Cell

A Sequence Motif within Chromatin Entry Sites Directs MSL Establishment on the *Drosophila* X Chromosome

Artyom A. Alekseyenko,^{1,2,5} Shouyong Peng,^{1,3,5} Erica Larschan,^{1,2} Andrey A. Gorchakov,^{1,2} Ok-Kyung Lee,¹ Peter Kharchenko,³ Sean D. McGrath,⁴ Charlotte I. Wang,² Elaine R. Mardis,⁴ Peter J. Park,^{1,3,*} and Mitzi I. Kuroda^{1,2,*} ¹Harvard-Partners Center for Genetics and Genomics, Division of Genetics, Department of Medicine, Brigham & Women's Hospital

²Department of Genetics, Harvard Medical School

³Children's Hospital Informatics Program, Children's Hospital

Boston, MA 02115, USA

⁴Washington University School of Medicine, Department of Genetics and Genome Sequencing Center, St. Louis, MO 63108, USA ⁵These authors contributed equally to this work

*Correspondence: peter_park@harvard.edu (P.J.P.), mkuroda@genetics.med.harvard.edu (M.I.K.)

DOI 10.1016/j.cell.2008.06.033

SUMMARY

The Drosophila MSL complex associates with active genes specifically on the male X chromosome to acetylate histone H4 at lysine 16 and increase expression approximately 2-fold. To date, no DNA sequence has been discovered to explain the specificity of MSL binding. We hypothesized that sequence-specific targeting occurs at "chromatin entry sites," but the majority of sites are sequence independent. Here we characterize 150 potential entry sites by ChIP-chip and ChIP-seq and discover a GA-rich MSL recognition element (MRE). The motif is only slightly enriched on the X chromosome (\sim 2fold), but this is doubled when considering its preferential location within or 3' to active genes (>4-fold enrichment). When inserted on an autosome, a newly identified site can direct local MSL spreading to flanking active genes. These results provide strong evidence for both sequence-dependent and -independent steps in MSL targeting of dosage compensation to the male X chromosome.

INTRODUCTION

The occurrence of sequence motifs and their utilization appear to be well correlated in organisms with small genomes, such as bacteria (Wade et al., 2005). However, genomic analyses in eukaryotes have revealed that short sequence-specific DNA-binding motifs alone are insufficient to explain the selective occupancy or specificity of regulatory factor function (Carroll et al., 2005; Sekinger et al., 2005). Therefore, a very important question in transcriptional regulation is how binding factors identify the motifs at which they will function, selecting them from an excess of sites of equal or greater predicted affinity in the genome. An additional layer of complexity is introduced when considering the factors that organize chromatin into active and silent domains and play a crucial role in the fidelity of gene expression in development (Goldberg et al., 2007; Li et al., 2007; Schneider and Grosschedl, 2007).

Dosage compensation provides a model system to study how critical chromatin regulatory decisions are initially targeted to genes and local chromatin domains. This is a process in which X chromosomal genes are differentially regulated to equalize expression in males (XY) and females (XX). In the three model systems studied in depth, *Drosophila*, *C. elegans*, and mammals, this occurs through sex-specific, X-specific targeting of chromatin regulatory complexes (Lucchesi et al., 2005).

The signal for X chromosome targeting is most clearly understood in mammals, where one of the female X chromosomes is inactivated and acquires characteristics of silenced heterochromatin. In X inactivation, expression of the noncoding *Xist* RNA appears to be necessary and sufficient to initiate the process of silencing a chromosome in cis (Lee et al., 1996; Herzing et al., 1997; Wutz and Jaenisch, 2000). *Xist* RNA associates with the chromosome that encodes it (Clemson et al., 1996), and this is linked to the silencing event (Okamoto et al., 2004). Additional X chromosome sequences might play a role as "boosters" for spreading of *Xist* RNA (Lyon, 2003), but autosomes can be at least partially inactivated if expressing *Xist* RNA in early development (Lee and Jaenisch, 1997), suggesting that X-specific booster sequences are not obligatory.

In *C. elegans*, the dosage compensation complex (DCC) downregulates genes on both hermaphrodite X chromosomes by approximately half to achieve dosage compensation (Lucchesi et al., 2005). In this model system, there is evidence for hundreds of autonomous recognition sites along the length of the X chromosomes that contribute to targeting (Csankovszki et al., 2004). Recently, specific DNA sequences have been correlated with this function (McDonel et al., 2006) and found in locations of high DCC occupancy by ChIP-chip (Ercan et al., 2007). Although these sequences are only enriched ~1.25- to 1.5-fold on X, nonrandom clustering may explain how the X chromosome is initially identified (McDonel et al., 2006; Ercan et al., 2007).

A second step of sequence-independent spreading has been postulated to explain the full pattern of X chromosome binding (Csankovszki et al., 2004), but characterized recognition sites have not yet been shown to possess this ability when moved to autosomes.

In Drosophila, the MSL complex upregulates X-linked genes in males by approximately 2-fold to achieve dosage compensation (Lucchesi et al., 2005). No DNA sequence motif uniquely identifies the X chromosome, although X chromosomes throughout the Drosophila genus show enrichment for simple sequence repeats when compared to autosomes (Stenberg et al., 2005; Gallach et al., 2007). Although the MSL-binding pattern on the X chromosome has been mapped at high resolution by ChIPchip, no sequence within the binding sites has been identified to account for the striking X chromosome specificity (Aleksevenko et al., 2006; Gilfillan et al., 2006; Legube et al., 2006). MSL complex binds broadly within the bodies of active X-linked genes, with a 3' bias (Alekseyenko et al., 2006; Gilfillan et al., 2006). Reminiscent of mammals, there are at least two noncoding RNAs that participate in dosage compensation in Drosophila and are encoded by the X chromosome (Amrein and Axel, 1997; Meller et al., 1997). These roX RNAs act like nucleation sites for MSL spreading when ectopically expressed from autosomes, suggesting that they, like Xist, are autonomous targeting elements (Kelley et al., 1999; Park et al., 2002). However, unlike Xist. roX RNAs expressed from other chromosomes can still recognize a roX mutant X chromosome in trans, suggesting that there are additional characteristics that distinguish the X chromosome from the autosomes (Meller et al., 1997; Park et al., 2002).

