Title: Non-canonical *Drosophila* X chromosome dosage compensation and repressive topologically-associated domains.

Authors: Hangnoh Lee and Brian Oliver

Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Kidney and Digestive Diseases, National Institutes of Health, Bethesda, Maryland, USA

Email addresses:

hangnoh.lee@nih.gov (HL), briano@niddk.nih.gov (BO)

Abstract

Background: In animals with *XY* sex chromosomes, *X*-linked genes from a single *X* chromosome in males are imbalanced relative to autosomal genes. To minimize the impact of genic imbalance in male *Drosophila*, there is a dosage compensation complex (MSL), that equilibrates *X*-linked gene expression with the autosomes. There are other potential contributions to dosage compensation. Hemizygous autosomal genes located in repressive chromatin domains are often de-repressed. If this homolog-dependent repression occurs on the *X*, which has no pairing partner, then de-repression could contribute to male dosage compensation.

Results: We asked whether different chromatin states or topological associations correlate with *X* chromosome dosage compensation, especially in regions with little MSL occupancy. Our analyses demonstrated that male *X* chromosome genes that are located in repressive chromatin states are depleted of MSL occupancy, however they show dosage compensation. The genes in these repressive regions were also less sensitive to knockdown of MSL components.

Conclusions: Our results suggest that this non-canonical dosage compensation is due to the same trans-acting de-repression that occurs on autosomes. This mechanism would facilitate immediate compensation during the evolution of sex chromosomes from autosomes. This mechanism is similar to that of *C. elegans*, where enhanced recruitment of *X* chromosomes to the nuclear lamina dampens *X* chromosome expression as part of the dosage compensation response in *XX* individuals.

Keywords

Dosage Compensation, Drosophila melanogaster, MSL complex, Topologically-associated domain, Lamina-associated domain

Background

Genes come in pairs and large-scale deviation from this state is detrimental, most probably as a result of disrupted gene expression balance [1,2]. Sex chromosomes are a peculiar exception to this general rule. In *XY* systems, males have what amounts to a heterozygous deletion of an entire chromosome, bearing ~20% of the genes in the case of Drosophila, with no impact on fitness. In such systems, compensation often rectifies gene dose effects as a way to maintain gene balance [3–5].

In *Drosophila melanogaster*, a male-specific complex called the Male Specific Lethal (MSL) complex, plays a role in increasing expression of genes from the single *X* chromosome relative to autosomes. MSL and perhaps other unidentified sources of compensation ultimately achieve remarkably equalized levels of *X*-linked gene expression in males with one *X* and females with two *X*s, as well as balancing *X* expression with the autosomes [6,7]. This boosting of expression by MSL complex is primarily achieved via enhanced elongation of transcription [8], but there also is evidence that RNA polymerase II (PolII) binding is increased by 1.2 fold at male *X* chromosome promoters [9–11]. The complex includes MSL-1, MSL-2, and MSL-3 proteins, Maleless (MLE), and Males absent on the first (MOF) proteins, and two non-coding RNAs, *roX1* and *roX2 [12]*. MOF has a histone acetyltransferase activity and functions in enhanced elongation of X chromosome gene transcription by acetylating histone H4K16 (H4K16Ac) [8]. Binding of MSL complex to the male *X* chromosome occurs at Chromosome Entry Sites (CES), also referred to as High-Affinity Sites (HAS) [13,14]. The sites contain GA-rich sequences, called the MSL-recognition element (MRE) [14].

A related scheme for *X* chromosome dosage compensation occurs in *C. elegans*, where *XX* worms are hermaphrodites and *X0* worms are males. In *XO* males, yield of *X* chromosome gene products is increased using various mechanisms (e.g. increased PolII recruitment, mRNA stability, or translation rate) in both males and hermaphrodites [15–18]. However, solving the gene production difference between autosomes and *X* chromosomes in males results in over-expression in *XX* animals. To manage this increased activity, *XX* hermaphrodite *C. elegans* have a dosage compensation complex that represses gene expression from both *X* chromosomes [5,16]. The *C. elegans* Dosage Compensation Complex (DCC) targets the *X* chromosomes and spreads from recruitment sites on the *X* [19]. Recruitment of DCC on *X* chromosome is linked to increased mono-methylation of histone H4K20 (H4K20me1) [20,21], as

well as depletion of histone modifications that mark active transcription, such as H4K16Ac [16,22,23], and H2A.Z variant histone [24]. These epigenetic changes accompany topological remodeling of the *X* chromosomes [25] and reduced PolII recruitment at *X*-linked promoters in hermaphrodites. [3,5,26]. This remodeling includes nuclear sub-localization of the *X* chromosomes to the lamina, which is generally repressive. Disruption of the anchoring between heterochromatin and nuclear lamina re-localized *X* chromosomes more centrally in the nucleus and results in partial de-repression of the *X*-linked genes [27]. The modulation of H4K16Ac in animals with a single *X* is a conserved characteristic between *D. melanogaster* and *C. elegans* [22] although the *XX* mechanisms differ [23].

Intriguingly, nuclear architecture-level de-repression also occurs in autosomal dosage compensation in *D. melanogaster*. Genes within repressive "topologically associated domains" (TADs), which include Lamina-associated domains (LADs) involved in dosage compensation in *C. elegans* hermaphrodites, show better autosomal dosage compensation in *Drosophila* hemizygotes [28]. The effect of autosomal deletions is de-repression of the non-deleted genes *in trans*, as well as a spreading of de-repression into flanking regions within the LAD. This suggests that these repressive domains are built based on additive or synergistic cooperation between gene homologs. This observation is of particular interest for two reasons. First, the necessity of two homologs for the repression is reminiscent of chromosomal pairing-dependent events, such as pairing-sensitive silencing [29,30], or transvection [31,32]. In transvection, the existence of homologous chromosome in proximity leads to enhancer action *in trans* or insulator bypass *in cis* [31]. As such, chromosomal pairing may provide a mechanistic basis of how autosomal deletions result in the de-repression of non-deleted genes [28]. The absence of a pairing partner for the single *X* in males might therefore be consequential. Second, the repression at the two-dose state, and de-repression at one-dose state, is analogous to *X* chromosome dosage compensation in *C. elegans*. This led us to ask if the de-repression of one-dose genes from repressive domains occurs on *D. melanogaster X* chromosomes. If so, this would contribute to dosage compensation in males.

Results

X-linked repressive TADs genes display low expression levels, but are dosage compensated in males.

To determine the overall structure of chromatin domains on the *X*, we used results from three previous studies that divided genome into repressive vs. non-repressive chromatin domains/TADs and LADs vs. non-LADs. LAD and DamID (DNA adenine methyltransferase Identification) based chromatin occupancy information was from *Kc* cells [33,34]. TAD information was from Hi-C conformation capture from mixed sex embryos [35]. From the Hi-C study, "Null" TADs were characterized by general lack of chromatin marks, except for a weakly enriched binding of an insulator protein, Suppressor of Hairy-wing [SU(HW)]. The LAD and "Null" TADs correspond and largely overlap with "Black" domain DamID work. The "Black" domain has increased signals of Histone H1, Effete (EFF), Suppressor of Under-Replication (SUUR) and Lamin B protein binding. These repressive TADs are known to share various characteristics [35], and there are significant overlaps among the identified gene sets (**Figure 1A**). For example, 63% of genes that are in LADs are also in Black domains, and 78% of genes that are in Black domains are in Null domains. We collectively refer to these overlapping domains as "repressive TADs".

Each of these three repressive TADs covered 23 to 47% of the protein-coding genes in the *Drosophila* genome. To describe which genes on the *X* were in repressive TADs, we parsed by chromosome (**Figure 1B**). Collectively, genes within LADs included 38% of *X* chromosome genes and 29% of autosomal genes (p = 1.19e-06, Fisher's exact test). Genes within Null domains included 45% of *X* chromosome genes and 47% of autosomal genes (p = 0.25). Genes within Black domains included 33% of *X* chromosome genes and 35% of autosomal genes (p = 0.076). Clearly, a large fraction of the genome, including the *X*, are in repressive domains. If these genes are simply "off", then asking if they are dosage compensated is a futile effort (2 x 0 = 0). Therefore, we carefully examined expression levels from genes that are within the repressive domains to see if we could reliably detect expression. We used previously reported expression data for this analysis [36,37]. Expression levels in repressive domains were reassuringly lower than in non-repressive domains. We found these trends of lower expression in repressive-TAD genes when we investigated different cell lines (**Figure 1C-E**) and sexed salivary glands (**Figure 1F-H**), but there was clear evidence of expression.

Determining the difference between low and off is critical for this analysis. We measured the biological and technical noise levels by measuring intergenic signals (**Figure 1C-E**). The 99th percentiles for intergenic signal were 0.87 Fragments Per Kilobase of transcript per Million mapped reads (FPKM) in *S2* cells (male) and 0.98 FPKM in *Kc* cells (female). This is in stark

contrast to expression levels in the repressive TADs from *Kc* cells, where LAD and Black domains were determined (**Figure 1C-E**, the top panel). The median gene expression level was 9.1 FPKM for genes within LADs and 16.1 FPKM for genes within Null domains. Genes in Black domains showed lower expression at a median of 2.8 FPKM, but all of these expression levels far exceed our estimates of noise. In *Kc* cells, approximately 19.6% and 40.4% of the *X*-linked genes demonstrate gene expression above the cutoff levels from LAD and Null domains, respectively. Only 5.7% of the *X*-linked genes were expressed from Black domains, indicating that the Black domain has the most repressive characteristics among three different calls of repressive TADs. Autosomal genes from repressive TADs also displayed lower gene expression levels compared to non-repressive TAD genes with 9.7 (LAD), 16.1 (Null), and 2.7 FPKMs (Black), which are not significantly different from repressive TAD genes on the *X* ($p > 0.2$, Mann-Whitney U test). In male *S2* cells, the repressive TAD genes demonstrated 9.8 (LAD), 15.4 (Null), and 5.1 FPKMs (Black), underscoring that repressive TAD genes on the *X* have comparable expression levels between *S2* and *Kc* cells ($p > 0.05$, **Figure 1C-E**), and are thus dosage compensated. We made a similar observation from sexed salivary glands. A large fraction of genes from repressive TADs showed higher expression than the technical noise level, which we determined based on backgrounds signals from the control probes of microarrays (**Figure 1F-H**, normalized intensities of approximately 2.3 in both sexes). For example, about 27.4% and 26.6% of the total *X*-linked LAD genes showed gene expression above the background levels in female and male salivary glands, respectively. Thus, we were confident that we could make meaningful measurements of dosage compensation among those genes. Briefly, a substantial portion of the genes in repressive TAD domains showed detectable levels of gene expression. We used these genes in our analysis.

