

Factor cooperation for chromosome discrimination in *Drosophila*

Christian Albig^{1,2}, Evgeniya Tikhonova³, Silke Krause¹, Oksana Maksimenko³, Catherine Regnard^{1,*} and Peter B. Becker^{1,*}

¹Molecular Biology Division, Biomedical Center, Faculty of Medicine and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität Munich, 82151 Martinsried, Germany, ²Graduate School for Quantitative Biosciences (QBM), Ludwig-Maximilians-Universität Munich, 81377 Munich, Germany and ³Group of Molecular Organization of Genome, Department of the Control of Genetic Processes, Institute of Gene Biology, Russian Academy of Sciences, 119334 Moscow, Russia

Received September 28, 2018; Revised November 05, 2018; Editorial Decision November 26, 2018; Accepted November 29, 2018

ABSTRACT

Transcription regulators select their genomic binding sites from a large pool of similar, non-functional sequences. Although general principles that allow such discrimination are known, the complexity of DNA elements often precludes a prediction of functional sites. The process of dosage compensation in *Drosophila* allows exploring the rules underlying binding site selectivity. The male-specific-lethal (MSL) Dosage Compensation Complex (DCC) selectively binds to some 300 X chromosomal ‘High Affinity Sites’ (HAS) containing GA-rich ‘MSL recognition elements’ (MREs), but disregards thousands of other MRE sequences in the genome. The DNA-binding subunit MSL2 alone identifies a subset of MREs, but fails to recognize most MREs within HAS. The ‘Chromatin-linked adaptor for MSL proteins’ (CLAMP) also interacts with many MREs genome-wide and promotes DCC binding to HAS. Using genome-wide DNA-immunoprecipitation we describe extensive cooperativity between both factors, depending on the nature of the binding sites. These are explained by physical interaction between MSL2 and CLAMP. *In vivo*, both factors cooperate to compete with nucleosome formation at HAS. The male-specific MSL2 thus synergises with a ubiquitous GA-repeat binding protein for refined X/autosome discrimination.

INTRODUCTION

The rules according to which transcription factors select only a small subset of genomic binding sites from a large excess of similar sequences remain largely elusive. Typically,

the binding sites for transcription factors consist of short sequence motifs, yet only a few percent of all genomic sites that conform to a consensus motif are functional and bound *in vivo*. This selectivity may be explained by cooperativity between factors binding complex DNA elements (1,2), the instructive role of DNA conformation (3–6) and sequence context as well as the role of chromatin organization, which may occlude non-functional sites (7,8).

A striking example is the process of sex chromosome dosage compensation in *Drosophila melanogaster*, which doubles the transcription output of most genes on the single X chromosome in males to match the two active X chromosomes in females. This sophisticated regulation is brought about by the male-specific-lethal (MSL) dosage compensation complex (MSL-DCC, or just DCC). The DCC consists of five MSL protein subunits (MSL1, MSL2, MSL3, MOF [males-absent-on-the-first] and MLE [maleless]) and two long, non-coding *roX* RNAs (RNA-on-the-X) [reviewed in (9–11)].

Dosage compensation is genetically encoded on the X chromosome in the form of ~300 ‘High Affinity Sites’ (HAS) for the DCC, which are also referred to as ‘chromosomal entry sites’ (CES). The current model poses that the DCC first interacts with HAS on the X chromosome and then transfers to active genes in its vicinity [(12) and reviewed in (11,13)]. These genes are epigenetically marked by methylation of histone H3 at lysine 36 (H3K36me3), a mark that is placed co-transcriptionally. The DCC subunit MSL3 contains a chromo-barrel domain that serves as a ‘reader head’ to scan the chromatin for the active methylation mark (14,15). Upon binding, the associated ‘writer’ subunit MOF acetylates histone H4 at lysine 16 (H4K16) (16–18), which somehow boosts the production of functional mRNA through unfolding of the chromatin fiber (19). Any gene integrated on the X chromosome is subject to this regulation. Understanding dosage com-

*To whom correspondence should be addressed. Tel: +49 89 2180 75427; Fax: +49 89 2180 75425; Email: pbecker@bmc.med.lmu.de
Correspondence may also be addressed to Catherine Regnard. Tel: +49 89 2180 71602; Fax: +49 89 2180 75425; Email: cregnard@bmc.med.lmu.de

pensation, therefore, requires understanding the nature of X-specific DCC binding.

The HAS harbor a low-complexity, GA-rich consensus motif, referred to as 'MSL recognition element' (MRE) (20,21), which is indispensable for DCC binding. However, the genome contains several thousand MREs on the X chromosome outside of HAS and on autosomes, therefore only ~2% of MREs are functional and bound by the DCC (20,21). The direct MSL2 binding sites have been experimentally determined by *in vitro* genome-wide DNA immunoprecipitation assays (22). MSL2 binds to DNA via a C-terminal CXC domain followed by a region rich in prolines (23,24). Remarkably, the CXC domain recognizes a subset of MREs whose consensus motif has a notable 5' extension characterized by a particular DNA shape (22). These CXC-dependent sites are named 'Pioneering-sites-on-the-X' (PionX), as they (i) are the first to be bound upon *de novo* induction of dosage compensation in females, (ii) are preferentially contacted by an MSL2-MSL1 sub-complex and (iii) are enriched on the evolutionary young neo-X chromosome of *Drosophila miranda* (22,25). The PionX motif is superior over the MRE motif in predicting which genomic sites function as HAS. The PionX motif is up to ~10-fold enriched on the X chromosome, providing a first clue about how MSL2 distinguishes the X chromosome from autosomes (22). In general, however, the interaction of MSL2 with PionX sites does not fully explain HAS targeting, since only a small fraction of the MSL2 *in vitro* binding sites (mostly containing a PionX signature) overlap with functional HAS *in vivo*.

A solution to the problem was suggested by Larschan *et al.*, who found a zinc finger protein that associates with about half of MREs throughout the genome. They termed this protein CLAMP (Chromatin Linked Adaptor for MSL Proteins) since depletion of the protein leads to dissociation of the DCC from the X chromosome, as comprehensively shown at the resolution of polytene chromosomes (26,27). CLAMP is an essential protein in *Drosophila* independent of sex, which binds thousands of GA-rich sequences genome-wide (27–29) and therefore does not qualify as a determinant of X-specificity. Remarkably, CLAMP binds to HAS only in male cells, suggesting a functional relationship with the DCC (27).

It is possible that CLAMP facilitates MSL2 binding to MREs by keeping these elements nucleosome-free, in analogy to early observations that the GAGA factor (GAF) keeps promoters and polycomb response elements clear of nucleosomes to allow other regulators to bind (30–33). Indeed, Urban *et al.* recently found that CLAMP promotes the accessibility of DNA in chromatin over long distances surrounding its binding sites (34). In this study, the authors probed chromatin accessibility by Micrococcal Nuclease (MNase) digestion in a titration series. In addition, the authors suggested that CLAMP leads to a global decompaction of the X chromosome in males.

To explore the relationship between CLAMP and MSL2 we integrated data from several approaches. We monitored how the two factors influenced each other's binding to genomic sequences *in vitro* by DNA immunoprecipitation (22,35,36). We observed mutual recruitment, explained by direct interaction between both proteins and

shared affinity for long GA-repeat sequences. This DNA binding cooperativity improved reliable selection of functional MREs which are located within HAS, however at the expense of binding to additional, non-functional sites. To explore whether the chromatin organization of the genome plays a role, we monitored DNA accessibility genome-wide in S2 and Kc cells by ATAC-seq (Assay for Transposase Accessibly Chromatin with high-throughput sequencing) (37,38) and observed how the pattern of nucleosome-free chromatin changed upon RNA interference with CLAMP or MSL2 expression. We integrated these data with direct measurements of the *in vivo* interactions of both proteins by ChIP-seq (Chromatin Immunoprecipitation with high-throughput sequencing) (39). The data do not support the hypothesis of a hierarchical relationship between the two proteins. Rather, both factors synergize to keep common binding sites nucleosome-free and stabilize each other's binding. We conclude that the correct targeting of X chromosomal HAS involves synergistic action of the male-specific MSL2 and the general GA-repeat binding CLAMP to compete with nucleosome formation.

MATERIALS AND METHODS

Cell culture and RNAi

S2-DRSC (DGRC stock # 181), Kc167 (DGRC stock # 1) and L2-4 (S2 subclone, provided by P. Heun) cells were cultured in Schneider's *Drosophila* Medium (Thermo Fisher), supplemented with 10% heat-inactivated fetal bovine serum (Sigma-Aldrich), 100 units/ml penicillin and 0.1 mg/ml streptomycin (Sigma-Aldrich) at 26°C. RNAi against target genes in S2, L2-4 and Kc cells was performed as previously described (22). In brief, double-stranded RNA fragments (dsRNA) were generated with MEGAscript kit (Thermo Fisher) from polymerase chain reaction (PCR) products obtained using the following forward and reverse primers (separated by comma):

- *clamp* RNAi #1: TAATACGACTCACTATAGGGAC GTCCAAACCCTTCAGTTGT, TAATACGACTCACTATAGGGATTGAGTGCAAACGATCAGC;
- *clamp* RNAi #2 (DRSC29935): TTAATACGACTCAC TATAGGGAGAGAAGACCTTACCAAAAACAT, TTAATACGACTCACTATAGGGAGAGCTTATGT TGGATATGGTGT;
- *trl* RNAi: TTAATACGACTCACTATAGGGAGAAT GTCGCTGCCA, TTAATACGACTCACTATAGG GAGATTGCCTGGA;
- *gst* RNAi: TTAATACGACTCACTATAGGGAGAAT GTCCCTATACTAGGTTA, TTAATACGACTCAC TATAGGGAGAACGCATCCAGGCACATTG; The sequence of *Schistosoma japonicum* GST was amplified from pGEX-6P-1 (GE Healthcare).
- *gfp* RNAi: TTAATACGACTCACTATAGGGTGCTC AGGTAGTGGTTGTCG, TTAATACGACTCACTA TAGGGCCTGAAGTTCATCTGCACCA;
- *msl2* RNAi: TTAATACGACTCACTATAGGGAGA ATGGCCCAGACGGCATA, TTAATACGACTC ACTATAGGGAGACAGCGATGTGGGCATGTC.

Cells were washed with serum-free medium and 10 μg dsRNA per 10⁶ cells at a concentration of 10 $\mu\text{g}/\text{ml}$ in serum-free medium (10⁶ cells in 6-well plate for ATAC-seq and immunofluorescence microscopy) or 4.2 μg dsRNA per 10⁶ cells at 6.3 $\mu\text{g}/\text{ml}$ serum free medium (12 * 10⁶ cells in 75 cm² flask for ChIP-seq) was added, incubated for 10 min at room temperature (RT) with slight agitation and further 50 min at 26°C. Two volumes of complete growth medium were added and cells were incubated for 5 days at 26°C. Then, 1 \times of initial volume (6-well plate) or 0.75 \times of initial volume (75 cm² flask) of growth medium was added and cells were incubated for further 2 days at 26°C.

Sf21 cells (Thermo Fischer) were cultured in SF900 II SFM (Thermo Fisher), supplemented with 10% heat-inactivated fetal bovine serum (Sigma-Aldrich), 0.1 mg/ml gentamicin (Sigma-Aldrich) at 26°C.

Immunofluorescence microscopy

For immunofluorescence microscopy (IFM), 0.5 * 10⁶ L2-4 cells after RNAi treatment in 200–500 μl growth medium were seeded into each well of a 3-well (14 mm) object slide (Thermo Fisher) and incubated for 2 h at 26°C. Immunofluorescence staining was performed as described in (40) with slight modifications. Briefly, cells were washed in phosphate-buffered saline (PBS), fixed for 7.5 min with 2% (v/v) formaldehyde (FA) in PBS on ice and permeabilized for 7.5 min with 1% (v/v) FA in PBS with 0.25% (v/v) Triton-X 100 on ice. Cells were washed twice in PBS and blocked for 1 h with 200 μl Blocking Buffer (3% (w/v) bovine serum albumin (BSA), 5% (v/v) heat-inactivated fetal bovine serum (Sigma-Aldrich) in PBS) at RT in a humid chamber. Cells were incubated for 1 h with primary antibody diluted in 30 μl Blocking Buffer with 0.1% (v/v) Triton-X 100 at RT in a humid chamber. Cells were washed twice for 5 min in 200 μl PBS and incubated for 1 h with suitable secondary antibody diluted in 30 μl Blocking Buffer with 0.1% (v/v) Triton-X 100 at RT in a humid chamber. Cells were washed twice for 5 min in 200 μl PBS, stained for 2 min with 0.125 $\mu\text{g}/\text{ml}$ DAPI in 200 μl PBS at RT and washed twice for 5 min in 200 μl PBS. Cells were mounted with 8 μl Vectashield (Vector Laboratories) and a coverslip was sealed to object slide with nail polish. Images were acquired with Axiovert 200 epifluorescence microscope (Zeiss) equipped with AxioCamMR CCD Camera (Zeiss).

