

Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex

Gregor D. Gilfillan,^{1,3} Tobias Straub,^{1,3} Elzo de Wit,² Frauke Greil,² Rosemarie Lamm,¹ Bas van Steensel,² and Peter B. Becker^{1,4}

¹Adolf-Butenandt-Institut, Molekularbiologie, Ludwig-Maximilians-Universität München, 80336 München, Germany;

²Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

The dosage compensation complex (DCC) of *Drosophila melanogaster* is capable of distinguishing the single male X from the other chromosomes in the nucleus. It selectively interacts in a discontinuous pattern with much of the X chromosome. How the DCC identifies and binds the X, including binding to the many genes that require dosage compensation, is currently unknown. To identify bound genes and attempt to isolate the targeting cues, we visualized male-specific lethal 1 (MSL1) protein binding along the X chromosome by combining chromatin immunoprecipitation with high-resolution microarrays. More than 700 binding regions for the DCC were observed, encompassing more than half the genes found on the X chromosome. In addition, several rare autosomal binding sites were identified. Essential genes are preferred targets, and genes binding high levels of DCC appear to experience the most compensation (i.e., greatest increase in expression). DCC binding clearly favors genes over intergenic regions, and binds most strongly to the 3' end of transcription units. Within the targeted genes, the DCC exhibits a strong preference for exons and coding sequences. Our results demonstrate gene-specific binding of the DCC, and identify several sequence elements that may partly direct its targeting.

[**Keywords:** Transcription; histone modification; chromatin remodeling; MSL complex; microarray; gene expression]

Supplemental material is available at <http://www.genesdev.org>.

Received December 7, 2005; revised version accepted February 1, 2006.

Sexual determination in many animals is governed by sex chromosomes, with one sex being homogametic (e.g., XX) and the other heterogametic (e.g., XY) (for review, see Charlesworth 1996). The resulting chromosome imbalance in the heterogametic sex requires the evolution of a dosage compensation mechanism, to avoid the lethality usually associated with chromosomal aneuploidy. Different species address this problem using opposing strategies, using chromosome-wide activation or inactivation to balance expression between the sexes (Lucchesi et al. 2005). In *Drosophila*, dosage compensation operates in the male, where transcription of the single male X is up-regulated, approximately twofold, to equal expression of the two female X chromosomes (Hamada et al. 2005; Straub et al. 2005b). This process is mediated by the dosage compensation complex (DCC, also known as the male-specific lethal complex), which

binds to the male X. The DCC is a ribonucleoprotein complex consisting of at least five proteins (MSL1–MSL3, MOF and MLE, usually referred to as the male-specific lethal or MSL proteins) and two noncoding RNAs (RNA on the X, *roX1* and *roX2*) (for review, see Taipale and Akhtar 2005). The interaction of the DCC with other proteins required for dosage compensation, which also have wider roles in the cell, is a matter of current research; for example, JIL-1 is an essential histone H3 kinase enriched on the X chromosome that has been shown to interact with the MSL proteins (Jin et al. 2000). At least one important role of the DCC is to target the histone acetyltransferase activity of MOF to the male X, where it acetylates histone H4 Lys 16 (H4K16) (Bone et al. 1994; Smith et al. 2000). Acetylation of H4K16 by MOF has been shown to cause de-repression of transcription from chromatin in vivo and in vitro (Akhtar and Becker 2000).

The mechanism by which the DCC recognizes a single chromosome in the nucleus remains an enigma (Oh et al. 2004a). Historically, three models of the mechanism for MSL targeting to the X have been proposed. Initially, the majority of genes were expected to lie in proximity to an

³These authors contributed equally to this work.

⁴Corresponding author.

E-MAIL pbecker@med.uni-muenchen.de; FAX 49-89-2180-75-425.

Article published online ahead of print. Article and publication date are at <http://www.genesdev.org/cgi/doi/10.1101/gad.1399406>.

enhancer-like sequence that attracted the DCC (Baker et al. 1994). However, the discovery that MSL1 and MSL2 could bind to a reproducible subset of 35–70 sites (including the *roX* genes) in the absence of the other MSL proteins led to the proposal of the “entry and spreading” model (for review, see Kelley and Kuroda 2000). Accordingly, the MSL proteins were proposed to recognize the X via the ~35 high-affinity “entry sites,” from where they could spread to bind the remaining target sites by an unknown mechanism (Lyman et al. 1997; Kelley et al. 1999). Recent studies of flies with altered DCC concentrations, X to autosome translocations, and analysis of new “entry sites” suggest that the entry sites are simply high-affinity sites in a continuum of affinity sites dispersed along the X chromosome (Demakova et al. 2003; Fagegaltier and Baker 2004; Oh et al. 2004b; Dahlsveen et al. 2006). The most recent model therefore proposes that X-chromosome binding is governed largely or even solely by DNA elements of a degenerate nature (for review, see Straub et al. 2005a). The weakest elements, incapable of attracting the DCC alone when moved to autosomal locations, are able to attract the DCC in their natural X-linked situation, where they benefit from an increased local concentration of DCC due to their proximity to higher-affinity sites. This recent “affinity model” therefore bears similarity to the first model originally proposed to explain X-specific targeting; namely, that every compensated gene would be close to an enhancer-like DNA element responsible for attracting the DCC.

At the resolution afforded by immunostaining of polytene chromosomes, the DCC is seen to bind to the gene-rich interbands (Demakova et al. 2003), opening the possibility that the DCC coats entire chromosomal domains containing several genes. Evidence suggests that genes with similar transcriptional profiles tend to be clustered in *Drosophila* (Spellman and Rubin 2002), making compensation of entire gene clusters an interesting possibility. In principle, the DCC could also control clusters of genes by binding to sequences behaving like locus control regions or matrix attachment regions, possibly invoking or regulating chromatin looping (for reviews, see Hart and Laemmli 1998; Sipos and Gyurkovics 2005).

Alternatively, it has been proposed that the DCC may be attracted by components of the transcription apparatus (Lucchesi 1998). The DCC may aid transcription elongation rather than promoter accessibility, as H4K16 acetylation was found by chromatin IP in the middle and 3' end of two compensated genes rather than at the promoter (Smith et al. 2001). In addition, on polytene chromosomes, Sass et al. (2003) were able to discern MSL1 binding to the transcribed portion of a gene from a GAL4 activator presumably binding to the promoter. However, whether the DCC-binding pattern changes during development is unclear, as both subtle changes (Sass et al. 2003) and perpetual binding of MSL3 and MSL2 (Kotlikova et al. 2006) have been reported throughout larval development. It is therefore currently unclear whether the DCC binds to a programmed set of sites, or is targeted by a more flexible set of commands.

Thus, while recognition and targeting of the X is poorly understood, how the DCC finds and binds to its sites of action remains even more of a mystery. To study the wild-type-binding pattern of the DCC and attempt to identify the targeting cues, we performed chromatin immunoprecipitation (ChIP) with MSL1 from *Drosophila* embryos. Hybridization of the ChIP probes to DNA microarrays spanning the X chromosome at high resolution (ChIP on chip) allowed analysis of MSL1 binding along the entire X chromosome. In contrast to the banded “coating” of MSL1 seen at the low resolution of the polytene chromosome, the binding pattern observed by ChIP demonstrated targeting to many single genes. In addition, the high resolution allowed analysis of MSL1-binding profiles within individual genes, revealing specificity for exons and coding sequences, with a bias toward the 3' end of genes. Correlation of MSL and RNA polymerase II (pol II)-binding profiles allowed models based on domain-wide binding of the DCC, or transcription-based targeting of the DCC, to be excluded. Our results rather suggest a model whereby individual genes under selective pressure to equalize dosage between the sexes have evolved targeting signals to attract the DCC. In support of this, we have identified a complex set of targeting motifs from DCC-binding coding sequences. Using these motifs, we demonstrate for the first time a limited prediction of DCC binding.

