & development*, **27**, 1551-1556.

26. Kuzu, G., Kaye, E.G., Chery, J., Siggers, T., Yang, L., Dobson, J.R., Boor, S., Bliss, J., Liu, W., Jogl, G. *et al.* (2016) Expansion of GA Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to Promote Drosophila Dosage Compensation. *PLoS Genet*, **12**, e1006120.

27. Urban, J., Kuzu, G., Bowman, S., Scruggs, B., Henriques, T., Kingston, R., Adelman, K., Tolstorukov, M. and Larschan, E. (2017) Enhanced chromatin accessibility of the dosage compensated Drosophila male X-chromosome requires the CLAMP zinc finger protein. *PLoS One*, **12**, e0186855.

28. Albig, C., Tikhonova, E., Krause, S., Maksimenko, O., Regnard, C. and Becker, P.B. (2019) Factor cooperation for chromosome discrimination in Drosophila. *Nucleic Acids Res*, **47**, 1706-1724.

29. Tikhonova, E., Fedotova, A., Bonchuk, A., Mogila, V., Larschan, E.N., Georgiev, P. and 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**.

30. Li, F., Parry, D.A. and Scott, M.J. (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.

31. Straub, T., Zabel, A., Gilfillan, G.D., Feller, C. and Becker, P.B. (2013) Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq. *Genome Res*, **23**, 473-485.

32. Chlamydas, S., Holz, H., Samata, M., Chelmicki, T., Georgiev, P., Pelechano, V., Dundar, F., Dasmeh, P., Mittler, G., Cadete, F.T. *et al.* (2016) Functional interplay

between MSL1 and CDK7 controls RNA polymerase II Ser5 phosphorylation. *Nature structural & molecular biology*, **23**, 580-589.

33. Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F. *et al.* (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science*, **330**, 1787-1797.

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

35. Schwartz, Y.B., Linder-Basso, D., Kharchenko, P.V., Tolstorukov, M.Y., Kim, M., Li, H.B., Gorchakov, A.A., Minoda, A., Shanower, G., Alekseyenko, A.A. *et al.* (2012) Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Res*, **22**, 2188-2198.

36. Li, J. and Gilmour, D.S. (2013) Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *The EMBO journal*, **32**, 1829-1841.

37. Zolotarev, N., Fedotova, A., Kyrchanova, O., Bonchuk, A., Penin, A.A., Lando, A.S., Eliseeva, I.A., Kulakovskiy, I.V., Maksimenko, O. and Georgiev, P. (2016) Architectural proteins Pita, Zw5,and ZIPIC contain homodimerization domain and support specific long-range interactions in Drosophila. *Nucleic Acids Res*, **44**, 7228-7241.

38. Zolotarev, N., Maksimenko, O., Kyrchanova, O., Sokolinskaya, E., Osadchiy, I., Girardot, C., Bonchuk, A., Ciglar, L., Furlong, E.E.M. and Georgiev, P. (2017) Opbp is a new architectural/insulator protein required for ribosomal gene expression. *Nucleic Acids Research*, **45**, 12285-12300.

39. Maksimenko, O., Bartkuhn, M., Stakhov, V., Herold, M., Zolotarev, N., Jox, T., Buxa, M.K., Kirsch, R., Bonchuk, A., Fedotova, A. *et al.* (2015) Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Research*, **25**, 89-99.

40. Maksimenko, O., Kyrchanova, O., Klimenko, N., Zolotarev, N., Elizarova, A., Bonchuk, A. and Georgiev, P. (2020) Small Drosophila zinc finger C2H2 protein with an N-terminal zinc finger-associated domain demonstrates the architecture functions. *Biochimica et biophysica acta. Gene regulatory mechanisms*, **1863**, 194446.

41. Bischof, J., Maeda, R.K., Hediger, M., Karch, F. and Basler, K. (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 3312-3317.

42. Chang, K.A. and Kuroda, M.I. (1998) Modulation of MSL1 abundance in female Drosophila contributes to the sex specificity of dosage compensation. *Genetics*, **150**, 699-709.

43. Murawska, M. and Brehm, A. (2012) Immunostaining of Drosophila polytene chromosomes to investigate recruitment of chromatin-binding proteins. *Methods Mol Biol*, **809**, 267-277.

44. Maksimenko, O., Bartkuhn, M., Stakhov, V., Herold, M., Zolotarev, N., Jox, T., Buxa, M.K., Kirsch, R., Bonchuk, A., Fedotova, A. *et al.* (2015) Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Res*, **25**, 89-99.

45. Zolotarev, N., Maksimenko, O., Kyrchanova, O., Sokolinskaya, E., Osadchiy, I., Girardot, C., Bonchuk, A., Ciglar, L., Furlong, E.E.M. and Georgiev, P. (2017) Opbp is a new architectural/insulator protein required for ribosomal gene expression. *Nucleic Acids Res*, **45**, 12285-12300.

46. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10-12.

47. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**, 357-359.

48. Sabirov, M., Kyrchanova, O., Pokholkova, G.V., Bonchuk, A., Klimenko, N., Belova, E., Zhimulev, I.F., Maksimenko, O. and Georgiev, P. (2020) Mechanism and functional role of the interaction between CP190 and the architectural protein Pita in Drosophila melanogaster. *bioRxiv*, 2020.2010.2026.355016.

49. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*, **42**, W187-191.

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

51. Team, R.C. (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

52. Zhu, L.J., Gazin, C., Lawson, N.D., Pages, H., Lin, S.M., Lapointe, D.S. and Green, M.R. (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC bioinformatics*, **11**, 237.

53. Yu, G., Wang, L.G. and He, Q.Y. (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382-2383.

54. Yu, G., Wang, L.G., Han, Y. and He, Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*, **16**, 284-287.

55. Ramirez, F., Lingg, T., Toscano, S., Lam, K.C., Georgiev, P., Chung, H.R., Lajoie, B.R., de Wit, E., Zhan, Y., de Laat, W. *et al.* (2015) High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X Chromosome in Drosophila. *Mol Cell*, **60**, 146-162.

56. Straub, T., Grimaud, C., Gilfillan, G.D., Mitterweger, A. and Becker, P.B. (2008) The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. *PLoS Genet*, **4**, e1000302.

57. Valsecchi, C.I.K., Basilicata, M.F., Semplicio, G., Georgiev, P., Gutierrez, N.M. and Akhtar, A. (2018) Facultative dosage compensation of developmental genes on autosomes in Drosophila and mouse embryonic stem cells. *Nature communications*, **9**, 3626.

58. Copps, K., Richman, R., Lyman, L.M., Chang, K.A., Rampersad-Ammons, J. and Kuroda, M.I. (1998) Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. *The EMBO journal*, **17**, 5409-5417.

59. Palmer, M.J., Richman, R., Richter, L. and Kuroda, M.I. (1994) Sex-specific regulation of the male-specific lethal-1 dosage compensation gene in Drosophila. *Genes & development*, **8**, 698-706.

60. Lyman, L.M., Copps, K., Rastelli, L., Kelley, R.L. and Kuroda, M.I. (1997) Drosophila male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics*, **147**, 1743-1753.

61. Demakova, O.V., Kotlikova, I.V., Gordadze, P.R., Alekseyenko, A.A., Kuroda, M.I. and Zhimulev, I.F. (2003) The MSL complex levels are critical for its correct targeting to the chromosomes in Drosophila melanogaster. *Chromosoma*, **112**, 103-115.

62. Dahlsveen, I.K., Gilfillan, G.D., Shelest, V.I., Lamm, R. and Becker, P.B. (2006) Targeting determinants of dosage compensation in Drosophila. *PLoS Genet*, **2**, e5.

63. Zhou, S., Yang, Y., Scott, M.J., Pannuti, A., Fehr, K.C., Eisen, A., Koonin, E.V., Fouts, D.L., Wrightsman, R., Manning, J.E. *et al.* (1995) Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex-specific regulation and encodes a

protein with a RING finger and a metallothionein-like cysteine cluster. *The EMBO journal*, **14**, 2884-2895.

64. Li, F., Schiemann, A.H. and Scott, M.J. (2008) Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex. *Mol Cell Biol*, **28**, 1252-1264.

65. Morra, R., Yokoyama, R., Ling, H. and Lucchesi, J.C. (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.

66. Bone, J.R., Lavender, J., Richman, R., Palmer, M.J., Turner, B.M. and Kuroda, M.I. (1994) Acetylated histone H4 on the male X chromosome is associated with dosage compensation in Drosophila. *Genes & development*, **8**, 96-104.

67. Sural, T.H., Peng, S., Li, B., Workman, J.L., Park, P.J. and Kuroda, M.I. (2008) The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome. *Nature structural & molecular biology*, **15**, 1318-1325.

68. Larschan, E., Alekseyenko, A.A., Gortchakov, A.A., Peng, S., Li, B., Yang, P., Workman, J.L., Park, P.J. and Kuroda, M.I. (2007) MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell*, **28**, 121-133.

69. Esposito, E., Lim, B., Guessous, G., Falahati, H. and Levine, M. (2016) Mitosis-associated repression in development. *Genes & development*, **30**, 1503-1508.

70. Rembold, M., Ciglar, L., Yanez-Cuna, J.O., Zinzen, R.P., Girardot, C., Jain, A., Welte, M.A., Stark, A., Leptin, M. and Furlong, E.E. (2014) A conserved role for Snail as a potentiator of active transcription. *Genes & development*, **28**, 167-181.

71. Jaeger, J. (2011) The gap gene network. *Cellular and molecular life sciences : CMLS*, **68**, 243-274.

72. Burglin, T.R. (2008) The Hedgehog protein family. *Genome biology*, **9**, 241.

73. Von Ohlen, T., Lessing, D., Nusse, R. and Hooper, J.E. (1997) Hedgehog signaling regulates transcription through cubitus interruptus, a sequence-specific DNA binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 2404-2409.

74. Coulter, D.E., Swaykus, E.A., Beran-Koehn, M.A., Goldberg, D., Wieschaus, E. and Schedl, P. (1990) Molecular analysis of odd-skipped, a zinc finger encoding segmentation gene with a novel pair-rule expression pattern. *The EMBO journal*, **9**, 3795-3804.

75. Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J. *et al*. (2006) Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. *Mol Cell*, **21**, 811-823.