Genes in repressive TADs demonstrated comparable expression levels between female (*Kc*) and male (*S2*) cells from the *X* (**Figure 1C-E**), indicating that they are dosage compensated. To confirm that male *X*-linked genes are dosage compensated in repressive TADs, we compared expression profiles from salivary glands, where female and male were matched siblings. From microarray results, we observed that male *X*-linked genes from LAD regions demonstrated comparable expression levels to that of females (**Figure 1I**). The median signal intensity from male *X*-linked genes was 3.41, which did not differ from that of female (3.58, $p = 0.917$) despite the 50% difference in *X* gene dose. We obtained similar equilibrated expression of the *X* from other repressive TADs. *X* chromosome genes in the Null domains showed a median of 5.51 signal intensity in *X* males when it was 5.31 in *XX* females. Black domain genes had medians

of 2.68 and 2.64 signal intensities in *X* males and *XX* females, respectively. Overall gene expression signals from autosomes were consistent between two sexes (6.52, $p > 0.996$ for differential expression). Therefore, the repressive TAD genes are dosage compensated in males.

Repressive TAD genes lack MSL complex binding

X specific dosage compensation has canonical and non-canonical components. If canonical dosage compensation is active in repressive domains, the MSL complex should occupy those regions. To address this possibility, we first investigated chromatin occupancy by MOF, the key writer of the H4K16Ac mark [38] in the MSL complex [12]. MOF also has an MSL-independent role in regulating a smaller subset of genes in both sexes by participating in Non-Specific Lethal (NSL) complex [39]. We analyzed genome-wide Chromatin Immunopreciptation (ChIP) results [36,40] to determine occupancy of the MSL complex as well as H4K16Ac levels in tissue culture cells and salivary glands (we measured MOF and H4K16Ac enrichment within gene bodies because both MOF and H4K16Ac display broad enrichment patterns over these features [39]). Strikingly, in male *S2* cells, MOF binding in *X* chromosome repressive TADs was significantly lower than elsewhere on the *X* ($p < 6.01e-4$) and was comparable to occupancy levels on the *X* in female *Kc* cells (**Figure 2A-C**). H4K16Ac enrichment concurred with MOF occupancy. In all classes of *X* chromosome repressive TADs, H4K16Ac levels were significantly lower than in other domains ($p < 6.96e-09$, **Figure 2D-F)**. H4K16Ac levels on *X*-linked genes was still higher in *S2* cells than *Kc* cells even within repressive TADs ($p < 3.02e-09$). However, the differences in H4K16Ac levels between male and female cells were significantly smaller in LAD and Null domains than non-repressive TADs on the *X* ($p < 7.14e-04$, Permutation test). The Black domain genes showed the same trend, although this was not statistically significant ($p = 0.27$). Consistent with the MOF occupancy and H4K16Ac enrichment, we observed that MSL-1 occupancy was lower in genes within repressive TADs on the *X* compared to non-repressive TADs ($p < 1.11e-12$, Mann-Whitney U test, **Figure 2G-I**). Thus, the occupancy and activity of MSL complex is reduced in the case of the dosage compensated *X*-linked genes in *S2* cell repressive TADs.

To examine MSL complex activity at the repressive TADs in tissues, we analyzed ChIP results from sexed larval salivary glands. In males, X chromosome MOF binding was significantly higher at gene bodies in non-repressive TADs, compared to repressive TADs (**Figure 2J-L**, $p <$

2.2e-16). If MOF binding is functional, then the H4K16Ac mark should follow a matching enrichment pattern. Indeed, H4K16Ac levels were higher at genes in non-repressive TADs compared to repressive TADs (**Figure 2M-O**, $p < 2.2e-16$). As we observed in the tissue culture cells, the basal level of H4K16Ac was higher in repressive TADs of the male salivary glands, compared to that of female glands, despite the fact that MOF occupancy was not significantly higher in LAD and Black domains ($p > 0.63$, **Figure 2P,Q)**. This observation indicates that regulation of repressive TAD genes on the *X* chromosome occurs with limited or transient access to MSL complex, but can involve modulations of H4K16Ac in a canonical manner. We explore the degree of function associated with this lower H4K16ac in a later section.

Since the genes within repressive TADs have low occupancy of MSL complex and lower H4K16ac, we wondered if repressive TADs lack genomic signatures that are required for MSL complex binding. Specifically, we asked if lower MOF activity correlates with lower density of the MSL complex entry sites in repressive domains. *Drosophila* MSL complex specifically binds to *X*, which occurs at CES [14]. CES contains GA-rich DNA sequence motif, called MRE, whose introduction to an autosome results in local recruitment of MSL complex to that site [14]. We identified 11,306 MRE motifs from the *X* chromosome of the reference genome (using an *E*-value < 10e-5 cutoff). The number of *X* chromosome MREs in repressive domains was not statistically different from random (**Figure 2R**, $p > 0.1$ Permutation test), indicating that the repressive TADs are not free of MRE motifs. However, when we investigated if genes in repressive TADs recruit MSL complex to their chromatin regions, we found only 20 overlaps between LADs and the 150 CES (approximately 57 expected, $p << 0.001$, Permutation test, **Figure 2S**) that recruit MSL [14]. We obtained consistent results from Null and Black domains (**Figure 2R,S**). These observations suggest that, on male *X* chromosomes, MSL complex does not efficiently bind genes within the repressive TADs.

X-linked repressive TADs genes are less sensitive to disruption of MSL-complex functions compared to the canonical dosage-compensation target genes.

If the repressive TAD genes are dosage-compensated in a non-canonical way on the male *X* chromosome, such genes might be indifferent to MSL complex function. If the low level of H4K16ac is matched to the generally low level expression of the genes in repressive domains, compensation of such genes should depend on MSL function. To investigate the impact of disrupted MSL complex function on *X*-linked genes in repressive TAD domains, we analyzed

gene expression profiles of S2 cells whose MSL components were selectively depleted via RNAi-mediated knockdown [36,41,42]. When *mof* mRNA was depleted, *X*-linked genes within LADs displayed significantly less gene expression reduction; approximately 1.2 fold higher than genes in non-LAD domains (p = 1.1e-13, Mann-Whitney U test, **Figure 3A**). We made similar observations from *X* chromosome genes that belong to Null and Black domains from the Hi-C study and occupancy study. They exhibited about 1.1 to 1.3 fold more expression upon the depletion of *mof* than other *X*-linked genes in non-repressive domains (p < 2.3e-08). As expected, those chromatin regions that lack MOF binding and H4K16Ac were less sensitive to the RNAi treatment as well. Genes from regions without enrichment of MOF or H4K16Ac showed 1.2 fold more expression than the enriched regions (p = 1.1e-14 for MOF, and 0.11 for the acetylation). MOF is also bound to sites on autosomes as a part of NSL complex, while it activates only a small subset of genes that the complex binds to [43]. Consistent with this idea, we saw little down-regulation of overall autosomal gene expression from the *mof* depleted *S2* cells (p > 0.05, **Figure 3B**).

We also asked if the expression of *X*-linked genes in repressive TADs was less sensitive to depletion of other MSL components. Our analysis showed significantly less reduction in expression in gene within repressive TADs, relative to non-repressive TADs, when *msl-1* mRNA was depleted (**Figure 3C,D**, p < 0.001). Similarly, *msl-2* and *msl-3* knockdown caused 1.1 to 1.2 fold more *X* chromosome gene expression from genes in repressive TADs, compared to non-repressive TADs (**Figure 3E-H**, p < 0.01). These results were not due to the inaccurate detection of low abundant transcripts in hybridization-based techniques (i.e. microarrays) [44,45]. When we analyzed an independent study that performed RNA-Seq analysis of either *mof* or *msl-2* depleted *S2* cells, we also observed about 1.1 fold more expression from the *X*-linked genes within repressive TADs compared to non-repressive TADs (p < 0.001, **Figure 3I-L**). Therefore, results from the MSL knockdown were consistent with our observation in **Figure 2** and suggest that repressive TAD genes on the *X* chromosome do not rely entirely on MSL complex for dosage compensation.

We inspected MOF occupancy and H4K16Ac enrichment at individual genes in repressive TADs on the *X* to determine the patterns of occupancy across the gene bodies. Genes that were clearly regulated by the canonical dosage compensation machinery (i.e. MSL-dependent) display broad enrichment signals of MOF and H4K16Ac across the gene body regions, whereas MSL-independent MOF target genes (e.g. MOFs in NSL complex) show promoter-enriched

MOF binding patterns [39]. We observed genes that were sensitive to the *msl* or *mof* knockdown, for example *CG9947* and *arm*, had broad ChIP signals of MOF and H4K16Ac in contrast to an autosomal gene, *RpL32*, which has MOF enrichment only at its promoter region (**Figure 4A-C**). Compared to those canonical MSL target genes, the genes in repressive TAD regions showed absent MOF binding (**Figure 4D,E**, *CG34330* and *CG9521*), or weak MOF occupancy (**Figure 4F,G**, *CG8675* and *CG2875*). In all four specific cases, the knockdown of *mof* or *msl-2* did not lead to statistically significant reduction of gene expression in males ($p > 0.7$, **Figure 4D-G**), additionally the genes were still fully compensated relative to females in the salivary glands [male/female expression ratios of 1.02 (*CG34330*), 1.02 (*CG9521*), 1.04 (*CG8675*), and 0.97 (*CG2875*)]. For the latter class of genes that have weak MOF occupancy (*CG8675* and *CG2875*), we noticed that MOF also bound at the 3' ends of genes. Furthermore, H4K16Ac enrichment had peaks at the 3' ends as well, which contrasted with the promoter-focused peaks from NSL-MOF target genes (*RpL32*, **Figure 4A**), suggesting that there was some residual MSL activity, rather than NSL. The overall insensitivity to MSL RNAi-depletion suggests that the dosage compensation of these genes does not rely on the MSL complex, but requires additional mechanisms, such as de-repression.

Non-canonical dosage compensation is more evident within TAD boundaries that are maintained between male and female cells.

TAD boundaries are stable across different cell types, and even display evolutionary conservation [46]. Hi-C studies from *Kc* and *S2* cells showed that approximately 74% of TADs are located at the identical positions between the two cell lines [47]. The overall organization of *X* chromosome TADs is highly similar between the two cell lines, and depletion of *msl-2* or *msl-3* does not alter chromatin conformation of the *X* [48]. Nevertheless, it is still possible that the compensation of repressive TAD genes, as well as the increase of histone H4K16Ac (**Figure 2D-F, M-O**), are due to topological differences of chromatin between the two cell lines because the repressive TADs are originally defined from *Kc* cells as well as mixed sex embryos. Testing the possibility is important in this study for two reasons. First, it is possible that our observation could result from erroneous mapping of *Kc* cell-based TAD calls to *S2* cells, although this seems unlikely as MOF occupancy signal is very weak over the repressive TADs in the both cell lines (**Figure 2A-C, J-L**). Second, considering that CESs are enriched at TAD boundaries, from where the MSL complex spreads out based on proximity [48], if TAD structures differ between

male and female cells, the architectural difference could contribute to the non-canonical dosage compensation.