Results

The DCC and acetylated H4K16 are found in gene-rich regions

The protein and RNA components of the DCC, and acetylated H4K16, colocalize on the male X at the limited resolution afforded by polytene chromosome immunofluorescence (Bone et al. 1994; Meller et al. 1997; Franke and Baker 1999). To examine their coincidence at higher resolution in nonpolytenized nuclei, we performed ChIP with antibodies raised against MSL1, MOF, and acetylated H4K16 in 12- to 14-h-old *Drosophila* embryos. Binding profiles for these antibodies are in good agreement, as visualized by Southern hybridization of ChIP DNA to nylon membranes containing restriction fragments covering 2.6 Mb of DNA from the distal tip of the X chromosome (Supplementary Fig. 1; Straub et al. 2005b). Notably, MSL1 and MOF binding, and acetylated H4K16, coincide with many gene clusters. While MOF is also expressed in females and thus may have an additional role outside of the DCC (Hilfiker et al. 1997), it catalyzes the acetylation of H4K16, which, in turn, is believed to be crucial for the doubling of gene expression required for dosage compensation (Akhtar and Becker 2000; Smith et al. 2000; Morales et al. 2004).

To verify the observed enrichments, sequence-specific quantitative PCR (Q-PCR) was used to confirm three “peaks” and one “trough” of DCC binding in three independent ChIP experiments (Straub et al. 2005b; data not shown). At the limited number of locations examined by Q-PCR, coincident binding of all tested DCC

members (MSL1, MSL2, MSL3, MOF, and MLE) was seen. In addition, MLE was also found at high levels in a trough between these peaks, consistent with previous observations that MLE is expressed in both sexes, has roles in addition to the DCC, and can be seen binding to chromosomal sites not bound by other DCC members (Kotlikova et al. 2006 and references therein).

Based on these observations, we chose to use MSL1 as a marker for DCC binding in a genome-wide cDNA microarray analysis, and at higher resolution on the X chromosome through the use of a tiling microarray. MSL1 is arguably the best marker of DCC binding, as MSL1 and MSL2 have been suggested to form the DNA-binding "core" of the DCC (Lyman et al. 1997) and MSL1 consistently gave the most robust ChIP signals. While MSL2 is the only DCC member with a strictly male expression (Bashaw and Baker 1995; Zhou et al. 1995), MSL1 is inherently unstable in the absence of MSL2 and the DCC does not form in females (Gilfillan et al. 2004). This permitted the use of mixed male and female wild-type embryos, as females are expected to contribute only background signals to any ChIP with MSL1.

Genome-wide binding of the DCC

Given the enrichment of the DCC in many gene-rich regions seen by Southern analysis, we examined the gene targets of the DCC by hybridizing the MSL1 ChIP probes to a cDNA array. Genome-wide binding of the DCC was examined by hybridizing four independent MSL1 ChIP probes to a cDNA array containing 12,144 features representing genes from all chromosome arms. The results show a marked specificity for the X chromosome (Fig. 1). Of 1389 X-linked genes, 773 (56%) were seen to bind the DCC during the 2-h developmental window examined, and are listed in Supplementary Table 1. A control hy-

bridization with an unrelated antibody raised against the insulator-binding protein CTCF showed a markedly different binding (data not shown).

In addition, a further 27 hits were found on the autosomes (Fig. 1), corresponding to only 0.4% of autosomal genes. The existence of autosomal DCC binding has been known for some time, but no target loci have been documented. Autosomal binding sites visualized on polytene chromosomes are often weak and inconsistent, even between nuclei of the same individual, but several reproducible sites have been documented (Kelley et al. 1999; Demakova et al. 2003). The autosomal MSL1-binding targets identified in the cDNA array hybridizations are listed in Supplementary Table 2. However, of these 27 genes, only seven map to cytological positions of previously described autosomal binding sites for the DCC (Demakova et al. 2003). In order to verify the cDNA results and study the precise binding profile within genes, we proceeded to examine MSL1 binding to the X chromosome at higher resolution by the use of a tiling array.

MSL1 targeting is gene specific

MSL1 ChIPs were hybridized to a custom microarray representing the entire X chromosome with oligonucleotides spaced at <100-base pair (bp) intervals, excluding repeat elements (~10% of X-chromosome sequences). For control purposes, a further 2.1 Mb of autosomal sequence from chromosome arm 3R, containing the bithorax complex, was included on the array. As expected, MSL1 binding is highly specific for the X chromosome. Example profiles of MSL1 binding to 200-kb segments of the X- and 3R-chromosome arms are displayed in Figure 2. Scatterplots detailing the experimental variability are available as Supplementary Figure 2, A–C. In general, the three independent ChIP hybridization experiments were

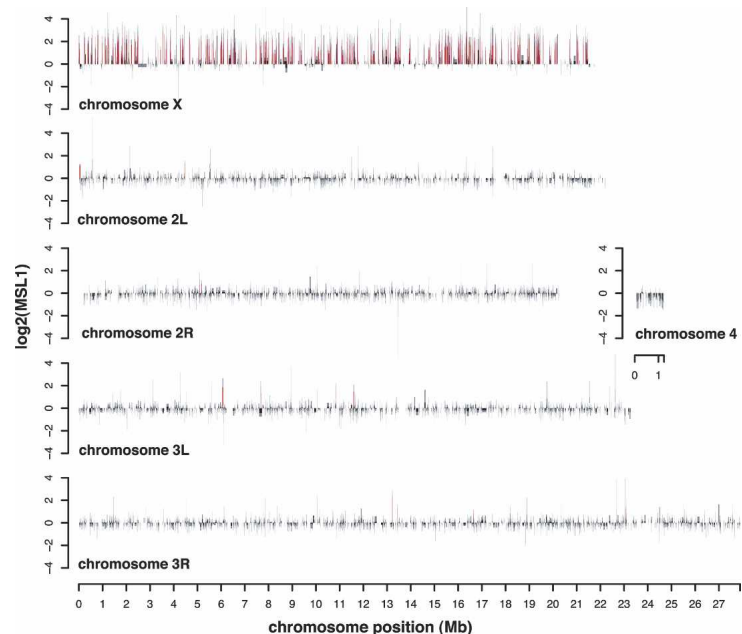


Figure 1. Genome-wide MSL1 binding as revealed by ChIP hybridization to a cDNA array. Arms of the *Drosophila* chromosomes are shown above a scale in megabases. The Y-axes represent the \log_2 of the MSL1 ChIP enrichment ratios over mock IP. Each bar represents a cDNA on the array. The standard deviation between the four replicate experiments is depicted by gray error bars. Genes that were significantly bound by MSL1 are colored red ($p < 0.01$).

