

Dosage compensation defects due to *roX* RNA loss are rescued by recalibration of X/autosome stoichiometry

Fotios Gkountromichos¹, George Yankson², Muhunden Jayakrishnan¹, Aline Campos Sparr¹, Marisa Müller¹, Patrick Heun², Peter B. Becker^{1,*}

¹Biomedical Center Munich, Molecular Biology Division, Ludwig-Maximilians-Universität, 82152 Planegg-Martinsried, Germany

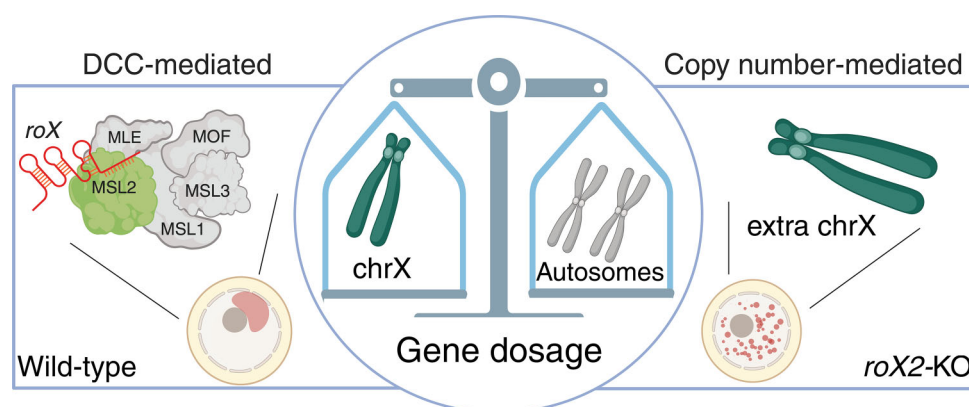
²Department of Biology, Molecular Genetics, Technical University Darmstadt, Darmstadt 64289, Germany

*To whom correspondence should be addressed. Email: pbecker@bmc.med.lmu.de

Abstract

Metazoa evolved regulatory networks to balance sex chromosome expression. In *Drosophila*, males have a single gene-rich X chromosome, whereas females have two. Balanced X/autosome expression is essential for viability, and in male flies is achieved by activation of genes on the X through the male-specific-lethal (MSL) dosage compensation complex (DCC). This ribonucleoprotein assembly contains long non-coding *roX* RNAs. To dissect the functional requirements of *roX* in a cell-based system, we deleted the *roX2* gene in male S2 cells and selected two independent lines lacking detectable *roX* RNA. In the absence of *roX*, the remaining MSL protein complex was unable to associate with known or newly identified binding sites and thus failed to activate transcription. Surprisingly, the X/autosome expression ratio appeared nevertheless compensated. Cytogenetic and genomic analyses revealed that both *roX*-deficient cell populations had acquired additional X chromosomes. We propose that uncompensated X expression compromises fitness and gives a selective growth advantage to cells that rebalance their genomes through chromosome mis-segregation. Remarkably, ectopic expression of *roX2* reversed this selection, restored DCC binding, and normalized the karyotype. These findings illustrate that X chromosome dosage compensation is critical for viability even in cultured cells, and provide a striking example of rapid evolution under stringent selection.

Graphical abstract



Introduction

Maintaining balanced gene expression between sex chromosomes and autosomes is of vital importance. In *Drosophila melanogaster*, the proteins encoded by male-specific-lethal (MSL) genes form a ribonucleoprotein complex (the dosage compensation complex, DCC) that transcriptionally compensates the X chromosome monosomy in male flies. An epigenetic reader/writer module composed of MSL3 and MOF (males absent on the first) is connected via the scaffold protein MSL1 to the DNA-binding subunit, MSL2, which en-

sures the selective targeting of the DCC to the X chromosome. MSL3 binds to actively transcribed chromatin marked by methylated histone H3 at lysine 36 (H3K36me). The acetyltransferase MOF then acetylates histone H4 lysine 16 (H4K16ac), which prevents chromatin compaction. This activity contributes to an approximate doubling of transcription from the male X chromosome, thus matching the combined output of the two female X chromosomes [1, 2]. In the absence of a functional DCC-mediated dosage compensation system, male *Drosophila* die in the late larval stage [3]. Whether this is

Received: March 4, 2026. Revised: May 13, 2026. Accepted: May 16, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

caused by haploinsufficient genes that are critical for further development or because of a generally compromised fitness due to unbalanced genome-wide expression is not known.

The functions of the protein components of the DCC are, in principle, well characterized, but the role of the long non-coding *roX* (RNA on X) RNAs that associate with the MSL complex *in vivo* remains poorly understood. The realization that the long *roX1* and the much shorter *roX2* RNAs tightly co-localized with the MSL proteins on the X chromosome suggested an involvement in dosage compensation, which was then confirmed [4–8]. The *roX1* and *roX2* genes are located on the X chromosome. The discovery of X-linked non-coding RNAs involved in dosage compensation in *Drosophila* resonated with the critical involvement of *Xist* RNA in mammalian X inactivation, suggesting a common principle underlying diverse dosage compensation strategies [9]. In contrast to *Xist*, *Drosophila roX* RNAs can trigger X chromosome activation even if transcribed ectopically from an autosome [5, 10]. *roX* RNA genes evolved fast in *Drosophilids*, suggesting that they may have driven the evolution of the dosage compensation system as we know it today [10, 11].

The two *roX* genes serve distinct functions during development, but each RNA alone can sustain male life, if put to the test. *roX1* is transcribed in both sexes up to stage 12, from which point onwards it is male-specific. The male-specific transcription of *roX2* RNA begins at stage 10 and its expression mirrors the profile of gradual X chromosome activation [6, 12]. Deletion of either *roX* gene does not compromise male viability, but impairment of both causes male lethality, with 5% of males emerging as developmentally delayed escapers [13, 14]. Strangely, in these organismic studies, male larvae lacking both *roX* genes appear healthy. Overexpression of MSL proteins can partially overcome a *roX* deficiency, pointing to a role for *roX* RNA in targeting the DCC to the X chromosome, rather than being required for activation of its histone acetyltransferase activity [10, 15]. The association of either *roX* RNA with the MSL proteins is mediated by the DexH helicase MLE (maleless) [16]. The helicase unwinds conserved stem-loop structures at the 3' end of *roX* RNAs, exposing linear motifs to which MSL2 can bind [17–20]. This interaction has been proposed to “trap” the DCC specifically on the X chromosome, potentially explaining the exclusive X chromosome binding despite genome-wide distribution of the DNA sequences that resemble MSL recognition elements (MREs) [10, 21, 22].

Currently, the highly selective association of the DCC with the X is explained by two non-exclusive models. The first model assumes that the DCC associates with the X chromosome in a two-step process. Accordingly, the initial recognition of a limited number of DNA elements exclusive to the X, such as ‘high affinity sites’ (HASs), ‘chromosome entry sites’ (CESs), and ‘Pioneering on the X’ (PionX) sites, is independent of RNA. From these sites, the complex “spreads” to neighboring genes in a *roX*-dependent manner [1, 2]. An alternative model poses that the X chromosome evolved numerous DCC-binding sites with varying affinities, where higher-affinity sites are less dependent on *roX* RNA for DCC recruitment [12, 23–25]. Traditional methods such as polytene chromosome immunostaining and standard chromatin immunoprecipitation sequencing (ChIP-seq) primarily detect only the highest occupancy sites and cannot easily quantify *roX* dependency at sites of varying affinity, making it difficult to distinguish between the models.

The complex developmental phenotypes of *roX* gene mutations motivated us to explore the role of *roX* RNA in a simple proliferative cell culture model. Immortalized hematopoietic *Drosophila* S2 cells [26] previously served as a useful tool to study dosage compensation mechanisms [10, 27–29]. Despite significant aneuploidy, S2 cells compensate for the reduced dose of the X chromosome (relative to autosomes) by two additive mechanisms: approximately half of the required transcription boost is achieved through a generic, mechanistically unclear feedback response to chromosomal imbalance, while the remaining activation depends on the DCC [28]. Curiously, RNA interference (RNAi)-mediated depletion of MSL proteins disrupts localization of the DCC but does not impair cell proliferation over the course of the experiment, typically 6–10 days [27–29]. It is possible that dosage compensation is critical for *Drosophila* development but dispensable for proliferation of these cells in culture.

We used CRISPR/Cas9 [clustered regularly interspaced palindromic repeats /CRISPR-associated protein 9] genome editing to generate S2 cell lines lacking all *roX* expression, in order to investigate the structural and functional requirements of this long non-coding RNA (lncRNA) for dosage compensation. In the absence of *roX* RNAs, the MSL proteins form a complex, yet fail to localize to the X chromosome [20]. The *roX*-deficient single-cell model enables quantitative evaluation of *roX* function through phenotypic rescue. While full-length *roX2* complements all deficiencies, shorter derivatives do not, establishing a tractable system for comprehensive structure–function dissection of this highly evolved lncRNA.

Using CUT&RUN with improved informatic csaw-based analysis, we expanded the MSL2-binding site catalog to >1800 sites, the vast majority X-chromosomal, including most known MREs and many novel sites, half of which are promoter-proximal. Contrary to the two-step model, which posits *roX*-independent DCC binding at PionX sites and HASs, we found that *roX* is required for DCC occupancy at all site classes. This supports a quantitative affinity hierarchy model, in which DCC spreading is achieved by successive filling of progressively lower-affinity sites rather than through qualitatively distinct classes of sites that differ in their *roX*-dependency.