Protein purification

Recombinant MSL2-FLAG was expressed in Sf21 cells and purified by FLAG-tag affinity chromatography, as previously described in (23) with minor modification. In brief, Sf21 cells at 10⁶ cells/ml (250 * 10⁶ cells) were infected 1:1000 (v/v) with baculovirus, expressing MSL2-FLAG. After 72 h, cells were harvested and washed once in PBS, frozen in liquid nitrogen and stored at –80°C. For lysis, cells were rapidly thawed, resuspended in 1 ml Lysis Buffer per 10 ml of culture (50 mM HEPES pH 7.6, 300 mM KCl, 1 mM MgCl₂, 5% (v/v) glycerol, 0.05% (v/v) IGEPAL CA-360, 50 μM ZnCl₂) supplemented with 0.5 mM TCEP and cOmplete ethylenediaminetetraacetic acid (EDTA)-free Protease

Inhibitor Cocktail (Sigma-Aldrich) (PI). The suspension was passed thrice through Microfluidizer LM10 (Microfluidics). Cell extract was adjusted with Lysis Buffer containing 0.5 mM TCEP and PI to 2 ml per 10 ml of culture and incubated with end-over-end rotation for 1 h at 4°C. Cell debris were spun down at 4°C for 30 min at 50 000 *g*. The resulting supernatant was used for FLAG-tag affinity purification with 4 μl of a 50% slurry of FLAG-M2 beads (Sigma-Aldrich) per 1 ml of culture. Beads were first washed thrice in 20 bed volumes of Lysis Buffer, subsequently supernatant was added and incubated with end-over-end rotation for 3 h at 4°C. Beads were pelleted (4°C, 5 min, 500 *g*) and supernatant was removed. Beads were washed once with 20 bed volumes each of Lysis Buffer, Wash Buffer (Lysis Buffer with 1000 mM KCl), Lysis Buffer again, and finally twice with 20 bed volumes Elution Buffer (Lysis Buffer with 100 mM KCl). For protein elution, beads were incubated with 0.2 bed volumes of Elution Buffer containing 5 mg/ml FLAG peptide for 10 min at 4°C and subsequently 1.8 bed volumes of Elution Buffer with 0.5 mM TCEP and PI were added and incubated with end-over-end rotation for 30 min at 4°C. The elution step was repeated once. Elution fractions were combined, remaining beads were removed by passing through Corning Costar Spin-X centrifuge tube filters (Sigma-Aldrich) and the eluate was concentrated with 10 MWCO Amicon Ultra 0.5 ml (Merck). Protein concentration was determined using BSA standards on sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie brilliant blue G250 staining. To store at –80°C until use, concentrated protein was aliquoted and flash-frozen in liquid nitrogen.

For purification of recombinant CLAMP by FLAG-tag affinity chromatography, the coding sequence of CLAMP (CG1832-RA) was fused to a C-terminal coding sequence of FLAG affinity tag and cloned into pFast-Bac1, using the following forward and reverse primer: AAGGATCCATGGAAGACCTTACCAAAAAC, CCTTTCTCGAGTTACTTGTTCATCGTCTGTCCTTGT AGTCTTCCCCGTCTGTATGCATCCG.

CLAMP-FLAG was expressed in Sf21 cells and purified by FLAG-tag affinity chromatography, like MSL2-FLAG, but with the following modifications. Lysed cells were resuspended in 1 ml Buffer C per 10 ml of culture (50 mM HEPES pH 7.6, 1 M KCl, 1 mM MgCl₂, 5% (v/v) Glycerol, 0.05% (v/v) IGEPAL CA-360, 50 μM ZnCl₂, 375 mM L-Arginine (according to (41)) supplemented with 0.5 mM TCEP and cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) (PI) and passed thrice through Microfluidizer LM10 (Microfluidics). Subsequently, the extract was adjusted with Buffer C containing PI to 2 ml per 10 ml of culture and supplemented with 0.1% (v/v) polyethylenimine by adding 2% (v/v) polyethylenimine (neutralized with HCl to pH 7.0) drop-by-drop while string in an ice bath, according to (42). Cell debris were spun down at 4°C for 30 min at 50 000 *g*. To the resulting supernatant 20 U Benzonase was added per 10 ml of culture and incubated with end-over-end rotation for 1 h at 4°C. The resulting extract was used for FLAG-tag affinity purification with 4 μl of a 50% slurry of FLAG-M2 beads (Sigma-Aldrich) per 1 ml of culture. Beads were washed thrice in 20 bed volumes of Buffer C, subsequently supernatant was

added and incubated with end-over-end rotation for 3 h at 4°C. Beads were pelleted at 4°C for 5 min at 500 g and supernatant was removed. Beads were washed five times with 20 bed volumes of Buffer C. For protein elution, beads were incubated with 0.2 bed volumes of Buffer C containing 5 mg/ml FLAG peptide for 10 min at 4°C and subsequently 1.8 bed volumes of Buffer C containing 0.5 mM TCEP and PI were added and incubated with end-over-end rotation for 30 min at 4°C. The elution step was repeated once. The combined elution fractions were processed on flash-frozen in liquid nitrogen as for MSL2-FLAG.

Genomic DNA preparation

For genomic DNA (gDNA) extraction, 5×10^7 S2 cells were harvested, washed in PBS and gDNA was extracted with NucleoSpin Tissue kit (Macherey-Nagel). Remaining RNA contaminants in eluted gDNA were digested with 0.1 mg/ml RNaseA (Sigma-Aldrich) for 1 h at 37°C. Subsequently, gDNA was sonicated with Covaris AFA S220 using microTUBEs at 175 W peak incident power, 10% duty factor and 200 cycles per burst for 430 s at 5°C to generate ~150–200 bp fragments, as described in (22). Sheared gDNA was purified with MinElute kit (QIAGEN), concentration was determined using Qubit (Thermo Fisher) and fragment size was verified using 2100 Bioanalyzer (Agilent).

Antibodies

For immunoblotting, affinity-purified polyclonal α -H3 C-term antibody (Abcam, ab1791), affinity-purified polyclonal α -CLAMP antibody (Novus, 49880002), polyclonal rabbit α -MSL2^{Gilfillan} serum [α -MSL2^{Gilfillan}, (43)], polyclonal rabbit α -MSL2 serum [generated by Pineda Antikörper-Service here termed ' α -MSL2^{Pineda}' against the MSL2-fragment (amino acids 296–608) similar to the one used for α -MSL2^{Gilfillan}, polyclonal rabbit α -GAF serum [α -GAF' (31)] and affinity-purified monoclonal α -FLAG M2 antibody (Sigma-Aldrich, F3165) were used. For immunostaining, α -MSL2^{Gilfillan} and culture supernatant containing monoclonal α -MSL3 (21) were used. For ChIP, α -MSL2^{Gilfillan}, α -GAF and α -CLAMP antibodies were used. For DIP, supernatant containing monoclonal α -MSL2 [generated by E. Kremmer (Helmholtz Zentrum Munich, Germany) against the same MSL2-fragment as used for α -MSL2^{Gilfillan}] and supernatant containing monoclonal α -CLAMP [generated by E. Kremmer (Helmholtz Zentrum Munich, Germany) against the peptide LATTTDDNKT-CYI] were used. For IP, serum containing polyclonal α -MSL2^{Pineda} and affinity-purified polyclonal α -CLAMP* antibody described in (44) (kind gift of E. Larschan) were used.

Yeast two-hybrid assay

Yeast two-hybrid assay was carried out using yeast strain pJ69-4A (MATa trp1-901 leu2-3,112 ura3-52 his3-200 gal4 Δ gal80 Δ GAL2-ADE2 LYS2::GAL1-HIS3 met2::GAL7-lacZ), 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 with some modifications. All cells were grown at 30°C in an orbital shaker at 250–300 rpm. Yeast colonies were transferred into a 15 ml culture tube with 6–7 ml of YPDA medium and grown for 1 day. The culture was 10-fold diluted in a 0.5 l culture flask with YPDA medium and cultivated for 3 h. Aliquots of 1.5 ml cell suspension were pelleted by centrifugation at 4000 g for 10 s, and the supernatant was removed. Pelleted cells were resuspended in 1 ml 0.1 M LiAcO, incubated for 30 min and pelleted as above. To the pellet, 240 μ l 50% (v/v) PEG 3380, 36 μ l 1 M LiAcO and 50 μ l of a mixture of two plasmids (the amount of each plasmid in a mixture of 400–800 ng) was added sequentially and the cells suspended to homogeneity. The tube was incubated at 30°C for 30 min, then at 42°C for 5 min and then placed on ice for 1–2 min. Cells were pelleted at 4000 g for 15–20 s and resuspended in 100 μ l ddH₂O. The resuspended cells were plated on selective medium lacking Leu and Trp ('medium-2'). The plates were incubated at 30°C for 2–3 days. Afterward, the colonies were streaked out on plates on selective medium lacking either Leu, Trp and His ('medium-3'), or lacking adenine in addition to the three amino acids ('medium-4'), or lacking the three amino acids but containing 5 mM 3-amino-1,2,4-triazole ('medium-3 + 5mM 3AT'). The plates were incubated at 30°C for 3–4 days and growth assessed. Each assay was prepared as three independent biological replicates with three technical repeats. Fusion proteins were cloned into pGBT9 and pGAD424 vectors from Clontech (Supplementary Table S1) and verified by sequencing.

CLAMP-MSL2 co-immunoprecipitation

Per reaction, 100 nM recombinant MSL2-FLAG and CLAMP-FLAG in 50 μ l Binding Buffer (2 mM Tris/HCl pH 7.5, 100 mM KCl, 2 mM MgCl₂, 10% (v/v) Glycerol, 10 μ M ZnCl₂) were incubated with end-over-end rotation for 30 min at 26°C. For α -MSL2 immunoprecipitation, the reaction was added to 25 μ l magnetic Dynabeads Protein G (Thermo Fisher) pre-coupled with antibodies. For α -CLAMP immunoprecipitation, the reaction was added to 5 μ l beads pre-coupled with antibodies. For pre-coupling, beads were washed thrice with 1 ml Binding Buffer and incubated with end-over-end rotation for over-night at 4°C with 1 μ l α -MSL2 antibody per 10 μ l beads or with the corresponding pre-immune serum as control and with 20 μ l α -CLAMP* antibody complemented with 0.75 μ l of an irrelevant pre-immune serum per 10 μ l beads or with 1 μ l irrelevant pre-immune serum as control. Beads were washed thrice with 1 ml Binding Buffer, the co-IP reaction was added and incubated with end-over-end rotation for 15 min at RT. Beads were washed thrice in 100 μ l Binding Buffer, resuspended in SDS-Sample Buffer and analyzed together with the corresponding input and unbound sample by SDS-PAGE with Coomassie brilliant blue G250 staining, quantified using a ChemiDoc Touch Imaging System (Bio-Rad). Co-IP reactions were performed with three independent CLAMP and MSL2 preparations.

Mapping of interaction domain

Sf21 cells at 10^6 cells/ml (20×10^6 cells) were infected 1:1000 (v/v) with each baculovirus stock separately. While MSL2^{ΔCXC}-FLAG and MSL2-FLAG were described previously in (23), new constructs have been generated for CLAMP-FLAG (see above) and deletion mutants of MSL2 and CLAMP, cloned into pFastBac1 using the corresponding primer (Supplementary Table S2). After 72 h, cells were harvested and lysed in 1 ml Lysis Buffer (50 mM HEPES pH 7.6, 300 mM KCl, 1 mM MgCl₂, 5% (v/v) glycerol, 0.05% (v/v) IGEPAL CA-360, 50 μM ZnCl₂) with cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) (PI) per 20 ml of culture. The relative amounts of expressed protein were quantified by α-FLAG western blot. Per co-IP reaction, MSL2-FLAG and CLAMP-FLAG (or their mutants) containing extracts were mixed in 1:1 to 1:3 ratio and if required adjusted with cell extract from uninfected cells. Cell extracts were supplemented with 12.5 U Benzonase, 1 mM DTT, 0.2 mM MgCl₂ and PI, incubated with end-over-end rotation for 10 min at RT and cell debris were spun down at 4°C for 10 min at 20 000 g. For α-MSL2 immunoprecipitation, the reaction was added to 15 μl magnetic Dynabeads Protein G (Thermo Fisher) pre-coupled with antibodies. For pre-coupling, beads were washed thrice with 1 ml Lysis Buffer and incubated with end-over-end rotation for over-night at 4°C with 1 μl α-MSL2 antibody per 10 μl beads or with the corresponding pre-immune serum as control. Beads were washed thrice with 1 ml Lysis Buffer, the extract mixture was added and incubated with end-over-end rotation for 45 min at 4°C. Beads were washed thrice in 500 μl Binding Buffer, resuspended in SDS-Sample Buffer and analyzed together with the corresponding input sample by SDS-PAGE and α-FLAG western blot.

ATAC-seq

ATAC-seq was performed as described earlier (37,38) with some modification for *Drosophila* cells. In brief, 50 000 S2 or Kc cells, also with prior RNAi, were used per reaction. Cells were washed in 100 μl PBS, resuspended in 100 μl Lysis Buffer (10 mM Tris/HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% (v/v) IGEPAL CA-630) with cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) and incubated for 3 min on ice. Nuclei were spun down at 4°C for 10 min at 600 g, resuspended in 50 μl 1× TD buffer with 2.5 μl TD Enzyme (Illumina) and incubated with slight agitation for 30 min at 37°C. Tagmented DNA was purified with MinElute kit (QIAGEN) and eluted in 15 μl H₂O. To determine the cycle number for library amplification, quantitative PCR was performed in triplicates with 0.5 μl sample in 10 μl reaction containing 1× NEB-Next HiFi PCR mix (NEB), 1.25 μM Ad1 and Ad2 primer each and 0.5x SYBRGreen (Thermo Fisher). Cycle number at quarter maximal intensity was used for library amplification (typically 11–13 cycles). Libraries were amplified by using 12.5 μl sample in 50 μl PCR reaction containing 1× NEBNext HiFi PCR mix (NEB), 1.25 μM Ad1 and Ad2 primer each. ATAC libraries were purified with MinElute kit (QIAGEN) and concentration determined using 2100 Bioanalyzer with High Sensitivity DNA kit (Agilent). Libraries were sequenced on HiSeq 1500 (Illumina)

instrument yielding typically 25–35 million 50 bp paired-end reads per sample.