One model is that the subset of MSL-binding sites that remain in the absence of either the MSL3 chromodomain protein, the MLE helicase, or the MOF acetyltransferase defines the set of core MSL1/MSL2 initiation sites, and that the majority of sites are recognized in a second, sequence-independent step in cis (Kelley et al., 1999; Larschan et al., 2007). Cytologically, there are ~30-70 sites seen on polytene chromosomes in ms/3 mutant males (Palmer et al., 1994; Demakova et al., 2003). However, the existence of this limited set of sites has not been sufficient to explain the apparently autonomous recognition of many X-linked genes in translocation and transposition stocks (Fagegaltier and Baker, 2004; Oh et al., 2004). This has led to an alternative model in which degenerate sequences define all sites on the X chromosome, with a hierarchical recognition mechanism. However, in this case the sequences have been degenerate enough to confound definition (Dahlsveen et al., 2006; Gilfillan et al., 2006, 2007; Kind and Akhtar, 2007).

Here we use chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq) to identify a large set of candidate chromatin entry sites (CES) in order to test the two-step model for MSL binding. These sites are proposed to be recognized in a sequence-dependent manner, and indeed we found a GA- or TC-rich motif common to the majority of sites. Mutagenesis studies reveal the importance of this motif for MSL recognition in a cell culture reporter assay and when placed on autosomes. Interestingly, a newly identified site can direct local MSL spreading to flanking active genes on an autosome, supporting its proposed activity as a CES. These results provide strong evidence for both sequence-dependent and -independent steps in the model for MSL targeting of dosage compensation and provide a general framework for the identification of initiation sites for other chromatin regulators that function broadly over chromatin domains.

RESULTS

MSL3-Independent High-Affinity Sites Represent a Subset of the Wild-Type Binding Pattern and Are Predominantly Located within Genes

We previously proposed that the incomplete MSL-binding pattern seen on polytene chromosomes in the absence of MSL3 might represent specific initiation or chromatin entry sites for MSL complex recognition of the X (Kelley et al., 1999). To test this hypothesis, we combined genetic analysis with ChIP-chip. We chose to analyze msl3 mutant embryos to avoid potential complications from working with dying male larvae. Construction of a genetic stock in which all male embryos were msl3 mutant was accomplished by crossing msl3 mutant males carrying an ms/3⁺ transgene on their X chromosome to ms/3 mutant females. All sons from such a cross are msl3 mutant, while all daughters receive the msl3⁺ transgene. Females, however, do not express MSL2 (Lucchesi et al., 2005), so MSL2 ChIP of the population results in analysis of only the male embryos. We performed two independent biological replicates of the ChIP-chip experiment (Figure 1), calculated mean values for each probe on the microarrays, and then determined significant clusters of binding on the X and the autosomes.

As expected, our positive clusters were highly enriched on the X chromosome. Using the relative enrichment on the X compared to an autosome, we determined that there were at least 150 sites on the X (and 5 sites on 2L, the autosome represented on the array). There are no absolute criteria for setting the parameters that determine the number of sites; however, this was a reasonable compromise between the enrichment on the X and coverage. One hundred and fifty is a higher number of MSL3-independent sites than originally predicted based on immunostaining of msl3 mutant polytene chromosomes. However, we recently demonstrated that banding patterns on polytenes do not provide single site resolution and thus can lead to both underestimation of the number of individual binding sites as well as an inability to distinguish relatively large changes in binding site occupancy (Larschan et al., 2007). We named the sites CES followed by their cytological position, based on the polytene chromosome map (e.g., CES11D1) (Table S1 available online). Among the 150 sites, we recovered roX1, roX2, and 18D11, previously identified as MSL3-independent sites in our lab (Kelley et al., 1999; Oh et al., 2004), as well as DBF3 (CES3A8), DBF7 (CES10F2), DBF9 (CES18D3), and DBF12 (CES1B14), previously characterized as high-affinity sites by P. Becker and colleagues (Dahlsveen et al., 2006).

Next we asked whether these 150 candidate chromatin entry sites were a subset of wild-type MSL targets. On average, the clusters observed in our CES ChIP-chip experiments were approximately 1.5 kb in size, and virtually all (98%) were contained within the full set of sites in wild-type. Consistent with this observation, we found that CES overlap with genes (135 of 150).



Figure 1. MSL3-Independent Chromatin Entry Sites Are a Subset of the Wild-Type Binding Pattern for MSL Complex and Coincide with the Strongest Enrichment Peaks Detected by ChIP-seq

Two representative chromatin entry sites, CES11D1 (A) and CES15A8 (B), are shown. ChIP-chip profiles were generated from *y w; MSL3-TAP; msl3* embryonic chromatin (WT) using IgG to IP the TAP epitope, or *msl3* mutant embryos (CES) using anti-MSL2 antibodies. DNA resulting from ChIP was hybridized to custom NimbleGen tiling arrays (Alekseyenko et al., 2006). The y axis shows the log₂ ratio of IP/Input signal. The ChIP-seq tag profile (Solexa) was obtained from an MSL3-TAP transformed male cell line, Clone 8, using IgG to IP the TAP epitope. The ChIP-seq profile displays broad distribution along WT MSL targets and high peaks that correspond to entry sites. The y axis shows the tag density. Gray lines within ChIP-chip and ChIP-seq panels indicate the regions identified as bound clusters (See Experimental Procedures for details). Genes are color-coded based on their transcriptional status (transcribed, red; nontranscribed, black; genes that are differentially transcribed in S2 and Clone 8 cells, salmon; and genes without transcriptional data, gray). Genes on the top row are transcribed left to right, and genes on the bottom row are transcribed from right to left. Numbers along the x axis refer to chromosomal position (bp) (Dm1 release coordinates). Polytene map cytological locations are indicated below.

ChIP-seq of the Wild-Type MSL Pattern Detects Strong Peaks of Binding within CES Candidates

Concurrent with the ChIP-chip analysis of msl3 mutants, we performed Solexa sequencing of MSL3-TAP ChIP from male tissue culture cells. We were testing whether this nonarray and nonhybridization-based approach would yield similar or higher-resolution results for wild-type MSL3 binding. When we profiled wildtype MSL3-TAP in Clone 8 cells, we were able to observe the broad binding patterns covering active genes on the X previously seen in ChIP-chip analysis (Alekseyenko et al., 2006), with 91% of the positive genes in ChIP-seq overlapping with 89% of bound genes in our previous ChIP-chip data set. Using ChIP-seq we found 619 binding sites on X and only 2 on 2L (our negative control), demonstrating the high quality of the experiment. Notably, we also discovered a set of sharp peaks extending far above the broadly bound regions, which were not evident in ChIP-chip analysis. These peaks are of such striking amplitude that when plotted in Figure 1, the broad and significant signal of overall binding surrounding them is not readily evident at this scale (see gray bars that indicate significant signal). Furthermore, there was a remarkable coincidence of the ChIP-seq peaks from MSL3-TAP in wild-type with our newly identified chromatin entry sites found genetically, with 91% inclusion of the 150 entry sites in the top 309 Solexa peaks (see Figure S2D and Experimental Procedures for details). The set of top peaks covering the entry sites also were distinguished by their magnitude (Figure S2E), therefore we used the top 309 peaks for further analysis.