To test the effects of potentially different TAD structure between lines and/or sexes, we investigated *X*-linked gene dosage compensation where TAD locations did not match between the two cell lines. For this purpose, we subclassified repressive TADs into four groups [47], which divided regions that are: within TAD boundaries in both *Kc* and *S2* cell lines (TT); either at boundaries or interspace between two large TADs in both cell lines (II); and within large TADs from only one of the cell lines (TI or IT, **Figure 5A**). The proportion of mismatches in TADs and their boundary locations (e.g. TT and II versus TI and IT) did not show bias based on autosome and *X* chromosome location, even in repressive TAD regions ($p > 0.1$ for all repressive TADs, Fisher's exact test with Bonferroni correction, **Figure 5B-D**). This observation indicates that sex-specific alteration of TAD structures are unlikely to drive non-canonical male *X* chromosome dosage compensation. We also investigated the possible function of sex-specific TADs, by asking if knockdown of MSL components had a greate effect on repressive TADs domains specific to male *S2* cells (IT). However, this was not the case. Instead, we observed that the genes within repressive TAD domains and boundaries whose TAD locations were well-matched between the two cell lines (TT and II) were better compensated than the others (TI and IT) after depleting MSL components (**Figure 5E-G**). For example, LAD-associated genes from TT or II class regions demonstrated excellent dosage compensation after *msl-2* knockdown compared to the same genes from the control RNAi samples (0.94 fold, $p = 0.047$, Mann-Whitney U test, **Figure 5E**). We observed the same trend when we looked at genes from Null and Black domains (**Figure 5F,G**). In conclusion, these results suggest that our observation of non-canonical dosage compensation is not due to sex-specific modification of large TAD structures between male and female cells. TAD that differ between female *Kc* and male *S2* cells may be due to the rearrangements in these highly aneuploid cells, not differences in sexual identity.

Discussion

A subset of *X* chromosome genes that are unbound by MOF still dosage compensate [49]. We have studied *X* chromosome dosage compensation of genes within repressive TADs in *Drosophila*, and their association with MSL dosage compensation complexes and activities. Our results revealed that genes from such repressive TADs are compensated with minimal

339 contributions from MSL. We suggest those regions are able to achieve dosage compensation
340 due to the weaker repressive TADs on the unpaired male chromosome.
341
342 This non-canonical compensation may be the same as observed in the case of autosomal
343 deletions within or at the boundaries of repressive TADs [28]. These deletions have a dominant
344 de-repressing effect, which results in partial dosage compensation for the hemizygous segment,
345 and over-expression of genes in flanking two-dose regions (**Figure 6A,B**). These data suggest
346 that repressive domains are established, strengthened, or stabilized by the existence of
347 homologous pairs of chromosomes. There is strong precedent for pairing-dependent
348 mechanisms in *D. melanogaster* that are known to activate or repress genes when homologous
349 chromosomes are proximally located [29–32]. We suggest that the unpaired *X* chromosomes of
350 males have weaker repressive domains than the same domains in the paired *X* chromosomes
351 of females (**Figure 6C,D**). Thus, one can think of this as dosage compensation mediated by
352 partial *X* inactivation in females, with de-repression in males. This model hinges on the
353 reorganization of the nuclear lamina-DNA interaction, which can clearly regulate gene activities
354 during cell differentiation even in the absence of global changes of the nuclear architecture [50].
355 For example, in mouse embryonic stem cells, loss of the tethering in the *Hdac3* deletion
356 releases genomic regions of lineage-specific genes from nuclear lamina resulting in precocious
357 expression of those genes [51].
358
359 De-repression of one-dose genes in males is reminiscent of the *C. elegans* dosage
360 compensation mechanism (**Figure 6E,F**). In *C. elegans*, *XX* individuals are hermaphrodites and
361 *XO* individuals are males. Both *X* chromosomes in hermaphrodites are subjected to dosage
362 compensation control by repression [3,5,16]. The process involves DCC complex-dependent
363 chromatin remodeling in *XX* hermaphrodites [20–22], that includes enrichment for H4K20me1
364 and depletion for H4K16Ac. In *XO* worms, the *X* shows decondensation [23]. In addition to the
365 chromatin remodeling, there is local positioning of both *X* chromosomes of hermaphrodites to
366 the LADs at the nuclear periphery which contributes to the repression of *X*-linked gene
367 expression; the loss of this tethering results in de-repression of *X*-linked genes in
368 hermaphrodites [27]. The de-repression of *X*-linked genes in tethering mutants of *cec-4* or *lem-*
369 *2*, which encode a chromodomain protein or a component of nuclear lamina, respectively,
370 results in a less extreme compensation phenotype than DCC mutants, raising the possibility that
371 tethering to the nuclear lamina is an additional or supplemental mechanism to achieve dosage

372 compensation by repression in *XX* individuals [27]. Thematically, this is identical to the non-
373 canonical model for *Drosophila* dosage compensation that we propose.
374
375 *X* chromosome dosage compensation by de-repression appears to rely on a general feature of
376 repressive domains, requiring very little evolutionary innovation. As sex chromosomes evolve
377 from an autosomal pair, the sex chromosome specific to the hetergametic sex, becomes
378 recombinationally silent and accumulates inversions, insertions, and pseudogenes that further
379 disrupt pairing [52–54]. As this process occurs, partial dosage compensation by de-repression
380 would be an immediate response, not requiring the evolution of any specific machinery.
381 Improved dosage compensation can evolve to boost gene expression in *XY* males and by
382 enhancing repression in *XX* females. This could account for some of the commonality between
383 *D. melanogaster* and *C. elegans* despite their divergence ~ 1 billion of years ago [55,56].
384
385 The MSL complex does not function specifically on the *X* chromosome in the male germline of
386 *D. melanogaster* [57,58], although they may be dosage compensated [59] (but also see [60]).
387 There is a clear depletion of genes with male biased expression in regions of high MSL
388 occupancy [61], but given that these specific MSL sites do not appear to be used in the male
389 germline, the suggestion that MSL drives these genes to other locations seems spurious. We
390 have shown that the regions without MSL entries sites correspond to the repressive TADs.
391 Thus, we propose that *X*-linked genes with male germline functions are more likely to be in
392 repressive TADs, where they can show increased expression as a result of de-repression.
393 Indeed, in our previous results from gene expression profiling of hemizygote files with
394 autosomal deletions [28], we observed that genes with male-biased expression are de-
395 repressed in females. There has been strong evolutionary pressure to relocate genes with male
396 germline function off the *X* chromosomes [62–64]. Those that remain might use de-repression
397 to achieve high expression even on the single *X*.
398
399 **Conclusion**
400
401 Our results collectively suggest that MSL complex-independent *X* chromosome dosage
402 compensation exists in *Drosophila melanogaster*. We suggest that this non-canonical dosage
403 compensation mechanism involves de-repression of one-dose *X* chromosome genes in males,
404 which are repressed in their two-dose state in females. Our results have an implication for the *X*
405 chromosome dosage compensation mechanism before the evolution of the MSL complex.

Materials and Methods

TADs information used in this study

We obtained LAD information from [65], HiC domains from [35], and DamID-based chromatin domains from [34]. All these results were generated based on *Drosophila* reference genome release 5. We used Flybase 5.57 gene model [66] in describing genes within such TADs. We defined genes to belong to TADs only when both boundaries of a gene locate in a TAD region. We performed our gene ontology analysis in FlyMine version 45.1 [67]. Results in the Additional File 1 represents significantly enriched terms, adjusted p value < 0.05, after Holm-Bonferroni correction. Hi-C based TAD boundary information for *S2* and *Kc* cells were obtained from [47].

Drosophila cell line data from modENCODE studies

We used our previous results on RNA-Seq expression profiles of *Drosophila Kc* and *S2* cells [37] for this study after updating gene IDs to Flybase 5.57. We used FPKM > 1 as an expression cutoff based on the top 99th percentile of the intergenic FPKM signals (0.87 and 0.98 for *Kc* and *S2* cells, respectively). We used following chromatin immunoprecipitation (ChIP)-on-chip results from modENCODE study (model organism ENcyclopedia of DNA Elements, [40]. modENCODE submission IDs 3043 and 3044 for MOF binding in *Kc* and *S2* cells, respectively, ID 318 for Histone H4K16 acetylation in *Kc* cells, and IDs 319 and 320 for *S2* cells. In our description of H4K16 acetylation levels in *S2* cells in **Figure 2**, we used median values from these two different submissions. We obtained MSL-1 binding results from modENCODE submission ID 3293. These datasets can also be obtained from Gene Expression Omnibus (GEO, [68] with these accession IDs: GSE27805-6, GSE20797-9, and GSE32762. modENCODE study [40] provided smoothed log-intensity values between ChIP signal and the input signal, called M values, whose processed mean is shifted to 0. We used median M values within gene boundaries in describing MOF/MSL-1 binding or H4K16 acetylation in **Figure 2A-I**. MOF binding and H4K16 acetylation enriched/not-enriched regions in **Figure 3** directly followed peak calls from the original study.

Salivary gland expression profiles and ChIP-Seq results

We obtained microarray expression profiling and ChIP-Seq results from the 3rd instar larva salivary glands for MOF binding and Histone H4K16 acetylation from [36]. The gene

expression profiles were provided as GCRMA (GC Robust Multi-array Average, [69]-normalized signal intensities, and we used the top 95 percentiles of signals from non-*Drosophila* control probes as an expression cutoff. We demonstrated the median values from three replicates in **Figure 1C-E**. The original results can be found from ArrayExpress [70] with accession ID of E-MEXP-3506. ChIP-Seq results for MOF binding and H4K16 acetylation, from the same study, can be accessed with ArrayExpress ID E-MTAB-911. In the result, the authors performed analysis with DESeq [71] to calculate log2 fold changes between ChIP and input samples for non-overlapping 25 bp windows across the genome. We used median values of such log2 fold changes within gene boundaries in describing the ChIP results in **Figure 2J-O**.