Surprisingly, X-chromosomal gene expression in these *roX*-deficient cells matched wild-type (WT) levels. We discovered that up to 50% of the cells in two independent knockout cell populations had gained extra X chromosomes, effectively compensating for the loss of DCC function through karyotypic adjustment. We assume that the lack of dosage compensation leads to reduced fitness and a strong selection for cells that rebalanced their X/autosome complement through chromosome mis-segregation that happens frequently in cultured cells. Remarkably, stable re-expression of a *roX* transgene rescued dosage compensation and reversed the karyotypic changes. These observations strongly suggest that lack of dosage compensation severely compromises the fitness of aneuploid cells in culture and provide an instructive example of rapid karyotypic evolution under stringent selection.

Materials and methods

Cell culture and manipulations

Drosophila melanogaster S2 cells were obtained from Stefan Ameres (IMBA, Vienna). They were cultured at 26°C in Schneider’s medium (Thermo-Fisher, Cat. No. 21 720 024)

supplemented with 10% fetal bovine serum (FBS; PAN Biotech, Cat. No. P30-3031 or Capricorn Scientific, Cat. No. FBS-12A) and 1% penicillin–streptomycin (Sigma-Aldrich, Cat. No. P-4333). Cells were routinely counted with a CASY cell counter (Roche Innovatis).

CRISPR/Cas9 genome editing

Genome editing was done using the CRISPR/Cas9 technology as described in [30] with some alterations. Candidate guide RNAs (gRNAs) targeting the promoter region of *roX2* were designed using Benchling [31]. For the 3' end targeting of *roX2*, single guide RNAs (sgRNAs) were initially designed using the GPP sgRNA designer (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgRNA-design> [32]). All gRNAs were tested for off-target effects using the CRISPRfly “Optimal target finder” (<http://targetfinder.flycrispr.neuro.brown.edu>, guide length = 20, stringency = high, PAM = NGG only) [33]. The 20 bp gRNAs were fused to a tracrRNA backbone during synthesis and cloned downstream of the *Drosophila* U6 promoter. These constructs were synthesized as gBlocks (Integrated DNA Technologies) and PCR-amplified before transfection using the Phusion polymerase (NEB, Cat. No. M0530L). Transfection of cells for genome editing was done with XtremeGENE HP transfection reagent (Sigma-Aldrich, Cat. No. 6366244 001). S2 cells (10^6) were seeded in 500 μ l of medium per well of a 24-well plate and allowed to attach for 3–4 h. Cells were transfected with a mixture containing 190 ng of a plasmid expressing Cas9 and blasticidin resistance (pIB_Cas9_Blast) [30] and 110 ng of total sgRNA plasmids (36.6 ng each of sgRNA1, sgRNA3, and sgRNA5) in 100 μ l of serum-free medium. XtremeGENE HP transfection reagent (1.5 μ l) was added to the DNA mixture, vortexed briefly, and incubated for 35 min at room temperature (RT) before dropwise addition to cells. After 24 h, medium was replaced with selection medium containing blasticidin (final concentration 25 μ g/ml, Gibco, Cat. No. A1113903). After 3 days in selection medium, cells were taken out of selection, and polyclonal pools were screened by PCR for the *roX2* deletion. Genomic DNA was isolated using a PCR clean-up kit (Macherey-Nagel, REF 740609.50), and PCR amplification was performed with Phusion polymerase (Thermo, Cat. No. F-530XL) using the following conditions: 98°C for 1 min; 35 cycles of 98°C for 20 s, 65°C for 30 s, 72°C for 1.5 min; final extension at 72°C for 10 min. PCR products were analyzed on 2% agarose gels. For more information on the sequences and primers used, see [Supplementary Table S1](#). For clonal isolation, edited polyclonal pools were seeded at low density of 1000, 2000, or 3000 cells per 6 cm plate and allowed to attach for at least 1 h. Cells were overlaid with a 1:1 mixture of freshly prepared 2 $\times$ Schneider's medium (Serva, Cat. No. 47521.04) and 0.8% low melting agarose (BioRad, Cat. No. 161-3111) equilibrated to 37°C (3.5 ml per plate). Plates were sealed with parafilm, placed in a humidified chamber, and incubated at 26°C for 2–3 weeks to allow clonal outgrowth. Individual colonies were picked, expanded for ~2 weeks, and genotyped by PCR with sequencing confirmation of deletion junctions.

Generation of stable transgenic cell lines

Stable transgenic lines expressing *roX* variants were generated by transfection of *roX2* knockout S2 cells (KO-B) using Effectene transfection reagent (Qiagen, Cat. No. 301425) according to the manufacturer. Briefly, 8×10^5 cells were seeded

in 350 μ l of medium per well of a 24-well plate and allowed to adhere for 3–4 h. Transfection complexes were prepared by mixing 200 ng of *roX* expression plasmid (*roX* being expressed under the HSP70 promoter) and 10 ng of blasticidin resistance plasmid in 60 μ l of Buffer EC, followed by addition of 1.6 μ l of Enhancer, vortexing, and incubation at RT for 2–5 min. Effectene reagent (5 μ l) was then added, vortexed for 10 s, and incubated at RT for 5–10 min. The transfection mixture was diluted with 350 μ l of medium and added dropwise to cells. After 3 days, medium was replaced with selection medium containing 10 μ g ml⁻¹ blasticidin (Gibco, Cat. No. A1113903). Cells were maintained under blasticidin selection and monitored for transgene expression for 3–4 weeks, then taken out of selection and passaged several times before use in experiments.

Biochemical methods

Immunoblotting

Cells (1.5×10^6) were harvested, washed once with phosphate-buffered saline (PBS), and resuspended in 60 μ l of Laemmli sample buffer. Extracts were boiled at 95°C for 10 min and stored at -20°C until use. Proteins were separated by sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS–PAGE) on either 8% acrylamide gels (Serva, Cat. Nos 43260.01, 43262.01) for larger proteins or 14% gels for histones and histone marks, loading ~2.5 $\times 10^5$ cell equivalents per lane. Electrophoresis was carried out in Running Buffer (25 mM Tris, 192 mM glycine, 0.1% SDS), and proteins were transferred to nitrocellulose membranes (Cytiva, Cat. No. GE10600002) by wet transfer at 350 mA on ice for ~90 min. Transfer Buffer (25 mM Tris, 192 mM glycine) was supplemented with 20% methanol. Membranes were blocked in 5% bovine serum albumin (BSA) in Tris-buffered saline (TBS) for 1 h at room temperature, then incubated with primary antibodies diluted in TBS overnight at 4°C. Primary antibodies were used to blot for MSL2 (guinea pig α -MSL2), MOF, MLE, H4K16ac, Lamin, and histone H3. Following three 10 min washes in TBS-T (0.1% Tween-20, Sigma, Cat. No. P9416), membranes were incubated with LI-COR IRDye secondary antibodies in TBS for 1 h at room temperature. For a detailed list of antibodies and their dilutions used, see [Supplementary Table S1](#). After three additional washes in TBS-T, signals were detected using the LI-COR Odyssey CLx Imaging System. Band intensities were quantified using the Image Studio Lite Software (Version 5.2.5, LI-COR Biosciences) and normalized to Lamin (for non-histone proteins) or H3 (for histone modifications) and compared with WT levels. Western blot protein ratios were quantified from biological replicates (cells harvested on different days), with technical replicates representing independent gel images of each biological sample. The data were processed using R (v4.5.1) [34]. For normalization, technical replicates were normalized to the mean WT value within the same gel image, then averaged to yield a single fold change (FC) value per biological replicate. Outliers among biological replicates were identified using the median absolute deviation (MAD) method, and excluded if at least three replicates remained per genotype. Statistical comparisons between genotypes and the WT were performed using two-sided *t*-tests on log₂-transformed FC values of biological replicates. Significance levels: ***P* < 0.001, **P* < 0.01, *P* < 0.05, ns = not significant. Plots display individual biological replicates (larger points), technical replicates (smaller

faint points), means, and standard deviation (SD) error bars.

RNA isolation, cDNA synthesis, and real-time quantitative PCR

Total RNA from 4.5×10^6 cells was isolated using the RNeasy kit (Qiagen, Cat. No. 74 104) according to the manufacturer's instructions. RNA concentrations were measured, and samples were stored at -80°C . For DNase treatment, 2 μg of RNA was incubated with 1 μl of recombinant DNase (Roche, Ref. 04 716 728 001) in $1\times$ DNase buffer with 1 μl of recombinant RNase inhibitors (RNasin, Promega, Cat. No. N2511) in a final volume of 20 μl for 1 h at 37°C . cDNA from 1 μg of DNase-treated RNA was synthesized using the SuperScript III reverse transcriptase kit and random hexamers (Invitrogen, Cat. No. 18 080 051). The cDNA samples were treated with 0.5 μl of RNase H (NEB, Cat. No. M0297L) at 37°C for 20 min, and finally cDNA was stored at -20°C . Quantitative PCR (qPCR) was performed in 384-well plates using Fast SYBR Green Master Mix (Thermo, Cat. No. 4385610) on a LightCycler 480 instrument (Roche). Each reaction contained $1\times$ SYBR Green Master Mix, 1 μM final concentration for each of the forward and reverse primers, and 1:10 diluted cDNA for all primers used. For a detailed list with oligos used in this study, see [Supplementary Table S1](#). Reactions were performed in technical triplicates, with biological replicates representing cells harvested and processed on different days. Data were analyzed in R (v4.5.1) using the $\Delta\Delta\text{Cp}$ method, with glyceraldehyde phosphate dehydrogenase (GAPDH) as the reference gene after averaging the technical replicates, and relative expression was calculated as $2^{-\Delta\Delta\text{Cp}}$, and expressed as the mean FC over the WT of $n = 4$ biological replicates. Error bars represent the SD of the biological replicates.