To determine whether or not a bias for recovery of these fragments was intrinsic to the general chromatin composition of those regions rather than being a consequence of the specific ChIP experiments, we sequenced sonicated and amplified input DNA. We found that any intrinsic biases in the input could not account for the sharp peaks observed after MSL pulldown and recovery (data not shown).

To determine whether the peaks were somehow enhanced by the ChIP-seq technique, or quenched in ChIP-chip analysis, we assayed MSL3-TAP chromatin IP samples directly by QPCR for regions encompassing the ChIP-seq peaks. We found an excellent correlation between fold differences seen by QPCR and those measured by ChIP-seq around 3 of 3 peaks analyzed (Figure S1). We conclude that the key difference between the ChIP-seq and ChIP-chip platforms appears to be in the dynamic range that can be measured. Several studies have reported a general agreement between the two platforms (Euskirchen et al., 2007; Mikkelsen et al., 2007; Boyle et al., 2008), but those comparisons were mostly based on visual inspection rather than a statistical comparison. Also, if the methods are compared for transcription factor binding, which typically occurs in sharp peaks, the differences would not be so apparent. However, the MSL complex binds guite broadly over transcription units, and before the ChIP-seq analysis, it was less evident that there



Figure 2. A GA-Rich Motif Is a Common Feature in MSL Chromatin Entry Sites

(A) Identification of the MRE motif enriched in CES. The motif logo is shown with representative examples below. This motif occurs once in CES11D1 and three times in CES5C2. The GA-rich core is highlighted in red. We are arbitrarily showing the GA- rather than the TC-rich strand. This motif was identified using the chromatin entry sites as determined by the ChIP-chip data; a nearly identical motif is obtained from the WT Solexa data (Figure S2).

(B) Solexa peaks center on the MRE motif in chromatin entry sites. Shown is a heatmap of Solexa profiles (tag densities increase from red to white), aligned by the MRE motif in each of the 137 individual CES that contain the motif. In nearly all cases, the location of the sharp peaks coincides with the motif location.

(C) The MRE motif tends to localize to the center of strong Solexa peaks. The top 500 Solexa peaks were ranked by height on the x axis and then MRE motifs within a 2 kb window flanking each peak were mapped on the y axis. Some peaks contain multiple motifs; to avoid showing the same region twice, only the strongest peaks in any 3 kb region were considered. Most motifs map near the center, but motif centering decreases as peak intensity weakens.

(D) The MRE motif is enriched on the X chromosome, compared to autosomes. Motif frequency for each chromosome was normalized to its frequency on all autosomes. Colors correspond to different thresholds for defining the occurrence of the motif. At all thresholds, the motif is enriched on the X. As the stringency increases, the enrichment increases but a smaller number of motifs are identified.

were strong quantitative differences within the broad binding profiles. For MSL complex binding, ChIP-seq clearly reveals a subpattern of high-affinity sites within the wild-type binding pattern. The strong coincidence of the peaks of wild-type MSL3-TAP binding detected by ChIP-seq with the subset of sites of MSL2 binding in *msl3* mutants provides strong evidence that this subset of MSL-binding sites may have a significant biological role.

Discovery of a Sequence Motif in Entry Sites and ChIPseq Peaks that Is Enriched on the X Chromosome

Encouraged by the strong correlation between ChIP-seq and ChIP-chip results, we used MEME (Bailey et al., 2006), a robust searching algorithm, to analyze the 150 candidate entry sites and the top 309 ChIP-seq peaks for a sequence motif. We found the same highly significant 21 bp sequence centered around a core GAGA (or TCTC) repeat using both data sets (Figures 2A and S2). This sequence is compatible with the conserved MSL-binding motif we previously identified by direct comparison of the *roX1*

the high-affinity sequences characterized by Gilfillan et al. (2007). Other motif searching algorithms including WEEDER, and MDscan (Liu et al., 2002; Pavesi et al., 2004), gave a very similar but shorter top motif. We note that while the chromatin entry sites form a subset of the wild-type sites, previous attempts to recover a motif from wild-type data failed because of the excess of extraneous DNA sequences included in those regions. For convenience we will refer to our newly discovered motif as the "MSL recognition element" (MRE). Interestingly, the MRE includes the GAGA factor (GAF) consensus sequence GAGAGA, but to date significant coincidence of MSL and GAF binding has not been evident (Sun et al., 2003; van Steensel et al., 2003). Similarly, we wondered whether a CT-rich logo might indicate involvement of CTCF, named initially for its affinity for CT-rich sequences (Lobanenkov et al., 1990). However, the logo discovered for Drosophila CTCF from ChIP-chip studies of the Adh and Bx-C regions bears no resemblance to our GA motif (Holohan et al., 2007).

and roX2 genes from multiple species (Park et al., 2003) and



Figure 3. Test for CES Function in a Luciferase Reporter Assay

(A) Experimental design and structure of the constructs used for a dual Firefly/ Renilla luciferase reporter assay. S2 cells cotransfected with the Firefly luciferase construct containing each 1.5 kb CES fragment and the control Renilla luciferase plasmid were treated with dsRNA against either GFP (control) or MSL2 (experimental).

(B) Examples of Firefly/Renilla luciferase activity conferred by nine chromatin entry sites normalized to the value obtained after *ms/2* RNAi. CES18D11 is a rare example of a CES that failed to show MSL2-dependent upregulation in this assay. Error bars represent standard deviations based on four replicates.

In Figure 2B, we display a heatmap of the ChIP-seq profiles around the chromatin entry sites containing MREs. Strikingly, in nearly all cases the sharp wild-type ChIP-seq peaks are centered on the MRE within the CES, with a rapid decrease away from the motif. To determine the relationship between the peaks and the occurrence of MREs, we ordered the ChIP-seq peaks by their height and plotted all the motifs (p value < 10^{-5}) in the neighborhood (1 kb in each direction) (Figure 2C). Clearly, MREs are concentrated at the center of the strongest peaks; they are more dispersed as the peaks get weaker. If we are to base the likelihood of a peak being a CES on the clustering of MREs within the peak, the estimate of 150 chromatin entry sites appears to be somewhat conservative. But beyond ~300, the alignment degenerates, suggesting that weaker peaks are not dependent on MREs.