MSL entry sites

We used 150 CES that were characterized by ChIP-chip and ChIP-Seq studies [14] to generate a position weight matrix for DCC binding using MEME (Multiple EM for Motif Elicitation) suite version 4.11.2 [72]. We set the length of the motif to be 21 bp to match with the original CES study. Using the position weight matrix, we identified locations with MREs across the *Drosophila* genome release 5. We used FIMO 4.11.2 (Find Individual Motif Occurrences, [73] in this identification with *Expect* value (*E*-value) threshold of 1.0e-05. In our description of MRE/CES occurrence in **Figure 2**, we randomly shuffled positions of TADs on *X* chromosome genome using Bedtools 2.26.0 [74] while preserved the sizes of TADs. The results in **Figure 2R,S** demonstrate overlap between such shuffled TADs and MRE/CES from 2,000 randomizations.

***S2* cell RNAi results for MSL knockdown**

We used *mof*, *msl-1*, *msl-3* knockdown results from a microarray study [36], ArrayExpress E-MEXP-1505). For the estimation of gene expression changes, we used Robust Multi-array Average (RMA, [75] method for background adjustment and normalization, and filtered out genes of which FPKM value is less than 1 from the *S2* cell RNA-Seq result [37]. We use R limma package version 3.28.21 [76] as in the official manual for our differential expression analysis. We obtained the microarray study of the *msl-2* knockdown data from [41]. We conducted same data handling process as above. We also re-analyzed RNA-Seq results from [42] (GEO GSE16344). We used HISAT 2.0.4 [77] for the mapping of sequencing reads to *Drosophila* genome release 5. We used a parameter for unpaired sequencing (-U) in running HISAT. We measured gene-level read abundances with HTSeq 0.6.1 [77] with the default

setting. From the counting result, we used polyA$^+$ protein coding genes that have more than 1 count per million mapped reads from any of the four samples (two controls and two RNAi) in our differential expression analysis. We performed differential expression analysis using DESeq2 [78]. In **Figure 3** and **Figure 5**, we demonstrated genes of which expression is more than 1 FPKM, which we also used to filter microarray results from MSL knockdown.

List of abbreviations

CES:Chromosome Entry Sites, ChIP: Chromatin Immunoprecipitation, DCC: Dosage Compensation Complex, FPKM: Fragments Per Kilobase of transcript per Million mapped reads, GEO: Gene Expression Omnibus, GO: Gene Ontology, H4K16Ac: Histone H4 Lysine 16 Acetylation, HAS: High-Affinity Sites, LAD: Lamina-Associated Domains, MLE: Maleless, modENCODE: model organism ENcyclopedia of DNA Elements, MOF: Males absent on the first, MRE: MSL-recognition element, MSL: Male specific lethal complex, NAR: Nucleoporin-Associated Region, NSL: Non-Specific Lethal, TAD: Topologically Associated Domain.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

The datasets analysed during the current study are available in the GEO and ArrayExpress repositories. We used modENCODE ChIP-chip results that are available in GEO with these accession IDS: GSE27805-6, GSE20797-9, and GSE32762. The salivary glands results are in ArrayExpress (E-MTAB-911 and E-MEXP-3506). We re-analyzed MSL complex knockdown results from GEO GSE16344 and ArrayExpress E-MEXP-1505.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by the Intramural Research Programs of the National Institutes of Health (NIH), National Institute of Diabetes and Digestive and Kidney Diseases, to BO, and Korean Visiting Scientist Training Award (KVSTA, HI13C1282) to HL.

Authors' contributions

HL and BO conceived of the idea and designed the analyses. HL performed computational analysis on the presented results. HL and BO interpreted and wrote the manuscript.

Acknowledgements

We thank the members of the Oliver lab for their helpful discussions and Dr. Per Stenberg and Dr. Sergey V. Razin for kindly sharing processed results from their studies. We utilized the high-performance computational capabilities of the Biowulf linux cluster at the NIH, Bethesda, MD. This research was supported by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

Figure Legends

Figure 1. Repressive TAD genes display lower gene expression levels and are dosage-compensated in male cells. (A) Venn-diagram displays overlap among the three repressive TADs that are described in this study. **(B)** Pie charts demonstrate proportion of repressive TAD genes (gray) vs. non-repressive TAD genes (white) in *Drosophila* genome. In (A) and (B), only protein-coding polyA$^+$ genes are counted. The numbers do not directly indicate numbers of "expressed" genes in each TAD. **(C-H)** Gene expression levels from the repressive TAD genes (gray) and non-repressive TAD genes (white) based on LAD (C,F), Hi-C (D,G), and chromatin occupancy studies (E,H). The top two rows show RNA-Seq results from *Drosophila* cell lines (unit: log2 FPKM, C-E), and the bottom two rows are from microarray study done with larval salivary glands (unit: normalized signal intensity, E-H). Intergenic signals from the 99th percentiles and below in RNA-Seq analyses, as well as background signals from the 95th percentiles and below in microarray results are indicated. **(I)** Comparisons of X chromosome gene expression levels from the repressive TADs between female and male salivary glands. Boxplots indicate the distribution of gene expression levels above expression cutoffs. Middle lines in box display medians of each distribution. Top of the box. 75th percentile. Bottom of the box. 25th percentile. Whiskers indicate the maximum, or minimum, observation within 1.5 times of the box height from the top, or the bottom of the box, respectively. Notches show 95% confidence interval for the medians. *** $p < 0.001$ from Mann-Whitney U test. The same format and test have been used for all boxplots appeared in this study.

Figure 2. Repressive TAD genes have a limited binding of MSL complex. (A-I) Chromatin immunoprecipitation (ChIP) results from MOF binding (A-C), Histone H4K16 acetylation (D-F), and MSL-1 binding (G-I) are summarized as boxplots for *Drosophila* cell lines (*Kc* and *S2*). Gene level ChIP signals are separately shown based on LAD (A,D,G), Hi-C (B,E,H) and chromatin occupancy (C,F,I) study results. **(J-O)** ChIP results from the 3rd instar larval salivary glands. ** $p < 0.01$, *** $p < 0.001$. **(P, Q)** Direct comparisons of MOF binding (P) and H4K16Ac enrichment (Q) between female and male salivary glands from (J-O). (**R, S**) The histogram represents expected numbers overlaps between repressive TADs and MRE (R), or CES (S). We performed random shuffling of the X chromosome genome 2,000 times and demonstrated the frequencies of the numbers of overlaps. Red lines. the actual number of overlaps between LADs, and MREs or CES's. *p* values are from permutation tests.

559 **Figure 3. Different responses from the repressive vs. non-repressive TAD genes upon**
560 **knockdown of a MSL complex component.** Boxplots represent gene expression changes in
561 log2 scale from depletion of MSL components in *Drosophila S2* cells. Plots are based on three
562 independent studies [36,41,42], which used either microarray (A-H) or RNA-Seq technology (I-
563 L). **(A, B)** Differential gene expression from *mof* knockdown cells. Changes from the repressive
564 TADs (left three columns, LAD, Null, and Black) as well as MOF binding, or Histone H4K16
565 acetylation regions are presented. **(C-H)** Results from *msl-1*, *msl-2*, or *msl-3* knockdown. **(I, J)**
566 Results from *mof* knockdown, measured by RNA-Seq analysis. **(K, L)** *msl-2* knockdown.
567 **(A,C,E,G,I,K)** Changes from *X* chromosome genes. **(B,D,F,H,J,L)** Changes from autosomal
568 genes. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
569
570 **Figure 4. Repressive TAD genes that are less sensitive to *msl* or *mof* knockdown lack**
571 **MOF enrichment at their gene bodies.** Top four panels demonstrate normalized ChIP signals
572 of MOF and H4K16Ac from *S2* cells as well as male salivary glands. The bottom three panels
573 display RNA-Seq read coverages from our re-analysis of Zhang et al [42]. The plots are scaled
574 based on their maximum coverage from one of the three samples: control RNAi, *mof* RNAi, and
575 *msl-2* RNAi (indicated in the square brackets). Note that there is no sample-to-sample
576 normalization because the total number of reads are similar across the three samples; 7.2, 7.5
577 and 8.1 million mapped reads for the control, *mof* RNAi, and *msl-2* RNAi samples, respectively.
578 **(A)** An autosomal gene (*RpL32*). **(B, C)** Canonical MSL targets genes. *CG9947* (B) and *arm*
579 (C). **(D-E)** Repressive target genes that are compensated via the non-canonical dosage
580 compensation. *CG34430* (D), *CG9521* (E), *CG8675* (F), and *CG2875* (G).
581
582 **Figure 5. Repressive TAD-based dosage compensation occurs at where X chromosome**
583 **TAD structures are maintained between female and male cells. (A)** Schematic illustration of
584 classification of TAD differences between *Kc* and *S2* cells based on [47] (modified with
585 permission by the authors) **(B-D)** Percentage of the number of 2kb windows from four different
586 classes of TAD difference between *Kc* and *S2* cells (TT, II, TI, and IT); windows that are found
587 from within TAD boundaries in both *Kc* and *S2* cells (TT), that are at boundaries or interspace
588 between TADs in both *Kc* and *S2* cells (II), that are found within the TAD boundaries in *Kc* cells
589 but not in *S2* cells (TI), and that are not in TAD boundaries in Kc cells but within TAD
590 boundaries in *S2* cells (IT). Top. distribution of 2kb windows from all autosomes. Bottom.
591 Distribution of 2kb windows from X chromosomes. **(E-G)** Gene expression changes of *X*-linked
592 genes in *S2* cells upon RNAi knockdown of *mof*, *msl-1*, *msl-2*, and *msl-3* as appeared in Figure

3 but based on TAD difference classes (TT, II, TI, and IT). Changes from repressive TADs (grey) and non-repressive TADs (white) are displayed. *P* values indicate differences from the median gene expression fold changes of non-LAD associated genes. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 6. Models demonstrating the parallelism among dosage compensation of autosomal dosage compensation in hemizygous *D. melanogaster*, and *X* chromosome dosage compensation in *C. elegans* and *D. melanogaster*. **(A, B)** A proposal of de-repression mediated compensation of one-dose autosomal genes in hemizygous *D. melanogaster* based on our previous study [28]. **(C, D)** A model of *X* chromosome dosage compensation in D. melanogaster based on the current study as well as other references [5,7,14,39,41,48,79,80]. **(E, F)** A model of *X* chromosome dosage compensation in *C. elegans* based on the references [16,20–23,26,27,81].

References

1. Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc Natl Acad Sci U S A. 2012;109:14746–53.

2. Sheltzer JM, Amon A. The aneuploidy paradox: costs and benefits of an incorrect karyotype. Trends Genet. 2011;27:446–53.

3. Ercan S. Mechanisms of x chromosome dosage compensation. J Genomics. 2015;3:1–19.

4. Disteche CM. Dosage compensation of the sex chromosomes. Annu Rev Genet. 2012;46:537–60.

5. Ferrari F, Alekseyenko AA, Park PJ, Kuroda MI. Transcriptional control of a whole chromosome: emerging models for dosage compensation. Nat Struct Mol Biol. 2014;21:118–25.