CUT&RUN and chromatin immunoprecipitation after micrococcal nuclease treatment sequencing

CUT&RUN was performed as described previously [25, 35]. For normalizing MSL2, pre-immune serum (PPI) was used as a negative control. For H4K16ac, rabbit IgG was used as negative control for normalization. Chromatin immunoprecipitation sequencing (ChIP-seq) on micrococcal nuclease (MNase)-digested chromatin and sonicated chromatin was performed as previously described in [36] and [37]. DNA yielded by both ChIP-MNase and CUT&RUN was quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Cat. No. Q32851). Libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645) according to the manufacturer for ChIP and with subtle modifications for CUT&RUN (<https://doi.org/10.17504/protocols.io.bagaibse> [38]). Quality control and analysis of all libraries was done with the TapeStation system (Agilent). Libraries were sequenced on the NextSeq2000 (Illumina) instrument, yielding typically 20–25 million paired-end reads per sample for ChIP-MNase or ~ 5 million for CUT&RUN. All sequencing was performed at the Laboratory of Functional Genomic Analysis (LAFUGA, Gene Center Munich, LMU).

Microscopy

Immunofluorescence microscopy

Immunofluorescence microscopy (IFM) analysis of S2 cells was performed as described in [10, 37] with slight alterations. Briefly, coverslips were placed in 12-well plates and

UV-sterilized for 10 min. After sterilization, coverslips were coated with poly-L-lysine (Sigma #P8920, 0.01% w/v final concentration) and air-dried under sterile conditions. Cells ($0.8\text{--}1 \times 10^6$) were seeded on the dried coverslips and allowed to attach for at least 1 h. After adhering to the coverslips, the cells were washed briefly with PBS at room temperature, fixed in ice-cold 2% formaldehyde (FA; Polysciences, Cat. No. 04018-1) in PBS for 7.5 min, and permeabilized in 1% FA and 0.25% Triton X-100 (Sigma, Cat. No. T8787) in PBS for 7.5 min on ice. Following two washes with ice-cold PBS, cells were blocked in 3% BSA (Serva, Cat. No. 11930.03) in PBS for 1 h at room temperature. Primary antibodies were diluted in 0.1% Triton X-100 and 1.2% normal donkey serum (NDS; Johnson Immuno, Cat. No. 017-000-121) in PBS and incubated overnight at 4°C . After two PBS washes at room temperature, cells were incubated with secondary antibodies diluted in 0.1% Triton X-100 and 1.2% NDS in PBS for 1 h at room temperature. For more information on the antibodies and fluorophores used, see [Supplementary Table S1](#). Following two PBS washes, nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI; Sigma, D-9542) in PBS for 2 min. Coverslips were washed once with PBS and once with water, then mounted using Vectashield antifade mounting medium for fluorescence (Vector Labs, Cat. No. H-1000) and sealed with nail polish.

Confocal microscopy

Confocal microscopy was performed at the Core Facility Bioimaging of the Biomedical Center with an inverted Leica SP8X WLL microscope, equipped with a 405 nm laser, a WLL2 laser (470–670 nm), and an acusto-optical beam splitter. Images were recorded with a $\times 100/1.4$ NA STED white objective with zoom factor of 4.5, yielding a field of view of $25.8 \times 25.8 \mu\text{m}$. The image format was set to 512×512 pixels, resulting in a pixel size of 50.4 nm. Z-stacks were acquired in xyz scan mode using galvo-based z-scanning (z-galvo) with a constant step size of 0.24 μm and unidirectional x-scanning at a speed of 200 Hz (pixel dwell time: 7.7 μs). The confocal pinhole was set to 1.0 Airy units using 580 nm as the reference wavelength. Fluorescence signals of DAPI, MSL2 (Cy2), MSL3 (Cy3), and H4K16ac (Alexa Fluor 647) were recorded frame sequentially to avoid channel cross-talk, with the following excitation and detection settings (excitation wavelength; detector: detection window): DAPI (405 nm; PMT1: 415–480 nm), Cy2 (488 nm; HyD2: 498–535 nm), Cy3 (551 nm; HyD4: 561–620 nm), and Alexa Fluor 647 (650 nm; HyD5: 660–710 nm). Detector gain was set to 700 V (PMT1) and 50% (HyDs in standard mode). Images were acquired at 8-bit depth and subsequently deconvolved using the Huygens Software (v17.04) and stored at 16-bit depth ($512 \times 512 \times 20$ z-planes; z-step $\sim 0.24 \mu\text{m}$). Maximum intensity projections and scaling of image intensity were performed using Fiji (ImageJ, 2.16.0) [39].

Widefield microscopy

Widefield microscopy was performed at the Core Facility Bioimaging of the Biomedical Center with an inverted Leica THUNDER Imager 3D Live Cell TIRF with DMi8 stand equipped with LED5 multi-LED illumination (390/475/560/635 nm) and a fast filter wheel with emission filters centered at 460, 535, 590, and 642 nm. Images were recorded with an HC PL APO $\times 100/1.47$ oil CORR TIRF objective, yielding a field of view of $133 \times 133 \mu\text{m}$ per tile

(2048 × 2048 pixels at ~65 nm pixel⁻¹ sampling, Leica K8 sCMOS camera, 16-bit depth). Z-stacks were acquired sequentially across four channels using ~0.3 µm z-steps and a region spanning 3 × 4 fields of view. Images were acquired with the following settings: DAPI for DNA (Ex 391/32 nm, Em 435/30, 8% LED, 100 ms), Cy2 for MSL2 (Ex 479/33 nm, Em 519/25 nm, 12% LED, 100 ms), Cy3 for MSL3 (Ex 554/24 nm, Em 594/32 nm, 15% LED, 300 ms), and Alexa Fluor 647 for H4K16ac (Ex 638/31 nm, Em 695/58 nm, 20% LED, 200 ms). Gain was set to 1 (HDR Combined Gain-16 bit mode) and aperture diaphragm to 8 for all channels. Further processing of the images was performed using Fiji and macro scripts.

Image processing

Images were processed (maximum intensity projections and scaling) uniformly across conditions for display purposes. Image analysis was performed in R (v4.5.1) with the following packages: EBImage (v4.50.0, for image processing and feature extraction) [40], tidyverse suite of tools (v2.0.0, for pre-processing, statistical analysis, and visualization) [41], corr (v0.4.5, correlation analysis), cluster (v2.1.8.1, clustering validation), and umap (v0.2.10.0, dimensionality reduction) [42]. Briefly, single-cell images were read as multi-channel TIFF files, and the MSL2 channel (channel 2) was extracted for analysis. Images were normalized to a 0–1 intensity range prior to feature extraction. For image-based profiling, MSL2 immunofluorescence intensity distributions and Haralick texture features were extracted from single cells [43]. For each cell, 46 quantitative features were extracted. For segmentation, images were Gaussian-blurred and thresholded at 0.85 to identify high-intensity regions corresponding to MSL2-enriched foci. Prior to dimensionality reduction, highly correlated variables were removed. Uniform Manifold Approximation and Projection (UMAP) was performed using the umap R package with parameters $n_neighbors = 100$ and $min_dist = 0.1$. Statistical differences in feature values between clusters were assessed using Kruskal–Wallis tests with effect size filtering [$\epsilon^2 > 0.1$, false discovery rate (FDR)-adjusted $P < 0.05$], followed by pairwise Wilcoxon rank-sum tests with FDR correction to identify cluster-specific features. The proportion of cells assigned to each cluster was quantified for each genotype and rescue condition. For interpretability, we describe the UMAP clusters by their enriched genotypes and dominant image characteristics: the WT/*roX2*-FL-enriched cluster corresponds to nuclei with a more spatially defined (chromosome territory-like) MSL2 signal and stronger local texture/contrast, consistent with the presence of high-intensity MSL2-enriched regions detected after Gaussian blurring and thresholding. In contrast, KO and partially or non-rescuing *roX* derivatives map predominantly to clusters characterized by more diffused nuclear MSL2 patterns. These terms refer to quantified image intensities and texture properties. Data for one out of four independent replicates are shown, a cumulative UMAP colored by genotype for each other replicate can be found in [Supplementary Fig. S1A](#).