ChIP-seq

ChIP-seq was performed as previously described (39) with slight modification. S2 cells ($\sim 10^8$ cells) after RNAi were harvested and chilled on ice. Cells were cross-linked with 1% formaldehyde for 55 min on ice by adding 1 ml volume 10× fixing solution (50 mM HEPES pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) with 10% formaldehyde per 10 ml culture and reaction was stopped by adding 125 mM glycine and incubating for 10 min on ice. For nuclei isolation, cells were washed in PBS and resuspended in Buffer A (10 mM Tris/HCl pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25% (v/v) Triton-X 100) with cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) (PI) at 10^8 cells/ml and incubated with end-over-end rotation for 10 min at 4°C. The cells were pelleted at 4°C for 10 min at 2300 g and resuspended in Buffer B (10 mM Tris/HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% (v/v) Triton-X 100) with PI at 10^8 cells/ml and incubated with end-over-end rotation for 10 min at 4°C. For chromatin fragmentation, nuclei were spun down at 4°C for 10 min at 2300 g, resuspended in RIPA (10 mM Tris/HCl pH 8.0, 140 mM NaCl, 1 mM EDTA, 1% (v/v) Triton-X 100, 0.1% (v/v) SDS, 0.1% (v/v) DOC) with PI at 10^8 cells/ml in 1 ml for shearing with Covaris AFA S220 using 12×12 tubes at 100 W peak incident power, 20% duty factor and 200 cycles per burst for 25 min at 5°C to generate 150–200 bp fragments. Protein A and Protein G (GE Healthcare) beads (mixed in a 1:1 ratio) were washed thrice with 30 bed volumes RIPA. To remove cell debris, sheared chromatin was centrifuged at 4°C for 15 min at 15 000 g and 100 μl soluble chromatin in the supernatant was pre-cleared with 3 μl (6 μl 50% slurry) Protein A and Protein G beads mix by incubating with end-over-end rotation for 1 h at 4°C. Beads were pelleted at 4°C for 5 min at 500 g and supernatant was directly used for immunoprecipitation by adding antibody to 200 μl chromatin adjusted to 500 μl with RIPA including PI and incubating with end-over-end rotation for 16 h at 4°C. To remove precipitates, chromatin was centrifuged at 4°C for 15 min at 15 000 g and 100 μl supernatant was added to 3 μl Protein A and Protein G beads mix (RIPA-equilibrated as above) by incubating with end-over-end rotation for 4 h at 4°C. Beads were spun down and washed five times with 60 bed volumes RIPA including PI by incubating with end-over-end rotation for 10 min at 4°C. For DNA recovery, beads were spun down, resuspended in 6.7 bed volumes TE Buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA), RNA was digested with 50 μg/ml RNaseA (Sigma-Aldrich) for 30 min at 37°C, and after addition of 0.5% (m/v) SDS, proteins were digested with 0.5 μg/ml Proteinase K (Sigma-Aldrich) for 16 h at 65°C with agitation. DNA was purified with 1.8× AMPure XP beads (Beckmann Coulter). Quantitative PCR was performed using SYBR Green PCR Master Mix (Thermo Fisher) and 0.5 nM forward and reverse primer each (Supplementary Table S3), analyzed on Light-Cycler 480 (Roche). Libraries were prepared with NEBNext ChIP-Seq Library Prep Kit for Illumina (NEB) and analyzed with 2100 Bioanalyzer with DNA 1000 kit (Agilent).

Libraries were sequenced on HiSeq 1500 (Illumina) instrument yielding typically 25–30 million 50 bp single-end reads per sample.

DIP-seq

DIP-seq was performed accordingly to (22), with modifications for combined DIP of two proteins. Sheared gDNA was diluted to 4 mg/ml in Binding Buffer (2 mM Tris/HCl pH 7.5, 100 mM KCl, 2 mM MgCl₂, 10% (v/v) Glycerol, 10 μM ZnCl₂). Ten microliter (corresponding to 10%) was taken as input material and adjusted to 50 μl with Binding Buffer for DNA purification. Per reaction, 100 nM recombinant MSL2 and/or 50 nM recombinant CLAMP was added in 100 μl diluted gDNA and incubated with end-over-end rotation for 30 min at 26°C. For immunoprecipitation, the reaction was added to 7.5 μl (15 μl 50% slurry) Protein G beads (GE Healthcare) pre-coupled with antibodies. For pre-coupling, beads were washed thrice with 100 μl Binding Buffer and incubated with end-over-end rotation for 3–4 h at 4°C with 1 ml culture supernatant containing monoclonal antibody or culture medium as control. Beads were spun down for 1 min at 500 g, washed thrice with 100 μl Binding Buffer, the DIP reaction was added and incubated with end-over-end rotation for 15 min at RT. Beads were spun down for 1 min at 500 g, washed twice in 100 μl Binding Buffer and resuspended in 50 μl Binding Buffer. Five microliter were taken for Western blot analysis. Proteins were digested by adding Proteinase K at 0.5 mg/ml and incubated for 1 h at 56°C with agitation and DNA was purified with 1.8× AMPure XP beads (Beckmann Coulter). Libraries were prepared with MicroPlex Library Preparation Kit v2 (Diagenode) and analyzed with 2100 Bioanalyzer with DNA 1000 kit (Agilent). Libraries were sequenced on HiSeq 1500 (Illumina) instrument yielding typically 20–40 million 50 bp single-end reads per sample. DIP reactions were performed with three independent CLAMP and MSL2 preparations.

Data analysis

Sequencing data were processed using SAMtools (45) version 1.3.1, BEDtools (46) version 2.26.0, R version 3.4.2 (<http://www.r-project.org>) and Bioconductor (<http://www.bioconductor.org>) using default parameters for function calls, unless stated otherwise.

Read processing

Sequence reads were aligned to the *D. melanogaster* release 6 reference genome (BDGP6) using Bowtie (47) version 1.1.2 (parameter `-m 1`) for ChIP-seq and DIP-seq samples and Bowtie2 (48) version 2.2.9 (parameters `-very-sensitive`, `-no-discordant`, `-no-mixed`, `-X 100`) for ATAC-seq samples considering read-pairs as non-nucleosome if ≤ 100 bp according to (37).

Peak calling, robust peak sets and HAS definition

Peaks were called using Homer (49) version 4.9.1 calling function `findPeaks` for DIP- and ChIP-seq samples using

the corresponding input sample as control (parameters `-style factor`, `-size 200`, `-fragLength 200`, `-inputFragLength 200` and `-C 0`) and for ATAC-seq samples without control (parameters `-style dnase`, `-C 0`, `-gsize 137e6`, `-minDist 50`). Peaks were defined as robust if the region was called in at least two replicate samples. HAS regions were used as defined by (22) with 309 HAS in total, of which 304 are located on the X chromosome and 5 on autosomes.

X chromosomal enrichment

The X chromosomal enrichment was defined as the ratio of X chromosomal peak density over autosomal peak density and peak density was calculated as the number of peaks divided by chromosome length as defined by (22).

De novo motif discovery and MRE definition

Enriched motifs in peak region were discovered using MEME (50) version 4.11.4 (parameters `-dna`, `-mod zoops` and `-revcomp`). For discovering the PionX motif, the list of PionX sites in (22) was used. As described in reference (22), we refer to MRE as the motif discovered within HAS, which is highly similar to the originally defined motif (20,21).

Motif search

Motif search in peak region was performed with FIMO (51) version 4.11.4 (parameters `-qv-thresh`, `-thresh 0.2`, `-max-stored-scores 1e6`) applying a fifth-order background model.

Browser profiles

Browser profiles were generated using tsTools (R) (<https://rdrr.io/github/musikutiv/tsTools/>) using mean per base read count per million mapped reads (rpm) of biologicals replicates, after aligned reads were extended to 200 bp fragments.

Assignment of peak regions to chromatin states

Peak regions within one of the chromatin states were assigned to the corresponding chromatin state defined by (52) and peaks overlapping with multiple chromatin states were assigned as 'none'.

ChIP-seq analysis

For calculating ChIP-seq signal enrichment as log₂ ratio of IP over input, aligned reads were extended to 200 bp fragments and reads overlapping with target regions requiring a minimal overlap of half read length were counted and normalized to million mapped reads (rpm). For generating heat maps by calling the function `pheatmap` (R) and average profiles in a 2 kb window around HAS, coverage vectors were calculated as mean per base read count of IP samples and normalized to rpm. To test for statistical significant differences between *gst* RNAi and *trl* RNAi condition in CLAMP ChIP-seq, signal enrichment for the merged set of robust peak sets of both samples was calculated. Testing for difference was performed using `limma` (53) (R) including

batch variables as random effect and calling the functions `lmFit`, `eBayes` (parameters `trend = T`, `robust = T`) and `topTable` (parameter `adjust.method = 'fdr'`). Regions were defined as statistical significant different between conditions with false discovery rate (FDR) < 0.05.

DIP-seq analysis

As proxy for MSL2-FLAG *in vitro* genomic binding we used the previously published DIP-seq experiment by (22) (GSE75033), where MSL2 was immunoprecipitated using α -FLAG M2 beads (Sigma-Aldrich) directed against the FLAG-tag, here referring to as 'MSL2 (α -FLAG) – Villa'. For calculating DIP-seq signal enrichment as \log_2 ratio of IP over input, aligned reads were extended to 200 bp fragments and reads overlapping with target regions requiring a minimal overlap of half read length were counted and normalized to rpm. For generating heat map by calling the function `phatmap` (R), euclidean distance between the enrichment for all biological replicates at combined robust peak set was measured by calling the function `dist` (R) and samples were hierarchical clustered using the 'complete' method by calling the function `hclust` (R). The row dendrogram was divided into 18 clusters by calling the function `cutree` (R) (parameter `k = 18`). For predicting the DNA roll between the base pairs at position +1 and +2 of the PionX motif, the best matching motif found by FIMO within the peak region was extended by 2 bp at each end. The roll was predicted using `DNashapeR` (54) (R) by calling the function `getShape`. The DNA roll between the first two base pairs was assigned to the first base pair for simplicity, therefore referring to as 'roll at position +1'. The number of GA:TC dinucleotides in the peak region was counted using `rDNAse` (R) by calling the function `kmer` considering also the reverse complement. We calculated the GA:TC dinucleotide density as the number of GA:TC-dinucleotides divided by the peak length. The length of GA:TC repeats was counted using `Biostrings` (R) by calling the function `vcountPattern` considering also the reverse complement.

ATAC-seq analysis

For calculating ATAC-seq signal intensities, fragments on the + strand were moved by +4 bp and on the – strand by –5 bp according to (37). Signal intensities in target regions were calculated by calling `summarizeOverlaps` (parameter `ignore.strand = TRUE`) from `GenomicAlignments` (55) (R). Further analysis was performed using `DESeq2` (56) (R), including batch variables as random effect. Sites were considered as statistical significant different between conditions with absolute \log_2 fold-change > 0.5 and FDR < 0.1 by calling the function `results` (parameters `lfcThreshold = 0.5`, `altHypothesis = 'greaterAbs'`). For analyzing ATAC-seq signal at HAS, the size factor obtained by using the combined robust peak sets of the corresponding samples was used. For comparing ATAC-seq signals in Kc cells with S2 cells, only X chromosomal peak regions were considered.

RESULTS

MSL2 requires CLAMP for efficient binding to HAS *in vivo*

Larschan and colleagues previously reported that binding of MSL3 to the polytenic X chromosome and the crosslinking of MSL2 to some HAS required the presence of the CLAMP protein, which recognizes GA-rich sequences matching the MRE motif (26,27). On the other hand, recombinant MSL2, the male-specific DNA-binding subunit of the DCC, can select MRE sequences and is able to enrich for X chromosomal sites, especially those with PionX signature, in genomic DNA *in vitro* (22).

To explore the relationship between CLAMP and MSL2 in our system, we depleted CLAMP by RNA interference (RNAi) in male S2 cells and monitored the X chromosome binding of MSL3 and MSL2 by immunostaining. RNAi against *clamp* or *mssl2* is efficient and does not affect each other's protein levels, excluding indirect effects (Figure 1A). An irrelevant RNAi against green fluorescent protein (*gfp*) or *Schistosoma japonicum* glutathione-S-transferase (*gst*) served as control ('control' in all figures). Upon *clamp* RNAi, MSL2 and MSL3 no longer localize at a coherent X chromosome territory, but redistribute in smaller speckles, indicating a targeting deficiency (Figure 1B). The fact that both subunits still co-localize suggests that the DCC is still intact and rather mis-targeted as a complex.

To further investigate MSL2 binding *in vivo* in absence of CLAMP, we performed ChIP-seq of MSL2 in male S2 cells after depletion of CLAMP by RNAi (Supplementary Figure S1A). Upon depletion, the CLAMP ChIP-seq signal was robustly reduced and only minor residual binding was observable (Supplementary Figure S1B). Under those conditions MSL2 binding to HAS was massively reduced (Figure 1C and D). HAS with a PionX signature are the primary contacts for MSL2 *in vivo* (22). MSL2 binding to the subset of HAS which overlap with PionX sites (HAS-PionX) seemed to be slightly less affected by *clamp* RNAi, however the small amounts of remaining CLAMP after RNAi also tended to be enriched at HAS-PionX sites (Supplementary Figure S1C and D). The dependence of MSL2 on CLAMP for binding to all HAS *in vivo* was somewhat unexpected as MSL2 can target to PionX sites without the help of any additional factor *in vitro*.

As a control for these experiments we depleted the GAF. GAF (encoded by the *trithorax-like [trl]* gene) also binds to GA-rich sites genome-wide and co-localizes with CLAMP at many sites (27). Contrasting CLAMP, GAF is absent from most HAS (21,57,58). As expected, the depletion of GAF did not affect MSL2 binding (Figure 1C and D).