A motif utilized for X-specific MSL binding on the X ideally should be enriched on X. Therefore, we surveyed the *D. melanogaster* genome for the occurrence of potential MREs on X and autosomes. We found that the motif was approximately 2fold enriched on X versus autosomes for a wide range of thresholds for defining a sequence match (see Figure 2D). At a p value of 10^{-5} , there are 3652 and 8770 motifs on the X and autosomes, respectively, for an enrichment of 1.8-fold in terms of motif density (167.7 copies/Mb on X versus 92.3 copies/Mb on autosomes). At this stringency, 137 of the 150 candidate entry sites contain an MRE (91%), with 51 having 2 or more MREs within 500 bp (34%).

The genome sequences of twelve *Drosophila* species became available recently, and it has been of considerable interest to determine how appropriate dosage compensation evolved following sex chromosome translocation events in the various species. We found that the MRE discovered in *D. melanogaster* was similarly enriched on the X chromosomes of other sequenced species. Figure S2 shows results for two species (*D. simulans* and *D. yakuba*).

We define chromatin entry sites functionally, as the binding sites for partial MSL complex in the absence of MSL3. In contrast, the MRE is a consensus sequence that is important within CES but clearly is found elsewhere in the genome. The simple presence of an MRE is not sufficient to predict MSL binding, as autosomes carry thousands of matches that are not bound. In addition, among 3652 predicted MREs on X, only ~6% (224/3652) are located within the 150 selected entry sites. At this threshold, ~91% of the entry sites have motifs, but many motifs are ignored. Therefore, we continued our analysis on two fronts, first asking whether the MRE sequence is implicated functionally in MSL binding and second, identifying additional parameters that would better define the motifs that are actually chosen on X.

Functional Analysis of the MRE within Candidate Entry Sites with a Luciferase Reporter Assay

To determine whether the MRE is important for MSL recognition at the newly identified sites, and whether it can function when removed from its normal chromosomal context, we employed a luciferase-based reporter assay in male S2 tissue culture cells. The luciferase assay (Figure 3) is based on the observation that MSL binding downstream of the major roX2 transcription unit positively regulates expression of roX2 (Bai et al., 2004). The MSL-binding site in roX2 displays DNase hypersensitivity and thus has been called the roX2 DHS (Kageyama et al., 2001). We constructed a plasmid containing the roX2 promoter region and the roX2 DHS, with the intervening roX2 transcription unit replaced by the Firefly luciferase gene (Figure 3A). Upon transfection into S2 cells, the activity of the Firefly luciferase is highly dependent on the presence of MSL complex, as activity is depressed ~7-fold by RNAi treatment for MSL2 (Figure 3B). The DHS region is required to confer MSL dependence to the reporter gene. To provide accurate quantitation of reporter gene activity, we used a dual luciferase system in which normalization was conducted using a control plasmid that contains the Renilla luciferase gene fused to an MSL-independent promoter. The Firefly and Renilla luciferase reporter genes are quantitatively assayed using independent substrates.

We found that 1.5 kb wild-type DNA segments containing candidate entry sites gave a clear response in the reporter assay (Figure 3 and Table S2). We tested 29 of the 150 candidate MSL-binding sites in the luciferase assay, and 21 of these responded positively (Table S2 and data not shown). To exclude *roX2* promoter-specific effects, we cloned three different autosomal promoters upstream of the luciferase gene. Derivatives of these constructs carrying various CES also demonstrated dependency on MSL2, similar to the activity observed with the *roX2* promoter. Interestingly, the effect inversely correlated with the

Α			Luciferase	ChIP
	CES11D1		3.2 +/- 0.1	17.4 +/-2.1
	CES11D1-2		3.3 +/- 0.2	18.2 +/-0.8
	CES11D1-3		1.0 +/- 0.1	1.3 +/-0.1
	CES11D1-M2R		1.6 +/- 0.1	1.4 +/-0.1
	CES11D1-M3R		1.2 +/- 0.1	1.8 +/-0.7
	CES11D1-150	1 initial and the second se	1.1 +/- 0.1	5.2 +/-0.7
	CES11D1-scr150		1.1 +/- 0.1	5.7 +/-0.4
в				
5	CES5C2		7.1 +/- 2.0	n/d
	CES5C2-1		4.3 +/- 0.6	n/d
	CES5C2-3	i i i i i i i i i i i i i i i i i i i	5.0 +/- 0.6	n/d
	CES5C2-4		1.1 +/- 0.3	1.4 +/-0.2
	CES5C2-6		5.2 +/- 0.3	28.3 +/-4.3
	CES5C2-mut1		2.1 +/- 0.1	9.1 +/- 1.3
	CES5C2-mut2	II	1.1 +/- 0.1	21.2 +/- 3.0
	CES5C2-mut3		1.3 +/- 0.1	3.2 +/-0.5
	CES5C2-2mut1	min	0.93 +/- 0.1	1.4 +/-0.1
	CES5C2-2mut2	min -	1.0 +/- 0.1	1.1 +/-0.2
	CES5C2-3mut	шiш	1.1 +/- 0.1	1.4 +/-0.1
	CES5C2-150	, min	5.3+/- 0.1	26.6 +/-5.2
	CES5C2-scrArt	min	1.02 +/- 0.1	3.1 +/-0.4
	CES5C2-scrSY		3.6 +/- 0.1	57 +/- 3.2

absolute promoter strength (Table S2). Next we performed deletion and mutational analysis to test whether or not key residues in the MRE were necessary for the observed MSL-dependent transcriptional response.

We initially narrowed down the 1.5 kb positive fragments for five putative entry sites to 500 bp, and in four cases the positive activity tracked unambiguously with the mapped MREs (Figures 4 and S3 and Table S2). In subsequent deletion analyses, we were able to identify a 150 bp fragment of CES5C2 with three MREs that retained function, but a 150 bp subclone of CES11D1, with one MRE, did not (Figures 4 and S3). We created point mutants in the CES5C2 (150 bp) and CES11D1 (500 bp) fragments to examine the requirement for core residues (Figures 4 and S3). We found that the MREs were absolutely required for reporter gene activity in all cases and combinations. In order to further ask whether the MREs were not only necessary but might also be sufficient in this context, we retained the motif location but randomized the rest of the surrounding sequence, keeping the GC content constant. Surprisingly, we found that for CES5C2, such randomized fragments were still functional, demonstrating that the central sequence motif plays the primary role in MSL recognition and function. These results (Figure 4) provide strong evidence for the importance of key residues in the MRE.