76. Cubenas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P. and Corces, V.G. (2017) Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res*, **45**, 1714-1730.

77. Gortchakov, A.A., Eggert, H., Gan, M., Mattow, J., Zhimulev, I.F. and Saumweber, H. (2005) Chriz, a chromodomain protein specific for the interbands of Drosophila melanogaster polytene chromosomes. *Chromosoma*, **114**, 54-66.

78. Hochheimer, A., Zhou, S., Zheng, S., Holmes, M.C. and Tjian, R. (2002) TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. *Nature*, **420**, 439-445.

79. Fedotova, A.A., Bonchuk, A.N., Mogila, V.A. and Georgiev, P.G. (2017) C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors. *Acta naturae*, **9**, 47-58.

80. Heger, P., George, R. and Wiehe, T. (2013) Successive gain of insulator proteins in arthropod evolution. *Evolution; international journal of organic evolution*, **67**, 2945-2956.

81. Toups, M.A. and Hahn, M.W. (2010) Retrogenes reveal the direction of sex-chromosome evolution in mosquitoes. *Genetics*, **186**, 763-766.

82. Pease, J.B. and Hahn, M.W. (2012) Sex chromosomes evolved from independent ancestral linkage groups in winged insects. *Molecular biology and evolution*, **29**, 1645-1653.

83. Jiang, X., Biedler, J.K., Qi, Y., Hall, A.B. and Tu, Z. (2015) Complete Dosage Compensation in Anopheles stephensi and the Evolution of Sex-Biased Genes in Mosquitoes. *Genome biology and evolution*, **7**, 1914-1924.

84. Park, Y., Mengus, G., Bai, X., Kageyama, Y., Meller, V.H., Becker, P.B. and Kuroda, M.I. (2003) Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex. *Mol Cell*, **11**, 977-986.

85. Kelley, R.L., Wang, J., Bell, L. and Kuroda, M.I. (1997) Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism. *Nature*, **387**, 195-199.

86. Kageyama, Y., Mengus, G., Gilfillan, G., Kennedy, H.G., Stuckenholz, C., Kelley, R.L., Becker, P.B. and Kuroda, M.I. (2001) Association and spreading of the Drosophila dosage compensation complex from a discrete roX1 chromatin entry site. *The EMBO journal*, **20**, 2236-2245.

87. Park, Y. and Kuroda, M.I. (2001) Epigenetic aspects of X-chromosome dosage compensation. *Science*, **293**, 1083-1085.

88. Schauer, T., Ghavi-Helm, Y., Sexton, T., Albig, C., Regnard, C., Cavalli, G., Furlong, E.E. and Becker, P.B. (2017) Chromosome topology guides the Drosophila Dosage Compensation Complex for target gene activation. *EMBO reports*.

89. Prayitno, K., Schauer, T., Regnard, C. and Becker, P.B. (2019) Progressive dosage compensation during Drosophila embryogenesis is reflected by gene arrangement. *EMBO reports*, **20**, e48138.

90. Lv, M., Yao, Y., Li, F., Xu, L., Yang, L., Gong, Q., Xu, Y.Z., Shi, Y., Fan, Y.J. and Tang, Y. (2019) Structural insights reveal the specific recognition of roX RNA by the dsRNA-binding domains of the RNA helicase MLE and its indispensable role in dosage compensation in Drosophila. *Nucleic Acids Res*, **47**, 3142-3157.

91. Ankush Jagtap, P.K., Muller, M., Masiewicz, P., von Bulow, S., Hollmann, N.M., Chen, P.C., Simon, B., Thomae, A.W., Becker, P.B. and Hennig, J. (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.

92. Rieder, L.E., Jordan, W.T., 3rd and Larschan, E.N. (2019) Targeting of the Dosage-Compensated Male X-Chromosome during Early Drosophila Development. *Cell reports*, **29**, 4268-4275 e4262.

93. Pal, K., Forcato, M., Jost, D., Sexton, T., Vaillant, C., Salviato, E., Mazza, E.M.C., Lugli, E., Cavalli, G. and Ferrari, F. (2019) Global chromatin conformation differences in the Drosophila dosage compensated chromosome X. *Nature communications*, **10**, 5355.

94. Renschler, G., Richard, G., Valsecchi, C.I.K., Toscano, S., Arrigoni, L., Ramirez, F. and Akhtar, A. (2019) Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. *Genes & development*, **33**, 1591-1612.

FIGURE LEGENDS

Figure 1. Structure organization and protein-protein interactions of MSL1.

(A) Scheme of MSL1, including regions examined in the analysis. CC, coiled-coil domain. PEHE, domain interacting with MSL3 and MOF. **(B)** Mapping the MSL1 domains that interacted with zinc-finger proteins. **(C)** Detailed mapping of the N-terminal domain regions in MSL1 that interact with Zinc-finger proteins. **(D)** Alignment of the N-terminal domain (1-85 aa) of MSL1 among Drosophilidae.

Figure 2. Testing the functional activity of MSL1 mutants.