Since Larschan and colleagues suggested some competition in DNA-binding between CLAMP and GAF (59), we explored whether the two proteins compete for binding to HAS *in vivo*. Upon RNAi depletion of GAF in male S2 cells the CLAMP ChIP-seq signal was unchanged at HAS (Supplementary Figure S1B and D). Genome-wide, at an FDR < 0.05, only 93 out of 5983 CLAMP binding sites (1.55%) showed a statistically significant difference in CLAMP binding (Supplementary Figure S1E). These sites were mostly located in chromatin with enhancer- or

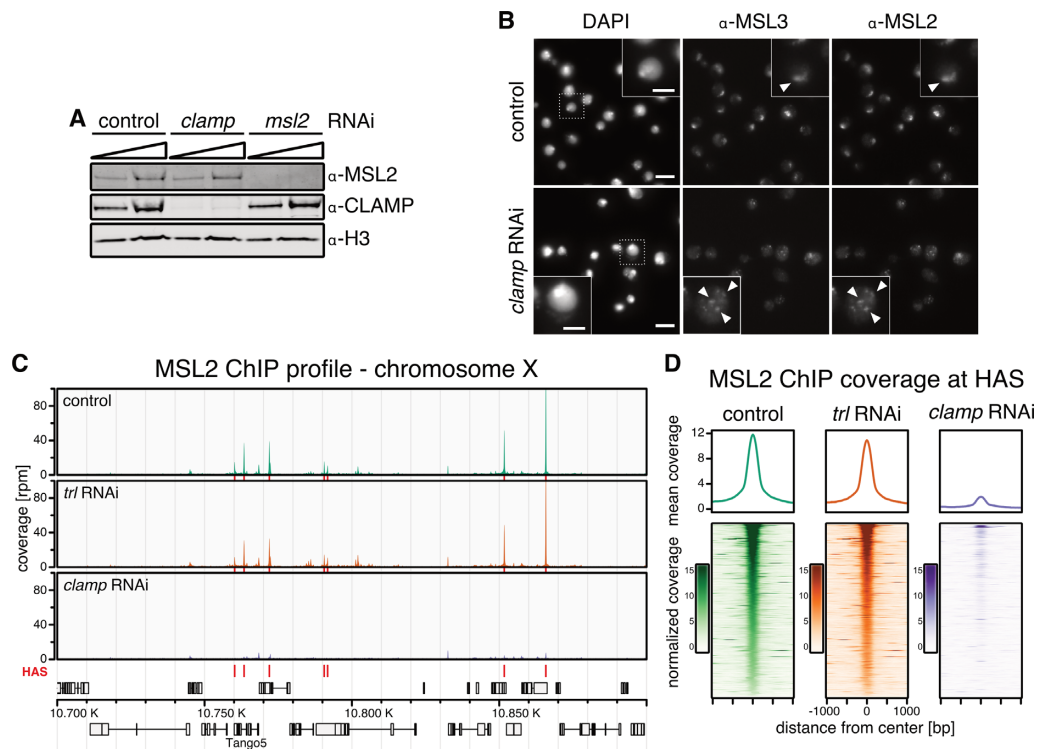


Figure 1. Binding of MSL2 to HAS in male S2 cells depends on CLAMP. (A) Western blot detection of MSL2, CLAMP and H3 in whole cell extracts from S2 cells after *msl2* or *clamp* RNAi. An irrelevant RNAi directed against green fluorescent protein (*gfp*) or *Schistosoma japonicum* glutathione-S-transferase (*gst*) sequences serve as control for these and further experiments. (B) Immunofluorescence microscopy of MSL2 and MSL3 in control cells and upon *clamp* RNAi. A region of zoom-in is marked by dashed rectangle. Scale bar: 10 μ m (5 μ m in inset). White arrow heads indicate the X chromosomal territory in control cells and speckles of MSL2 and MSL3 co-localization in *clamp* RNAi. (C) Genome browser profile of MSL2 ChIP-seq showing mean coverage (control cells $n = 3$, *trl* RNAi: $n = 3$, *clamp* RNAi: $n = 4$) along a representative 200 kb window on the X chromosome. Red bars above the gene models indicate location of HAS. (D) Average profile and heat map of mean MSL2 ChIP-seq coverage (control cells $n = 3$, *trl* RNAi: $n = 3$, *clamp* RNAi: $n = 4$) in a 2 kb window centered on 309 HAS as indicated. Heat maps are sorted by decreasing MSL2 enrichment in peak region in the control data.

promoter-related histone marks and CLAMP binding was increased or decreased (Supplementary Figure S2).

To analyze potential competition between GAF and CLAMP, we also monitored GAF binding at CLAMP binding sites after depletion of CLAMP (Supplementary Figure S1F). Nine robust CLAMP binding sites were monitored by ChIP-qPCR including three HAS (Supplementary Figure S1G). At none of these sites did GAF binding change upon removal of CLAMP.

We conclude that MSL2/DCC binding to HAS depends on CLAMP but not on GAF and that CLAMP and GAF do not generally compete for binding to GA-rich sequences.

Genomic binding sites of CLAMP defined *in vitro* by DNA immunoprecipitation

We previously assayed the DNA binding sites of MSL2 *in vitro* through genome-wide DNA immunoprecipitation (DIP) (22,35,36). In such experiments, purified and fragmented genomic DNA is incubated with protein of interest under conditions where competitive DNA binding occurs. The protein is immunoprecipitated and the bound DNA sequenced (22,35,36). For CLAMP, such an analysis had not been done. We expressed and purified recombinant full-length CLAMP via a baculovirus expression system

(Supplementary Figure S3) and assayed its *in vitro* DNA-binding property by genome-wide DIP-seq.

In vitro, CLAMP bound to 4037 sites in the *Drosophila* genome under our DIP conditions (Figure 2A), comparable in number to 5214 sites determined by our ChIP-seq approach. About one third (32.3%, $n = 1307$) of the *in vitro* CLAMP binding sites overlapped with *in vivo* binding sites (Figure 2B). Sequence analyses within *in vitro* binding sites yielded the core of the GA-repeat consensus motif defined *in vivo* by ChIP-seq, replicating the previously identified CLAMP motif, which has high similarity to the MRE motif (Figure 2C) (27). Furthermore, the CLAMP *in vitro* binding motif is highly similar to the previously identified motif for the CLAMP DNA-binding domain *in vitro* by protein-binding microarray, confirming the DNA binding specificity for our full-length recombinant protein (27,29). As expected, CLAMP bound sites on all chromosomes *in vitro* with a 2.4-fold enrichment of X chromosomal sequences similarly to *in vivo* with 1.7-fold enrichment (Figure 2D), reflecting the \sim 2-fold enrichment of the MRE motif on the X chromosome (20,21). However, the binding intensities *in vitro* and *in vivo* were uncorrelated (Figure 2E). Whereas the *in vitro* binding sites solely reflect the intrinsic binding affinity of CLAMP to DNA, the *in vivo* interactions are obviously modulated by other factors (such as MSL2, see below).

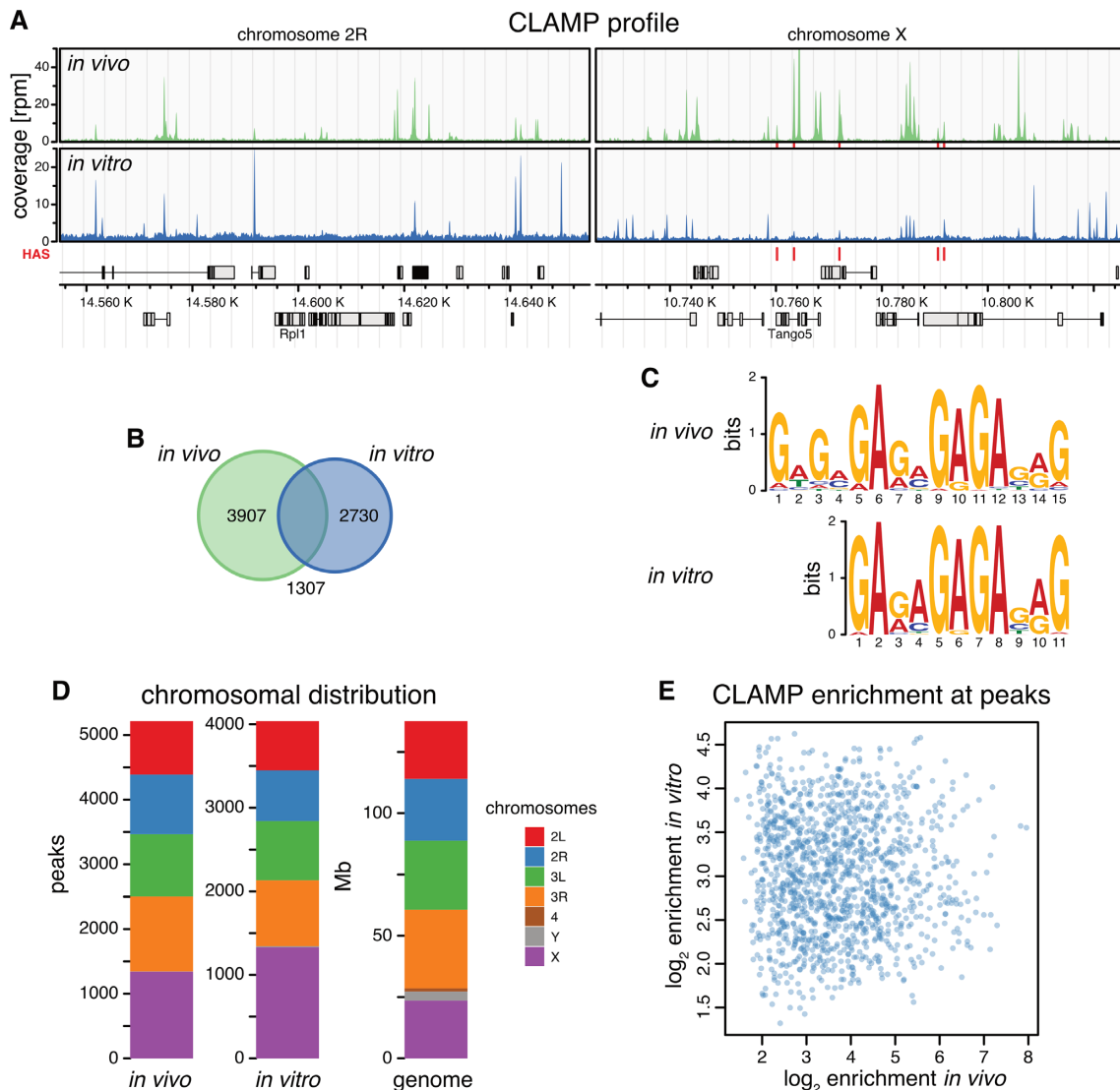


Figure 2. CLAMP selects GA-rich consensus sequence motifs *in vitro*. (A) Genome browser profile of *in vivo* CLAMP ChIP-seq (upper panel) and *in vitro* CLAMP DIP-seq (lower panel) (mean coverages, $n = 3$ each) along representative 200 kb windows on chromosome 2R and X. Red bars above the gene models indicate positions of HAS. (B) Venn diagram of robust peak sets from *in vivo* CLAMP ChIP-seq ($n = 5214$) and *in vitro* CLAMP DIP-seq ($n = 4037$). (C) *De novo* discovered motifs from robust peak sets as in (B). (D) Bar chart of chromosomal distribution of robust peak sets as in (B). The chromosome sizes serve as reference for uniform distribution (genome). (E) Scatterplot of *in vitro* CLAMP DIP-seq mean \log_2 enrichment ($n = 3$) against *in vivo* CLAMP ChIP-seq mean \log_2 enrichment ($n = 3$) at 1307 overlapping peak regions displayed on (B).

Cooperation between CLAMP and MSL2 increases the efficiency of HAS recognition in the context of the genome

We recently described the *in vitro* genomic binding sites of MSL2 by DIP-seq (22). This study revealed the capacity of MSL2 to identify PionX sites and to enrich X chromosomal sequences, but the protein missed many HAS and in addition pulled out autosomal GA-rich sites that do not correspond to physiological binding sites (22). We now explored whether cooperation with CLAMP may improve the binding specificity and/or capacity of MSL2. Therefore, we expressed and purified recombinant full-length CLAMP and MSL2 via a baculovirus expression system (Supplementary Figure S3) and assayed their DNA binding properties by genome-wide DIP-seq.

In our previous study, MSL2 protein was immunoprecipitated with an α -FLAG antibody recognizing the epitope of its FLAG-tag. Because CLAMP protein is also FLAG-tagged to facilitate its purification, we could not use the α -FLAG antibody, but performed the DIP experiments using α -CLAMP and α -MSL2 antibodies specific for the proteins. The analysis is complicated by the fact that the α -MSL2 antibody available to us yielded a DIP profile for MSL2 without significant enrichments of sites under our conditions (Figure 3A). Remarkably however, the α -MSL2 antibody retrieved a robust profile for MSL2 in the presence of CLAMP. For reference, we included our previously published MSL2 DIP-seq data using the α -FLAG antibody for immunoprecipitation as proxy for MSL2 binding *in vitro* (22).