6. Lucchesi JC, Kelly WG, Panning B. Chromatin remodeling in dosage compensation. Annu Rev Genet. 2005;39:615–51.

7. Lucchesi JC, Kuroda MI. Dosage compensation in Drosophila. Cold Spring Harb Perspect Biol [Internet]. 2015;7. Available from: http://dx.doi.org/10.1101/cshperspect.a019398

8. Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, et al. X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature. 2011;471:115–8.

9. Conrad T, Cavalli FMG, Vaquerizas JM, Luscombe NM, Akhtar A. Drosophila dosage compensation involves enhanced Pol II recruitment to male X-linked promoters. Science. 2012;337:742–6.

628 10. Straub T, Becker PB. Comment on “Drosophila dosage compensation involves enhanced
629 Pol II recruitment to male X-linked promoters.” Science. 2013;340:273.

630 11. Ferrari F, Jung YL, Kharchenko PV, Plachetka A, Alekseyenko AA, Kuroda MI, et al.
631 Comment on “Drosophila dosage compensation involves enhanced Pol II recruitment to male X-
632 linked promoters.” Science. 2013;340:273.

633 12. Gelbart ME, Kuroda MI. Drosophila dosage compensation: a complex voyage to the X
634 chromosome. Development. 2009;136:1399–410.

635 13. Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB. The chromosomal high-affinity
636 binding sites for the Drosophila dosage compensation complex. PLoS Genet. 2008;4:e1000302.

637 14. Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee O-K, Kharchenko P, et al. A
638 sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X
639 chromosome. Cell. 2008;134:599–609.

640 15. Disteche CM. Dosage compensation of the sex chromosomes and autosomes. Semin Cell
641 Dev Biol. 2016;56:9–18.

642 16. Lau AC, Csankovszki G. Balancing up and downregulation of the C. elegans X
643 chromosomes. Curr Opin Genet Dev. 2015;31:50–6.

644 17. Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, et al. Evidence for
645 compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans
646 and Drosophila melanogaster. Nat Genet. 2011;43:1179–85.

647 18. Gupta V, Parisi M, Sturgill D, Nuttall R, Doctolero M, Dudko OK, et al. Global analysis of X-
648 chromosome dosage compensation. J Biol. 2006;5:3.

649 19. McDonel P, Jans J, Peterson BK, Meyer BJ. Clustered DNA motifs mark X chromosomes
650 for repression by a dosage compensation complex. Nature. 2006;444:614–8.

651 20. Vielle A, Lang J, Dong Y, Ercan S, Kotwaliwale C, Rechtsteiner A, et al. H4K20me1
652 Contributes to Downregulation of X-Linked Genes for C. elegans Dosage Compensation. PLoS
653 Genet. Public Library of Science; 2012;8:e1002933.

654 21. Kramer M, Kranz A-L, Su A, Winterkorn LH, Albritton SE, Ercan S. Developmental
655 Dynamics of X-Chromosome Dosage Compensation by the DCC and H4K20me1 in C. elegans.
656 PLoS Genet. 2015;11:e1005698.

657 22. Wells MB, Snyder MJ, Custer LM, Csankovszki G. Caenorhabditis elegans dosage
658 compensation regulates histone H4 chromatin state on X chromosomes. Mol Cell Biol.
659 2012;32:1710–9.

660 23. Lau AC, Zhu KP, Brouhard EA, Davis MB, Csankovszki G. An H4K16 histone
661 acetyltransferase mediates decondensation of the X chromosome in C. elegans males.
662 Epigenetics Chromatin. 2016;9:44.

663 24. Petty EL, Collette KS, Cohen AJ, Snyder MJ, Csankovszki G. Restricting dosage
664 compensation complex binding to the X chromosomes by H2A.Z/HTZ-1. PLoS Genet.
665 2009;5:e1000699.

666 25. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, et al. Condensin-driven
667 remodelling of X chromosome topology during dosage compensation. Nature. 2015;523:240–4.

668 26. Kruesi WS, Core LJ, Waters CT, Lis JT, Meyer BJ. Condensin controls recruitment of RNA
669 polymerase II to achieve nematode X-chromosome dosage compensation. Elife.
670 2013;2:e00808.

671 27. Snyder MJ, Lau AC, Brouhard EA, Davis MB, Jiang J, Sifuentes MH, et al. Anchoring of
672 Heterochromatin to the Nuclear Lamina Reinforces Dosage Compensation-Mediated Gene
673 Repression. PLoS Genet. 2016;12:e1006341.

674 28. Lee H, Cho D-Y, Whitworth C, Eisman R, Phelps M, Roote J, et al. Effects of Gene Dose,
675 Chromatin, and Network Topology on Expression in Drosophila melanogaster. PLoS Genet.
676 2016;12:e1006295.

677 29. Kassis JA. Unusual properties of regulatory DNA from the Drosophila engrailed gene: three
678 “pairing-sensitive” sites within a 1.6-kb region. Genetics. 1994;136:1025–38.

679 30. Kassis JA. 14 - Pairing-Sensitive Silencing, Polycomb Group Response Elements, and
680 Transposon Homing in Drosophila. In: Dunlap JC, Wu C-T, editors. Advances in Genetics.
681 Academic Press; 2002. p. 421–38.

682 31. Morris JR, Chen JL, Geyer PK, Wu CT. Two modes of transvection: enhancer action in trans
683 and bypass of a chromatin insulator in cis. Proc Natl Acad Sci U S A. 1998;95:10740–5.

684 32. Lee AM, Wu C-T. Enhancer-promoter communication at the yellow gene of Drosophila
685 melanogaster: diverse promoters participate in and regulate trans interactions. Genetics.
686 2006;174:1867–80.

687 33. van Bemmel JG, Pagie L, Braunschweig U, Brugman W, Meuleman W, Kerkhoven RM, et
688 al. The insulator protein SU(HW) fine-tunes nuclear lamina interactions of the Drosophila
689 genome. PLoS One. 2010;5:e15013.

690 34. Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic
691 protein location mapping reveals five principal chromatin types in Drosophila cells. Cell.
692 2010;143:212–24.

693 35. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-
694 dimensional folding and functional organization principles of the Drosophila genome. Cell.
695 2012;148:458–72.

696 36. Conrad T, Cavalli FMG, Holz H, Hallacli E, Kind J, Ilik I, et al. The MOF chromobarrel
697 domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. Dev Cell.
698 2012;22:610–24.

699 37. Lee H, McManus CJ, Cho D-Y, Eaton M, Renda F, Somma MP, et al. DNA copy number
700 evolution in Drosophila cell lines. Genome Biol. 2014;15:R70.

701 38. Akhtar A, Becker PB. Activation of transcription through histone H4 acetylation by MOF, an
702 acetyltransferase essential for dosage compensation in Drosophila. Mol Cell. 2000;5:367–75.

703 39. Kind J, Vaquerizas JM, Gebhardt P, Gentzel M, Luscombe NM, Bertone P, et al. Genome-
704 wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in

705 Drosophila. Cell. 2008;133:813–28.

706 40. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, et al.
707 Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature.
708 2011;471:480–5.

709 41. Hamada FN, Park PJ, Gordadze PR, Kuroda MI. Global regulation of X chromosomal genes
710 by the MSL complex in Drosophila melanogaster. Genes Dev. 2005;19:2289–94.

711 42. Zhang Y, Malone JH, Powell SK, Periwal V, Spana E, Macalpine DM, et al. Expression in
712 aneuploid Drosophila S2 cells. PLoS Biol. 2010;8:e1000320.

713 43. Feller C, Prestel M, Hartmann H, Straub T, Söding J, Becker PB. The MOF-containing NSL
714 complex associates globally with housekeeping genes, but activates only a defined subset.
715 Nucleic Acids Res. 2012;40:1509–22.

716 44. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the
717 transcriptome. BMC Biol. 2011;9:34.

718 45. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray
719 in transcriptome profiling of activated T cells. PLoS One. 2014;9:e78644.

720 46. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian
721 genomes identified by analysis of chromatin interactions. Nature. 2012;485:376–80.

722 47. Ulianov SV, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, et al. Active
723 chromatin and transcription play a key role in chromosome partitioning into topologically
724 associating domains. Genome Res. 2016;26:70–84.

725 48. Ramírez F, Lingg T, Toscano S, Lam KC, Georgiev P, Chung H-R, et al. High-Affinity Sites
726 Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X
727 Chromosome in Drosophila. Mol Cell. 2015;60:146–62.

728 49. Philip P, Stenberg P. Male X-linked genes in Drosophila melanogaster are compensated
729 independently of the Male-Specific Lethal complex. Epigenetics Chromatin. 2013;6:35.

730 50. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SWM, Solovei I, Brugman W, et al.
731 Molecular maps of the reorganization of genome-nuclear lamina interactions during
732 differentiation. Mol Cell. 2010;38:603–13.

733 51. Poleshko A, Shah PP, Gupta M, Babu A, Morley MP, Manderfield LJ, et al. Genome-Nuclear
734 Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. Cell [Internet]. 2017;
735 Available from: http://dx.doi.org/10.1016/j.cell.2017.09.018

736 52. Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex
737 chromosomes. Heredity . 2005;95:118–28.

738 53. Bachtrog D. Sex chromosome evolution: molecular aspects of Y-chromosome degeneration
739 in Drosophila. Genome Res. 2005;15:1393–401.

740 54. Ellegren H. Sex-chromosome evolution: recent progress and the influence of male and
741 female heterogamety. Nat Rev Genet. 2011;12:157–66.

742 55. Blair Hedges S. The origin and evolution of model organisms. Nat Rev Genet. Nature
743 Publishing Group; 2002;3:838–49.

744 56. Krylov DM, Wolf YI, Rogozin IB, Koonin EV. Gene loss, protein sequence divergence, gene
745 dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome
746 Res. 2003;13:2229–35.

747 57. Rastelli L, Richman R, Kuroda MI. The dosage compensation regulators MLE, MSL-1 and
748 MSL-2 are interdependent since early embryogenesis in Drosophila. Mech Dev. 1995;53:223–
749 33.

750 58. Franke A, Dernburg A, Bashaw GJ, Baker BS. Evidence that MSL-mediated dosage
751 compensation in Drosophila begins at blastoderm. Development. The Company of Biologists
752 Ltd; 1996;122:2751–60.

753 59. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. Noncanonical compensation
754 of zygotic X transcription in early Drosophila melanogaster development revealed through
755 single-embryo RNA-seq. PLoS Biol. 2011;9:e1000590.

756 60. Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, Presgraves DC. Sex chromosome-
757 specific regulation in the Drosophila male germline but little evidence for chromosomal dosage
758 compensation or meiotic inactivation. PLoS Biol. 2011;9:e1001126.