(A) Schematic presentation of the MSL1 deletion proteins expressed in flies. **(B)** Western blot analysis of protein extracts prepared from adult flies expressing different MSL1 protein variants ($MSL1^{WT}$, $MSL1^{\Delta 1\text{-}15}$, $MSL1^{\Delta 1\text{-}84}$, $MSL1^{\Delta 41\text{-}85}$, $MSL1^{\Delta 66\text{-}85}$). Immunoblot analysis performed using MSL1 and GAF antibodies (internal control). **(C)** Viability of males expressing MSL1 variants in the *msl1*-null background. Viability (as a relative percentage) of *msl1⁻;* MSL1*/MSL1* males to *msl1⁻/*CyO, GFP*;* MSL1*/ MSL1* males obtained in the progeny of cross between females and males with *msl1⁻/* CyO, GFP*;* MSL1*/MSL1* genotype. MSL1* is any transgene expressing one of the tested MSL1 variants. At least 500 flies were scored in three independent experiments. The results are expressed as the mean of three independent crosses, with error bars showing standard deviations. **(D)** Viability (as a relative percentage) of female adult flies simultaneously expressing MSL1* and MSL2 to female adult flies expressing only MSL1*. Scored flies were obtained from the progeny of crosses between MSL1*/MSL1* females and MSL2-3xFLAG/TM6,Tb males. MSL2-3xFLAG is transgene (86Fb) expressing the MSL2 cDNA under the control of the Ubi promoter at the 86Fb site. At least 300 females were scored in three independent experiments. The results are expressed as the mean of three independent crosses with error bars showing standard deviations.

Figure 3. Testing the functional activity of MSL1 mutants in female model systems.

Distribution of the MSL complex on the polytene chromosomes from 3rd-day female larvae expressing both MSL1 variants and the MSL2-3xFLAG protein. Panels show the immunostaining of proteins using rabbit anti-MSL1 antibody (green) and mouse anti-FLAG antibody (red). DNA was stained with DAPI (blue).

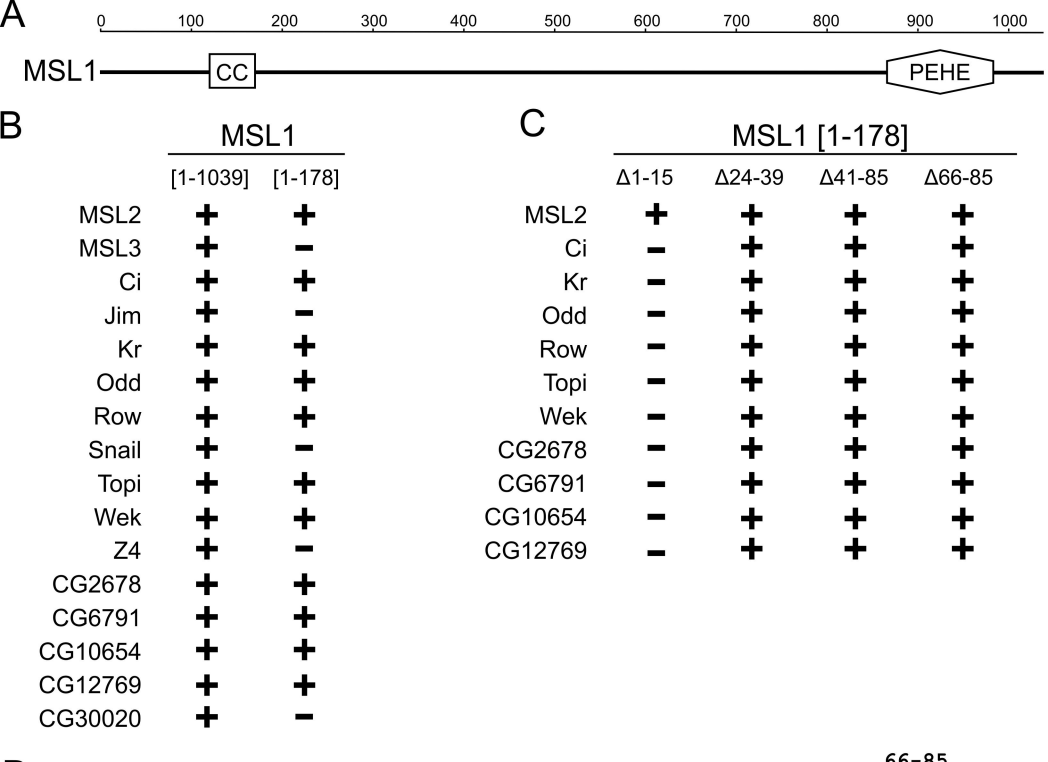
Figure 4. Comparisons of the binding patterns of MSL proteins between 2-3-day adult males and females expressing MSL2