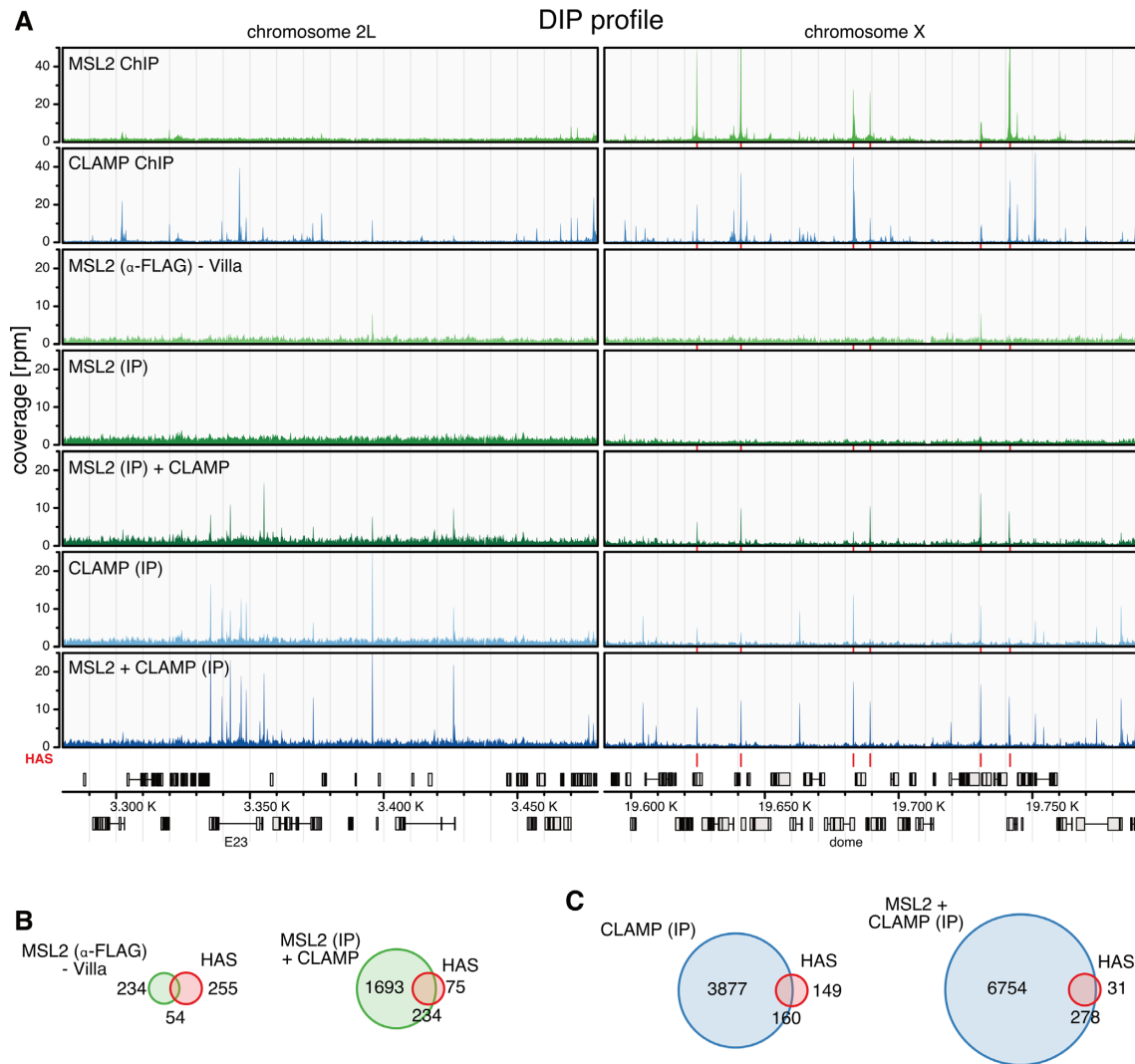


Figure 3. Intrinsic DNA binding cooperativity between CLAMP and MSL2 *in vitro*. (A) Genome browser profile showing genomic MSL2 and CLAMP binding profiles as indicated. The MSL2 and CLAMP *in vivo* ChIP-seq (top 2 profiles) and *in vitro* DIP-seq profiles (bottom five profiles) represent mean coverages ($n = 3$) along representative 200 kb windows on chromosome 2L and X. *In vitro* DIP-seq panels depict the following conditions from top to bottom: MSL2 with α -FLAG IP from Villa *et al.* (22) as proxy for MSL2 *in vitro* binding, MSL2 with α -MSL2 IP, MSL2 plus CLAMP with α -MSL2 IP, CLAMP with α -CLAMP IP and MSL2 plus CLAMP with α -CLAMP IP. Red bars above the gene model and between the panels indicate positions of HAS. (B) Venn diagrams relating robust peak sets from *in vitro* DIP-seq (green) to HAS (red, $n = 309$). Left panel: MSL2 with α -FLAG IP from Villa *et al.* (22) as proxy for MSL2 *in vitro* binding ($n = 288$, overlapping $n = 54$). Right panel: MSL2 plus CLAMP with α -MSL2 IP ($n = 1972$, overlapping $n = 234$). (C) Venn diagrams relating robust peak sets from *in vitro* DIP-seq (blue) to HAS (red, $n = 309$). Left panel: CLAMP with α -CLAMP IP ($n = 4037$, overlapping $n = 160$). Right panel: MSL2 plus CLAMP with α -CLAMP IP ($n = 7032$, overlapping $n = 278$).

MSL2 alone binds *in vitro* to 288 sites of which 54 are HAS (out of 309 HAS in total), as reported previously (22) (Figure 3B). In the presence of CLAMP, MSL2 bound now to 1927 sites including 234 HAS. A similar situation was observable for CLAMP binding *in vitro*. CLAMP alone bound 4037 sites including 160 HAS. In presence of MSL2, CLAMP bound 7032 sites including 278 HAS (Figure 3C). This reveals an intrinsic cooperativity between MSL2 and CLAMP through stabilizing each other's interactions at many common binding sites in the genome under these conditions. This cooperativity allows both factors together to select most HAS from genomic DNA *in vitro*, but at the cost of an increased number of non-physiological binding sites. However, those additional binding sites all bear sequence

determinants similar to the physiological ones showing that the intrinsic binding properties of both proteins contribute to an increased affinity for GA-rich sequences in general (see below, Supplementary Figure S4). Expectedly, many of those sites are autosomal and thus lead to a lower X chromosomal enrichment of bound sequences (see 'Discussion' section).

To explore whether different modes of binding cooperativity exist we performed hierarchical clustering of the *in vitro* binding intensities at the combined robust peak set of all DIP-seq samples (Figure 4). We identified 12 major clusters defined by the different binding behaviors of both proteins. We categorized the clusters into four binding scenarios (Figure 5A and B; Supplementary Figures S5 and 6A):

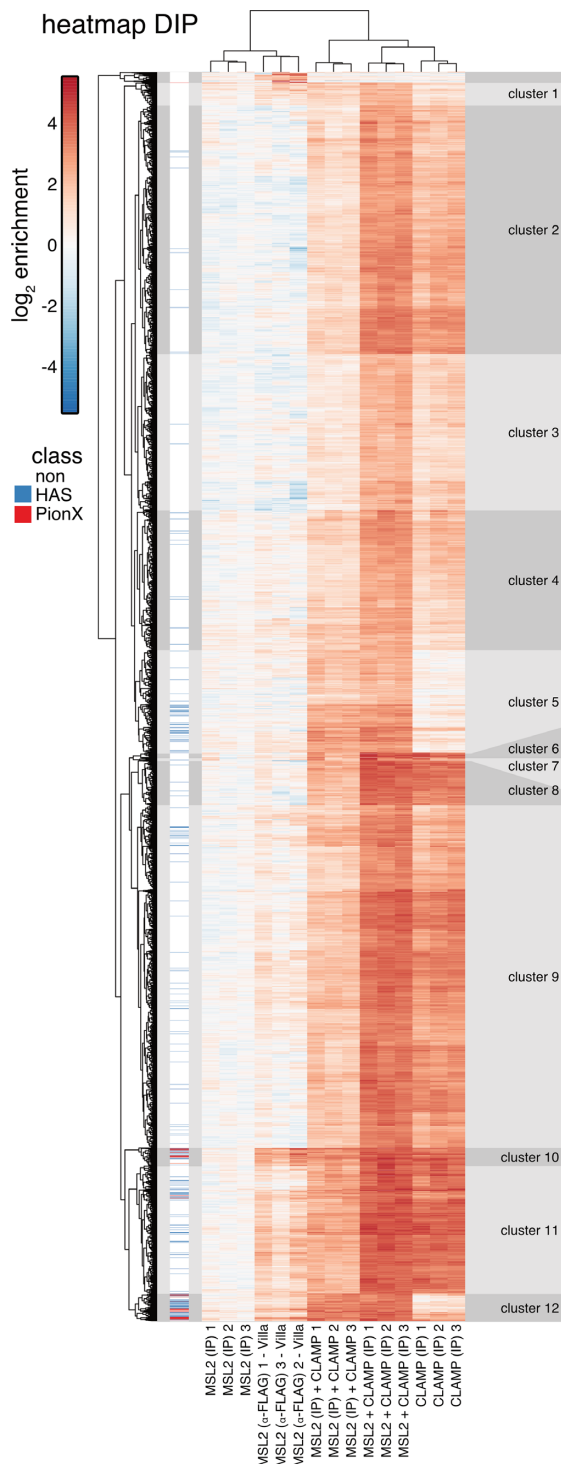


Figure 4. Genome-wide DNA binding of CLAMP and MSL2 *in vitro*. Clustered heat map of *in vitro* DIP-seq signal enrichment from reactions containing either MSL2 or CLAMP alone, or both proteins, as follows—the target of immunoprecipitation (IP) is indicated in brackets: MSL2 (IP α -FLAG) from Villa *et al.* (22) as proxy for MSL2 *in vitro* binding; MSL2 (IP α -MSL2), MSL2 and CLAMP (IP α -MSL2); CLAMP (IP α -CLAMP); MSL2 and CLAMP (IP α -CLAMP) at all combined robust peaks ($n = 7119$). For each reaction three independent replicates are shown. Hierarchical clustering revealed 18 clusters. Clusters 1–12 had distinct MSL2 and CLAMP binding properties, the remaining six clusters at top of the heat map are small and show inconsistent enrichment between MSL2 with α -FLAG IP replicates.

(i) independent: CLAMP and MSL2 bind alone and the presence of the other factor makes little difference [clusters 10 and 11]; (ii) CLAMP-dependent: CLAMP binds largely independently of MSL2 and recruits MSL2 [clusters 6–9 (clusters 2–4 show similar behavior with lower signal enrichment)]; (iii) MSL2-dependent: MSL2 binds largely independently of CLAMP and recruits CLAMP [cluster 12]; (iv) interdependent: CLAMP and MSL2 do not bind alone, but show cooperative binding [cluster 5 (cluster 1 shows similar behavior with lower signal enrichment)]. As expected, MSL2 binding to PionX sites did not depend on CLAMP (Figures 4 and 5A). CLAMP alone is absent from many PionX sites and is recruited by MSL2 to these sites.

De novo discovery of most-enriched sequence motifs within each cluster yielded variations of the GA-rich MRE motif of variable length and regularity (Supplementary Figure S4). We attempted to stratify the binding sites further by extracting additional sequence features within each binding site (Figure 5C; Supplementary Figures S4 and 6B): (i) the score of the best-matching MRE motif; (ii) the score of the best matching PionX motif; (iii) the DNA roll at position +1 of the best-matching PionX motif (high roll at position +1 is a defining feature of the PionX signature (22); for simplicity we assigned the inter-base structural feature roll to the lead base pair); (iv) density of GA:TC-dinucleotides; (v) length of the longest (GA:TC)_x-repeat (as x repeats).

Interestingly, the ‘independent’ sites have the highest MRE scores, the highest GA:TC-dinucleotide density and the longest GA:TC-repeats (Figure 5C and Supplementary Figure S6B). This is in good agreement with previous findings, as the first motif discovered in MSL2 *in vitro* binding sites is a long, low-complexity GA:TC-repeat, which was suggested to be bound through the MSL2 C-terminus (22), and CLAMP *in vitro* binding intensities correlate with GA:TC-repeat length (29). The ‘CLAMP-dependent’ sites tend to have higher MRE scores, higher GA:TC-dinucleotide density, longer GA:TC-repeats, but low PionX scores as well as low roll at position +1. By contrast, the ‘MSL2-dependent’ sites tend to have lower MRE scores, lower GA:TC-dinucleotide density and shorter GA:TC-repeats, but the highest PionX scores including the ‘high roll at position +1’ (22). Interestingly, ‘interdependent’ sites lack the features that characterize CLAMP or MSL2 binding sites *in vitro* (high PionX and MRE scores, high GA:TC-dinucleotide density and long GA:TC-repeats length), with the exception of ‘high roll at position +1’. Apparently, neither MSL2 nor CLAMP alone bind well to these sites, but together they mount an interaction surface able to recognize a degenerate GA-rich MRE consensus motif (Supplementary Figure S4). We speculate that the CXC domain of MSL2 reads out the DNA shape at the 5' end of these binding sites.