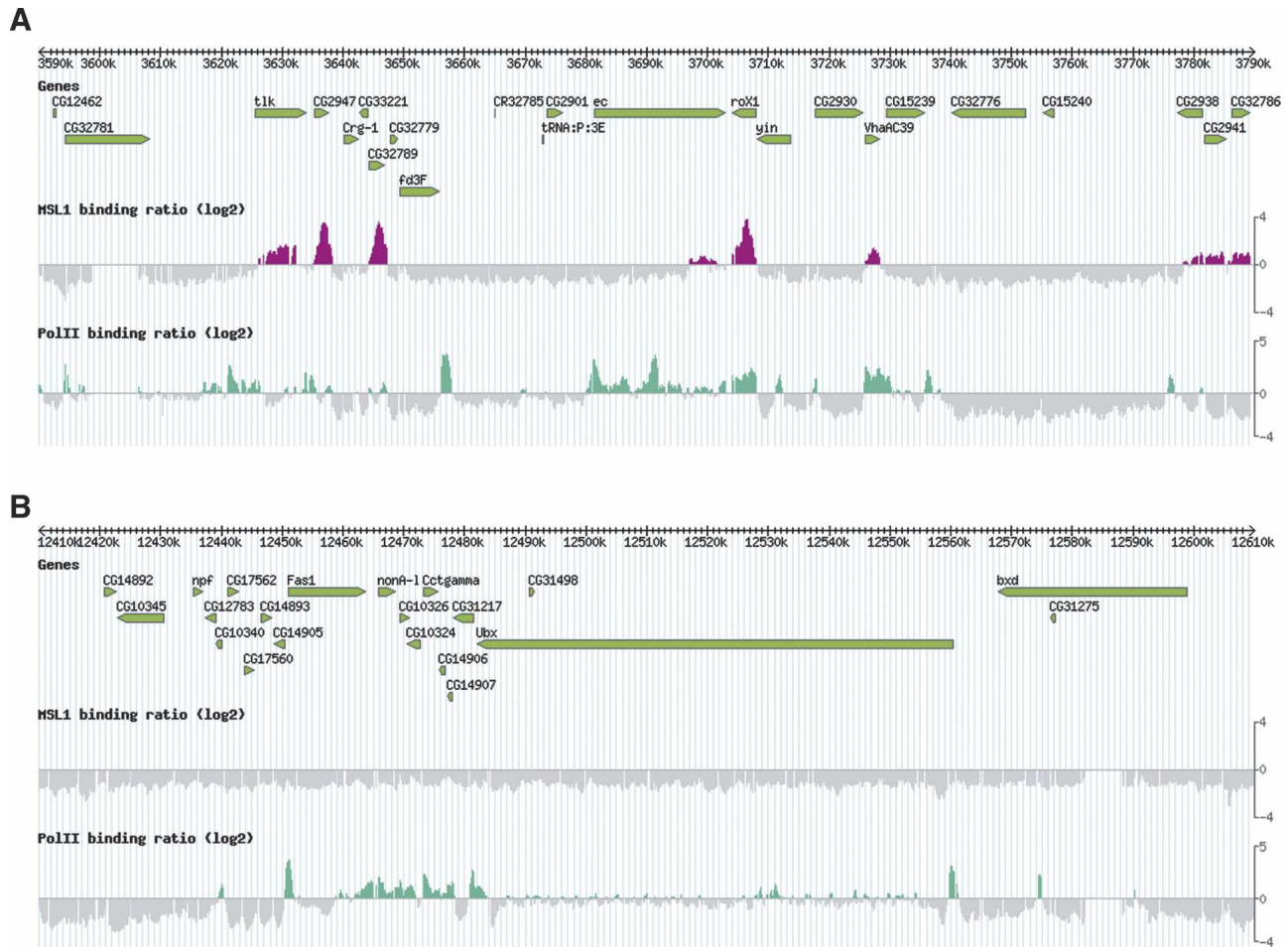


Figure 2. MSL1 and RNA pol II ChIP hybridization to tiling array. Representative 200-kb sections of the X chromosome (A) and chromosome 3R (B). MSL1 and RNA pol II data are the log₂ signal ratio of specific IP/input for each oligonucleotide on the array. Each oligonucleotide of the array is represented by a single vertical bar and colored when signal rises above threshold. Data are the mean of three independent ChIP experiments. The entire data set can be browsed at <http://wpl054.bio.med.uni-muenchen.de/cgi-perl/browse/genome1>.

very similar (all pairwise correlation coefficients are >0.9). For the tiling array, all MSL1 signals were normalized to input DNA, unlike in the cDNA arrays, which were normalized to mock IP DNA. Therefore, a single mock IP was also hybridized to the tiling array and normalized to input DNA. Comparing the mock IP to its paired MSL1 hybridization reveals, as expected, a random distribution (correlation coefficient 0.01) (Supplementary Fig. 2D).

Examination of the total binding of MSL1 on the X chromosome reveals a population of oligonucleotides that bind MSL1, clearly discernible from those that do not (Fig. 3A). As expected, the control autosomal sequence is almost entirely free of MSL1. The total amount of the X chromosome covered by MSL1 is surprisingly low, ~25%. Given the resolution of the ChIP technique, which creates a spreading effect around a binding site due to the random shearing of DNA by sonication (average chromatin size was ~700 bp), the amount of DNA actually bound by MSL1 in the cell is likely to be lower still.

We found >700 separable regions of MSL1 binding, substantially more than can be resolved by indirect immunostaining of polytene chromosomes (Demakova et al. 2003). The median size of these regions was 2.9 kb, but MSL1 generally does not coat entire genes. In spite of this, we found examples of MSL1 covering clusters of neighboring genes, and the longest uninterrupted region of MSL1 binding covered 52 kb. Notably, such regions do not appear evenly coated in MSL1, but rather contain multiple “peaks” of binding.

Importantly, many sites of MSL1 binding are single peaks within individual genes (Fig. 2A). MSL1 clearly binds to genes in favor of intergenic regions, with >90% of MSL1 binding within genes (Fig. 3B). Similar to the results obtained with the cDNA array, 1183 of the 2309 genes (51%) represented on the X-chromosome tiling array were found to bind MSL1. Genes binding MSL1 from analysis of both cDNA and tiling array data sets are listed in Supplementary Table 1. The cDNA and tiling arrays show good agreement on common genes, as illustrated by correlating the two data sets (Supplementary Fig. 3).

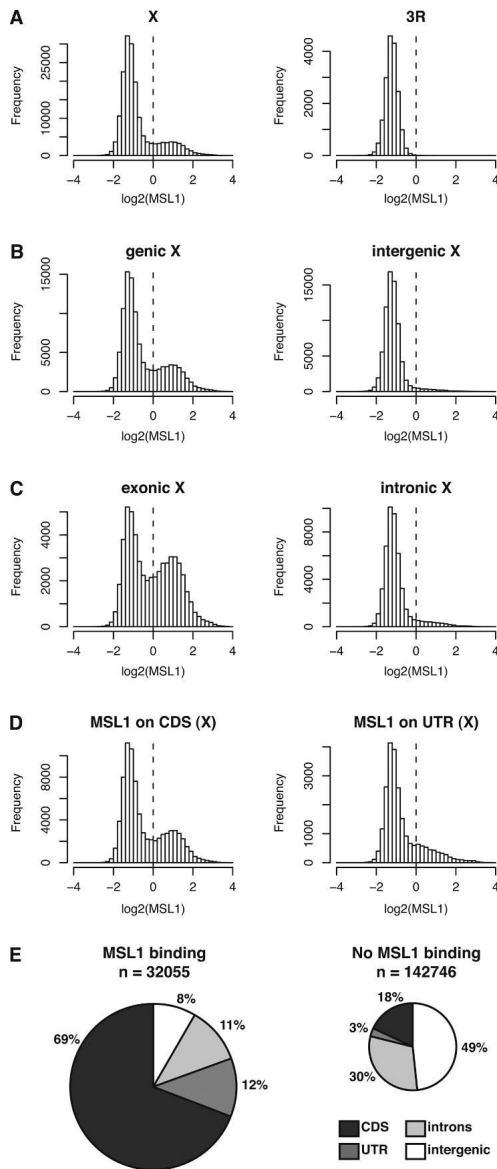


Figure 3. MSL1 binds primarily to coding sequences. Signal distribution (array oligonucleotides) of MSL1 enrichment in defined genomic sequences on custom tiling array. MSL1 binding to X and 3R chromosomes (A), genes and intergenic sequences on the X chromosome (B), exons and introns on the X chromosome (C), and coding sequences and 5'/3' untranslated regions on the X chromosome (D). Oligonucleotides to the right of the dotted line were considered to be MSL1 binding. (E) Summary of genomic features in MSL1-binding and nonbinding oligonucleotide categories. (CDS) Coding sequence; (UTR) 5' and 3' mRNA untranslated regions.

In addition to the 27 genes binding MSL1 in the cDNA array analysis, the autosomal sequence on the tiling array identified one further site of MSL1 binding. This 700-bp intergenic region upstream of the *ear* gene could conceivably represent the autosomal site previously reported at cytological position 88E (Demakova et al. 2003).

Two clusters of transfer RNA genes on the X at position 12DE, coding serine and arginine acceptor tRNAs,

bind MSL1, although in one such cluster the binding is weak and not directly within a gene. Dosage compensation of serine tRNA expression has been reported previously (Birchler et al. 1982), so it is possible that this is mediated by the DCC. In addition to these clusters, there exist only a handful of dispersed tRNA genes on the X, only one of which exhibits binding of MSL1 (CR32826), so tRNA gene dosage compensation does not appear to be universal.

The DCC preferentially accumulates in coding sequences and at the 3' end of genes

In addition to its preference for genic over intergenic regions, MSL1 shows a very clear bias within a gene for binding to exons rather than intron sequences (Fig. 3C). Within the exons themselves, there is a strong association with coding sequences instead of 5' or 3' untranslated regions (Fig. 3D). A summary of MSL1 binding to coding sequence features is shown in Figure 3E. Notably, the vast majority of genes show more DCC binding at the 3' end (Fig. 4A,B). In the accompanying paper by Alekseyenko et al. (2006), similar conclusions were reached regarding the number of MSL1-binding regions and 3' bias, as observed in two *Drosophila* cell lines and late-stage embryos. In the remainder of genes, we observed a variety of binding profiles, including coating and binding in the middle of the gene (see examples in Fig. 4C,D). Promoter binding is extremely rare, and in those instances where it was observed, is likely a result of signal "spill-over" from a neighboring gene. However, we were unable to discern any "rules" that would apparently govern placement of MSL1 within a gene.