Chromosome spreads and karyotype analysis

Chromosome spreads were prepared from Schneider S2 cells as described in [44], except that cells were fixed for 10 min using 3.7% formalin (Merck, 8187081000) in PBST (PBS + 0.1% Triton X-100) solution and only stained with 1 µg ml⁻¹

DAPI in PBST. Slides were imaged using a ×100 oil immersion objective on a Nikon-CREST widefield microscope. Image deconvolution was performed with the Huygens Essential Software from SVI (version 24.10.0p7 64b). Individual chromosome identification was done using the *Drosophila* S2 cell karyotype in supplementary fig. S3a of Olszak *et al.* (2011) [44]. For a comprehensive summary of the mitotic spreads analyzed in this study, see [Supplementary Table S2](#).

Data analysis

Sequencing data pre-processing

Paired-end CUT&RUN and ChIP-MNase sequencing reads were aligned to the *D. melanogaster* dm6 reference genome (UCSC) using Bowtie2 (v2.2.9, -local -very-sensitive-local -no-unal -no-mixed -no-discordant) [45]. Maximum fragment sizes were set to 700 bp for CUT&RUN and 1000 bp for ChIP-MNase to accommodate the distinct fragment size distributions of each method. SAM files were converted to BAM format retaining high-quality alignments (samtools v1.22.1, samtools view -q 2) [46], name-sorted, and converted to BED format using BEDtools (v2.31.1) [47] bamToBed to extract fragment coordinates. CUT&RUN data were normalized to total read counts. Scaling factors were calculated as 10⁶ divided by the total number of filtered reads per sample (excluding secondary, supplementary, and unmapped reads, and reads with mapping quality < 30). Normalized coverage tracks were generated using deepTools2 (v3.5.0) [48] bamCoverage at 10 bp resolution with target-specific smoothing (20 bp for MSL2/PPI; 80 bp for H4K16ac/IgG). Pairwise target-to-control comparisons were computed using bamCompare to generate ratio tracks with a pseudocount of 1. Pre-immune serum (PPI) and rabbit IgG were used as negative controls for the targets MSL2 and H4K16ac, respectively. For ChIP-MNase data, spike-in normalization was performed using reads aligned to the *Drosophila virilis* droVir3 genome. Scaling factors were calculated as 10⁶ divided by the spike-in read count for each sample. Normalized to Input coverage tracks were generated using BEDtools (v2.31.1) genomecov and converted to bigWig format using bedGraphToBigWig (UCSC utilities, 3.4.1) [49]. Biological replicates were merged ($n = 3$ or $n = 2$ for stable lines) using Wiggletools (1.2.2) [50] and visualized using the IGV genome browser (v2.19.4) [51] and the dm6 *D. melanogaster* genome release. For correlation analysis between the CUT&RUN biological replicates across targets and genotypes, after total reads and negative control normalization, see [Supplementary Fig. S2A](#).

Window-based binding site identification using csaw

DCC-binding sites were identified from CUT&RUN data derived from WT cells using a window-based differential binding approach implemented in the csaw R/Bioconductor package (v1.42.0 [52]). Analysis parameters were optimized separately for the transcription factor MSL2 and the histone modification H4K16ac. For MSL2, reads were counted in 20 bp sliding windows with fragment extension of 100 bp and maximum fragment length of 250 bp. For H4K16ac, reads were counted in 120 bp sliding windows with fragment extension of 150 bp and maximum fragment length of 500 bp; windows were retained if signal was present in at least three replicates. Read counts were TMM-normalized (edgeR, v4.6.3) [53], and dispersions were estimated robustly. Differential binding between the target and negative control was assessed us-

ing quasi-likelihood *F*-tests. Significance thresholds were $FDR \leq 0.01$ and $|\log_2 FC| \geq 2$ for MSL2, and $FDR \leq 0.05$ and $|\log_2 FC| \geq 1$ for H4K16ac. Thresholds were chosen to reflect the biology of each target: stringent parameters for MSL2 to identify discrete, high-confidence binding sites, and more permissive parameters for H4K16ac to capture the broader landscape of histone acetylation across MSL-bound domains, while maintaining optimal signal-to-noise ratios for each dataset. Significant windows were merged within 50 bp to define MSL2 CUT&RUN csaw sites (MCCSs) and H4K16ac CUT&RUN csaw sites (HCCSs). Significant windows passing all filters were merged into contiguous binding regions using the GenomicRanges package (v1.60.0) [54] with a maximum gap of 50 bp. Merged regions were annotated to genomic features using the ChIPseeker package (v1.44.0) [55] with TxDb.Dmelanogaster.UCSC.dm6.ensGene (v3.12.0), defining promoters as 1 kb upstream from the transcription start site (TSS).

Analysis of CUT&RUN data

High-resolution merged MCCSs were compared with previously characterized HASs [29] and PionX sites [56]. Overlap analysis was performed by creating a union of all peak sets using GenomicRanges, then determining membership of each unified region in the original sets using countOverlaps. Overlaps were visualized using UpSetR (v1.4.0) [57]. Signal intensity at binding sites was quantified by extracting mean enrichment values from merged bigWig files (ratio of target/control) using rtracklayer (v1.68.0) [58].

Heatmap and metaplot visualization

Signal profiles at binding sites were visualized using deepTools2 (v3.5.0). For site-centered analysis, computeMatrix reference-point was used with flanking 2 kb windows centered on binding site midpoints. For gene body analysis, computeMatrix scale-regions was applied to MSL2-bound genes with 2 kb upstream/downstream flanks and 5 kb scaled gene body regions. Matrices were sorted by mean signal in descending order using the WT sample. Heatmaps were generated with plotHeatmap using color maps with z-score ranges of 0–4 and 0–8.

Motif analysis

Genomic sequences underlying DCC-binding sites were extracted from the *D. melanogaster* genome (dm6) using the BSgenome.Dmelanogaster.UCSC.dm6 R package (v1.4.1). Binding site regions (MCCSs, HASs, and PionX sites) were resized to 200 bp centered on the summit, and sequences were extracted using the Biostrings package (v2.76.0) [59]. Control background sequences were generated by randomly sampling genomic regions matched for chromosome composition and sequence length. For each binding site, a corresponding background region was selected from the same chromosome with matching width. Background regions overlapping true binding sites were excluded to ensure specificity. Random sampling was performed with a fixed seed [42] for reproducibility. STREME (Sensitive, Thorough, Rapid, Enriched Motif Elicitation; MEME Suite v5.5.8) [60] was used for discriminative motif discovery comparing target sequences against the matched background set. Parameters: `-dna -nmotifs 5 -minw 6 -maxw 20 -thresh 0.05`. STREME was applied to all binding site categories. Discovered motifs were compared with known *Drosophila* transcription factor-binding motifs using Tomtom

(MEME Suite v5.5.8) [61] against multiple databases: FlyReg v2, Fly Factor Survey, OnTheFly 2014, and dmmpmm2009. Parameters: `-min-overlap 5 -dist pearson -evalue -thresh 10.0`. Motif occurrences across MCCSs were subsequently quantified using FIMO (MEME Suite v5.5.8) with default parameters. Overlap between MCCSs and motif occurrences or gene set promoter regions was assessed using countOverlaps from the GenomicRanges package (v1.60.0). Gene categories were defined as follows: active genes [counts per million (CPM) ≥ 5 in S2 RNA-seq across all WT replicates], housekeeping genes [37], and active-regulated genes (active genes excluding housekeeping genes). Promoter regions (1 kb upstream of the TSS) were constructed for each gene set, and overlap with MCCSs was assessed independently, such that categories are non-exclusive. All analyses were performed in R (v4.5.1).

DNA copy number estimation from ChIP-MNase input samples

Chromosome copy number was estimated from ChIP-MNase-seq input samples, which represent chromatin that has been mildly sonicated, MNase digested, and pre-cleared (see ChIP-MNase sequencing above), and thus serves as a proxy for genomic DNA content. Replicate input BAM files were processed to generate read coverage across 50 kb non-overlapping genomic bins. For each bin, read counts were converted to an RPKM-like metric (reads per kilobase per million mapped reads) to normalize for bin length and library size differences. Spike-in normalization was applied to account for technical variation between samples: each sample was aligned to both the *D. melanogaster* dm6 reference genome and the *D. virilis* spike-in genome (droVir3). The ratio of *D. melanogaster* to spike-in reads was used to derive per-sample scaling factors, enabling quantitative comparison of absolute DNA content across conditions. Downstream processing, statistical analyses, and visualizations were performed in R (v4.5.1) using the GenomicRanges (v1.60.0), Rsamtools (v2.24.1) [62], and tidyverse (v2.0.0) packages.

Chromosome-level copy number analysis

Per-chromosome copy numbers were calculated by computing the median of bin-level copy number values for each chromosome arm (chr2L, chr2R, chr3L, chr3R, chr4, and chrX) within each biological replicate. To convert relative coverage values to absolute copy numbers, values were scaled to the WT autosomal baseline: the median of per-chromosome medians across autosomes (chr2L, chr2R, chr3L, and chr3R) in WT samples was set to the expected copy number of 4, reflecting the tetraploid nature of S2 cells. All chromosome values were then scaled proportionally using this baseline. For each condition (WT, KO-A, KO-B, and rescue), the mean and SD of scaled copy numbers were calculated across biological replicates ($n = 3$ per condition). Error bars in figures represent the SD of replicate-level medians converted to the absolute copy number scale. A comprehensive comparison of the scaled copy number per bin for the X chromosome, for all conditions (including replicates), can be found in [Supplementary Fig. S5](#).