The cooperativity between MSL2 and CLAMP expands each other's binding repertoire *in vitro*. Together both proteins are capable to identify nearly all physiological functional MRE sequencing (HAS) *in vitro*. They also bind to many autosomal sites, which may be occluded by nucleosomes *in vivo*. Accordingly and in agreement with earlier findings (22), the enrichment of X chromosomal sequences *in vitro* is almost entirely due to the action of MSL2. MSL2-dependent sites which harbor the PionX signature are 10.4-

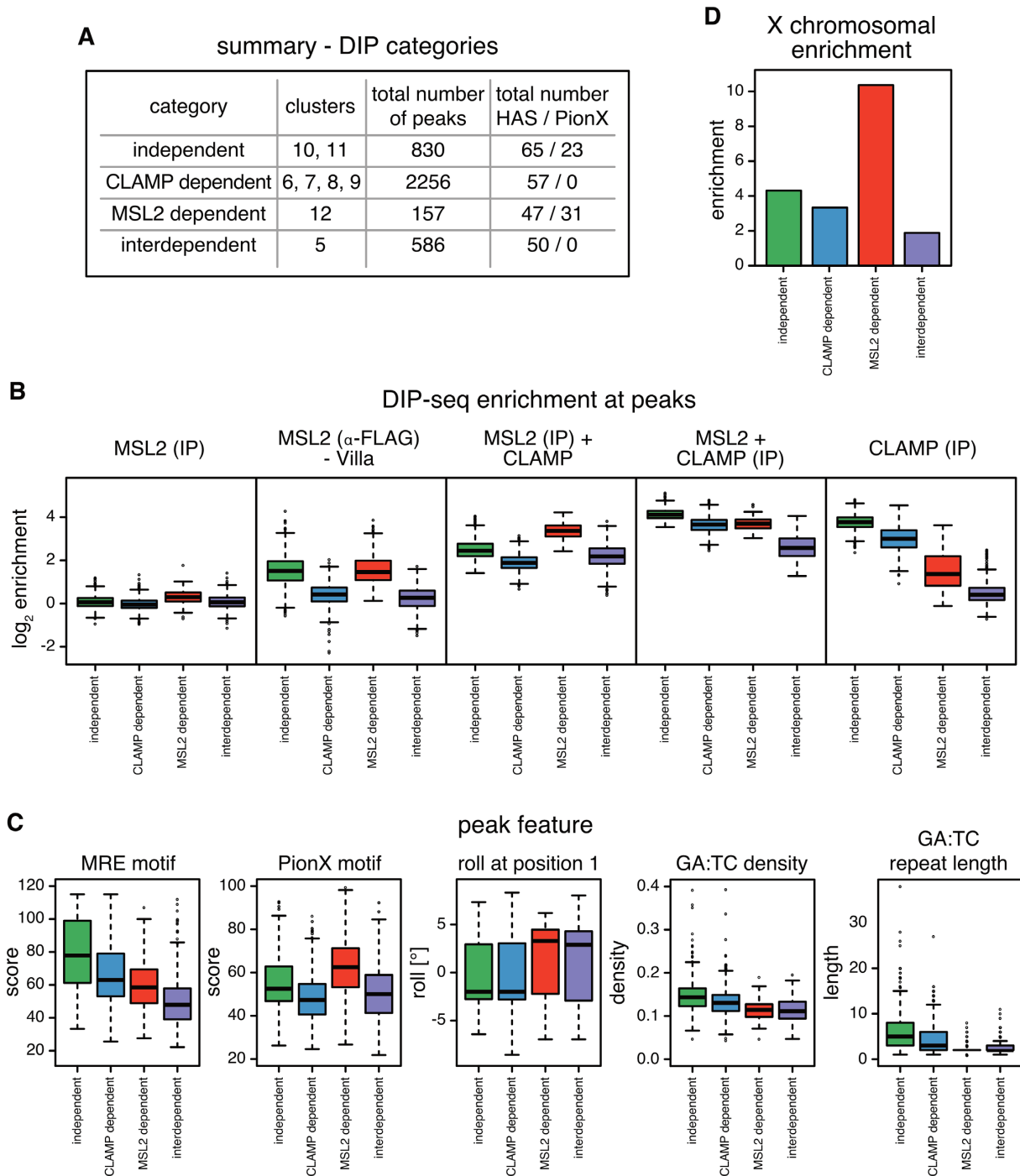


Figure 5. Cooperation between CLAMP and MSL2 in genome-wide DNA binding *in vitro*. (A) Summary of distinct binding properties discovered in the heat map (Figure 4). Clusters were assigned into four categories: independent, CLAMP-dependent, MSL2-dependent and interdependent. (B) Boxplot of mean \log_2 enrichment ($n = 3$) of *in vitro* DIP-seq at peaks grouped by the four categories described in (A) for MSL2 (IP α -MSL2), MSL2 (IP α -FLAG) from Villa *et al.*, MSL2 and CLAMP (IP α -MSL2); CLAMP (IP α -CLAMP); MSL2 and CLAMP (IP α -CLAMP). (C) Boxplot of peak features grouped by the four categories described in (A). Panels from left to right show: score for the best matching MRE motif; score for the best matching PionX motif; the roll at position 1 of the best matching PionX motif; the density of GA:TC dinucleotides and the length of GA:TC repeats. (D) Bar chart of X chromosomal enrichment of peaks grouped by the four categories described in (A).

fold enriched on the X chromosome, recapitulating the 9.8-fold enrichment of PionX sequences (Figure 5D and Supplementary Figure S6C) (22).

CLAMP and MSL2 form a stable complex

The results of the DIP experiments suggest cooperativity between CLAMP and MSL2. Such synergism may be due to physical interaction of the two proteins. Previously, CLAMP has been found associated with the DCC after crosslinking *in vivo* (60).

As a direct test for protein interactions we employed a Yeast Two-Hybrid assay (Y2H). MSL2 (and deletion derivatives) were expressed in fusion with a Gal4 DNA binding domain (DBD) along with CLAMP (and deletion derivatives) fused to the Gal4 activation domain (AD). If co-expressed in yeast, the association of the two proteins reconstitutes the function of the transcription factor GAL4 that activates the *his3* gene in the yeast strain pJ69-4A that is auxotroph for histidine (61). Yeasts in which DBD and AD fusion proteins interact can grow on plates lacking histidine, leucine and tryptophan. The assay revealed a robust and reproducible interaction of MSL2-DBD and AD-CLAMP (Supplementary Figure S7A) even in stringent conditions posed by absence of adenine and presence of 3-amino-1,2,4-triazole, a competitive inhibitor of the enzyme encoded by *his3*. MSL2-DBD or AD-CLAMP alone did not support growth in the presence of the unfused AD or DBD, respectively (Supplementary Figure S7A). Using appropriate deletion constructs we found that the first 153 amino acids of CLAMP are sufficient to interact with DBD-MSL2 but further deletion to the first 123 amino acids abolished the interaction (Supplementary Figure S7A). This assay locates the MSL2 interaction site to an N-terminal fragment of CLAMP, which harbors the first ZnF domain. The remaining six C-terminal ZnF domains are involved in binding to GA-dinucleotide repeats (27,29,59). Using various C-terminal deletion constructs of MSL2 showed that the ~200 C-terminal amino acids downstream of the CXC domain are required for interaction with CLAMP (Supplementary Figure S7A). As a control for proper MSL2 folding we confirmed that MSL2¹⁻⁵⁷³-DBD was still able to interact with MSL1 via the N-terminal RING domain in our assay (Supplementary Figure S8B) as reported previously (62).

To probe whether this interaction was direct we tested the recombinant purified proteins used in DIP in the absence of DNA. The two proteins were probed at equimolar concentration (100 nM) for interaction by co-immunoprecipitation (co-IP) with α -MSL2 and α -CLAMP antibodies. Both proteins were quantitatively (~90%) immunoprecipitated with the α -MSL2 antibody (Figure 6A and B). While with the α -CLAMP antibody only ~50% of each protein was immunoprecipitated, presumably because the antibody amount was limiting. Of note, the IP of CLAMP was more efficient in the presence of MSL2 (Supplementary Figure S9A and B), perhaps due to conformational stabilization.

To map the interaction surfaces within MSL2 and CLAMP, we co-expressed deletion mutants of both recombinant FLAG-tagged proteins in Sf21 cells and immunoprecipitated them from total cell extracts. To map the CLAMP-interaction site in MSL2, we used a series of C-

terminal deletions systematically lacking conserved regions identified within MSL2. In short, alignment of MSL2 sequences from 12 *Drosophila* species revealed five conserved regions (CR) (Supplementary Figures S9B and 10): the highly conserved RING and CXC domains; CR2 consisting of 66 amino acids with 50%-90% conservation; a conserved PAKKFR motif, as part of a stretch of 25 residues enriched in basic amino acids; and CR3, which corresponds to the 28 C-terminal amino acids of the *D. melanogaster* MSL2 (Supplementary Figure S9B). The 20 amino acid proline-rich region within MSL2's C-terminus separates the CR2 from the PAKKFR motif. Alignment of CLAMP sequences from 12 *Drosophila* species revealed that CLAMP is conserved throughout its entire length (Supplementary Figure S11).

MSL2 derivatives harboring CR2 (MSL2¹⁻⁷²⁶ and MSL2¹⁻⁶⁸⁸) showed quantitative binding of CLAMP, whereas fragments lacking CR2 (MSL2¹⁻⁶¹⁹ and MSL2¹⁻⁵⁶⁷) did not bind CLAMP (Figure 6C and D). CLAMP binding was unaffected by deletion of the CXC domain, but internal deletion of CR2 abolished all CLAMP binding. We therefore refer to the conserved region between amino acids 620–685 as 'CLAMP-binding domain' (CBD) of MSL2.

To interrogate the MSL2 interaction surface in CLAMP we used conditions where CLAMP was in excess over MSL2 to favor the interaction with MSL2 in α -MSL2 co-IP experiments. While full-length CLAMP reproducibly bound to MSL2, none of eight CLAMP deletion mutants, including the N-terminal CLAMP¹⁻¹⁵³ fragment, which interacted with MSL2 in the Y2H assay (Supplementary Figures S7A and 8A), bound to MSL2 above a background, even at protein concentrations in the extracts approaching 100 nM (Supplementary Figure S9C–F).

In summary, our data document a stable interaction between MSL2 and CLAMP and we map the interaction domains to the N-terminus of CLAMP including the first zinc finger domain and the CBD of MSL2 just downstream of the CXC domain.

CLAMP and MSL2 cooperate to keep HAS nucleosome-free

The genome-wide DIP revealed the potential for extensive cooperation between MSL2 and CLAMP to bind shared binding sites, but the physiological X/autosome discrimination of MSL2 was not improved in free DNA. Conceivably, exclusive X chromosome binding requires a chromatin environment.

To survey the contribution of either factor to HAS accessibility in chromatin, we performed ATAC-seq after RNAi against *clamp* or *msl2* in male S2 and female Kc cells (Figure 7A and Supplementary Figure S12A). Depletion of CLAMP in S2 and Kc cells by RNAi caused only few significant changes in accessibility genome-wide (258 of 8913 sites and 102 of 9767 sites in S2 and Kc cells, respectively; Figure 7B and Supplementary Figure S12B). This indicates that most sites bound by CLAMP are kept accessible by other factors. Most sites affected by CLAMP depletion became less accessible, showing that CLAMP contributes to keeping particular loci open. These sites, that depend on CLAMP to be accessible, include especially many HAS (61

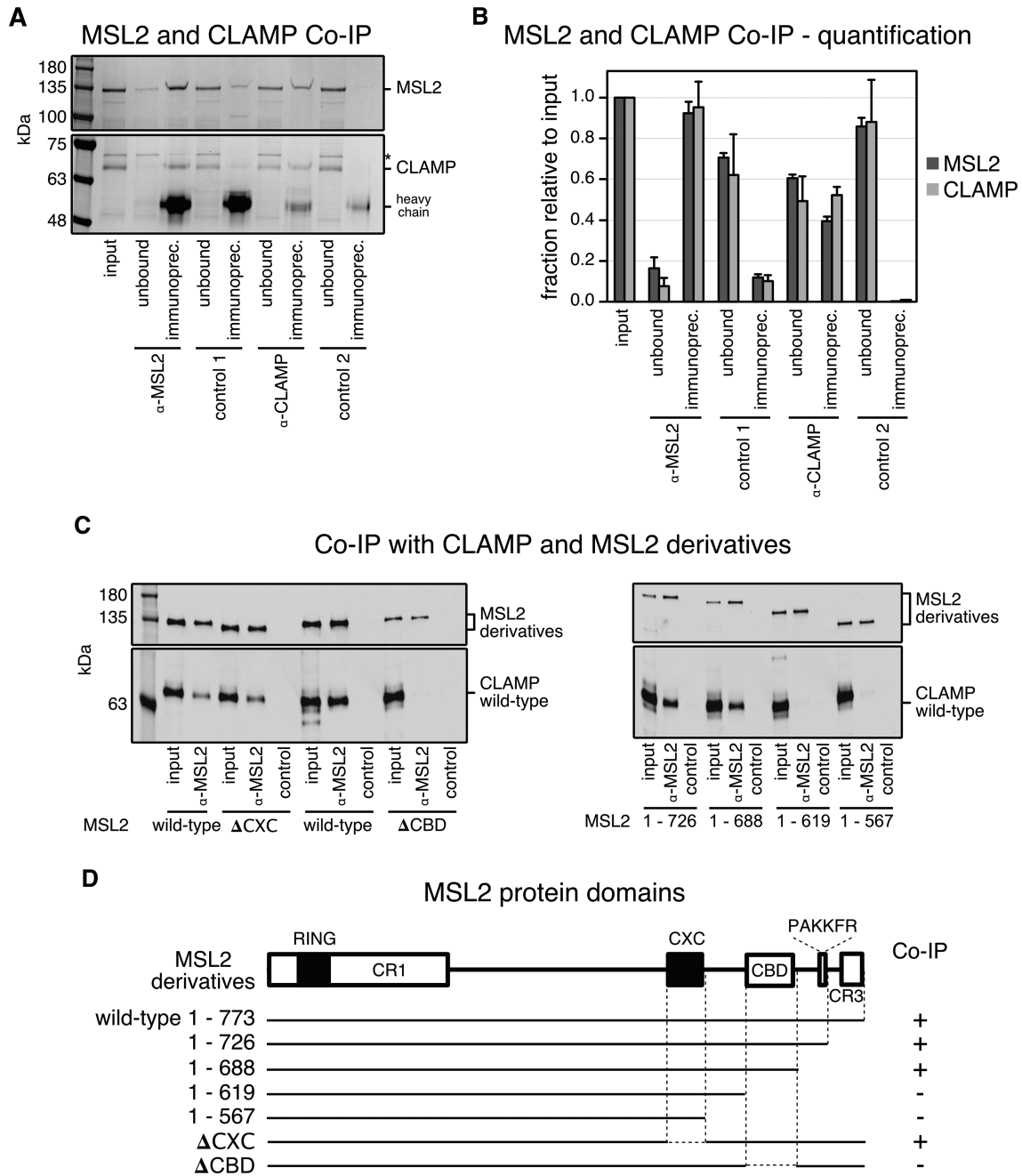


Figure 6. A conserved C-terminal region in MSL2 is responsible for CLAMP binding. **(A)** SDS-PAGE analysis with Coomassie staining of co-IP fractions. Purified wild-type recombinant MSL2-FLAG and CLAMP-FLAG were immunoprecipitated with α -MSL2 serum and the corresponding pre-immune serum as control (control 1), and with affinity-purified α -CLAMP antibody mixed into an irrelevant rabbit serum and with the irrelevant rabbit serum only as control 2. The corresponding unbound fractions are loaded next to each IP. A contaminant present in the MSL2 preparation is labeled with asterisk. Molecular weight markers are shown to the left. **(B)** Bar chart of the quantification from co-IP experiments as in (A), combining data from three independent MSL2-FLAG and CLAMP-FLAG purifications. The amount of each protein in the unbound fractions and IP's were quantified relative to the input. Error bars represent the standard deviation ($n = 3$). **(C)** Quantitative western blot analysis using α -FLAG antibody of co-IP experiments with extracts from Sf21 cells expressing wild-type CLAMP-FLAG and various MSL2-FLAG C-terminal deletion mutants. Co-IP was performed with α -MSL2 serum and the corresponding pre-immune serum as control [control 1 in (A)]. IP fractions were loaded next to each corresponding input. **(D)** Summary of MSL2 and CLAMP interaction from co-IP experiments presented in (C). Interactions with a CLAMP/MSL2 ratio in IP-FLAG western blot analysis between 0.3 and 1.7 are depicted by (+) and no detectable interaction by (-). The MSL2 domain architecture is drawn to scale. White rectangles represent the conserved regions: CR1, CBD [CR2] and CR3 (Supplementary Figures S7B and 10). Black rectangles represent the RING and CXC domains.