Transgenic Entry Sites Attract MSL Complex to Autosomes and Are MSL3 Independent

Based on the results of the sequence motif analysis and the functional luciferase assays, we proceeded to in vivo cytological and ChIP analyses of five candidates: CES11D1 (one motif), CES1C4 (one motif), CES9A3 (two motifs), CES5C2 (three motifs), and CES18D11 (one motif). Of these, all were positive in the luciferase assay except CES18D11, which was chosen for further anal-

Figure 4. The MRE Motif Is Required for MSL Complex **Binding to CES**

Structure of the DNA fragments tested in the context of CES11D1 (A) and CES5C2 (B) and the activities they display in the luciferase assay in S2 cells and when inserted in a specific location (37B7) on chromosome 2L in transgenic flies (MSL2 ChIP). Numerical values in red were scored positive, black were negative, and blue were considered intermediate. Positions of the motifs are shown as red boxes; when mutated, MRE motifs are shown as non-red boxes (please refer to Figure S3 for sequence changes within the motif). Randomized DNA sequences in the shortest subclones not affecting the MRE motifs are shown in purple or dark blue (with only the motif core remaining intact). Luciferase and ChIP data are based on at least three independent experiments. n/d. not determined.

ysis because of previous cytological and ChIP data supporting its MSL-binding capacity (Oh et al., 2004). We inserted these candidate chromatin entry sites and their subclones into two different genomic positions on chromosome 2 by site-specific recombination. Following replacement of a mini-white gene cassette, we determined that each insertion had occurred at the correct position in the genome by PCR. We inserted the roX1 DHS as a positive control and a GFP gene as a negative control.

We assayed our initial transgenic insertion lines at the level of polytene chromosome staining. We asked whether or not a new MSL3-independent binding site was created in polytene chromosomes by immunostaining for MSL2 in an ms/3 mutant background. For this, we used females ectopically expressing MSL2 but homozygous mutant for ms/3. The candidates with MREs demonstrated positive signals in all cases (Figures 5, S4, and S5). On average, the CES candidates integrated at 37B7 displayed better ability to recruit MSL complex than when inserted in 53F8 (Table S3). This result confirmed that the isolated sites were MSL3 independent and thus fit our criteria for their initial identification. Our immunostaining results were generally in agreement with the luciferase data. Interestingly though, the CES18D clone that failed to upregulate the luciferase reporter displayed consistent binding in our immunostaining assay (Table S3 and Figure S5), with improved detection compared to our previous work (Oh et al., 2004).

The MRE Motif Is Necessary for MSL Complex Binding to Entry Sites In Vivo

Does MSL binding within entry sites always follow the MRE? To address this question we used directed ChIP analysis as a highresolution alternative to conventional cytological analysis. We used a mixed population of embryos in which all individuals carried one copy of the transgene insertion, but only 50% (the males) would be expected to form the MSL complex. We designed primers specific for the hybrid attL flanking sites common to all insertions to create a ChIP assay in which individual strains could be treated identically following chromatin isolation. Following crosslinking, IP with anti-MSL2, isolation of DNA, and analysis by QPCR, we found that 0.5-1.5 kb segments of CES5C2 and CES11D1 were positive in all assays and could



Figure 5. MSL Complex Is Recruited to Entry Sites Placed on Autosomes

Polytene chromosomes from homozygous *CES-X{37B7}*; *H83M2-6l, msl3* females were immunostained with anti-MSL1 antibodies. (A) *roX1* DHS (300 bp). (B) CES11D1 (1.5 kb). (C) CES9A3 (1.5 kb). (D) CES5C2 (1.5 kb). (E) CES5C2-6 (268 bp, containing three motifs). (F) CES5C2-4 (268 bp, lacking a motif). White arrows point to the positive signals at the integration sites, and a green arrow points to the cytological location with no visible staining. The X chromosome in each nucleus displays the partial pattern typical of *msl3* mutants.

be subdivided into positive and negative subclones (e.g., CES5C2-6 versus CES5C2-4) based on the presence or the absence of the MRE (Figure S3). Indeed, all the subfragments encompassing MREs recruited MSL complex to the autosomal integration site and were bound by MSL2, as assayed by ChIP (Figure 4).

Next we tested the same sequences and mutations for CES11D1 and CES5C2 that were tested in the luciferase assay. These fragments contained one (CES11D1) or three (CES5C2) MREs. Three-nucleotide substitution in the core of the single



Figure 6. Histone H3 Is Depleted over MREs in Chromatin Entry Sites Average H3 profile centered over the best MRE within 137 CES (black) shows clear depletion. In contrast, the best (green) or a random set (blue) of 137 MRE motifs outside CES display a relatively flat profile. The histone occupancy data are from S2 and Clone 8 cells (Larschan et al., 2007).

motif in the positive subclone CES11D1-2 completely abolished MSL binding. When we decreased the subclone size to 150 bp (CES11D1-150) and the motif was present, we observed reduced but significant binding. This demonstrates that the ChIP assay was more sensitive than the luciferase assay, possibly because it is directly measuring binding rather than asking for transcriptional upregulation as a consequence of binding. In this minimized system, we randomized the DNA sequence, keeping the GC content and position of the MRE constant. Surprisingly, this new altered sequence (CES11D1-scr150) displayed pronounced binding similar to the original DNA, indicating that the MRE is sufficient for MSL recruitment at this ectopic location, without multimerization (Figure 4). Additional confirmation of the importance of the MRE came from subfragments of CES5C2, which carries three motifs. First we found that mutation of individual MREs reduced binding to varying degrees (Figure 4). Second, simultaneous mutation of two MREs completely abolished recruitment of the complex. Interestingly, the smallest fragment tested (CES5C2-150) carrying all three intact MREs displayed binding very similar to larger subclones. We subsequently asked whether it is the whole motif (21 bp) or just its central part (15 bp) that is needed for complex binding. To answer this question we created two different randomized sequences. In one case, scrambling affected all sequences except the three 25 bp segments containing the motifs (CES5C2-scrSY), whereas in CES5C2-scrArt, only the 15 bp core sequences in these motifs remained unaltered. Spacing between the motifs and the GC content were kept constant. We detected weak residual binding in CES5C2-scrArt, but CES5C2-scrSY showed binding that exceeded the original fragment. These results suggest that integrity of the full MREs in this entry site are essential for binding (Figure 4) with no evidence for a requirement for local flanking sequences.