Transcription is not sufficient to recruit the DCC

To investigate the relationship between dosage compensation and transcription, we also examined the distribution of RNA pol II by ChIP using an antibody raised against the N terminus and recognizing all forms of pol II. We found that ~65% of genes on the X are bound by pol II during the 2-h time window of embryonic development examined by these ChIPs. While some of these may represent genes with paused polymerases not actively transcribing, the majority are expected to represent transcriptionally active genes (Law et al. 1998). In addition, an earlier study using genomic microarrays found a good correlation between the hybridization pattern of isolated mRNA and ChIP probes generated using a pan-polymerase antibody (MacAlpine et al. 2004). Surprisingly, ~25% of the genes binding pol II (374/1412) did not also bind MSL1, suggesting that they may not be dosage compensated, or compensated post-transcriptionally, although we cannot exclude signals from genes expressed only in females. Thus, transcription alone appears insufficient to attract the DCC, in agreement with recent observations of pol II and DCC binding to polytene chromosomes (Kotlikova et al. 2006). Also of interest is the high number (140) of protein-coding genes that bind MSL1 but have no detectable pol II (~12% of genes binding MSL1). These may represent genes containing targeting elements important for X-chromosome recog-

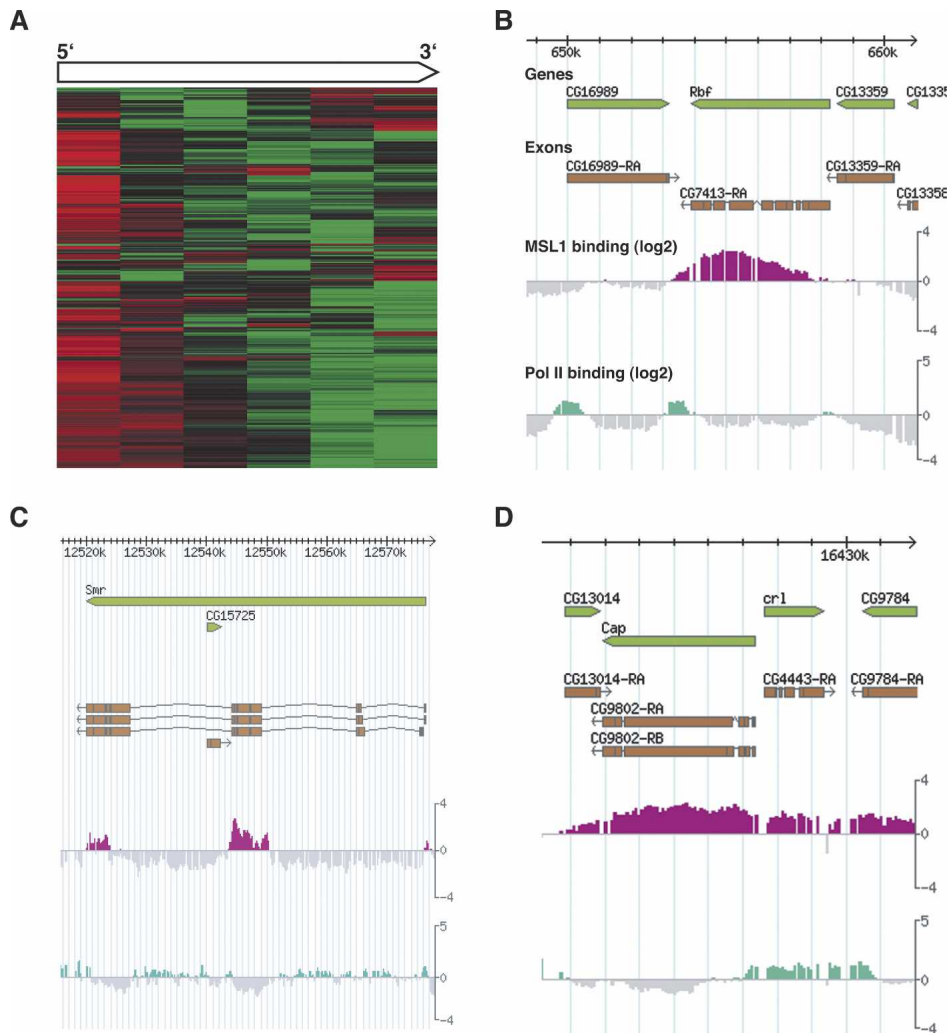


Figure 4. The majority of DCC-binding genes show a 3' accumulation of MSL1. (A) MSL1 binding along genes is skewed toward the 3' end. Three-hundred-thirty-four nonoverlapping genes with robust MSL1 binding (average log₂ ratio >0.5) and a length >2 kb were aligned and divided into six segments. The average MSL1 binding of each segment above (green) or below (red) the average binding of the gene is depicted as a heat map. The genes were hierarchically clustered and vertically ordered based on the signal distribution pattern. (B) The *Rbf* gene shows strong 3' MSL1 binding. (C) The bias for exons is easily seen on longer genes such as *Smr*. (D) Coating of genes exemplified by the *Cap* and *cr1* genes.

nitiation but not for dosage compensation, genes with tissue-specific expression, or genes that have been shut down or are awaiting transcription (discussed below). However, the absence of a chromosomal control region representing zero/background binding (equivalent to most autosomal sequences for the DCC) means that our threshold for pol II is difficult to verify. Thus, we cannot exclude the possibility that these genes are expressed below the level of detection of the ChIP hybridization technique.

The X chromosome does not bind more polymerase

Polymerase was found typically enriched at the promoters of genes, and in many cases also at the 3' end, in agreement with previous observations suggesting looping of the transcriptional unit (O'Sullivan et al. 2004). Although pol II is found on the majority of genes binding

MSL1 (1035/1183 = 87%), the binding profile of the two proteins is seldom similar (Fig. 2A). Importantly, there is no evidence that there is more pol II on the X chromosome (Fig. 5A). In addition, there is no correlation between the amount of polymerase bound by a gene and the amount of MSL1 (Fig. 5B). These observations support earlier suggestions that the DCC does not mediate dosage compensation by increasing the rate of transcription initiation (Lucchesi 1998), which would be expected to increase the amount of polymerase on the chromosome at any given time.

MSL1 does not participate in regulating expression of the DCC itself

To investigate potential auto-regulation of MSL gene expression, we included the autosomal loci of MSL1,

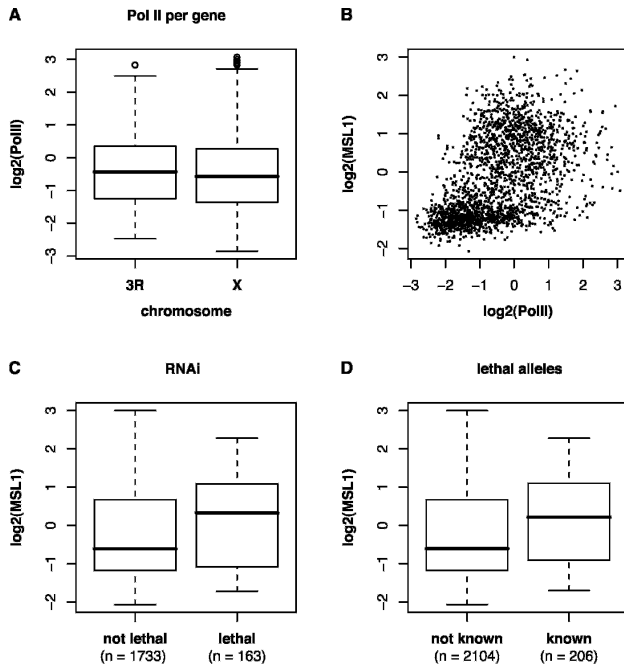


Figure 5. Relation of MSL1 to RNA pol II binding and gene function. (A) Boxplot comparing levels of pol II per gene on the X and 3R sequences from the tiling array. (B) Scatterplot of MSL1 binding data against pol II binding on X-chromosomal genes. Note that two populations are visible. In the population binding MSL1 (those with \log_2 values >0), there is no correlation between the amount of MSL1 and the amount of polymerase on a gene. (C) Boxplot showing increased MSL1 levels on X-chromosomal genes with growth defects identified on RNAi (RNAi lethal genes with z scores >1) (Boutros et al. 2004); two-sided t -test, p -value: 4×10^{-9} . (D) Boxplot showing MSL1 levels correlated to X-chromosomal genes with essential alleles listed on FlyBase (<http://www.flybase.org>); two-sided t -test, p -value: 2.2×10^{-16} .