RNA sequencing and analysis

RNA isolation and library preparation

Total RNA was isolated from S2 cells using the RNeasy Mini Kit (Qiagen, Cat. No. 74106) according to the manufacturer's instructions. Cells were counted, and $\sim 3\text{--}6 \times 10^6$ cells were

harvested per sample. RNA concentration was measured, and samples were stored at -80°C . RNA was treated with DNase as described earlier. DNase-treated RNA was cleaned-up using the RNeasy Mini Kit and quality-assessed by Qubit RNA quantification and TapeStation analysis to verify RNA integrity. Strand-specific RNA-seq libraries were prepared from 100 ng of DNase-treated total RNA using the QIAseq Fast-Select RNA Library Kit (Qiagen, Cat. No. 334302) according to the manufacturer's protocol with the following modifications: RNA fragmentation was performed at 95°C for 10 min and rRNA depletion was performed using $0.1\times$ FastSelect reagent. Libraries were amplified for a total of 20 PCR cycles, and concentrations were measured by Qubit HS DNA assay. Library size distribution was assessed by TapeStation, and a size selection step was performed using QIAseq beads (Qiagen, Cat. No. QC13.3) ($0.6\times$ ratio for removal of larger fragments, followed by $1\times$ for cleanup) to remove fragments with aberrant size distributions. Final libraries were eluted in 10 μl of nuclease-free water and submitted for sequencing. RNA-seq was performed in two batches. Batch 1 included WT S2 cells ($n = 4$ biological replicates) and *roX2* knockout clone KO-B ($n = 4$ biological replicates). Batch 2 included additional WT replicates ($n = 2$), *roX2* knockout clone KO-A ($n = 4$ biological replicates), and rescue cells expressing full-length *roX2* (res; $n = 2$ biological replicates). Biological replicates were derived from independently passaged cell populations.

RNA-seq data processing and analysis

Raw paired-end sequencing reads were processed using the snakePipes (v3.2.0) non-coding RNA-seq workflow [63]. Adapter sequences were trimmed using cutadapt via Trim Galore (v4.6/0.6.10). Quality control was performed using FastQC (v0.12.1) [64] and summarized with MultiQC (1.25) [65]. Trimmed reads were aligned to the *D. melanogaster* dm6 reference genome using STAR (v2.7.10b) [66] with parameters optimized for multi-mapping reads (`-outFilterMultimapNmax 1000 -outFilterMismatchNoverLmax 0.1 -outSAMstrandField intronMotif`). Gene-level read counts were quantified using the Ensembl gene annotation (release 96). Gene expression was quantified as transcripts per million (TPM) for visualization and as raw counts for normalization. RNA-seq differential expression analysis was performed using DESeq2 (v1.48.2) [67] on all annotated genes. Raw read counts were filtered to retain genes with at least 10 counts in a minimum number of samples equal to the smallest experimental group size. Differential expression was assessed using a batch-aware design with the Wald test in a single DESeq2 model including all samples. Genes were considered differentially expressed (DE) if they met the following criteria: adjusted *P*-value (*padj*) < 0.01 and $|\log_2\text{FC}| > 1$. Pairwise comparisons were performed between each KO condition, rescue, and WT (all replicates). Volcano plots were generated to visualize DE genes, with the top 20 genes by significance and top 20 by absolute $\log_2\text{FC}$ annotated (Supplementary Fig. S4C).

Gene expression and dosage compensation analysis

For visualization of individual gene expression (e.g. *roX1* and *roX2*), TPM values were $\log_{10}$ -transformed [$\log_{10}(\text{TPM} + 1)$] and plotted per condition with individual replicates shown. Error bars represent the SD of $\log_{10}$ -transformed values across replicates. To assess dosage compensation status, analysis was restricted to protein-coding genes located on ma-

ior chromosome arms (chr2L, chr2R, chr3L, chr3R, chr4, and chrX). Genes with < 10 counts in at least three samples (or two when not available) were excluded. Raw counts were normalized using DESeq2 median-of-ratios normalization. For each KO condition, mean normalized counts were calculated across replicates, and $\log_2$ FCs relative to the WT were computed per gene as $\log_2[(\text{KO} + 1)/(\text{WT} + 1)]$. $\log_2\text{FC}$ distributions were compared between X-chromosomal and autosomal genes for each genotype to evaluate whether X-linked gene expression was selectively affected by *roX2* loss.

Copy number-corrected expression analysis

To assess the relative contributions of DNA copy number and transcriptional regulation to X chromosome expression, copy number-corrected expression values were calculated using R with DESeq2 (v1.48.2), tidyverse (v2.0.0), and patchwork (v1.3.2) [68] packages. For each condition, per-gene RNA-seq expression values (DESeq2-normalized counts) were divided by the corresponding chromosome copy number estimated from ChIP-MNase sequencing input samples. X:A (X/autosome) expression ratios were calculated both with and without copy number correction to distinguish canonical DCC-mediated dosage compensation from other mechanisms. This method enables the decomposition of dosage compensation into copy number and transcriptional contributions. The buffering effect of elevated X chromosome copy number was quantified as the ratio of observed X:A expression to the expected ratio based on copy number alone.

Results

Generation and validation of *roX2* knockout S2 cell lines

We sought to investigate the functional requirements of the long non-coding *roX* RNA for dosage compensation in *Drosophila* in a simple, proliferative cell model. Descendants of the S2 cell line have proven very useful for studying dosage compensation. These aneuploid cells are mainly tetraploid, having two X chromosomes, but four copies of each autosome. They compensate the 2-fold reduced number of X chromosomes relative to autosomes transcriptionally through about equal contributions of a generic feedback mechanism that works on all chromosomes and a fixed 1.4-fold transcription boost due to the action of the MSL DCC [28]. The role of *roX* RNAs had not been analyzed in this well-controlled system.

S2 cells do not express the *roX1* RNA to significant levels, and dosage compensation in these cells thus is effectively mediated by *roX2* [14]. To generate *roX*-deficient cell lines, we deleted the *roX2* gene using CRISPR/Cas9 genome editing (Fig. 1). We transfected S2 cells with plasmids expressing Cas9 and encoding three sgRNAs targeting the *roX2* locus. Single clones harboring the plasmids were selected with blasticidin, isolated, and expanded in a time course of 6 weeks (Fig. 1A). Only two independent KO clones were recovered, designated KO-A and KO-B, and validated by PCR and sequencing of deletion junctions. Genome browser visualization of DNA sequencing reads from our ChIP input samples (see below) across the *roX2* locus verified the deletion of the *roX2* region in both alleles (Fig. 1B).

RNA-seq and qPCR analysis confirmed the absence of *roX2* transcripts in both KO-A and KO-B (Figs 1C and 2B;