If MREs are both necessary and sufficient for binding in the two cases tested, why does the MSL complex bind some MRE-containing sites but skip many others? This is not a novel problem in genomics, as predicted sites for sequence-specific binding factors often greatly exceed the number of sites actually utilized. Increasing the stringency with which we call a motif a match did not improve our ability to predict entry sites on the



X chromosome (Figure S2). These results suggest that one or more additional characteristics guide the MSL complex to specific sites on the X. Therefore, we searched for secondary sequence motifs that might contribute to the specificity of the MRE. However, all additional motifs lacked predictive value, as they were no more enriched in CES than they were in the entire X chromosome (data not shown).

A logical aspect of recognition could be relative accessibility of the MRE within its chromatin environment. For example, functional MREs might be located in a particular position with respect to nucleosomes or reside in relatively depleted regions. Therefore, we analyzed histone H3 occupancy over the entry sites. We determined the average H3 signal after aligning the 150 entry sites by their best motifs and found a clear H3 depletion centered over these sites (Figure 6 and data from Larschan et al., 2007). This averaged profile was not observed over the remaining, unselected motifs on X (Figure 6). These results suggest that nucleosome depletion might be a key defining characteristic of functional chromatin entry sites.

We further examined the enrichment of MREs within active genes. Requiring MREs to be coincident with H3K36me3, a chromatin mark in the bodies of active genes, reduces the number of potential sites from 3652 to 449 on X and from 1977 to 105 on 2L. The enrichment on the X compared to 2L therefore increases

Figure 7. MSL Complex Spreads into Neighboring Genes from CES5C2 Inserted on an Autosome

(A) MSL3-TAP ChIP in the region surrounding the 37B7 integration site where either GFP (1.9 kb), CES5C2-4 (258 bp, no MREs), or CES5C2-6 (268 bp, three MREs) are placed (*y w; CES-X[37B7]/ MSL3-TAP; H83M2-6I, msl3* mixed sex embryos). Top panel: genetic map and location of probes; Dm3 release coordinates are shown. Bottom panel: MSL3-TAP data for GFP (green), CES5C2-4 (blue), and CES5C2-6 (red) indicate that only CES5C2-6 effectively targets the MSL complex and directs limited spreading to neighboring genes. Error bars represent standard deviations based on three replicates.

(B) Two-step model for MSL targeting of the X chromosome. X, X chromosome, A, autosome. First step: MSL complex targets >150 chromatin entry sites containing MRE motifs on the X chromosome. The autosome is normally ignored, unless a CES from the X is inserted on the autosome. Second step: Local spreading from entry sites leads to MSL binding to the majority of active genes on the X chromosome. In addition, an ectopic CES can lead to targeting of flanking active genes on an autosome, as seen in the results displayed in (A).

from 1.8 to 4.3 when H3K36me3 is considered. Furthermore, the proportion of motifs on the X that fall within our data set of 150 CES is \sim 6% overall (224/ 3652), but when only the motifs in the H3K36me3positive regions are considered, the proportion of motifs that belong to CES increases to \sim 30% (133/449). This is encouraging, as we believe that 150 is likely to be a conservative estimate for the actual number of functional CES. Furthermore, 122 of the 150 candidate CES are retained under these criteria, strongly suggesting that additional

identifying characteristics correlate with active gene expression. These results suggest that chromatin context could play a key role in selection of potential sites, and that our predictive power will improve as we learn more about the rules that govern the organization of complex genomes.

High-Affinity Sites Inserted on Autosomes Direct MSL Spreading to Flanking Expressed Genes

The two-step model for MSL binding to the X chromosome in Drosophila males proposes that after an initial, sequence-specific recognition step, the complex spreads in cis to active genes to achieve its complete pattern. We found that candidate entry sites inserted on chromosome 2L did not show spreading at the level of polytene chromosome staining. However, we recently observed that polytene chromosome analysis cannot distinguish whether or not individual genes within clusters are bound or not. Therefore, to ask whether entry sites might cause MSL binding to occur in flanking autosomal chromatin, we performed ChIP analysis of the local area surrounding the attP insertion site in transgenics carrying the GFP, CES5C2-4, and CES5C2-6 insertions. We probed the region up to 20 kb distant from the insertion site, testing locations in flanking active and inactive genes. We found that GFP and CES5C2-4 (lacking MREs) showed no MSL binding above background at any of the flanking regions tested, but that CES5C2-6 (containing three MREs) showed significant binding at specific locations (Figures 7A and S6). Interestingly, these locations correlated with the 3' ends of active genes. A signal at flanking regions could be due to IP of larger than expected chromatin fragments. However, this was clearly not the case when considering the *Nak* gene. Probes near the 5' end of *Nak* gave no significant signal, even though they were located between the transgene insertion site and the strongly bound 3' end of the active gene. Together, these results demonstrate that entry sites removed from the X chromosome have the ability to cause MSL binding in flanking chromatin on autosomes. These results strongly support both sequence-dependent and sequence-independent steps in MSL binding.

DISCUSSION

We previously proposed that MSL targeting of the X chromosome occurs in at least two steps (Figure 7B). In the first step, *roX* genes and other chromatin entry sites attract the complex to the X chromosome (Kelley et al., 1999). In the second step, MSL complex spreads from entry sites to genes in flanking chromatin by recognizing chromatin marks indicative of active transcription (Larschan et al., 2007). The mechanism governing the first step in MSL recognition of the X chromosome had remained elusive. The isolation of a large set of potential chromatin entry sites by ChIP-chip and ChIP-seq has allowed us to identify common features of these sites.

Our studies have revealed that they are enriched for a GA (or TC) repeat sequence, and that mutagenesis of this sequence leads to loss of autonomous MSL binding to individual sites. Our findings have several important implications. First, we suggest that there are more MSL entry sites than originally proposed from the analysis of MSL binding to polytene chromosomes in ms/3 mutants. Based on ChIP-chip of ms/3 mutant embryos using anti-MSL2, there are at least 150 sites. Furthermore, the striking correlation of these sites with strong peaks of wild-type MSL binding using ChIP-seq suggests that the number could approach 300. This is important, as previous studies that appeared to refute the entry site model found apparently autonomous binding of X chromosome genes lacking mapped sites (Fagegaltier and Baker, 2004; Oh et al., 2004). The existence of many more entry sites than were detected at the cytological level could, perhaps, explain this finding.