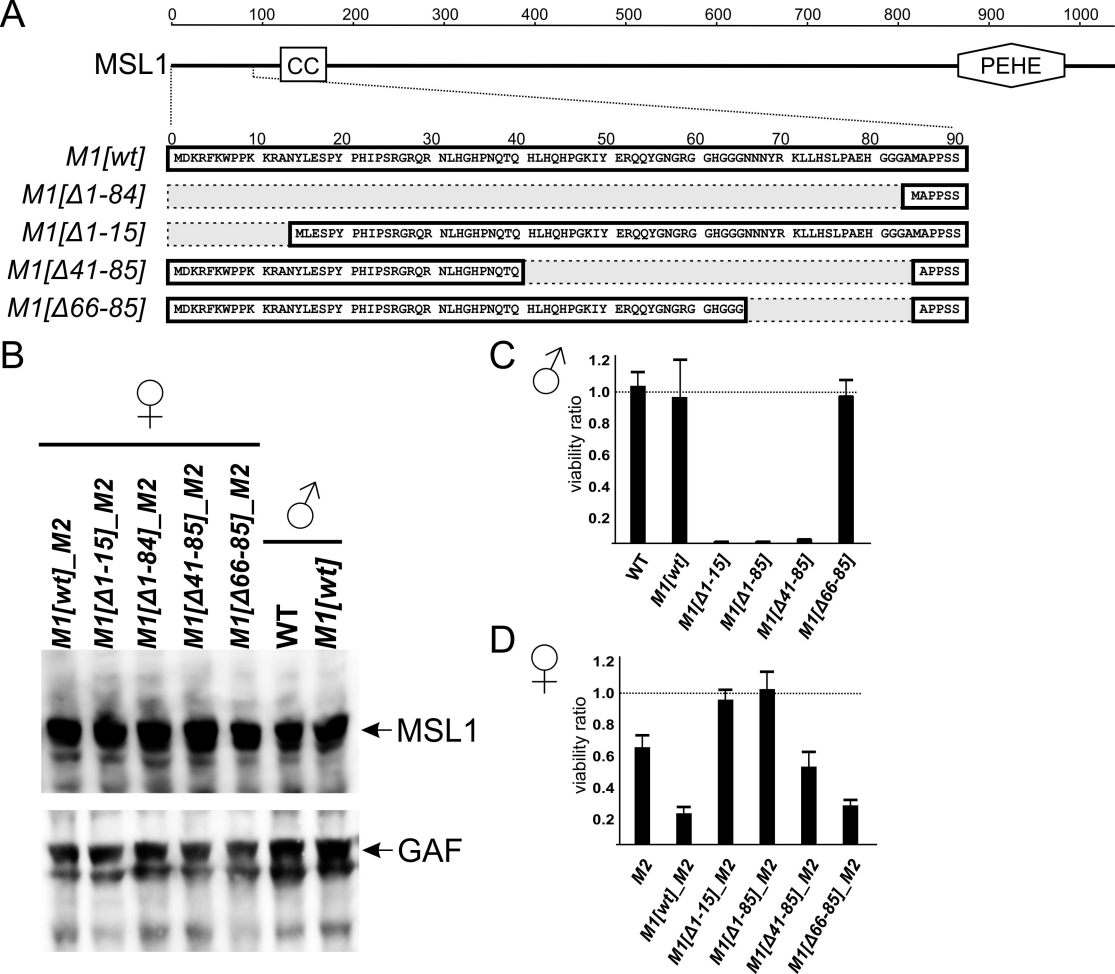
(A) Average signal (on the top) and heatmap (on the bottom) showing the log fold-change (FC) between normalized (RPKM) test signals and nonspecific IgG signals for HAS and autosomal regions (see Materials and Methods). LogFC was calculated after smoothing signals using Daniell kernel, with kernel size 50 and the addition of a pseudocount. On the heatmaps, the peaks are ranked according to the average logFC in the MSL2_M1[wt]♂ sample. (**B)** Enrichment of MSL1, MSL2, and MSL3 signals associated with the X-linked gene Pp2C1 and the non-coding RNAs roX1 and roX2. (**C)** Examples of MSL1, MSL2, and MSL3 signal enrichment associated with the autosomal genes ATPsynC, Men-b, and non-coding RNA lncRNA:CR43650.

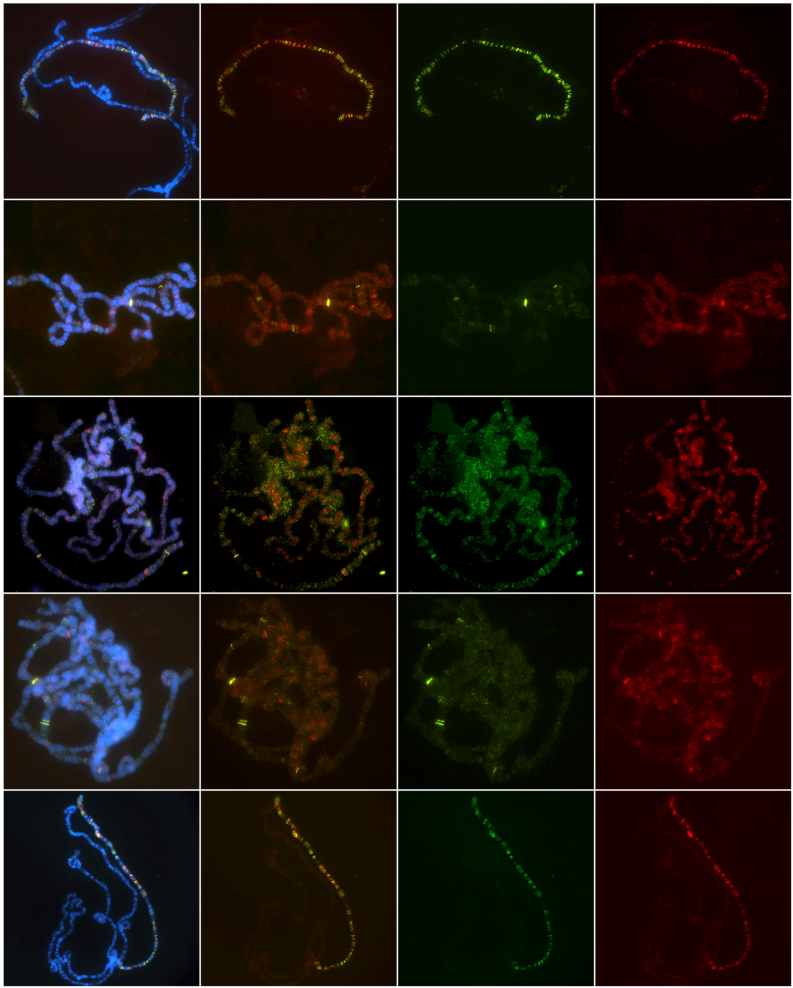
Figure 5. Role of the N-terminal regions in the recruitment of the dosage compensation complex to HAS regions.