759 61. Bachtrog D, Toda NRT, Lockton S. Dosage compensation and demasculinization of X
760 chromosomes in Drosophila. Curr Biol. 2010;20:1476–81.

761 62. Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, et al. Paucity of genes on
762 the Drosophila X chromosome showing male-biased expression. Science. 2003;299:697–700.

763 63. Sturgill D, Zhang Y, Parisi M, Oliver B. Demasculinization of X chromosomes in the
764 Drosophila genus. Nature. 2007;450:238–41.

765 64. Reinke V, Smith HE, Nance J, Wang J, Van Doren C, Begley R, et al. A global profile of
766 germline gene expression in C. elegans. Mol Cell. 2000;6:605–16.

767 65. van Bemmel JG, Pagie L, Braunschweig U, Brugman W, Meuleman W, Kerkhoven RM, et
768 al. The Insulator Protein SU(HW) Fine-Tunes Nuclear Lamina Interactions of the Drosophila
769 Genome. PLoS One. Public Library of Science; 2010;5:e15013.

770 66. McQuilton P, St Pierre SE, Thurmond J, FlyBase Consortium. FlyBase 101--the basics of
771 navigating FlyBase. Nucleic Acids Res. 2012;40:D706–14.

772 67. Lyne R, Smith R, Rutherford K, Wakeling M, Varley A, Guillier F, et al. FlyMine: an
773 integrated database for Drosophila and Anopheles genomics. Genome Biol. 2007;8:R129.

774 68. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO:
775 archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41:D991–5.

776 69. Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A Model-Based Background
777 Adjustment for Oligonucleotide Expression Arrays. J Am Stat Assoc. 2004;99:909–17.

778 70. Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, et al. ArrayExpress
779 update--simplifying data submissions. Nucleic Acids Res. 2015;43:D1113–6.

780 71. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol.
781 2010;11:R106.

782 72. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for
783 motif discovery and searching. Nucleic Acids Res. 2009;37:W202–8.

784 73. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif.
785 Bioinformatics. 2011;27:1017–8.

786 74. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
787 Bioinformatics. 2010;26:841–2.

788 75. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al.
789 Exploration, normalization, and summaries of high density oligonucleotide array probe level
790 data. Biostatistics. 2003;4:249–64.

791 76. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
792 expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.
793 2015;43:e47.

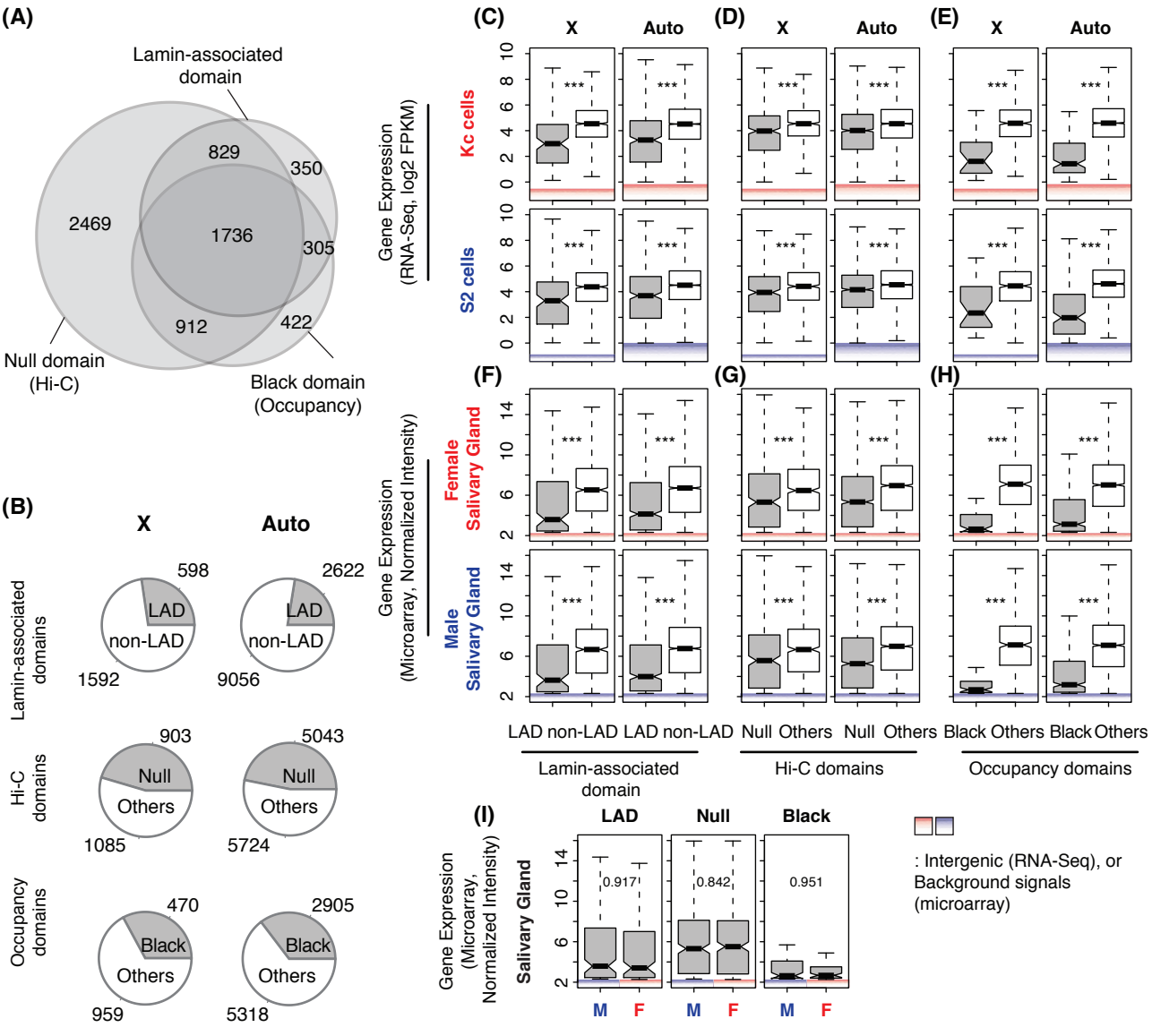
794 77. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
795 requirements. Nat Methods. 2015;12:357–60.

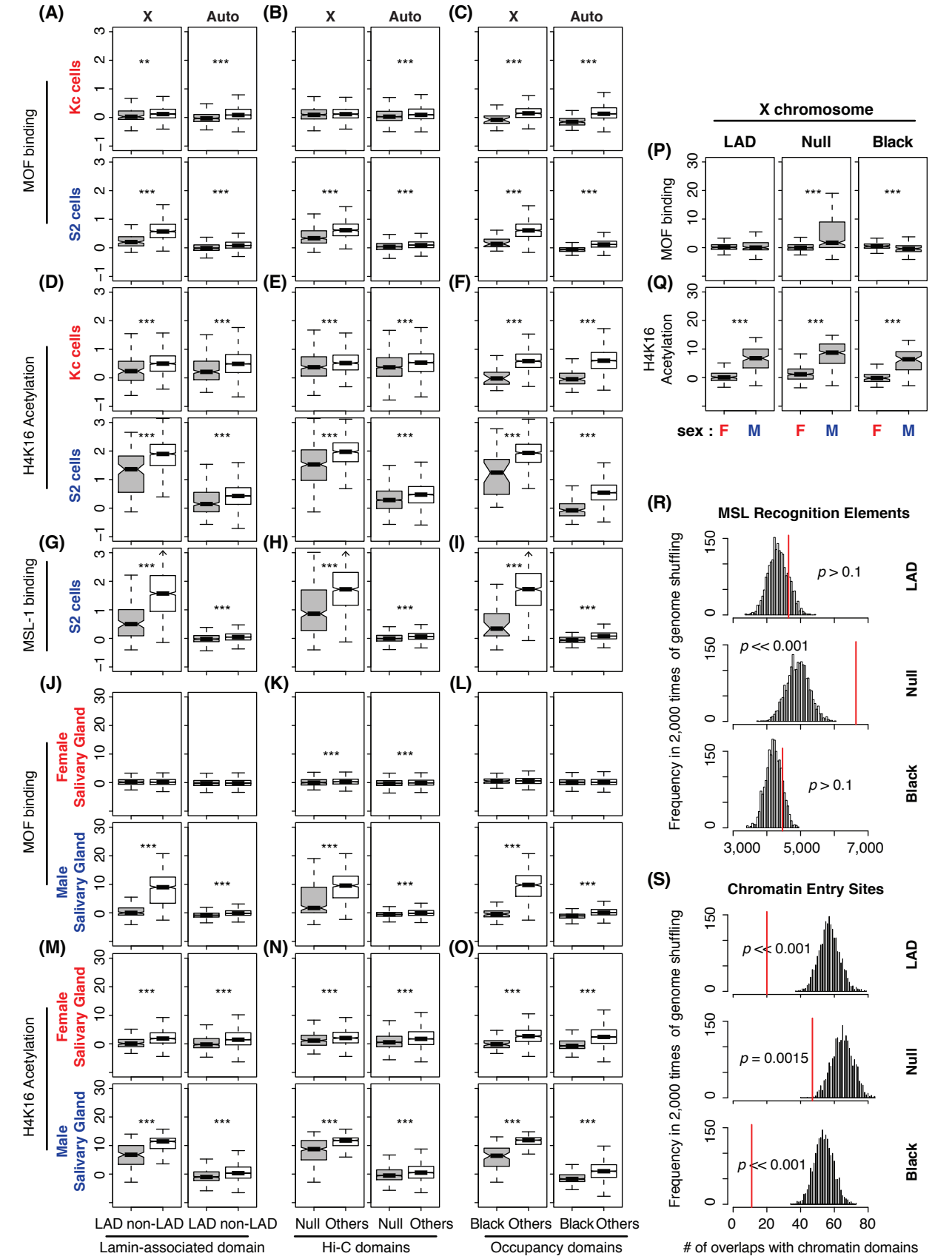
796 78. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
797 seq data with DESeq2. Genome Biol. 2014;15:550.