Interestingly, the X chromosome and autosomes of *D. mel-anogaster* and six other species were compared recently for the prevalence of all sequence words up to 13 nt long (Gallach et al., 2007). From this work it is clear that X chromosomes have evolved to be enriched for certain types of simple sequences including GA-rich repeats, which we implicate here in dosage compensation. However, additional parameters must be used to help restrict the recognition of many GA repeats, such as nucleosome depletion to leave these sites relatively open for binding, and colocalization within active gene clusters. We provide correlative evidence for these parameters and eagerly await additional insights as the *Drosophila* genome becomes increasingly tractable for analysis at the level of local physical properties and higher-order interactions. Importantly, the question of recognition site specificity of transcription factors in general faces a similar, as yet unsolved problem in metazoans. Of the many sites predicted to be favored by individual factors, how is it that only a small fraction of these are actually occupied? The fidelity of sequence motifs itself is clearly not sufficient to answer this question, and organization relative to genes, other binding factors, positioned nucleosomes, and three-dimensional (3D) location within the nucleus are all factors under investigation.

A second implication of our finding that chromatin entry sites are more numerous than previously expected is that spreading of MSL complex on the X chromosome may normally be quite modest. In particular, it is an appealing idea that entry sites are identified in part because of their location within expressed gene clusters and may not usually need to spread beyond such clusters (~10-30 kb). This would explain why spreading from candidate entry sites is rarely seen at the cytological level and requires ChIP for detection. Furthermore, this could explain previous cytological data, in which MSL spreading does not appear to extend into autosomes in X:A translocations (Fagegaltier and Baker, 2004; Oh et al., 2004). Alternatively, it is possible that spreading is normally quite robust on X but relies on the close proximity and 3D organization of other entry sites and roX genes. These conditions cannot be duplicated when individual chromatin entry sites are placed in ectopic locations for analysis. The next big step in addressing these possibilities will be to transpose large constructs, containing entry sites in combination with roX transgenes, to the autosomes.

Finally, our results suggest that the use of ChIP-seq technology may greatly enhance the ability to identify potential initiation sites for other chromatin-organizing complexes that act broadly on large genomes. The dynamic range measurable by ChIP-seq could prove to be a powerful tool for deciphering the currently unknown rules for higher-order organization of complex genomes.

EXPERIMENTAL PROCEDURES

Plasmids and Cloning

A 359 bp fragment from the *roX2* gene promoter (Bai et al., 2004) was cloned into the pGL3-Basic vector (Promega) to obtain pGL3-roX2prom. Predicted 1.5 kb entry site sequences, centered around the hybridization peaks on the microarray, were cloned in both orientations into pGL3-roX2prom. Three additional derivatives of pGL3-Basic were obtained by placing promoter regions from three autosomal genes: CG2937 (367 bp), CG2960 (668 bp), and CG7424 (833 bp) upstream of the luciferase gene. DNA fragments with multiple mutations were obtained by custom gene synthesis (IDT). CES fragments from pGL3-roX2prom-CES-X constructs were inserted in piB-GFP (Bateman et al., 2006) by replacing the hs70-GFP cassette.

Fly Genetics and Transgenesis

To obtain *msl3* mutant embryos for ChIP-chip experiments, *y w*; *msl3* females were crossed to *y w* $P\{w^+, msl3^+, cos8-1\}/Y$; *msl3* males (a gift of H. Oh). Females do not express MSL2; therefore only *msl3* mutant male embryos contributed to the anti-MSL2 ChIP. To generate transgenic flies with a series of sequences placed within the same genetic environment, we used the recombination-mediated cassette exchange approach (Bateman et al., 2006) and the source of phiC31 integrase described in Bischof et al. (2007) (see further details in the Supplemental Experimental Procedures).

To assay for MSL targeting to entry sites in an autosomal context, we used the *H83M2-6l, msl3* background where MSL2 is constitutively expressed from the *hs83* promoter (Kelley et al., 1995; Lyman et al., 1997). *CES-X; H83M2-6l, msl3* females were crossed to *MSL3-TAP; msl3* males (Alekseyenko et al., 2006). Analysis of dosage compensation targeting in embryos under conditions of wild-type, male-specific MSL2 expression was performed using chromatin isolated from mixed sex embryos of the following genotype: *y w; CES5C2-6/ MSL3-TAP; msl3*.

Dual Luciferase Reporter Assay

Luciferase reporter assays were conducted by cotransfecting SL2 cells with dsRNA and Firefly and Renilla luciferase plasmids in a 96-well plate (Dual glo, Promega). Calculations were performed by normalization of Firefly luciferase units to Renilla luciferase units and determination of standard deviation among four replicates (see further details in the Supplemental Experimental Procedures).

Chromatin IP

Embryo ChIP to Initially Identify CES

msl3 mutant embryos (1 g) were collected for chromatin preparation as described in Alekseyenko et al. (2006). Immunoprecipitation was performed with anti-MSL2 antibody, as in Larschan et al. (2007). Details on the smallscale and large-scale ChIP are provided in the Supplemental Experimental Procedures.

ChIP-chip and ChIP-seq Data Analysis

Immunoprecipitated DNA was isolated, amplified, and hybridized to Nimble-Gen tiling arrays as in Alekseyenko et al. (2006). We performed two replicates for each type of ChIP-chip experiment. For each array, we identified and corrected dye-bias for the log-ratios between ChIP and input using a normalization method that has been modified for ChIP-chip (for more details see the Supplemental Experimental Procedures). For ChIP-seq, the sequenced fragments were aligned against dm1 genome assembly. Tags with a unique alignment were accepted if the length was over 20 bp without mismatches, over 26 bp with one mismatch, over 29 bp with two mismatches, and 30 bp with three mismatches. This resulted in 2.9M mapped tags for the ChIP run and 1.1M mapped tags for the input run. Further details can be found in the Supplemental Experimental Procedures.

ChIP DNA Sample Prep for Solexa/Illumina

Immunoprecipitated DNA from Clone 8 cells was isolated as in Alekseyenko et al. (2006). DNA amplification was performed using the Illumina Sample Prep Kit with minor modifications. One microliter adaptor oligo mix was used per 100 ng ChIP DNA. The ligation mix was first amplified as detailed in the protocol and then the PCR products ($3-8 \mu g$) were gel-fractionated. Purified DNA was run on the Solexa Genome Analyzer. A detailed protocol for Solexa sample preparation is available in the Supplemental Experimental Procedures.

Immunocytochemistry

Polytene chromosomes were isolated from female *y w; CES-X; H83M2-6I, ms/3* larvae and processed for immunostaining as described in Larschan et al. (2007).

Real-Time PCR

Quantitation of immunoprecipitated DNA was performed using real-time PCR as described in Larschan et al. (2007). Primer sequences used are in Table S4.

ACCESSION NUMBERS

ChIP-chip and ChIP-seq data in this manuscript have been submitted to NCBI GEO public repository with the accession number GSE11485.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, six figures, and four tables and can be found with this article online at http://www.cell.com/cgi/content/full/134/4/599/DC1/.