MSL2, MSL3, MLE, and JIL1 (MOF is on the X) on our high-resolution tiling microarray. Despite including 5 kb of flanking sequence both upstream and downstream of each gene on the array, we found no binding of MSL1 to any of the genes. Polymerase was found on all of these genes. In contrast, the X-linked *mof* gene showed strong binding of MSL1. This implies that the DCC, as an entire complex, has no role in regulating its own mRNA transcription. Future ChIP experiments with the other members of the DCC may reveal such regulation by other, individual members. Interestingly, the loci of both *roX1* and *roX2* noncoding RNAs are sites of strong pol II binding, ending uncertainty over which polymerase transcribes these genes.

Essential genes and stably expressed genes are typical DCC targets

In order to examine the types of genes binding the DCC, we correlated MSL1 binding with previously published data on gene expression and phenotype. X-linked genes isolated in an RNA interference (RNAi) screen for cellu-

lar lethality (Boutros et al. 2004) associate with more MSL1 than genes that are not important for cell growth (Fig. 5C). In addition, genes that have been isolated as a lethal allele (as listed in FlyBase, <http://www.flybase.org>) show a stronger binding of MSL1 than genes for which no lethal allele is known (Fig. 5D).

Correlating MSL1 binding with previously published gene expression data (Arbeitman et al. 2002) revealed that many genes targeted for MSL1 in the 2-h time window studied here have a steady level of expression throughout fly development (Supplementary Fig. 4). This suggests that genes required in all developmental stages, most likely “housekeeping” genes, are targets for the DCC. Searching the Gene Ontology (<http://www.geneontology.org>) classifications did not reveal any particular class of gene enriched in the MSL1-binding targets (data not shown).

The observed gene-by-gene binding of MSL1 suggests that the targets should demonstrate dosage compensation. There currently exists no genome-wide data documenting the genes subject to dosage compensation in *Drosophila* embryos. The only comprehensive study to date was performed in SL2 tissue culture cells (Hamada et al. 2005), whereby the effect of MSL2 RNAi on gene expression was measured. Despite the differences between embryos and cells, correlating our MSL1 targets with the drop in expression of SL2 genes upon MSL2 RNAi revealed a good agreement of the two data sets, suggesting that DCC binding does, indeed, confer dosage compensation (Fig. 6). Notably, the results indicate that more MSL1 binding corresponds to a greater loss of expression on MSL2 RNAi (i.e., more MSL1

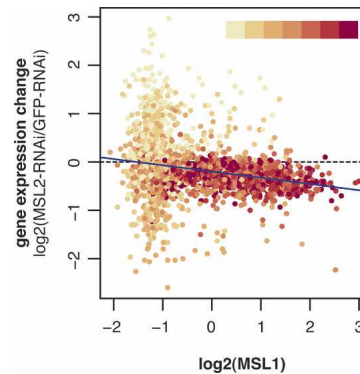


Figure 6. MSL1 binding correlates with dosage compensation. Gene expression data following RNAi of MSL2 in SL2 cells (Hamada et al. 2005) were correlated to embryonic MSL1 binding (tiling array data set). MSL1 targets (genes with mean MSL1 \log_2 ratios >0) show a drop in expression upon knockdown of MSL2 (Wilcoxon rank sum p -value: 2.2×10^{-16}). In addition, high levels of MSL1 binding correlate with the greatest loss of dosage compensation after MSL2 knockdown, as can be seen by the negative slope of the MSL1 targets (the least squares regression line for MSL1 targets is indicated in blue; slope -0.12 , p -value: 2×10^{-10}). Genes have been colored for absolute expression level measured in SL2 cells: (dark red) high expression; (yellow) low expression. No linear relation exists between MSL1 binding and absolute expression level.

binding = more dosage compensation). This effect is visible despite the subtle (twofold) changes of expression resulting from dosage compensation. Thus, weak recruitment of MSL1 may only provide partial dosage compensation, and full twofold compensation requires a relatively large amount of the DCC.

In addition, MSL1 targets have a higher expression level than nontargets (Supplementary Fig. 5). However, within the MSL1 target group, there is no correlation between the level of MSL1 binding and absolute gene expression levels, paralleling the lack of correlation between polymerase and MSL1 levels seen in our study.

Targeting the DCC

The peaks of MSL1 binding within single genes and poor association of the DCC with pol II suggest that individual genes contain the targeting information required to attract the DCC. Classical algorithms (see Materials and Methods) for identifying novel transcription factor-binding sites failed to uncover sequence elements that could explain the observed binding pattern. While these were often capable of distinguishing the X chromosome from autosomal sequences, they were unable to discriminate between MSL1-binding and nonbinding regions of the X (data not shown). We assume that most of the algorithms are misled by short repetitive elements that are strongly enriched on the X, coincide stochastically with binding elements, but are not binding determinants on their own.

Multivariate statistics based on sequence “word” frequencies have been used previously to uncover sequence signatures of different chromosomes (Stenberg et al. 2005). Using a similar but supervised approach, partial least squares (PLS) regression (Wold 1975), we were able to isolate several combinations of hexameric sequence motifs, partially describing the binding of MSL1 along sections of the X chromosome. Critically, the regression model could be used to predict a limited amount of MSL1 binding on further chromosome sections (Fig. 7A,B). The top-scoring combinations of hexamer sequences describing MSL1 binding are shown in Figure 7C. The complete list of hexamer loadings of the individual components is provided as Supplementary Table 3.

Discussion

Genes bound by the DCC

It has long been known that not all genes on the X chromosome are subject to dosage compensation, but those known examples were apparently exceptional cases: Loci also present on the Y chromosome, female-specific genes, and larval genes proposed to be members of redundant gene families (Baker et al. 1994 and references therein). Similar estimates of the number of bound genes were derived from the cDNA and tiling arrays, despite normalization to different control DNAs (mock IP and input DNA, respectively). In the 2-h time window of

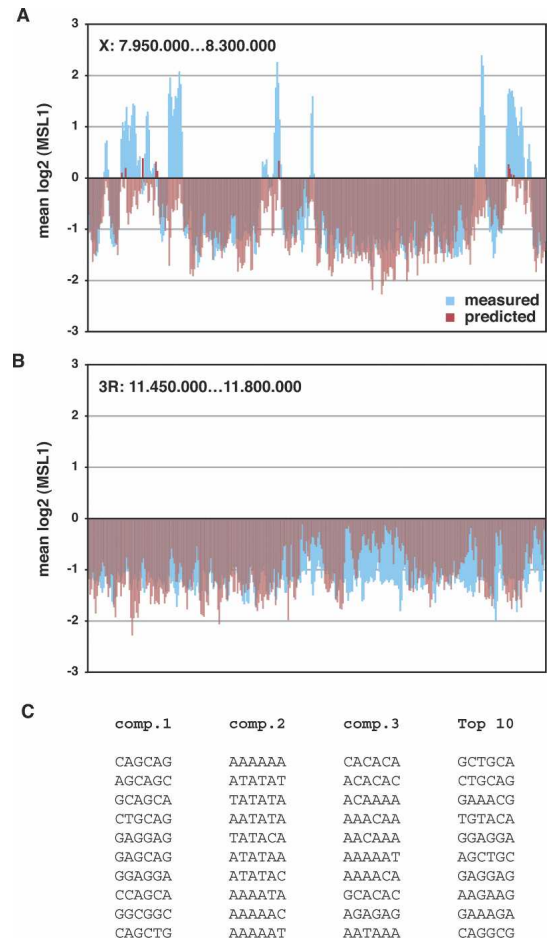


Figure 7. Sequence motifs have a role in DCC targeting. A 4-Mb section of the X chromosome and 1 Mb of 3R were used as a training set to describe DCC binding by a PLS regression. The regression model comprising three components was then applied to predict DCC binding on a further 350-kb section of the X chromosome (A) or 3R (B). Measured DCC binding from the tiling array is shown in the same figures for comparison. (C) The 10 top-scoring sequence motifs within the individual components (comp1–comp3) are listed. The 10 motifs with the most positive combined regression coefficients from all three components are shown (Top 10).

embryonic development studied in this analysis, we found that just over half of the annotated genes on the X chromosome were bound by the DCC. This may represent a slight overestimate of the genes actually binding the DCC, as the resolution of the ChIP analysis is defined by the average length of the input chromatin. Thus, “spill-over” of signal from genuine binding sites, due to our resolution of ~700 bp, may account for a number of genes considered targets for the DCC. The recent study by Kuroda and colleagues (Hamada et al. 2005) provided a substantial list of genes subject to dosage compensation by the DCC. However, difficulties in measuring twofold changes in gene expression by microarray analysis mean that this gene list is almost certainly an underestimate of the true number of compensated genes. Although our target list of DCC-binding genes is longer, we

nonetheless find that a large number of genes bind polymerase and yet do not bind the DCC; thus the number of compensated genes on the X may, in fact, be lower than previously assumed.