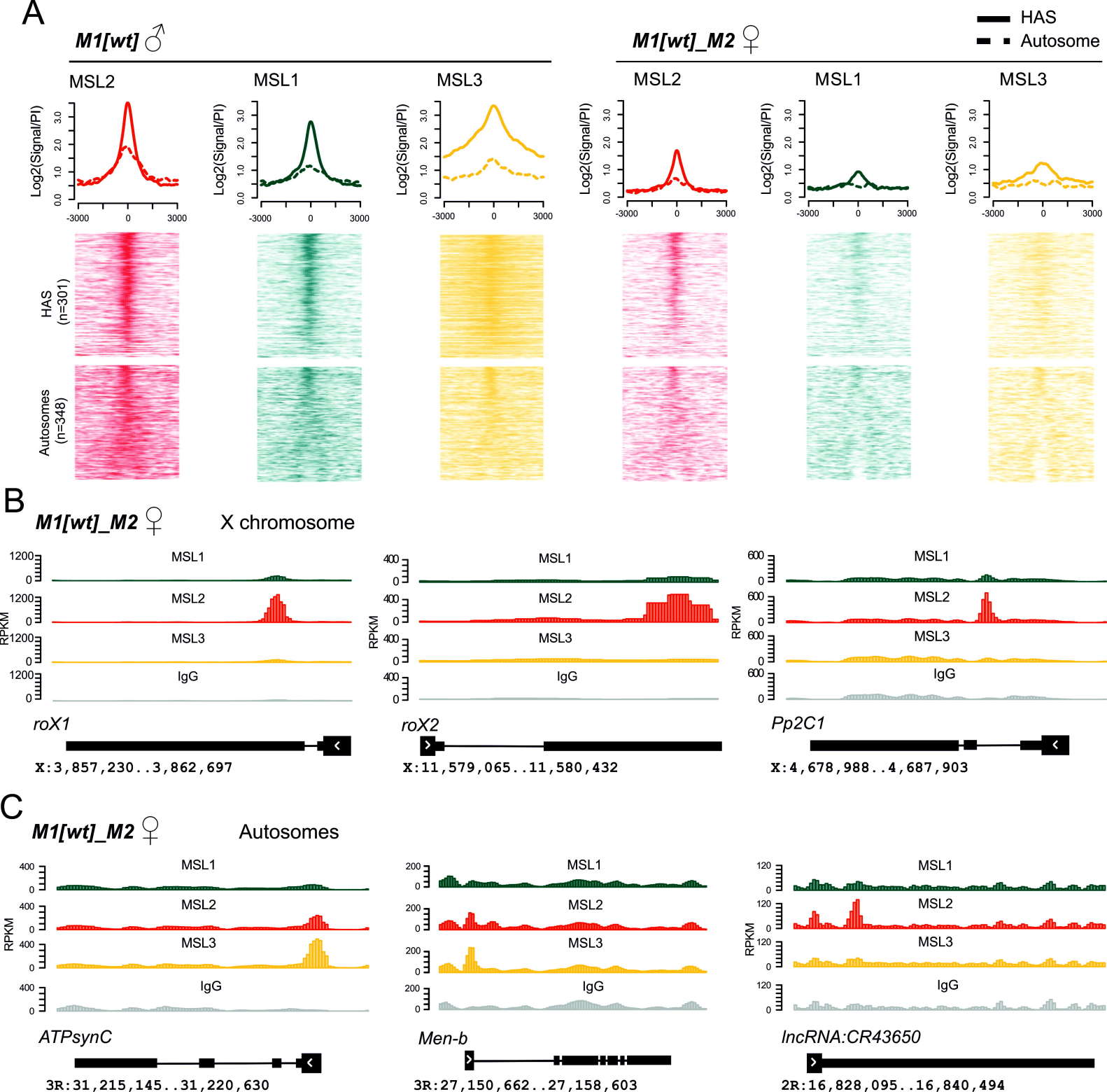
To study the functional role of the N-terminal region of MSL1 for the recruitment of the dosage compensation complex, we compared the profiles of MSL1, MSL2, and MSL3 binding in 3-day females expressing MSL2 and one of three MSL1 variants ($MSL1^{WT}$, $MSL1^{\Delta 1\text{-}15}$, $MSL1^{\Delta 41\text{-}85}$).