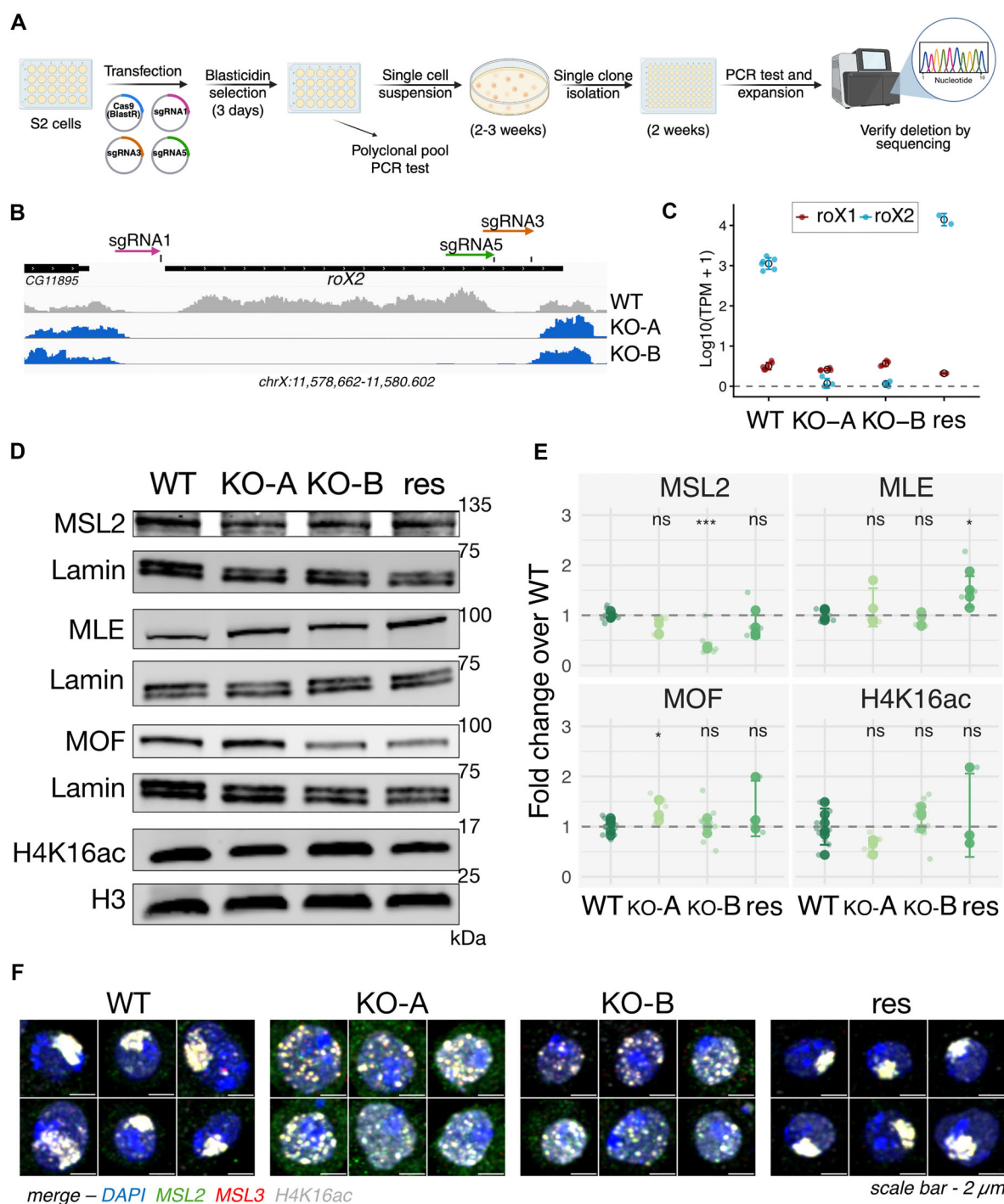


Figure 1. Generation and characterization of *roX2* KO S2 cell lines. **(A)** CRISPR/Cas9 strategy and workflow used to isolate *roX2* KO clones. Created in BioRender. Jayakrishnan, M. (2026) <https://BioRender.com/dwajdp>. **(B)** Deletion of the *roX2* gene. Top, schematic illustration of the positions of three sgRNAs relative to the *roX2* gene (black line). Bottom, genome browser view of ChIP-MNase input coverage tracks across the *roX2* locus (coordinates below) for the WT (gray) and the two *roX2* KO clones (KO-A and KO-B, blue). Loss of coverage reveals the deletion of the *roX2* region in both clones. **(C)** Quantification of *roX* RNAs in WT, KO, and rescue (res) lines. Total RNA-seq reads [$\log_{10}(\text{TPM} + 1)$ values] for *roX1* (blue) and *roX2* (red) are plotted for the indicated lines. Data points denote biological replicates. *roX2* expression is undetectable in KO-A and KO-B and restored in the rescue line; *roX1* expression remains very low across conditions. **(D)** Representative immunoblot probing DCC proteins and H4K16 acetylation (H4K16ac). Whole-cell extracts from WT, KO-A, KO-B, and rescue cells were probed for MSL2, MLE, MOF, and H4K16ac. Lamin and histone H3 serve as loading controls. Positions of size markers (kDa) are indicated. **(E)** Quantification of immunoblot signals from immunoblots as in (D). Signals are shown as FC relative to the WT after normalization to Lamin (non-histone proteins) or histone H3 (H4K16ac). Each data point represents an independent biological replicate; technical replicates are shown as smaller points, when applicable. The dashed line indicates WT protein expression (FC = 1). Statistical significance is annotated as indicated (ns, not significant). **(F)** Confocal microscopy to visualize DCC enrichment on the X chromosome. Galleries of six representative confocal images of nuclei of the indicated cell lines, stained for MSL2 (green), MSL3 (red), H4K16ac (gray), and DAPI (blue) marking nuclei. The images are deconvolved Z-maximum projections of multiples stacks; channels are merged and auto-scaled for display purposes. WT cells show the characteristic enrichment of MSL2/MSL3 and H4K16ac at the X-chromosome territory. In KO cells, the complex is delocalized. The rescue line restores X association. Scale bar, 2 μm.

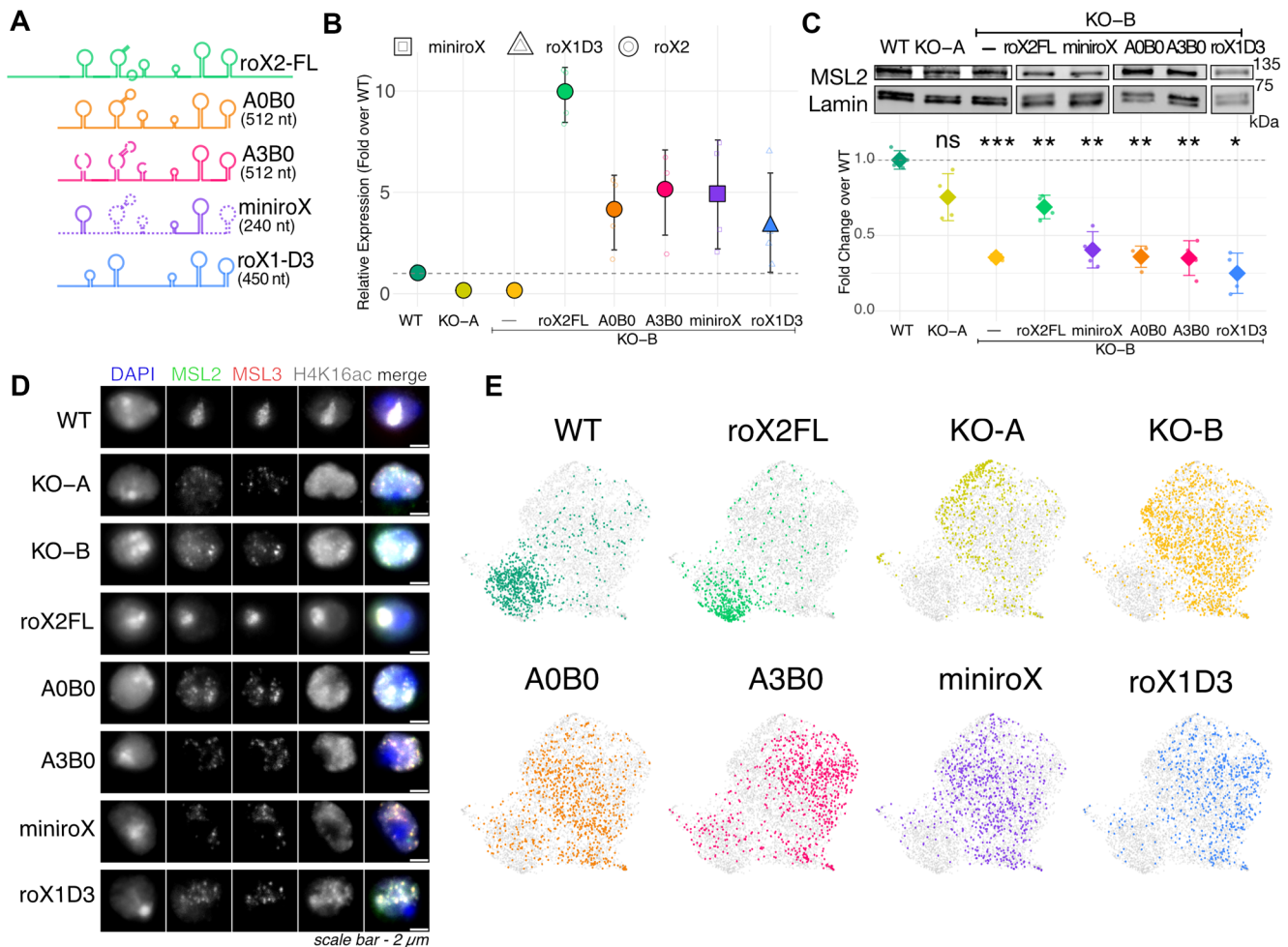


Figure 2. An experimental system to study *roX* requirements for X chromosome targeting. **(A)** Schematic depiction of *roX* constructs used for rescue experiments. Conserved stem-loop elements are indicated. RNA lengths are shown in parentheses. For details, see text. **(B)** Expression of *roX* transgenes. Quantitative reverse transcription-PCR (RT-qPCR) quantification of *roX* variant expression in stable cell populations using specific primer sets for *roX2* (circles), *miniroX* (squares), and *roX1-D3* (triangles). Values are normalized to *GAPDH* RNA and expressed relative to the WT using the $\Delta\Delta C_p$ method. The mean $\pm$ SD of four biological replicates is shown. The dashed line indicates WT expression. **(C)** MSL2 protein levels in cells expressing *roX* variants. Top, representative immunoblot of MSL2 across WT, KO clones, and rescue lines expressing the indicated constructs. Lamin serves as loading control. Bottom, quantification of MSL2 normalized to the internal control and relative to WT expression (dashed line). The mean signal of four biological replicates (data points) $\pm$ SD is shown. Statistical significance is indicated as annotated (ns, not significant). **(D)** Widefield microscopy images of representative cells illustrating DCC localization and H4K16ac upon *roX* RNA expression. Cells were stained for MSL2 (green), MSL3 (red), H4K16ac (gray), and DAPI (blue) marking nuclei. Individual channels are displayed in grayscale. All images were scaled uniformly for display. Scale bar, 2 μ m. **(E)** Classification of MSL2 localization states in single cells by image-based profiling. UMAP projection of single cells derived from widefield microscopy and quantified using MSL2 immunofluorescence intensity distributions and Haralick texture features. Each point represents one cell, colored by genotype/condition, as indicated. The WT/*roX2*-FL-enriched region corresponds to chromosome territory-like MSL2 enrichment, whereas KO and non-rescuing constructs map predominantly to diffuse/delocalized nuclear MSL2 patterns (Fig. 2D). One representative replicate of this experiment is shown; for UMAP projections of the other replicates, see [Supplementary Fig. S1A](#).