We also present several autosomal binding sites identified using the cDNA array. Notably, the majority of these do not compare with sites of autosomal DCC binding observed on polytene chromosomes (Demakova et al. 2003). Several reasons may explain these differences, including false-positive hits, inaccuracies of mapped polytene positions, the different developmental stages and tissues concerned, and the variable nature of the autosomal sites seen on polytene chromosomes. It is also worth noting that the only strong autosomal MSL1-binding site found in the tiling array reveals binding to an intergenic sequence. The X-chromosomal-binding sites for the DCC are very specific for genic sequences; therefore, if genuine autosomal sites for the DCC represent a different, perhaps nonfunctional, binding to intergenic sequences, they would not be recovered by the use of a cDNA array. Nonetheless, the autosomal sites are of interest because they may provide clues to the DNA sequences attracting the DCC.

DNA sequence elements may attract the DCC to target genes

The observed binding to many single genes and the obvious peaks within bound regions are incompatible with a model for DCC binding based on coating of large chromosomal domains, and suggest instead a gene-specific targeting. Furthermore, the observed specificity for genes over intergenic regions implies that the DCC binds directly to its target genes, rather than applying control of a domain analogous to that of the Locus Control Region regulating β -globin gene expression (Chakalova et al. 2005). Our results instead favor a model in which the DCC binds directly to the genes that are targets for dosage compensation. We found further evidence to support this, based on analysis of bound sequences, which suggests that DCC targeting is at least in part directed by DNA sequence.

Motif-finding algorithms commonly used to define transcription factor-binding sites were unable to isolate a targeting sequence that could direct DCC binding. Their failure suggests that DCC binding may be directed by a more complex combination of degenerate sequence motifs. A recent analysis of high-affinity sequences defined several paired motifs found enriched in MSL-binding sequences (Dahlsveen et al. 2006), but these were also insufficient to predict further DCC-binding sites. To examine more complex word combinations, we used PLS regression, with which we identified several sequences that to some extent explain the observed DCC binding. MSL1-binding sequences from a section of the X chromosome could be used to predict MSL1 binding on further stretches of the X chromosome. The diversity of motifs required to describe MSL1 binding suggests that combinations of short sequence motifs, dispersed through target sequence, are responsible for attracting

the DCC. The identified motifs notably contain GA and CA dinucleotides. Sequences containing the GAGA motif appear to have an important role in attracting the DCC to the *roX2* high-affinity site (Park et al. 2003), and have also been found in additional high-affinity sites (Dahlsveen et al. 2006). Furthermore, CA and GA dinucleotide repeats are enriched on the X chromosomes of *Drosophila* species exhibiting dosage compensation (Huijser et al. 1987; Pardue et al. 1987; Lowenhaupt et al. 1989). Whether sequence motifs, in complex combinations, are sufficient to explain all DCC targeting is currently unclear. However, our results strongly suggest that such combinations of sequence motifs have an essential role in DCC targeting, not limited to the high-affinity sites.

The finding that targeting signals encoded in the DNA sequence attract the DCC implies that binding would be identical in all cell types, unless access of the DCC to target regions is regulated. However, analysis of different cell types in the accompanying paper by Alekseyenko et al. (2006) found that a minority of genes display differential DCC binding. Clearly, the simplest targeting theory that the DCC recognizes a set of DNA motifs early in development and establishes an inflexible binding pattern is incorrect. In this context, the genes found in embryos to be binding MSL1 and not polymerase may be examples of genes already expressed and since silenced, or genes awaiting activation. In *Drosophila* larvae, conflicting reports observed either a sustained pattern of MSL1 and MSL3 binding to polytene chromosomes (Kotlikova et al. 2006), or subtle changes throughout larval development (Sass et al. 2003). The steady expression levels observed throughout development for many embryonic DCC-target genes in this study suggests that those genes will be expressed constitutively. Accordingly, we suggest that sustained DCC binding is the rule, and developmental changes the exception. Refinement of targeting motifs may allow a more accurate definition of the rules governing DCC targeting. Nonetheless, our current ability to predict DCC binding based on DNA sequence is limited, and it is therefore possible that other factors such as chromatin modifications or transcription have a role in targeting the DCC. For example, although transcription does not appear sufficient to attract the DCC, it may be a prerequisite for binding. Conceivably, chromatin changes induced by a "pioneer polymerase" (Orphanides and Reinberg 2000) could enable binding of the DCC, possibly explaining a limited amount of developmental regulation of DCC binding.

The striking specificity for DCC binding to exons and coding sequences is curious, as the sequences responsible for targeting the DCC must simultaneously perform the additional function of encoding functional protein, with accompanying constraints on sequence evolution. It has recently been reported that X-chromosomal genes have a higher codon bias than autosomal genes (Singh et al. 2005). We observed that DCC target genes have a higher expression than nontargets. Highly expressed genes typically have high codon bias (Hey and

Kliman 2002). We must therefore consider the possibility that the motifs identified by our PLS analysis may not direct DCC targeting, but instead may be the consequence of preferred codon usage in highly expressed, compensated genes. Some caution is also required in interpreting the observed specificity for exons, as small introns are a feature of *Drosophila* genes (Adams et al. 2000), and the resolution of the ChIP technique cannot exclude binding to many such features. However, many intron-less genes are targets of the DCC, confirming assertions that coding sequences can attract the DCC.

Evolution of DCC binding and dosage compensation

A requirement for our conclusion that the DCC binds to its site of action would be that bound genes demonstrate dosage compensation. Despite being limited to comparing our binding data from embryos with a published study from tissue culture cells (Hamada et al. 2005), the correlation between the data sets would suggest that this is, indeed, the case. Further correlation of MSL1 binding to published gene expression data demonstrates that many gene targets for the DCC have a sustained expression throughout development, and as such may perform “housekeeping” functions, as previously suggested (Sass et al. 2003). In keeping with this, we find that more MSL1 is bound to essential genes. Based on our observations, we propose that only those genes for which dosage is critical have evolved the ability to recruit the DCC. In this scenario, many genes not binding the DCC will exhibit lower expression in males than females. Furthermore, the finding that MSL levels positively correlate with the level of dosage compensation is important. This implies that many genes may not absolutely require an exact balancing of dosage between males and females, and partial compensation is sufficient to balance fitness between the sexes. An extension of this theory is that only those genes for which dosage is critical may have evolved the capacity to attract a large amount of DCC. Parallels to mammalian dosage compensation can be drawn, where leaky inactivation in the mammalian system (Carrel and Willard 2005) may be mirrored by incomplete activation in *Drosophila*.

Mechanism of dosage compensation

Our observation that there is no more polymerase on the X chromosome than autosomal sequences is consistent with previous suggestions that the DCC does not increase the amount of polymerase loading or promoter clearance, but rather the speed at which a transcript is completed (i.e., transcription elongation) (Smith et al. 2001). Our detection of negligible amounts of DCC on promoters further supports this conclusion. The observed 3' bias of DCC binding within many genes seen in this study also favors the idea that the DCC operates by assisting transcription elongation. Increased passage of polymerase through X-chromosomal genes is consistent with the elevated levels of H3.3 found on the X, purportedly

due to replication-independent replacement of canonical H3 by H3.3 during transcription (Mito et al. 2005). The observation that tRNA genes are bound by the DCC also suggests that the mechanism of dosage compensation may be applicable to both pol II and pol III. The acetylation of H4K16 by the DCC may serve to increase the rate of polymerase progression through chromatin, for example, by reducing polymerase pausing. Pausing has been noted for all RNA polymerases, and is exacerbated on chromatin templates (for review, see Sims et al. 2004). The acetylation of H4K16 in the middle and 3' end of two X-chromosomal genes suggests that H4K16 acetylation may follow a similar pattern to the DCC within genes themselves (Smith et al. 2001). We have shown that H4K16 acetylation is very similar to the binding of the DCC at intermediate, restriction fragment resolution. Chromosome-wide mapping of H4K16 acetylation at high resolution on a genome-wide scale is therefore a priority.