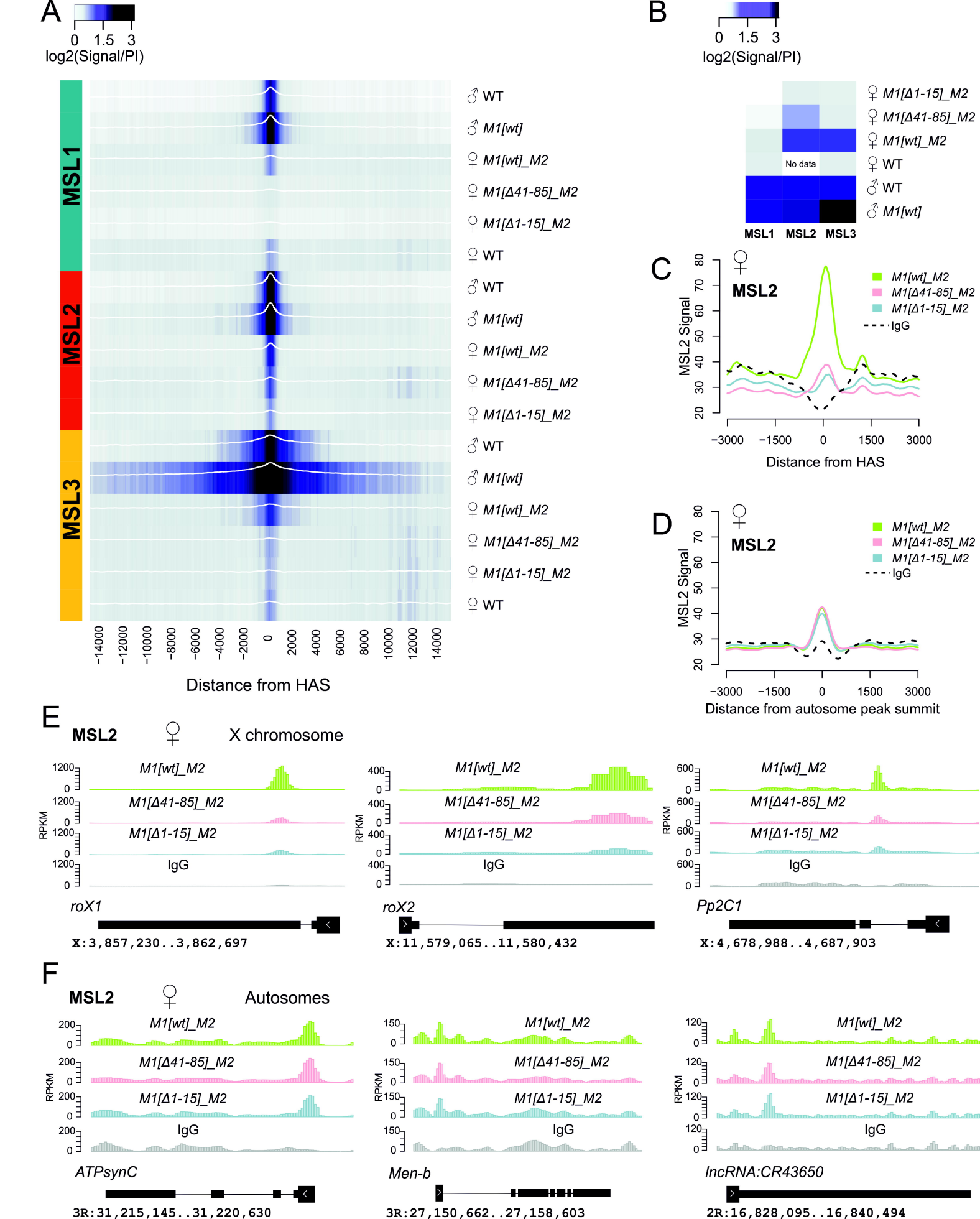
(A)Average log fold-change between normalized (RPKM) test signals and nonspecific IgG signals in HASs (see Materials and Methods). Average log fold-change was calculated after smoothing signals using Daniell kernel, with kernel size 50 and the addition of a pseudocount. Then, the tracks were additionally aggregated using 100-bp bins. (**B)** Heatmap showing average log fold-change between normalized (RPKM) test signals and nonspecific IgG signals in region ± 500 bps from HAS centers. (**C** and **D)** Average signal (RPKM) for MSL2 protein in different fly lines at HAS (C) and autosomal (D) regions (see Methods). (E) Enrichment of MSL2 protein signals in different fly lines at the X-linked gene Pp2C1 and non-coding RNAs roX1 and roX2. (**F)** Examples of MSL2 signal enrichment in different fly lines for the autosomal genes ATPsynC, Men-b, and non-coding RNA IncRNA:CR43650.