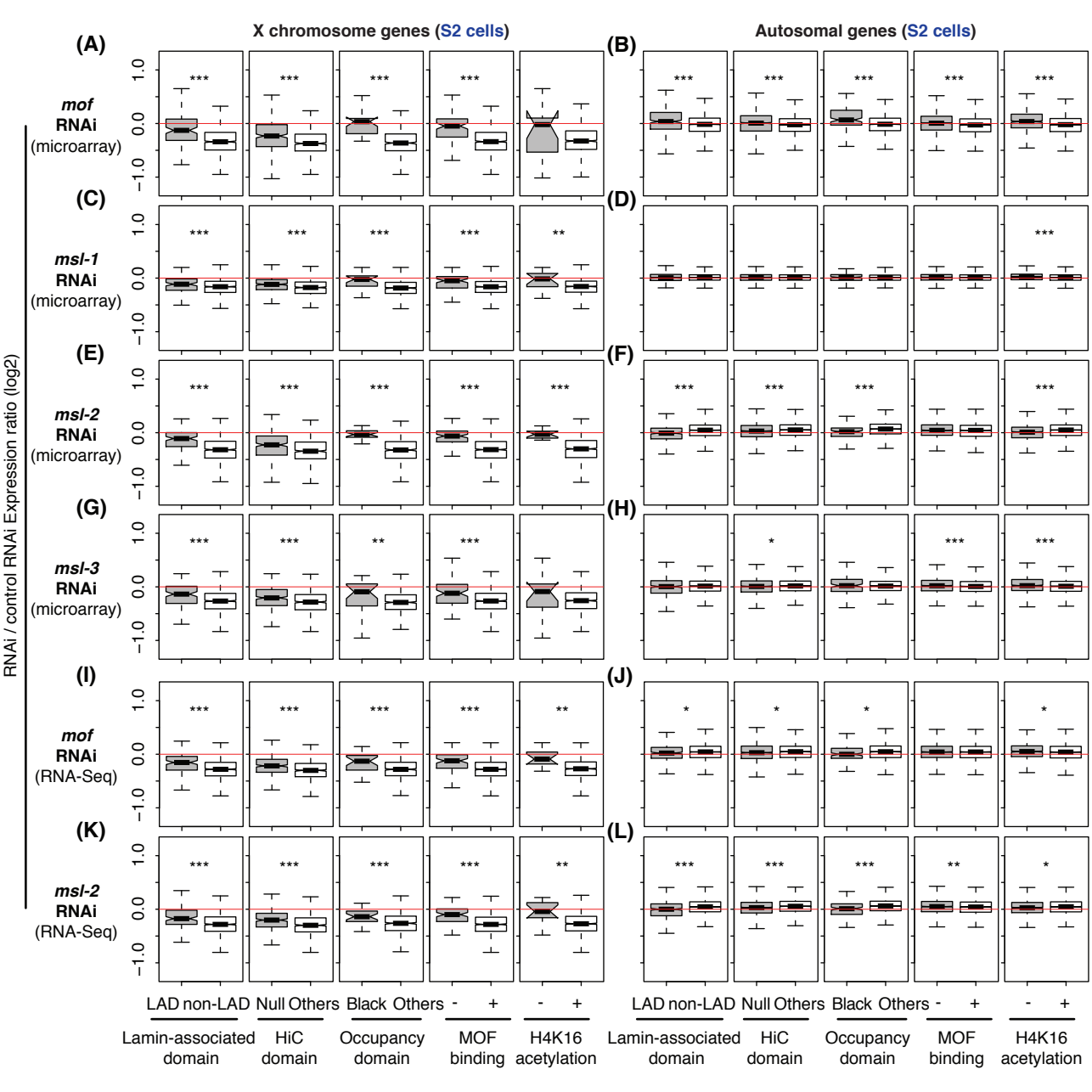
798 79. Vaquerizas JM, Suyama R, Kind J, Miura K, Luscombe NM, Akhtar A. Nuclear pore proteins
799 nup153 and megator define transcriptionally active regions in the Drosophila genome. PLoS
800 Genet. 2010;6:e1000846.

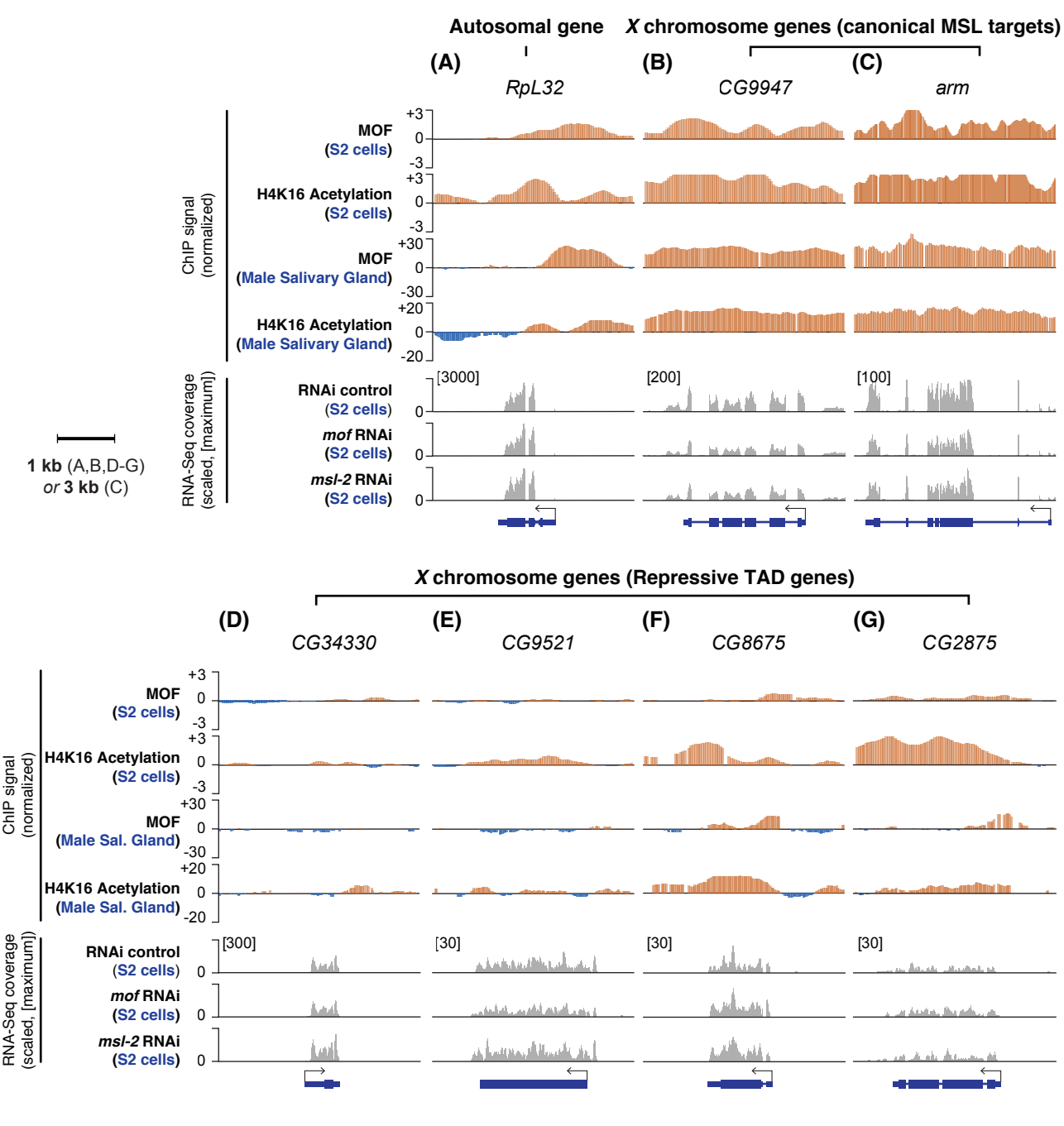
801 80. Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, et al. X chromosome
802 dosage compensation via enhanced transcriptional elongation in Drosophila. Nature. Nature
803 Research; 2011;471:115–8.

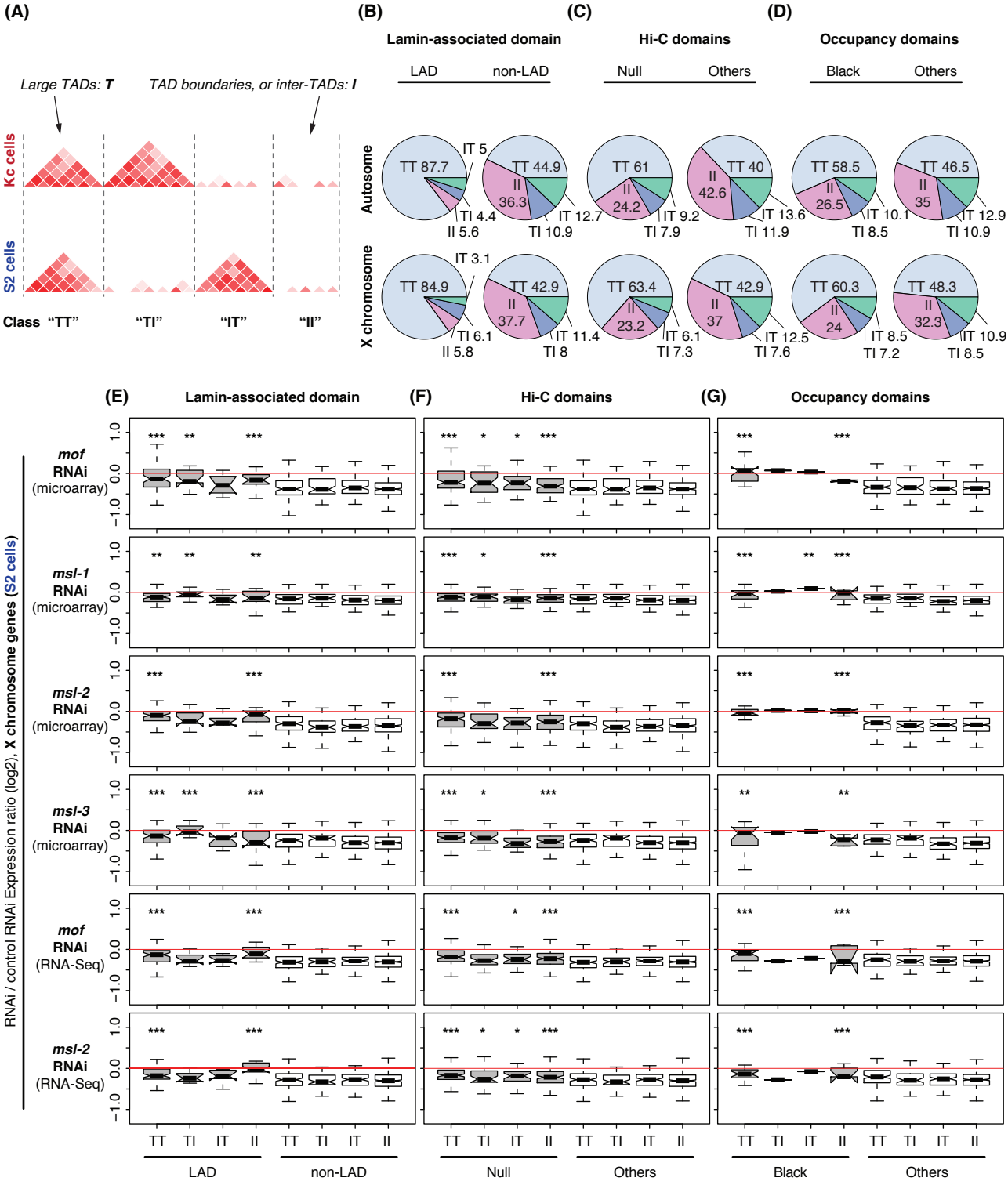
804 81. Albritton SE, Ercan S. Caenorhabditis elegans Dosage Compensation: Insights into
805 Condensin-Mediated Gene Regulation. Trends Genet. 2018;34:41–53.