A speculative model of DCC targeting based on a two-step (loading and spreading) mechanism

The DNA sequence elements capable of predicting DCC binding consist of many combinations of different motifs. To recognize such a variety of different sequences, the DCC must allow promiscuous protein–DNA sequence recognition. Binding sites would, in this scenario, be determined by the concentration of many recognition elements in a particular DNA sequence. Our observations are compatible with the “affinities” model for targeting of the DCC (Demakova et al. 2003; Fagegaltier and Baker 2004; Oh et al. 2004b; Dahlsveen et al. 2006). However, affinities alone fail to explain the recent observation that the MSL2 protein exhibits exceptionally stable binding to the male X chromosome, assayed by FRAP and FLIP (Straub et al. 2005c), as sites of low affinity would be expected to demonstrate highly dynamic binding to the MSL complex.

The motifs identified also do not explain the 3' bias. While the MSL1-binding profile suggests a 3' accumulation of the DCC, we cannot exclude that we observe instead a 5' depletion of DCC similar to histone depletion around promoters (Mito et al. 2005). Interestingly, similar protein gradients have been observed of yeast cohesins, which form a ring around the DNA helix (Ivanov and Nasmyth 2005). It has been suggested that these cohesin rings can slide along the DNA, and may be “pushed” by transcribing polymerase to the 3' ends of genes, with their resulting accumulation at sites of convergent transcription (Lengronne et al. 2004). However, we see no accumulation of DCC between convergently transcribed genes as reported for the cohesins.

We speculate that the DNA elements of highest affinity responsible for initially targeting the complex to the X (recognition elements) may be sites for loading of DCC onto the X chromosome. The DCC may form a topological linkage around DNA at these points, similar to that proposed for the cohesins. From these loading points, the DCC may spread to sites of lower affinity, its contact

with DNA stabilized by a ring-like structure, allowing the promiscuous yet stable binding.

Materials and methods

Clones

The cosmid and BAC clones used in Southern analysis are detailed in Supplementary Table 4. The details required to obtain clones can be found in the Supplemental Material. Clones were grown in LB agar or liquid media plus 25 µg/mL chloramphenicol (BACs) or 25 µg/mL kanamycin (cosmids). DNA was isolated using Qiagen maxi-prep kits.

ChIP

ChIPs were performed on chromatin prepared from 12- to 14-h-old mixed-sex Oregon-R embryos. ChIP was performed as described previously, including purification over a CsCl gradient (Kageyama et al. 2001). For control purposes, chromatin was also prepared using a CsCl-free protocol (Schwartz et al. 2005), but no differences to CsCl-prepared chromatin were seen (data available on request). Embryos were fixed in 4% formaldehyde at 18°C for 15 min. Affinity-purified antibodies against MSL1 (2 µL/IP) and MOF (5 µL/IP) were gifts from M. Kuroda (Howard Hughes Medical Institute, Boston, MA). Anti-CTCF (16 µL/IP) was a gift from R. Renkawitz (Justus-Liebig-Universität Gießen, Giessen, Germany). Anti-pol II H-224 SC-9001X (5 µL/IP) and control rabbit IgG antibodies SC-2027 (5 µL/IP) were purchased from Santa Cruz Biotech. Antibodies against acetyl H4K16 AHP417 (3 µL/IP) were purchased from Serotec. Following immunoprecipitation and reversal of cross-links, recovered DNA was resuspended in a final volume of 22 µL of H₂O. Seven microliters of this DNA was incubated with *Pfu* polymerase, ligated to linker, and subjected to linker-mediated PCR prior labeling for use in Southern or microarray hybridization.

Southern transfer and hybridization

BAC and cosmid clones were digested with two restriction enzymes, selected to yield a nonoverlapping digestion pattern as detailed in Supplementary Table 4. Approximately 1 µg of cosmid or 2 µg of BAC DNA was loaded per lane on 0.8% agarose gels. Loading was equalized to allow for differences in insert size by normalizing to a restriction fragment common to vector backbone. Electrophoresis, Southern transfer, and hybridization were performed as previously described (Straub et al. 2005b). Following hybridization, signals from individual bands were quantified using a Fuji FLA-3000 PhosphorImager. Low-intensity hybridization signals (less than three standard deviations above filter background) in the specific IP were discarded and assigned an arbitrary enrichment value of 1. Dividing the values from the specific MSL ChIP by those of the mock IP allowed calculation of fold enrichment.

Microarray design, hybridization, and data normalization

(1) X-chromosome tiling arrays. Array design, production, and hybridization were undertaken by NimbleGen Systems Inc. (<http://www.nimblegen.com>) as part of a ChIP Array Service. Experiments consisted of three biological replicates (three independent cross-linked chromatin preparations and IPs), each hybridized to a separate microarray containing two oligonucleotides (one forward and one reverse-complement) for each chromosomal position. For each hybridization, NimbleGen Systems returned raw data signal intensities for specific IP and input

DNA, plus a log₂ ratio of the IP to input signals. Forward and reverse oligo signals for each genomic position were averaged, then the data from all three hybridizations were normalized by quantile normalization using aroma (<http://www.maths.lth.se/help/R/aroma>) and the R statistical package (<http://www.R-project.org>). (2) cDNA arrays. Four independent MSL1 CHIPs and corresponding mock IP ChIP samples were labeled with either Cy3 or Cy5. One sample was labeled in the reversed dye orientation (i.e., MSL IP with Cy5 and mock IP with Cy3). DNA was hybridized to the fly 12k cDNA array (J. Delrow, Fred Hutchinson Cancer Research Center, Seattle, WA; for more information, see the Gene Expression Omnibus Accession viewer, <http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GPL1908>). Image analysis was done with ImaGene software (BioDiscovery). The data were normalized to the median log₂-ratio of the entire array. The Cyber-T algorithm (Baldi and Long 2001) was used to determine the *P*-value for every probe. Because of the large sample size (10k), the *P*-values need to be corrected for multiple hypothesis testing. Therefore, the False Discovery Rate (FDR) was determined (Benjamini et al. 1995). The cutoff for the FDR was set at 0.01.

Motif searches

PLS regression was performed using the PLS package: Ron Wehrens and Bjørn-Helge Mevik (2005), Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR), R package version 1.1-0 (<http://mevik.net/work/software/pls.html>). Four megabases of X chromosome and 1 Mb of chromosome 3R were divided into overlapping 2-kb segments. For each segment, average MSL1 binding and the frequency of all possible hexamers were calculated. PLS regression was performed using hexamer frequencies as predictor and the average MSL1 binding as response variables. Best results were obtained using the SIMPLS algorithm and a maximum of three components. In addition, we searched for putative dosage compensation DNA motifs, or combinations thereof, using MEME, AlignAce, and MOST (see Supplemental Material for references).

Acknowledgments

We thank M. Kuroda and R. Renkawitz for generous gifts of antibodies. Clones were kindly provided by V. Orlando, E. Madueno, J. Modolell, R.D.C Saunders, B. Minana, I. Siden-Kiamos, and L. Spanos. We also thank M. Boutros and M. Kuroda for sharing unpublished data; J. Delrow for providing cDNA arrays; and C. Regnard, G. Längst, M. Prestel, and S. Gilfillan for critical reading of this manuscript. Work in the laboratory of P.B.B. was supported by the Deutsche Forschungsgemeinschaft through Transregio5, the Leibniz Programme, and "The Epigenome" European Network of Excellence. Work in the B.v.S. laboratory is supported by "The Epigenome" Network of Excellence and a EURYI Award.