A
0
100
200
300
400
500
600
700
800
900
1000
MSL1
CC
PEHE
B
MSL1
[1-1039]
[1-178]
MSL2
MSL3
Ci
Jim
Kr
Odd
Row
Snail
Topi
Wek
Z4
CG2678
CG6791
CG10654
CG12769
CG30020
C
MSL1 [1-178]
Δ1-15
Δ24-39
Δ41-85
Δ66-85
MSL2
Ci
Kr
Odd
Row
Topi
Wek
CG2678
CG6791
CG10654
CG12769
D
1-15
24-39
41-85
66-85
D.simulans
D.sechellia
D.melanogaster
D.yakuba
D.erecta
D.ananassae
D.persimilis
D.pseudoobscura
D.willistoni
D.virilis
D.grimshawi


A
0 100 200 300 400 500 600 700 800 900 1000
MSL1
CC
PEHE
0 10 20 30 40 50 60 70 80 90
M1[wt] MDKRFKWPPK KRANYLESPY PHIPSRGRQR NLHGHPNQTQ HLHQHPGKIY ERQQYGNGRG GHGGGNNNYR KLLHSLPAEH GGGAMAPPSS
M1[Δ1-84] MAPPSS
M1[Δ1-15] MLESPY PHIPSRGRQR NLHGHPNQTQ HLHQHPGKIY ERQQYGNGRG GHGGGNNNYR KLLHSLPAEH GGGAMAPPSS
M1[Δ41-85] MDKRFKWPPK KRANYLESPY PHIPSRGRQR NLHGHPNQTQ APPSS
M1[Δ66-85] MDKRFKWPPK KRANYLESPY PHIPSRGRQR NLHGHPNQTQ HLHQHPGKIY ERQQYGNGRG GHGGG APPSS
B
♀
♂
M1[wt]_M2
M1[Δ1-15]_M2
M1[Δ1-84]_M2
M1[Δ41-85]_M2
M1[Δ66-85]_M2
WT
M1[wt]
MSL1
GAF
C
♂
viability ratio
1.2
1.0
0.8
0.6
0.4
0.2
WT
M1[wt]
M1[Δ1-15]
M1[Δ1-85]
M1[Δ41-85]
M1[Δ66-85]
D
♀
viability ratio
1.2
1.0
0.8
0.6
0.4
0.2
M2
M1[wt]_M2
M1[Δ1-15]_M2
M1[Δ1-85]_M2
M1[Δ41-85]_M2
M1[Δ66-85]_M2


DAPI +
merged
merged
MSL1
MSL2
WT
Δ1-84
Δ1-15
Δ41-85
Δ66-85


A
M1[wt] ♂
M1[wt]_M2 ♀
HAS
Autosome
MSL2
MSL1
MSL3
Log2(Signal/PI)
HAS (n=301)
Autosomes (n=348)
B
M1[wt]_M2 ♀
X chromosome
MSL1
MSL2
MSL3
IgG
RPKM
roX1
X:3,857,230..3,862,697
roX2
X:11,579,065..11,580,432
Pp2C1
X:4,678,988..4,687,903
C
M1[wt]_M2 ♀
Autosomes
ATPsynC
3R:31,215,145..31,220,630
Men-b
3R:27,150,662..27,158,603
lncRNA:CR43650
2R:16,828,095..16,840,494


A
0 1.5 3
log2(Signal/PI)
MSL1
MSL2
MSL3
♂ WT
♂ M1[wt]
♀ M1[wt]_M2
♀ M1[Δ41-85]_M2
♀ M1[Δ1-15]_M2
♀ WT
♂ WT
♂ M1[wt]
♀ M1[wt]_M2
♀ M1[Δ41-85]_M2
♀ M1[Δ1-15]_M2
♂ WT
♂ M1[wt]
♀ M1[wt]_M2
♀ M1[Δ41-85]_M2
♀ M1[Δ1-15]_M2
♀ WT
−14000 −12000 −10000 −8000 −6000 −4000 −2000 0 2000 4000 6000 8000 10000 12000 14000
Distance from HAS
B
0 1.5 3
log2(Signal/PI)
♀ M1[Δ1-15]_M2
♀ M1[Δ41-85]_M2
♀ M1[wt]_M2
♀ WT
♂ WT
♂ M1[wt]
No data
MSL1 MSL2 MSL3
C
♀
MSL2
M1[wt]_M2
M1[Δ41-85]_M2
M1[Δ1-15]_M2
IgG
MSL2 Signal
20 30 40 50 60 70 80
−3000 −1500 0 1500 3000
Distance from HAS
D
♀
MSL2
M1[wt]_M2
M1[Δ41-85]_M2
M1[Δ1-15]_M2
IgG
MSL2 Signal
20 30 40 50 60 70 80
−3000 −1500 0 1500 3000
Distance from autosome peak summit
E
MSL2 ♀ X chromosome
M1[wt]_M2
M1[Δ41-85]_M2
M1[Δ1-15]_M2
IgG
RPKM
1200 0
400 0
600 0
roX1
X:3,857,230..3,862,697
roX2
X:11,579,065..11,580,432
Pp2C1
X:4,678,988..4,687,903
F
MSL2 ♀ Autosomes
M1[wt]_M2
M1[Δ41-85]_M2
M1[Δ1-15]_M2
IgG
RPKM
200 0
150 0
120 0
ATPsynC
3R:31,215,145..31,220,630
Men-b
3R:27,150,662..27,158,603
lncRNA:CR43650
2R:16,828,095..16,840,494