[Supplementary Fig. S4C](#)). *roX1* expression remained negligible as in WT cells, indicating that *roX1* is not increased to compensate for the loss of *roX2*. In a functional rescue experiment, we generated a stable line expressing a full-length *roX2* transgene in the KO-B background (res), which restored *roX2* transcripts to levels comparable with the WT (Fig. 1C; [Supplementary Fig. 4C](#)).

The MSL proteins and *roX* RNA are stable as components of the DCC complex, but unstable if the integrity of the DCC is disrupted by depletion of a protein subunit [21, 69]. To explore whether *roX2* loss affected the MSL protein abundance, we monitored MSL protein levels in the KO lines by immunoblotting. We found that MSL2 was moderately reduced

in both KO clones, but more prominently in KO-B, while MLE and MOF levels remained largely unchanged (Fig. 1D, E). Re-expression of *roX2* in the rescue line partly restored WT MSL2 levels. The levels of H4K16ac, the epigenetic readout of DCC action, showed modest variation but overall were not substantially altered.

The hallmark of functional dosage compensation in male *Drosophila* cells is the selective binding of DCC to the X chromosome. Immunostaining of the MSL proteins visualizes their binding to a compact chromosomal territory. Depletion of any MSL protein abolishes the X chromosome “territory” staining of the complex [21, 27, 70]. Confocal microscopy of WT cells demonstrated the characteristic co-localization of MSL2,

MSL3, and H4K16ac at the dosage-compensated X chromosome (Fig. 1F). Strikingly, this subnuclear enrichment was lost in both *roX2* KO clones, where the complex was delocalized throughout the nucleus in speckles that may correspond to constitutive heterochromatin repeats [71, 72]. Expression of the *roX2* transgene in the rescue line restored X-chromosome association of the DCC, demonstrating that *roX2* is required for faithful targeting.

An experimental system for *roX* structure–function analyses

The successful rescue of DCC function by expression of *roX2* in the KO-B cell population suggested a strategy through which features of *roX* RNA function could be evaluated. As a proof of principle, we tested several *roX2* derivatives that in the systematic analyses of Akhtar and colleagues had been shown to support a certain level of viability during male fly development. Their *A0B0* construct ([17], Fig. 2A) represents “exon 3” of *roX2* with a total length of 512 bp, which contains all known binding sites for the helicase MLE. A derivative of this sequence, designated *A3B0*, has the three 5′ stem–loop structures mutated that attract MLE in an ATP-independent manner, but maintains critical 3′ stem–loops that are the target for MLE remodeling [17, 19, 20]. Ilik *et al.* also generated a “*miniX*” variant, in which central *roX2* sequences were deleted, leaving the 5′ and 3′ stem–loops intact [18]. Their data suggest that this might be the shortest *roX* derivative that can sustain male life. We selected the *A0B0*, *A3B0*, and *miniX* constructs to explore their ability to complement the *roX* deficiency in KO-B cells. Finally, we tested the hypothesis that the long “*roX1*” RNA contains a *roX2*-like module at its 3′ end. The 3′ *roX1-D3* region resembles *roX2* structurally (stem–loop structures occluding “*roX* boxes” [17]). Since *roX1* is functionally redundant with *roX2* for male viability, it is possible that the 3′ portion of *roX1* represented in *roX1-D3* can substitute for *roX2* in our system.

We generated stable cell populations expressing *roX2* sequences corresponding to *A0B0*, *A3B0*, *miniX*, and *roX1-D3*. RT-qPCR analysis confirmed expression of all transgenes at 3- to 5-fold elevated levels compared with endogenous *roX2* in WT cells (Fig. 2B). Immunoblot analyses revealed that expression of full-length (FL) *roX2* partially restored MSL2 levels, whereas none of the test constructs improved MSL2 levels (Fig. 2C). We assessed DCC localization by widefield microscopy, staining for MSL2, MSL3, and H4K16ac (Fig. 2D). *roX2FL* restored the characteristic X-chromosome territory enrichment pattern observed in WT cells. In contrast, expression of the truncated *roX* variants resulted in diffuse staining patterns lacking clear territory enrichment.

To assess the rescue capacity in an unbiased manner, we employed image-based profiling of single cells using MSL2 immunofluorescence intensity distributions and textural properties, followed by UMAP dimensionality reduction (Fig. 2E; Supplementary Fig. 1A). Unsupervised clustering identified two broadly distinct populations of cells. The WT cells (dark green) occupied a discrete region at the lower left of the UMAP space, whereas KO-A and KO-B cells were assigned to clearly separable areas within a larger cluster. Consistent with their restored territory staining, cells expressing *roX2-FL* predominantly co-localized with the WT cluster. In contrast, cells expressing *A0B0*, *A3B0*, *miniX*, or *roX1-D3* failed to co-

localize with the WT population. Minor fractions of cells expressing *roX* derivatives *A0B0* and *A3B0* were mapped to the “territory” area of the UMAP, predominantly occupied by WT and FL *roX2*.

MiniX and *roX1-D3* remained clearly distinct. Interestingly, *A0B0* and *A3B0* occupied partially overlapping but distinguishable patterns in the UMAP, indicating that the texture-based analysis could potentially distinguish subtle differences in DCC localization patterns. Independent repetitions of this experiment with distinct biological replicates showed similar results (Supplementary Fig. S1A). CUT&RUN in the stable transgenic cell lines confirmed the widefield microscopy results, i.e. constructs that partially rescue DCC targeting (*roX2-FL* and *A0B0*) show some degree of X-chromosome enrichment, whereas non-rescuing constructs (*A3B0*, *miniX*, and *roX1-D3*) do not support X-chromosome targeting (Supplementary Fig. S1B).

Genome-wide mapping of dosage compensation complex occupancy by CUT&RUN

The confocal images showed that the DCC was delocalized from the X chromosome in the absence of *roX*. To characterize DCC binding at a higher resolution genome-wide, we performed CUT&RUN for MSL2 and H4K16ac in WT and *roX2* KO cell lines (Supplementary Fig. S2A). We had shown earlier that CUT&RUN allows identification of a larger number of X-chromosomal binding sites compared with ChIP [25]. For a direct comparison, we mapped MSL2 by CUT&RUN and ChIP in WT, KO, and rescue lines. Representative genome browser views of sites of MSL2 and H4K16ac localization on the X chromosome are shown in Figure 3A. The comparison with matched ChIP data reveals that CUT&RUN indeed identifies more MSL2 binding sites. The ChIP peaks are much broader, presumably because of variable cross-linking of MSL2 to neighboring chromatin [12]. For the H4K16 acetylation, the correspondence between the two methods is excellent. This is also evident when we compare the signal of MSL2 and H4K16ac over gene bodies at MSL2-bound genes (Supplementary Fig. S2B, C).

To capture the unique characteristics of CUT&RUN more accurately than traditional peak-calling algorithms, we applied a window-based differential binding analysis (csaw) [73]. Our approach counts reads in sliding windows across the genome and identifies regions with statistically significant enrichment over background. Using stringent thresholds and optimized parameters for each target (see the Materials and methods), we identified statistically enriched regions and merged proximal windows to define MCCs and HCCs.

This analysis identified 1882 MCCs, with ~90% mapping to the X chromosome (Fig. 3B). As expected, MCCs encompassed many previously ChIP-derived HASs (309 sites) and PionX (56 sites) signature sites (Fig. 3A). However, UpSet plot analysis revealed that 1623 MSL2-binding sites, 86% of all MCCs, had not been described previously (Fig. 3C), indicating that ChIP predominantly detects only the most highly occupied sites. Consistent with this, the HAS and PionX signature sites within MCCs displayed significantly higher MSL2 and H4K16ac signals compared with novel MCCs (Wilcoxon rank-sum test, $P < 0.001$; Fig. 3D), confirming a hierarchical binding landscape where previously known sites represent the most robustly occupied DCC targets (Fig. 3D).