References

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- 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.
- 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.* (this issue).
- Arbeitman, M.N., Furlong, E.E., Imam, F., Johnson, E., Null, B.H., Baker, B.S., Krasnow, M.A., Scott, M.P., Davis, R.W., and White, K.P. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**: 2270–2275.
- Baker, B.S., Gorman, M., and Marin, I. 1994. Dosage compensation in *Drosophila*. *Annu. Rev. Genet.* **28**: 491–521.
- Baldi, P. and Long, A.D. 2001. A Bayesian framework for the analysis of microarray expression data: Regularized *t*-test and statistical inferences of gene changes. *Bioinformatics* **17**: 509–519.
- Bashaw, G.J. and Baker, B.S. 1995. The *msl-2* dosage compensation gene of *Drosophila* encodes a putative DNA-binding protein whose expression is sex specifically regulated by Sex-lethal. *Development* **121**: 3245–3258.
- Benjamini, Y., Hochberg, Y., Storey, J.D., and Tibshirani, R. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**: 289–300.
- Birchler, J.A., Owenby, R.K., and Jacobson, K.B. 1982. Dosage compensation of serine-4 transfer RNA in *Drosophila melanogaster*. *Genetics* **102**: 525–537.
- Bone, J.R., Lavender, J., Richman, R., Palmer, M.J., Turner, B.M., and Kuroda, M.I. 1994. Acetylated histone H4 on the male X chromosome is associated with dosage compensation in *Drosophila*. *Genes & Dev.* **8**: 96–104.
- Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Consortium, H.F., Paro, R., and Perrimon, N. 2004. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**: 832–835.
- Carrel, L. and Willard, H.F. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**: 400–404.
- Chakalova, L., Carter, D., Debrand, E., Goyenechea, B., Horton, A., Miles, J., Osborne, C., and Fraser, P. 2005. Developmental regulation of the β -globin gene locus. *Prog. Mol. Subcell. Biol.* **38**: 183–206.
- Charlesworth, B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**: 149–162.
- 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.
- 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.
- Fagegaltier, D. and Baker, B.S. 2004. X chromosome sites autonomously recruit the dosage compensation complex in *Drosophila* males. *PLoS Biol.* **2**: e341.
- Franke, A. and Baker, B.S. 1999. The *rox1* and *rox2* RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol. Cell* **4**: 117–122.
- Gilfillan, G.D., Dahlsveen, I.K., and Becker, P.B. 2004. Lifting a chromosome: Dosage compensation in *Drosophila melanogaster*. *FEBS Lett.* **567**: 8–14.
- Hamada, F.N., Park, P.J., Gordadze, P.R., and Kuroda, M.I. 2005. Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes & Dev.* **19**: 2289–2294.
- Hart, C.M. and Laemmli, U.K. 1998. Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.* **8**: 519–525.
- Hey, J. and Kliman, R.M. 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**: 595–608.
- Hilfiker, A., Hilfiker, K.D., Pannuti, A., and Lucchesi, J.C. 1997. *mof*, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J.* **16**: 2054–2060.
- Huiser, P., Hennig, W., and Dijkhof, R. 1987. Poly (dC-dA/dG-dT) repeats in the *Drosophila* genome: A key function for dosage compensation and position effects? *Chromosoma* **95**: 209–215.
- Ivanov, D. and Nasmyth, K. 2005. A topological interaction between cohesin rings and a circular minichromosome. *Cell* **122**: 849–860.
- Jin, Y., Wang, Y., Johansen, J., and Johansen, K.M. 2000. JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J. Cell Biol.* **149**: 1005–1010.
- Kageyama, Y., Mengus, G., Gilfillan, G., Kennedy, H.G., Stuckenholtz, 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. and Kuroda, M.I. 2000. The role of chromosomal RNAs in marking the X for dosage compensation. *Curr. Opin. Genet. Dev.* **10**: 555–561.
- 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.
- Kotlikova, I.V., Demakova, O.V., Semeshin, V.F., Shloma, V.V., Boldyreva, L.V., Kuroda, M.I., and Zhimulev, I.F. 2006. The *Drosophila* dosage compensation complex binds to polytene chromosomes independently of developmental changes in transcription. *Genetics* **172**: 963–974.
- Law, A., Hirayoshi, K., O'Brien, T., and Lis, J.T. 1998. Direct cloning of DNA that interacts in vivo with a specific protein: Application to RNA polymerase II and sites of pausing in *Drosophila*. *Nucleic Acids Res.* **26**: 919–924.
- Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. 2004. Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**: 573–578.
- Lowenhaupt, K., Rich, A., and Pardue, M.L. 1989. Nonrandom distribution of long mono- and dinucleotide repeats in *Drosophila* chromosomes: Correlations with dosage compensation, heterochromatin, and recombination. *Mol. Cell. Biol.* **9**: 1173–1182.
- Lucchesi, J.C. 1998. Dosage compensation in flies and worms: The ups and downs of X-chromosome regulation. *Curr. Opin. Genet. Dev.* **8**: 179–184.
- 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.
- MacAlpine, D.M., Rodriguez, H.K., and Bell, S.P. 2004. Coordination of replication and transcription along a *Drosophila* chromosome. *Genes & Dev.* **18**: 3094–3105.
- 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.
- Mito, Y., Henikoff, J.G., and Henikoff, S. 2005. Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **37**: 1090–1097.
- Morales, V., Straub, T., Neumann, M.F., Mengus, G., Akhtar, A., and Becker, P.B. 2004. Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. *EMBO J.* **23**: 2258–2268.
- Oh, H., Bai, X., Park, Y., Bone, J.R., and Kuroda, M.I. 2004a. Targeting dosage compensation to the X chromosome of *Drosophila* males. *Cold Spring Harb. Symp. Quant. Biol.* **69**: 81–88.
- Oh, H., Bone, J.R., and Kuroda, M.I. 2004b. Multiple classes of MSL binding sites target dosage compensation to the X chromosome of *Drosophila*. *Curr. Biol.* **14**: 481–487.
- Orphanides, G. and Reinberg, D. 2000. RNA polymerase II elongation through chromatin. *Nature* **407**: 471–475.
- O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. 2004. Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* **36**: 1014–1018.
- Pardue, M.L., Lowenhaupt, K., Rich, A., and Nordheim, A. 1987. (dC–dA)_n(dG–dT)_n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *EMBO J.* **6**: 1781–1789.
- 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.
- Sass, G.L., Pannuti, A., and Lucchesi, J.C. 2003. Male-specific lethal complex of *Drosophila* targets activated regions of the X chromosome for chromatin remodeling. *Proc. Natl. Acad. Sci.* **100**: 8287–8291.
- Schwartz, Y.B., Kahn, T.G., and Pirrotta, V. 2005. Characteristic low density and shear sensitivity of cross-linked chromatin containing polycomb complexes. *Mol. Cell. Biol.* **25**: 432–439.
- Sims III, R.J., Belotserkovskaya, R., and Reinberg, D. 2004. Elongation by RNA polymerase II: The short and long of it. *Genes & Dev.* **18**: 2437–2468.
- Singh, N.D., Davis, J.C., and Petrov, D.A. 2005. X-linked genes evolve higher codon bias in *Drosophila* and *Caenorhabditis*. *Genetics* **171**: 145–155.
- Sipos, L. and Gyurkovics, H. 2005. Long-distance interactions between enhancers and promoters. *FEBS J.* **272**: 3253–3259.
- 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.
- Smith, E.R., Allis, C.D., and Lucchesi, J.C. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J. Biol. Chem.* **276**: 31483–31486.
- Spellman, P.T. and Rubin, G.M. 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1**: 5.
- 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.
- Straub, T., Dahlsveen, I.K., and Becker, P.B. 2005a. Dosage compensation in flies: Mechanism, models, mystery. *FEBS Lett.* **579**: 3258–3263.
- Straub, T., Gilfillan, G.D., Maier, V.K., and Becker, P.B. 2005b. The *Drosophila* MSL complex activates the transcription of target genes. *Genes & Dev.* **19**: 2284–2288.
- Straub, T., Neumann, M.F., Prestel, M., Kremmer, E., Kaether, C., Haass, C., and Becker, P.B. 2005c. Stable chromosomal association of MSL2 defines a dosage-compensated nuclear compartment. *Chromosoma* **114**: 352–364.
- Taipale, M. and Akhtar, A. 2005. Chromatin mechanisms in *Drosophila* dosage compensation. *Prog. Mol. Subcell. Biol.* **38**: 123–149.
- Wold, H. 1975. Soft modeling by latent variables: The non-linear iterative partial least squares approach. Perspectives in Probability and Statistics. In *Papers in Honour of M.S. Bartlett* (ed. J. Gani), pp. 117–144. Academic Press, London.
- Zhou, S., Yang, Y., Scott, M.J., Pannuti, A., Fehr, K.C., Eisen, A., Koonin, E.V., Fouts, D.L., Wrightsman, R., Manning, J.E., et al. 1995. Male-specific lethal 2, a dosage compensation gene of *Drosophila*, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. *EMBO J.* **14**: 2884–2895.