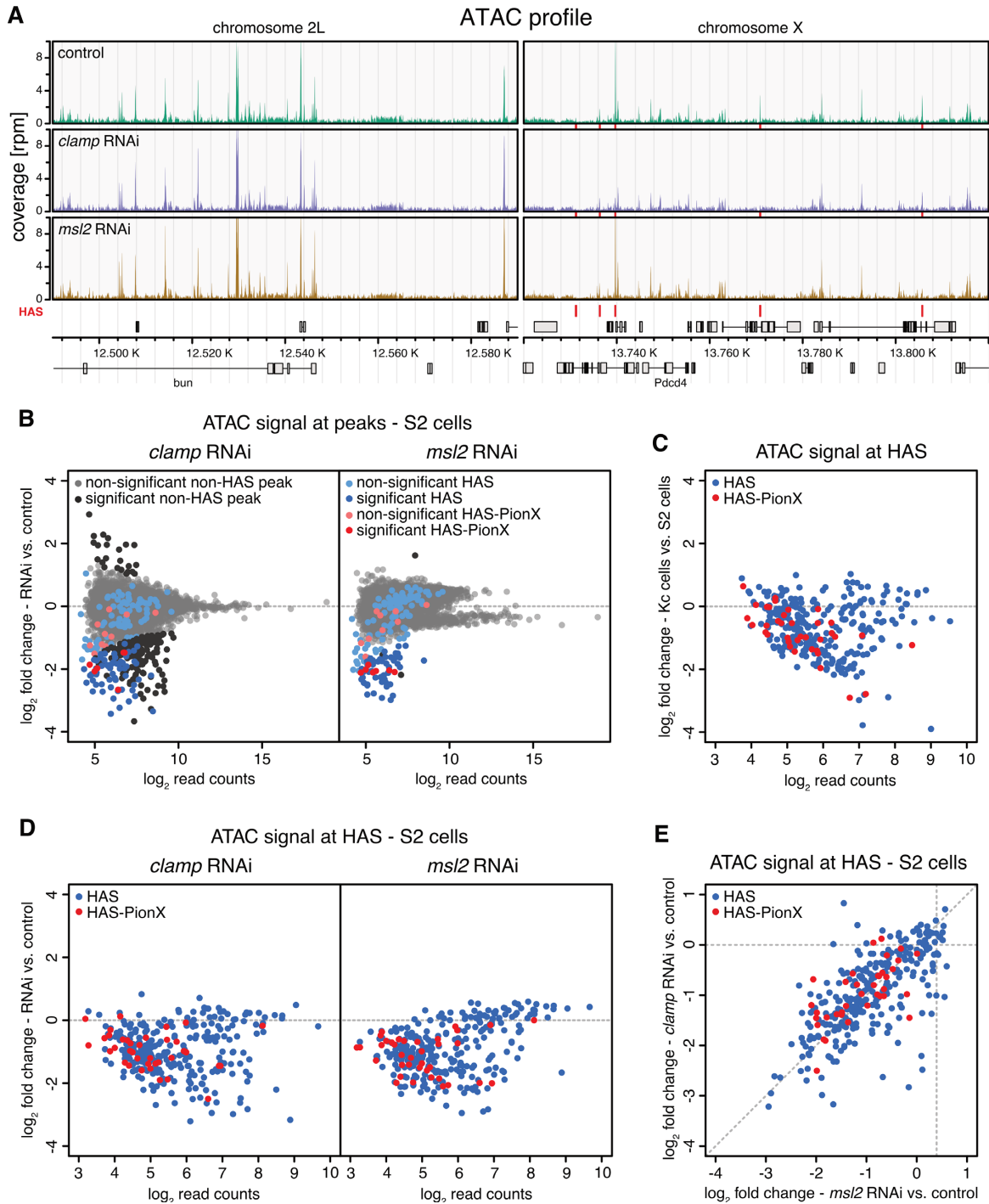


Figure 7. MSL2 and CLAMP synergize to keep HAS accessible in male cells. (A) Genome browser profile of ATAC-seq showing mean coverages ($n = 3$) along representative 100 kb windows on chromosome 2L and X. The panels show control S2 cells and cells after *clamp* RNAi or *msl2* RNAi as indicated. Red bars above the gene models and between the panels mark positions of HAS. (B) Scatter plot of mean \log_2 fold-change ($n = 3$) of ATAC-seq signal in S2 cells upon *clamp* RNAi (left panel) and *msl2* RNAi versus controls (right panel) against mean \log_2 read count in control at robust ATAC peaks ($n = 8913$). HAS non-overlapping with PionX sites (HAS) are marked in blue and HAS overlapping with PionX sites (HAS-PionX) are marked in red (the remaining sites are displayed in gray). Sites with statistically significant different ATAC signal between RNAi and control conditions ($|\text{lfcl}| > 0.5$ and $\text{fdr} < 0.1$) are marked in darker color. For *clamp* RNAi, 258 ATAC peaks are statistically significant different between conditions, including 61 HAS and 6 HAS-PionX. For *msl2* RNAi, 61 ATAC peaks are statistically significant different between conditions, including 49 HAS and 7 HAS-PionX. (C) Scatter plot of mean \log_2 fold-change ($n = 4$) of ATAC-seq signal in Kc cells versus S2 cells against mean \log_2 read count in S2 cells at HAS ($n = 309$). HAS non-overlapping with PionX sites (HAS, $n = 37$) and overlapping with PionX sites (HAS-PionX, $n = 37$) are marked in red. (D) Scatter plot of mean \log_2 fold-change ($n = 3$) of ATAC-seq signal in S2 cells upon *clamp* RNAi (left panel) and *msl2* RNAi versus controls (right panel) against mean \log_2 read count in control at HAS ($n = 309$). HAS non-overlapping with PionX sites (HAS, $n = 272$) are marked in blue and overlapping with PionX sites (HAS-PionX, $n = 37$) are marked red. (E) Scatter plot of mean \log_2 fold-change ($n = 3$) of ATAC-seq signal in S2 cells upon *clamp* RNAi (left panel) against *msl2* RNAi versus controls (right panel) at HAS ($n = 309$), as shown in (D).

HAS and 6 HAS-PionX) in S2 cells. Focusing on the 309 HAS showed that HAS are commonly accessible in male S2, but not in female Kc cells (Figure 7C and Supplementary Figure S12B) suggesting that the DCC is required for HAS accessibility. Indeed, depletion of MSL2 in S2 cells caused only few significant changes in accessibility genome-wide (61 of 8913 sites), but these sites were nearly exclusively HAS [49 HAS and 7 HAS-PionX] (Figure 7B).

Remarkably, most HAS are inaccessible in the absence of either CLAMP or MSL2 in S2 cells (Figure 7D). Apparently, CLAMP and MSL2 contribute equally to HAS accessibility (Figure 7E). The minority of HAS that remained accessible in the absence of MSL2 or CLAMP are also accessible in Kc cells (Figure 7C and D; Supplementary Figure S12B), indicating that these HAS are kept open by other factors.

Broadening the view, we found that CLAMP promotes access to other loci in the genomes of both cell lines (Figure 7B and Supplementary Figure S12B). According to the 9-state chromatin model (52) (Supplementary Figure S12C) many of these sites bear promoter or enhancer signatures. Unexpectedly, we found that sites marked by H3K27me3, the polycomb signature, are most enriched within the sites affected by CLAMP depletion.

These experiments identify a second layer of DNA binding cooperativity between CLAMP and MSL2. Faced with purified genomic DNA both proteins cooperate to bind to essentially all sites with appropriate DNA sequence features, regardless of chromosomal origin. By contrast, in a physiological chromatin environment extensive cooperativity between both factors is restricted to X chromosomal HAS.

DISCUSSION

The process of dosage compensation in *Drosophila* provides an excellent opportunity to study the principles of sequence-selective DNA binding. Male flies only survive if the regulatory DCC exclusively binds to the X chromosome. Earlier work suggested that this exclusivity is at least partly due to the functional cooperation between the DNA-binding subunit of the DCC, MSL2 (22–24,39), and the zinc finger protein CLAMP (26,27,29). Curiously, both factors share the intrinsic property to bind to GA-rich MRE sequences that are hallmarks of the X chromosomal high affinity sites (HAS) of the DCC. We recently found that the CXC domain of MSL2 can read the DNA signature of a prominent subset of MREs with notable 5' extension, termed PionX (22). Using chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) we now found that essentially all MSL2 interactions with HAS requires the presence of CLAMP (Figure 1), suggesting a tight cooperativity between both factors.

Cooperative DNA binding of transcription factors (TFs) for refined and stable recognition of complex DNA elements is a widely used principle of gene regulation (63). Such cooperativity may involve a direct contact between two factors. Their simultaneous contact with DNA and dimerization partner reduces the off-rate of each individual binder significantly (2). We indeed found a physical interaction between CLAMP and MSL2 and mapped the

CLAMP-binding domain on MSL2 just C-terminal to the DNA-binding CXC domain (Figure 6).

The dimerization of TFs may happen in solution or be promoted by a target DNA, which defines spacing and orientation of the proteins binding to adjacent sequences. Although we demonstrated stable, soluble complex formation between CLAMP and MSL2, we consider it unlikely that such diffusible complexes are abundant in cells. Interactions have so far only been observed in ChIP followed by mass spectrometry with prior crosslinking (60). Probing the dynamics of MSL2 by FRAP (fluorescence recovery after photobleaching) we found earlier (and recently confirmed) that MSL2 binds the X chromosome very tightly with no evidence for a freely diffusible component (64). Further, MSL2 is only expressed during S phase, when the newly synthesized X chromosome needs to be dosage compensated (65). We thus favor the idea that the proteins encounter each other on DNA and then 'lock in' to stabilize each other on the HAS binding site.

We speculate that this interaction may align the DNA-binding CXC domain of MSL2 with the GA-repeat binding zinc fingers of CLAMP to form a long, contiguous DNA interaction surface suited to read out the long (~20 bp) MRE/PionX sequences. In such an arrangement, the CXC domain may recognize the prominent DNA shape feature akin to the PionX signature at the 5' end of a binding site, whereas CLAMP may use its zinc fingers to interact with multiple GA-repeats in the 3' part.

Using *in vitro* DIP to monitor the influence of both CLAMP and MSL2 to each other's binding, we found both factors cooperate extensively to extend each other's binding repertoire. This cooperativity enables both factors together to bind nearly all HAS in the context of genomic DNA *in vitro* (Figure 3). At the same time, the cooperativity between CLAMP and MSL2 leads to increased selection of non-physiological binding sites with similar sequence determinants. In line with our previous findings, MSL2 is the main determinant for the X chromosomal enrichment (22). Here, we speculate that the protein amount of MSL2 is limiting *in vivo* (66) and therefore the cooperativity is restricted to genomic binding sites with highest affinity, in which physiological MREs in HAS/PionX are highly enriched (Figure 4).

Furthermore, our DIP analysis revealed several clusters of sequence elements that differ in MSL2 and CLAMP binding affinity and cooperativity (Figures 3 and 4). The different binding scenarios suggest that the cooperativity depends on the precise properties of each individual DNA sequence, such as length, composition and shape (Figure 5). The long GA-repeat that characterizes MREs suggests a flexible interaction of GA-recognition surfaces of either MSL2 or CLAMP to move sideways to accommodate binding of the partner. It is possible that CLAMP does not use all six GA-repeat binding zinc fingers simultaneously at shorter sites. Previously it was shown that the multiple-zinc finger protein CTCF makes flexible use of its zinc fingers to bind a diverse range of sequences (67).

Although the results suggest a certain flexibility of factor interaction with DNA, some simple trends can be seen. CLAMP binding correlates with GA:CT-dinucleotide density and -repeat length and MSL2 retrieves the PionX signa-

ture (Figure 5 and Supplementary Figure S4), in agreement with previous observations (22,29). Sites that are bound in an 'MSL2-dependent' mode contain PionX sequences with high roll at position +1 and contribute most to X chromosome specificity, in line with our earlier conclusion (22). Furthermore, both proteins can independently bind to sequences with high MRE score and long GA-repeats, in agreement with the previous finding that also MSL2 is capable to retrieve long GA-repeats most likely through interaction via its proline-rich C-terminus (22). Conversely, sites with low MRE scores (but high roll at position +1 as characteristic for the PionX signature) and short GA-repeats can only be bound if both factors cooperate (interdependence).