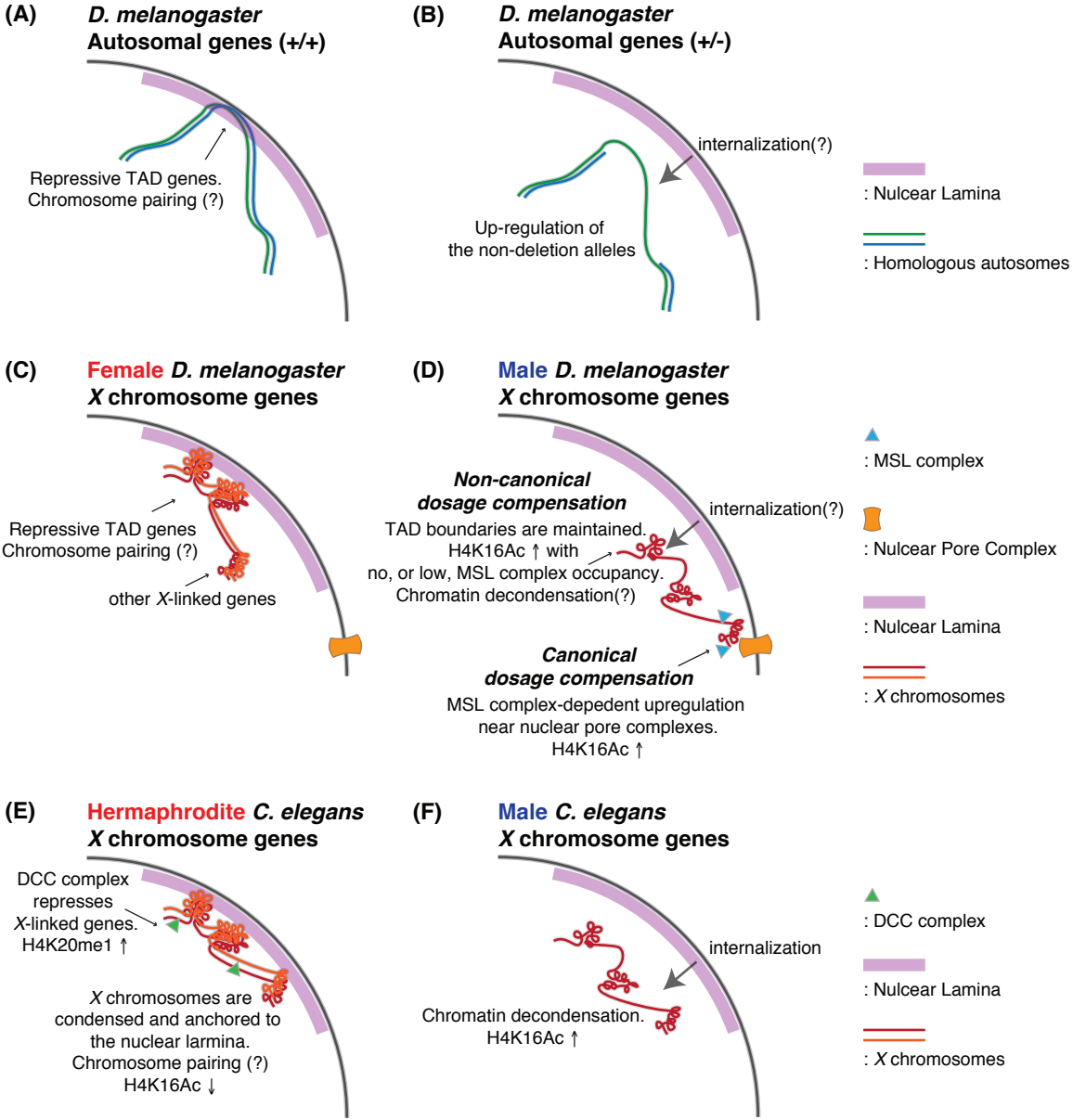

(A)
Lamin-associated domain
829
350
2469
1736
305
912
422
Null domain (Hi-C)
Black domain (Occupancy)
(B)
X
Auto
Lamin-associated domains
598
LAD
non-LAD
1592
2622
LAD
non-LAD
9056
Hi-C domains
903
Null
Others
1085
5043
Null
Others
5724
Occupancy domains
470
Black
Others
959
2905
Black
Others
5318
(C)
(D)
(E)
X
Auto
Gene Expression (RNA-Seq, log2 FPKM)
Kc cells
S2 cells
***
(F)
(G)
(H)
Gene Expression (Microarray, Normalized Intensity)
Female Salivary Gland
Male Salivary Gland
LAD non-LAD LAD non-LAD Null Others Null Others Black Others Black Others
Lamin-associated domain
Hi-C domains
Occupancy domains
(I)
LAD
Null
Black
Gene Expression (Microarray, Normalized Intensity)
Salivary Gland
0.917
0.842
0.951
M F M F M F
: Intergenic (RNA-Seq), or Background signals (microarray)


(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
(K)
(L)
(M)
(N)
(O)
(P)
(Q)
(R)
(S)
X
Auto
MOF binding
H4K16 Acetylation
MSL-1 binding
Kc cells
S2 cells
Female Salivary Gland
Male Salivary Gland
X chromosome
LAD
Null
Black
sex : F M F M F M
MSL Recognition Elements
Chromatin Entry Sites
Frequency in 2,000 times of genome shuffling
p > 0.1
p << 0.001
p = 0.0015
3,000 5,000 7,000
0 20 40 60 80
# of overlaps with chromatin domains
LAD non-LAD LAD non-LAD
Lamin-associated domain
Null Others Null Others
Hi-C domains
Black Others Black Others
Occupancy domains


X chromosome genes (S2 cells)
Autosomal genes (S2 cells)
(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
(K)
(L)
mof RNAi (microarray)
msl-1 RNAi (microarray)
msl-2 RNAi (microarray)
msl-3 RNAi (microarray)
mof RNAi (RNA-Seq)
msl-2 RNAi (RNA-Seq)
RNAi / control RNAi Expression ratio (log2)
1.0
0.0
−1.0
***
**
*
LAD non-LAD Null Others Black Others - + - +
Lamin-associated domain
HiC domain
Occupancy domain
MOF binding
H4K16 acetylation


Autosomal gene
X chromosome genes (canonical MSL targets)
(A)
RpL32
(B)
CG9947
(C)
arm
ChIP signal (normalized)
MOF (S2 cells)
H4K16 Acetylation (S2 cells)
MOF (Male Salivary Gland)
H4K16 Acetylation (Male Salivary Gland)
+3
0
-3
+30
-30
+20
-20
RNA-Seq coverage (scaled, [maximum])
RNAi control (S2 cells)
mof RNAi (S2 cells)
msl-2 RNAi (S2 cells)
[3000]
[200]
[100]
1 kb (A,B,D-G)
or 3 kb (C)
X chromosome genes (Repressive TAD genes)
(D)
CG34330
(E)
CG9521
(F)
CG8675
(G)
CG2875
[300]
[30]


(A)
Large TADs: T
TAD boundaries, or inter-TADs: I
Kc cells
S2 cells
Class "TT" "TI" "IT" "II"
(B)
Lamin-associated domain
LAD
non-LAD
(C)
Hi-C domains
Null
Others
(D)
Occupancy domains
Black
Others
Autosome
X chromosome
TT 87.7
IT 5
TI 4.4
II 5.6
TT 44.9
II 36.3
IT 12.7
TI 10.9
TT 61
II 24.2
IT 9.2
TI 7.9
TT 40
II 42.6
IT 13.6
TI 11.9
TT 58.5
II 26.5
IT 10.1
TI 8.5
TT 46.5
II 35
IT 12.9
TI 10.9
TT 84.9
IT 3.1
TI 6.1
II 5.8
TT 42.9
II 37.7
IT 11.4
TI 8
TT 63.4
II 23.2
IT 6.1
TI 7.3
TT 42.9
II 37
IT 12.5
TI 7.6
TT 60.3
II 24
IT 8.5
TI 7.2
TT 48.3
II 32.3
IT 10.9
TI 8.5
(E)
Lamin-associated domain
(F)
Hi-C domains
(G)
Occupancy domains
mof RNAi (microarray)
msl-1 RNAi (microarray)
msl-2 RNAi (microarray)
msl-3 RNAi (microarray)
mof RNAi (RNA-Seq)
msl-2 RNAi (RNA-Seq)
RNAi / control RNAi Expression ratio (log2), X chromosome genes (S2 cells)
1.0
0.0
−1.0
TT TI IT II TT TI IT II
LAD
non-LAD
Null
Others
Black
Others


(A)
D. melanogaster
Autosomal genes (+/+)
Repressive TAD genes.
Chromosome pairing (?)
(B)
D. melanogaster
Autosomal genes (+/-)
internalization(?)
Up-regulation of
the non-deletion alleles
: Nulcear Lamina
: Homologous autosomes
(C)
Female D. melanogaster
X chromosome genes
Repressive TAD genes
Chromosome pairing (?)
other X-linked genes
(D)
Male D. melanogaster
X chromosome genes
Non-canonical
dosage compensation
TAD boundaries are maintained.
H4K16Ac ↑ with
no, or low, MSL complex occupancy.
Chromatin decondensation(?)
internalization(?)
Canonical
dosage compensation
MSL complex-depedent upregulation
near nuclear pore complexes.
H4K16Ac ↑
: MSL complex
: Nulcear Pore Complex
: Nulcear Lamina
: X chromosomes
(E)
Hermaphrodite C. elegans
X chromosome genes
DCC complex
represses
X-linked genes.
H4K20me1 ↑
X chromosomes are
condensed and anchored to
the nuclear larmina.
Chromosome pairing (?)
H4K16Ac ↓
(F)
Male C. elegans
X chromosome genes
internalization
Chromatin decondensation.
H4K16Ac ↑
: DCC complex
: Nulcear Lamina
: X chromosomes