ACKNOWLEDGMENTS

Many thanks to M. Gelbart, T. Sural, P. Yang, and M. Tolstorukov for critical comments on the manuscript; H. Oh, J. Bateman, T. Wu, J. Bischof, F. Karch, and K. Basler for key fly stocks; S. Elgin for important discussions; Y. Schwartz and T. Kahn for sharing protocols; and J. Racine and A. Sarovschii for technical assistance. This work was supported by the National Institutes of Health: GM45744 to M.I.K., GM67825 to P.J.P., and HG003079 to E.R.M. (co-PI).

Received: March 29, 2008 Revised: May 22, 2008 Accepted: June 13, 2008 Published: August 21, 2008

REFERENCES

Alekseyenko, A.A., Larschan, E., Lai, W.R., Park, P.J., and Kuroda, M.I. (2006). High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. Genes Dev. 20, 848–857.

Amrein, H., and Axel, R. (1997). Genes expressed in neurons of adult male *Drosophila*. Cell 88, 459–469.

Bai, X., Alekseyenko, A.A., and Kuroda, M.I. (2004). Sequence-specific targeting of MSL complex regulates transcription of the *roX* RNA genes. EMBO J. *23*, 2853–2861.

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. *34*, W369–W373.

Bateman, J.R., Lee, A.M., and Wu, C.T. (2006). Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. Genetics *173*, 769–777.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. USA *104*, 3312–3317.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33–43.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275.

Csankovszki, G., McDonel, P., and Meyer, B.J. (2004). Recruitment and spreading of the C. elegans dosage compensation complex along X chromosomes. Science *303*, 1182–1185.

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. 10.1371/journal.pgen.0020005.

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

Ercan, S., Giresi, P.G., Whittle, C.M., Zhang, X., Green, R.D., and Lieb, J.D. (2007). X chromosome repression by localization of the C. elegans dosage compensation machinery to sites of transcription initiation. Nat. Genet. *39*, 403–408.

Euskirchen, G.M., Rozowsky, J.S., Wei, C.L., Lee, W.H., Zhang, Z.D., Hartman, S., Emanuelsson, O., Stolc, V., Weissman, S., Gerstein, M.B., et al. (2007). Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. Genome Res. *17*, 898–909. Fagegaltier, D., and Baker, B.S. (2004). X chromosome sites autonomously recruit the dosage compensation complex in Drosophila males. PLoS Biol. *2*, e341. 10.1371/journal.pbio.0020341.

Gallach, M., Arnau, V., and Marin, I. (2007). Global patterns of sequence evolution in Drosophila. BMC Genomics *8*, 408.

Gilfillan, G.D., Konig, C., Dahlsveen, I.K., Prakoura, N., Straub, T., Lamm, R., Fauth, T., and Becker, P.B. (2007). Cumulative contributions of weak DNA determinants to targeting the Drosophila dosage compensation complex. Nucleic Acids Res. *35*, 3561–3572.

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.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635–638.

Herzing, L.B.K., Romer, J.T., Horn, J.M., and Ashworth, A. (1997). Xist has properties of the X-chromosome inactivation centre. Nature 386, 272–275.

Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., and White, R. (2007). CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. PLoS Genet. *3*, e112. 10. 1371/journal.pgen.0030112.

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

Kelley, R.L., Solovyeva, I., Lyman, L.M., Richman, R., Solovyev, V., and Kuroda, M.I. (1995). Expression of MsI-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. Cell *81*, 867–877.

Kelley, R.L., Meller, V.H., Gordadze, P.R., Roman, G., Davis, R.L., and Kuroda, M.I. (1999). Epigenetic spreading of the *Drosophila* dosage compensation complex from *roX* RNA genes into flanking chromatin. Cell 98, 513–522.

Kind, J., and Akhtar, A. (2007). Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes. Genes Dev. 21, 2030–2040.

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.

Lee, J.T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature 386, 275–279.

Lee, J.T., Strauss, W.M., Dausman, J.A., and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell 86, 83–94.

Legube, G., McWeeney, S.K., Lercher, M.J., and Akhtar, A. (2006). X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in Drosophila. Genes Dev. *20*, 871–883.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

Liu, X.S., Brutlag, D.L., and Liu, J.S. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat. Biotechnol. *20*, 835–839.

Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene 5, 1743–1753. Lucchesi, J.C., Kelly, W.G., and Panning, B. (2005). Chromatin remodeling in dosage compensation. Annu. Rev. Genet. 39, 615–651.

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

Lyon, M.F. (2003). The Lyon and the LINE hypothesis. Semin. Cell Dev. Biol. 14, 313–318.

McDonel, P., Jans, J., Peterson, B.K., and Meyer, B.J. (2006). Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. Nature 444, 614–618.

Meller, V.H., Wu, K.H., Roman, G., Kuroda, M.I., and Davis, R.L. (1997). roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. Cell 88, 445–457.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Oh, H., Bone, J.R., and Kuroda, M.I. (2004). Multiple classes of MSL binding sites target dosage compensation to the X chromosome of *Drosophila*. Curr. Biol. *14*, 481–487.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644–649.

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

Park, Y., Kelley, R.L., Oh, H., Kuroda, M.I., and Meller, V.H. (2002). Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. Science 298, 1620–1623.

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

Pavesi, G., Mereghetti, P., Mauri, G., and Pesole, G. (2004). Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Res. *32*, W199–W203.

Schneider, R., and Grosschedl, R. (2007). Dynamics and interplay of nuclear architecture, genome organization, and gene expression. Genes Dev. *21*, 3027–3043.

Sekinger, E.A., Moqtaderi, Z., and Struhl, K. (2005). Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol. Cell *18*, 735–748.

Stenberg, P., Pettersson, F., Saura, A.O., Berglund, A., and Larsson, J. (2005). Sequence signature analysis of chromosome identity in three Drosophila species. BMC Bioinformatics *6*, 158.

Sun, L.V., Chen, L., Greil, F., Negre, N., Li, T.R., Cavalli, G., Zhao, H., van Steensel, B., and White, K.P. (2003). Protein-DNA interaction mapping using genomic tiling path microarrays in Drosophila. Proc. Natl. Acad. Sci. USA *100*, 9428–9433.

van Steensel, B., Delrow, J., and Bussemaker, H.J. (2003). Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding. Proc. Natl. Acad. Sci. USA *100*, 2580–2585.

Wade, J.T., Reppas, N.B., Church, G.M., and Struhl, K. (2005). Genomic analysis of LexA binding reveals the permissive nature of the Escherichia coli genome and identifies unconventional target sites. Genes Dev. *19*, 2619–2630.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *5*, 695–705.