TFs may cooperate for DNA binding even without direct interaction if their binding to DNA competes with nucleosome formation. In this scenario, binding of each individual TF hinders nucleosome formation and increases the likelihood that the second TF finds its close-by binding site accessible (1,68,69). Our ATAC-seq study suggests that this is an important aspect of MSL2/CLAMP function (Figure 7). TF cooperativity may be particularly important if the concentrations of a partner is too low to effectively compete with nucleosome formation by itself (7). The concentrations of MSL2 are suggested to be relatively low and tightly controlled because excess of MSL2 over X chromosomal binding sites will lead to binding of autosomal sites causing male lethality (66). Conceivably, the abundant CLAMP was coopted during the evolution of the MSL2-MRE/PionX interaction to increase the affinity of MSL2 to relevant binding sites. CLAMP binds longer GA-repeat sequences (Figure 2) (29) and because its binding has been reported to lead to regions of enhanced accessibility of chromatin in its neighborhood (34), we hypothesized that CLAMP may fulfill a 'chromatin opener' function for MSL2. This 'division of labor' model between an abundant 'chromatin opener' and a TF that profits from the accessible region was developed following the observation that binding of the GAF to GAGAG sequences in promoters, enhancers and polycomb response elements, leads to chromatin opening and facilitated the binding of other proteins in the neighborhood (30–33). Our data are not consistent with such a hierarchical model. Rather surprisingly, they reveal that both factors contribute equally to keep shared binding sites nucleosome-free (Figure 7). While CLAMP is required to stabilize the interaction of MSL2 at HAS (Figure 1), the reverse is also true: stable interaction of CLAMP at HAS critically depends on MSL2 (27). Our findings are reminiscent of the recently proposed cooperativity between the pioneer factor Zelda and the morphogen Bicoid during *Drosophila* preblastoderm development (70).

Given the many other CLAMP binding sites in the genome, we assume that CLAMP cooperates with other proteins elsewhere in the genome. Interestingly, we found that depletion of CLAMP leads to diminished accessibility at chromatin with hallmarks of polycomb repression (Supplementary Figure S12C) may pointing to a hitherto unappreciated cooperation with the polycomb machinery. The accessibility of most other CLAMP binding sites is independent of CLAMP, suggesting that other DNA-binding proteins promote accessibility of these binding sites.

Our DIP analysis clearly documents the potential for cooperative DNA binding genome-wide, but this alone is not sufficient to discriminate the physiological HAS on the X chromosome from similar sequences that are not *in vivo* targets of the DCC. *In vivo*, the synergism between MSL2 and CLAMP only manifests in chromatin at a relatively small number of X chromosomal HAS. Our study suggests that competing chromatin assembly may pose stringent demands on the cooperative action of MSL2 and CLAMP. Additional cooperativity may manifest at the level of MSL proteins and HAS DNA. *In vivo*, assembly of MSL2 with other MSL proteins and *roX* RNA involves dimerization via the MSL1 subunit. Such dimerization will bring two DNA-binding domains of MSL2 into proximity, increasing the potential for cooperative effects. The potential of DNA recognition by combinations of MSL2 and CLAMP DNA-binding domains is matched by the complexity of their DNA targets as many HAS contain several MRE sequences (20,21). Our approach of *in vitro* DNA immunoprecipitation to assess cooperativity of DNA-binding factors in the genomic context should be widely applicable.

We propose that the striking X chromosomal enrichment of the DCC *in vivo* relies on the limiting amount of MSL2, the combinatorial binding potential of MSL2 and CLAMP and on the balance between affinity and accessibility of target sequences in chromatin. Our systematic analysis of intrinsic DNA-binding properties of two key factors involved in X chromosome recognition and of their *in vivo* chromatin interactions sheds light on the sophistication of combinatorial DNA recognition that evolved to prevent male lethality.

DATA AVAILABILITY

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (71) and are accessible through GEO Series accession number GSE119708 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119708>).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank S. Krebs and the LAFUGA Genomics Facility for next generation sequencing, M. Müller for initial cloning of full-length CLAMP-FLAG, E. Kremmer for antibody generation and E. Larschan for the antibody against CLAMP.

Authors' contributions: C.A., C.R. and P.B.B. conceived the study; C.A. performed all experiments except for co-immunoprecipitation and Yeast Two-Hybrid experiments; S.K. expressed proteins and performed co-immunoprecipitation experiments; E.T. and O.M. conceived, performed and analyzed Yeast Two-Hybrid experiments; C.A. performed bioinformatics analyses; All authors analyzed data; C.R. and P.B.B. provided feedback and supervision; C.A., C.R. and P.B.B. wrote the manuscript; P.B.B. secured funding.

FUNDING

Deutsche Forschungsgemeinschaft [Be1140/8-1 to P.B.B.]; Russian Science Foundation [RSF #17-74-20155 to O.M.]; Graduate School for Quantitative Biosciences Munich, DFG Fellowship (to C.A.). The open access publication charge for this paper has been waived by Oxford University Press – *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.

Conflict of interest statement. None declared.

REFERENCES

- Mirny, L.A. (2010) Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 22534–22539.
- Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E. and Taipale, J. (2015) DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, **527**, 384–388.
- Watson, L.C., Kuchenbecker, K.M., Schiller, B.J., Gross, J.D., Pufall, M.A. and Yamamoto, K.R. (2013) The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. *Nat. Struct. Mol. Biol.*, **20**, 876–883.
- Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordan, R. and Rohs, R. (2014) Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.*, **39**, 381–399.
- Dror, I., Golan, T., Levy, C., Rohs, R. and Mandel-Gutfreund, Y. (2015) A widespread role of the motif environment in transcription factor binding across diverse protein families. *Genome Res.*, **25**, 1268–1280.
- Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H.J., Rohs, R. and Mann, R.S. (2015) Deconvolving the recognition of DNA shape from sequence. *Cell*, **161**, 307–318.
- Li, X.-Y., Thomas, S., Sabo, P.J., Eisen, M.B., Stamatoyannopoulos, J.A. and Biggin, M.D. (2011) The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biol.*, **12**, R34.
- Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M. and Zaret, K.S. (2015) Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, **161**, 555–568.
- Lucchesi, J.C. and Kuroda, M.I. (2015) Dosage compensation in Drosophila. *Cold Spring Harb. Perspect. Biol.*, **7**, a019398.
- Keller, C.I. and Akhtar, A. (2015) The MSL complex: juggling RNA-protein interactions for dosage compensation and beyond. *Curr. Opin. Genet. Dev.*, **31**, 1–11.
- 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.
- 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 Rep.*, **18**, 1854–1868.
- 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.
- 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.
- Prestel, M., Feller, C., Straub, T., Mitlohner, H. and Becker, P.B. (2010) The activation potential of MOF is constrained for dosage compensation. *Mol. Cell*, **38**, 815–826.
- Akhtar, A. and Becker, P.B. (2000) Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol. Cell*, **5**, 367–375.
- Smith, E.R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R.G., Allis, C.D. and Lucchesi, J.C. (2000) The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol. Cell Biol.*, **20**, 312–318.
- Gelbart, M.E., Larschan, E., Peng, S., Park, P.J. and Kuroda, M.I. (2009) Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat. Struct. Mol. Biol.*, **16**, 825–832.
- Ferrari, F., Alekseyenko, A.A., Park, P.J. and Kuroda, M.I. (2014) Transcriptional control of a whole chromosome: emerging models for dosage compensation. *Nat. Struct. Mol. Biol.*, **21**, 118–125.
- 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.
- 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.
- 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.
- Fauth, T., Muller-Planitz, F., König, 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.
- Zheng, S., Villa, R., Wang, J., Feng, Y., Wang, J., Becker, P.B. and Ye, K. (2014) Structural basis of X chromosome DNA recognition by the MSL2 CXC domain during Drosophila dosage compensation. *Genes Dev.*, **28**, 2652–2662.
- 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.
- 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.
- 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 Dev.*, **27**, 1551–1556.
- Urban, J.A., Doherty, C.A., Jordan, W.T. 3rd, Bliss, J.E., Feng, J., Soruco, M.M., Rieder, L.E., Tsiarli, M.A. and Larschan, E.N. (2017) The essential Drosophila CLAMP protein differentially regulates non-coding roX RNAs in male and females. *Chromosome Res.*, **25**, 101–113.
- Kuzu, G., Kaye, E.G., Chery, J., Siggers, T., Yang, L., Dobson, J.R., Boor, S., Bliss, J., Liu, W., Jögl, 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.
- Leibovitch, B.A., Lu, Q., Benjamin, L.R., Liu, Y., Gilmour, D.S. and Elgin, S.C.R. (2002) GAGA factor and the TFIID complex collaborate in generating an open chromatin structure at the Drosophila melanogaster hsp26 promoter. *Mol. Cell Biol.*, **22**, 6148–6157.
- Strutt, H., Cavalli, G. and Paro, R. (1997) Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.*, **16**, 3621–3632.
- Lu, Q., Wallrath, L.L., Granok, H. and Elgin, S.C. (1993) (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene. *Mol. Cell Biol.*, **13**, 2802–2814.
- Tsukiyama, T., Becker, P.B. and Wu, C. (1994) ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature*, **367**, 525–532.
- 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.
- Liu, X., Noll, D.M., Lieb, J.D. and Clarke, N.D. (2005) DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome Res.*, **15**, 421–427.
- Gossett, A.J. and Lieb, J.D. (2008) DNA Immunoprecipitation (DIP) for the determination of DNA-Binding specificity. *CSH Protoc.*, **2008**, doi:10.1101/pdb.prot4972.

37. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
38. Buenrostro, J.D., Wu, B., Chang, H.Y. and Greenleaf, W.J. (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.*, **109**, 21–29.
39. 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.
40. Schunter, S., Villa, R., Flynn, V., Heidelberger, J.B., Classen, A.K., Beli, P. and Becker, P.B. (2017) Ubiquitylation of the acetyltransferase MOF in *Drosophila melanogaster*. *PLoS One*, **12**, e0177408.
41. Leibly, D.J., Nguyen, T.N., Kao, L.T., Hewitt, S.N., Barrett, L.K. and Van Voorhis, W.C. (2012) Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. *PLoS One*, **7**, e52482.
42. Patel, A., Hashimoto, H., Zhang, X. and Cheng, X. (2016) Characterization of how DNA modifications affect DNA binding by C2H2 zinc finger proteins. *Methods Enzymol.*, **573**, 387–401.
43. Gilfillan, G.D., Straub, T., de Wit, E., Greil, F., Lamm, R., van Steensel, B. and Becker, P.B. (2006) Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev.*, **20**, 858–870.
44. Rieder, L.E., Koreski, K.P., Boltz, K.A., Kuzu, G., Urban, J.A., Bowman, S.K., Zeidman, A., Jordan, W.T. 3rd, Tolstorukov, M.Y., Marzluff, W.F. *et al.* (2017) Histone locus regulation by the *Drosophila* dosage compensation adaptor protein CLAMP. *Genes Dev.*, **31**, 1494–1508.
45. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
46. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
47. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
48. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
49. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
50. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
51. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
52. Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T. *et al.* (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*, **471**, 480–485.
53. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
54. Chiu, T.P., Comoglio, F., Zhou, T., Yang, L., Paro, R. and Rohs, R. (2016) DNashapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics*, **32**, 1211–1213.
55. Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.*, **9**, e1003118.
56. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
57. Alekseyenko, A.A., Ho, J.W., Peng, S., Gelbart, M., Tolstorukov, M.Y., Plachetka, A., Kharchenko, P.V., Jung, Y.L., Gorchakov, A.A., Larschan, E. *et al.* (2012) Sequence-specific targeting of dosage compensation in *Drosophila* favors an active chromatin context. *PLoS Genet.*, **8**, e1002646.
58. Greenberg, A.J., Yanowitz, J.L. and Schedl, P. (2004) The *Drosophila* GAGA factor is required for dosage compensation in males and for the formation of the Male-Specific-Lethal complex chromatin entry site at 12DE. *Genetics*, **166**, 279–289.
59. Kaye, E.G., Booker, M., Kurland, J.V., Conicella, A.E., Fawzi, N.L., Bulyk, M.L., Tolstorukov, M.Y. and Larschan, E. (2018) Differential occupancy of two GA-Binding proteins promotes targeting of the *Drosophila* dosage compensation complex to the Male X Chromosome. *Cell Rep.*, **22**, 3227–3239.
60. Wang, C.I., Alekseyenko, A.A., LeRoy, G., Elia, A.E., Gorchakov, A.A., Britton, L.M., Elledge, S.J., Kharchenko, P.V., Garcia, B.A. and Kuroda, M.I. (2013) Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nat. Struct. Mol. Biol.*, **20**, 202–209.
61. Feilolter, H.E., Harmon, G.J., Ruddell, C.J. and Beach, D. (1994) Construction of an improved host strain for two hybrid screening. *Nucleic Acids Res.*, **22**, 1502–1503.
62. 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.
63. Morgunova, E. and Taipale, J. (2017) Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.*, **47**, 1–8.
64. Straub, T., Neumann, M.F., Prestel, M., Kremmer, E., Kaether, C., Haass, C. and Becker, P.B. (2005) Stable chromosomal association of MSL2 defines a dosage-compensated nuclear compartment. *Chromosoma*, **114**, 352–364.
65. Lim, C.K. and Kelley, R.L. (2012) Autoregulation of the *Drosophila* Noncoding roX1 RNA Gene. *PLoS Genet.*, **8**, e1002564.
66. 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.
67. Nakahashi, H., Kieffer Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A. *et al.* (2013) A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.*, **3**, 1678–1689.
68. Polach, K.J. and Widom, J. (1996) A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.*, **258**, 800–812.
69. Adams, C.C. and Workman, J.L. (1995) Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell Biol.*, **15**, 1405–1421.
70. Hannon, C.E., Blythe, S.A. and Wieschaus, E.F. (2017) Concentration dependent chromatin states induced by the bicoid morphogen gradient. *Elife*, **6**, e28275.
71. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.