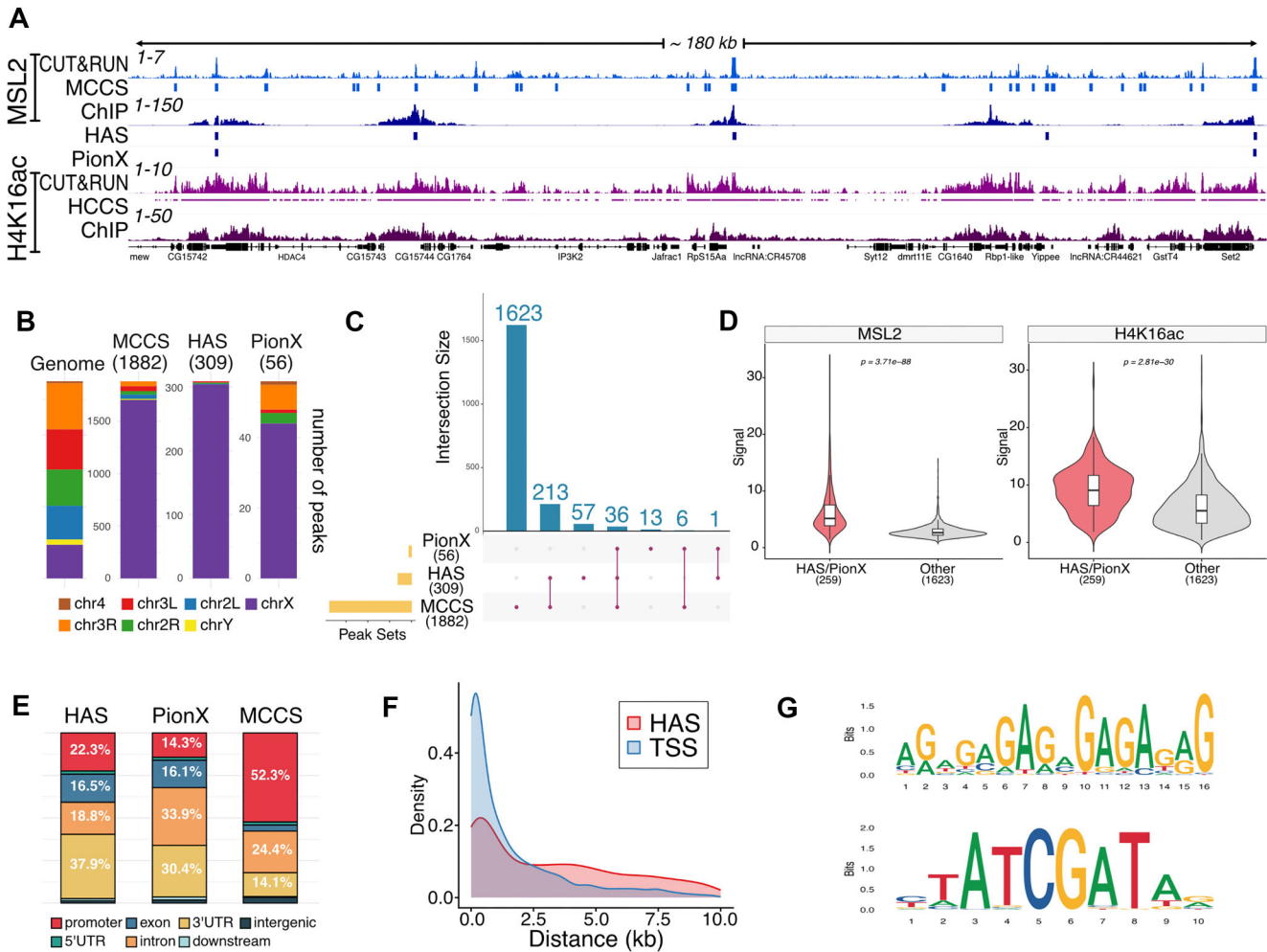


Figure 3. Genome-wide mapping of DCC occupancy by CUT&RUN and csaw. **(A)** Representative genome browser views of merged MSL2 and H4K16ac CUT&RUN profiles ($n = 3$, Supplementary Fig. S2A) across an ~ 180 kb region of the X chromosome, along with matched ChIP-MNase sequencing tracks ($n = 3$). Annotations of canonical DCC binding sites identified by ChIP (HAS and PionX) and newly identified CUT&RUN peak sets computed by csaw are indicated. Track scales are shown; gene names are displayed below. **(B)** Chromosomal distribution of ChIP-derived (HAS, PionX) and novel (MCCA) DCC-binding sites. Bars indicate the number of sites assigned to each chromosomal arm and the distribution of peaks among the chromosomes in a color code. The number of peaks in each category is provided in parentheses. **(C)** UpSet plot showing the overlap between MCCA and ChIP-derived HAS and PionX signature MSL2 binding sites. **(D)** CUT&RUN signal intensities at MCCA sites that overlap (HAS/PionX, red) or not (Other, gray) with HAS/PionX signature sites. Aggregate MSL2 and H4K16ac coverages at classified MCCA sites are shown. The numbers of sites in each group are indicated in parentheses. P -values: Wilcoxon rank-sum tests. **(E)** Genomic feature annotation of MSL2-binding sites. The bar graph shows the fraction of HASs, PionX sites, and MCCA sites assigned to genomic features indicated by the color code. Approximately half of MCCA sites map to promoter regions (red). **(F)** Proximity of MCCA sites to the TSSs and HASs. Density distributions showing distances from the MCCA to the nearest TSS and to the nearest HAS, as indicated. **(G)** Sequence logos of motifs enriched at MCCA-associated regions. The top scoring logo corresponds to the extended MRE. The bottom motif resembles DREF- (TATCGATA) or BEAF-32-binding sites (ATCGAT). Letter height denotes information content.

Genomic feature annotation revealed that approximately half of MCCA sites mapped to promoter regions (within 1 kb upstream of the TSS). In contrast, most of the ChIP-derived HASs and PionX sites ($\sim 73\%$ and 80% , respectively) are localized to gene bodies (Fig. 3E). Most of the newly identified MSL2 binding sites map to promoter regions, as visualized by plotting the distances of MCCA sites to TSSs as compared with HASs (Fig. 3F). Motif discovery analysis at MCCA peaks found the extended MRE as the top-scoring motif that had been identified in previous studies [25, 56, 74, 75, 76] (Fig. 3G). Interestingly, a palindromic motif with the consensus sequence YTATCGATAR (E-value = 4.8×10^{-4}) scored as second best. A search in the Tomtom database revealed highly significant matches to BEAF-32 (Boundary Element-Associated Factor;

P -value = 3.47×10^{-7} , E-value = 4.93×10^{-4}) and DREF (DNA replication-related element factor; P -value = 2.25×10^{-5}). Both proteins recognize the same core consensus sequence (TATCGATA). BEAF-32 binding sites have previously been shown to be enriched at MSL-bound genes [77]. We examined the prevalence of the MRE and the BEAF-32 binding motif across successive subsets of MCCA sites (Supplementary Fig. S3). Analysis of motif content across MCCA subsets revealed that BEAF-32-like motif prevalence was specifically elevated at chrX promoter-proximal MCCA sites (16.8%) compared with chrX MCCA sites overall (9.6%) or all MCCA sites genome-wide (8.7%), while MRE prevalence remained stable across these subsets ($\sim 30\%$; Supplementary Fig. S3). The majority of chrX promoter-proximal MCCA sites overlapped housekeep-

ing gene promoters (60.7%), and motif occupancy was comparable across all active gene categories (~22% MRE, ~21% BEAF-32). In contrast, sites with no overlap with active gene promoters showed the highest MRE prevalence (35.7%) but minimal BEAF-32 occupancy (5.0%), pointing to a distinct sequence composition at these sites.

In summary, CUT&RUN combined with csaw analysis identified 1882 MCCs, of which 1623 (86%) had not been previously characterized by ChIP. These newly identified sites are predominantly promoter-proximal, in contrast to the intragenic localization of HAs and PionX sites. CUT&RUN generated sharp, focal MSL2 peaks compared with the broad peaks observed by ChIP. The native chromatin conditions of CUT&RUN, which omits cross-linking, indicates that many of these sites correspond to direct MSL2–DNA contacts, rather than indirect associations stabilized by cross-linking to neighboring bound factors.

roX2 is essential for chromosome binding of the dosage compensation complex

Previous organismic studies suggested that in the absence of *roX* or the helicase MLE) the binding of MSL2 to the X chromosome is much reduced, but that a core complex of MSL2–MSL1 can bind to a small number of loci on the X chromosome [13, 24, 78]. Furthermore, recombinant MSL2 is able to bind MRE motifs including PionX signature sites *in vitro* in the absence of RNA [56, 76]. These observations suggested that *roX* RNA somehow modulates the intrinsic interactions of MSL2 with MREs *in vivo*. The availability of cell lines lacking *roX* RNA provides an opportunity to explore the *roX* dependence of MSL2 binding.

We displayed the CUT&RUN signals for MSL2 and H4K16ac at 56 PionX sites, 309 HAs, and all 1882 MCCs as separate heatmaps, ordered according to their intensities in WT cells (Fig. 4A). The cumulative signals are displayed above the columns. In both *roX2* KO clones, MSL2 and H4K16ac signals were dramatically reduced at all site classes, and this reduction was evident across the entire X chromosome in genome browser views (Fig. 4B). The compromised MSL2 binding and H4K16ac signals in KO-B cells were restored to WT levels upon expression of *roX2*-FL. Analysis of signal distributions across MSL2 target genes revealed the promoter-proximal enrichment of MCCs signals and the characteristic 3' enrichment pattern of H4K16ac accumulation in WT cells, which was lost in KO clones and restored upon rescue (Fig. 4C).

To further validate the global redistribution of the DCC upon *roX2* loss, we computed log₂FCs in MSL2 and H4K16ac signals relative to the WT, separately for autosomal and X chromosome sites (Fig. 4D). All chromosomal interactions of MSL2, whether on autosomes or the X chromosome, were diminished in both KO clones. H4K16ac was selectively lost from the X chromosome, as expected from loss of dosage compensation, but autosomes did not gain corresponding signals. As before, *roX2* expression in the KO-B line fully restored X chromosome enrichment, confirming that *roX2* is essential for proper genomic targeting of the DCC.

In summary, the analysis revealed a surprising requirement of *roX* RNA for all MSL2 chromosomal interactions in cells, independent of their sequence, shape, affinity, or even chromosomal location. Evidently, the intrinsic DNA sequence

readout of MSL2 does not suffice for stable chromosome interactions but requires additional input from non-coding RNA.

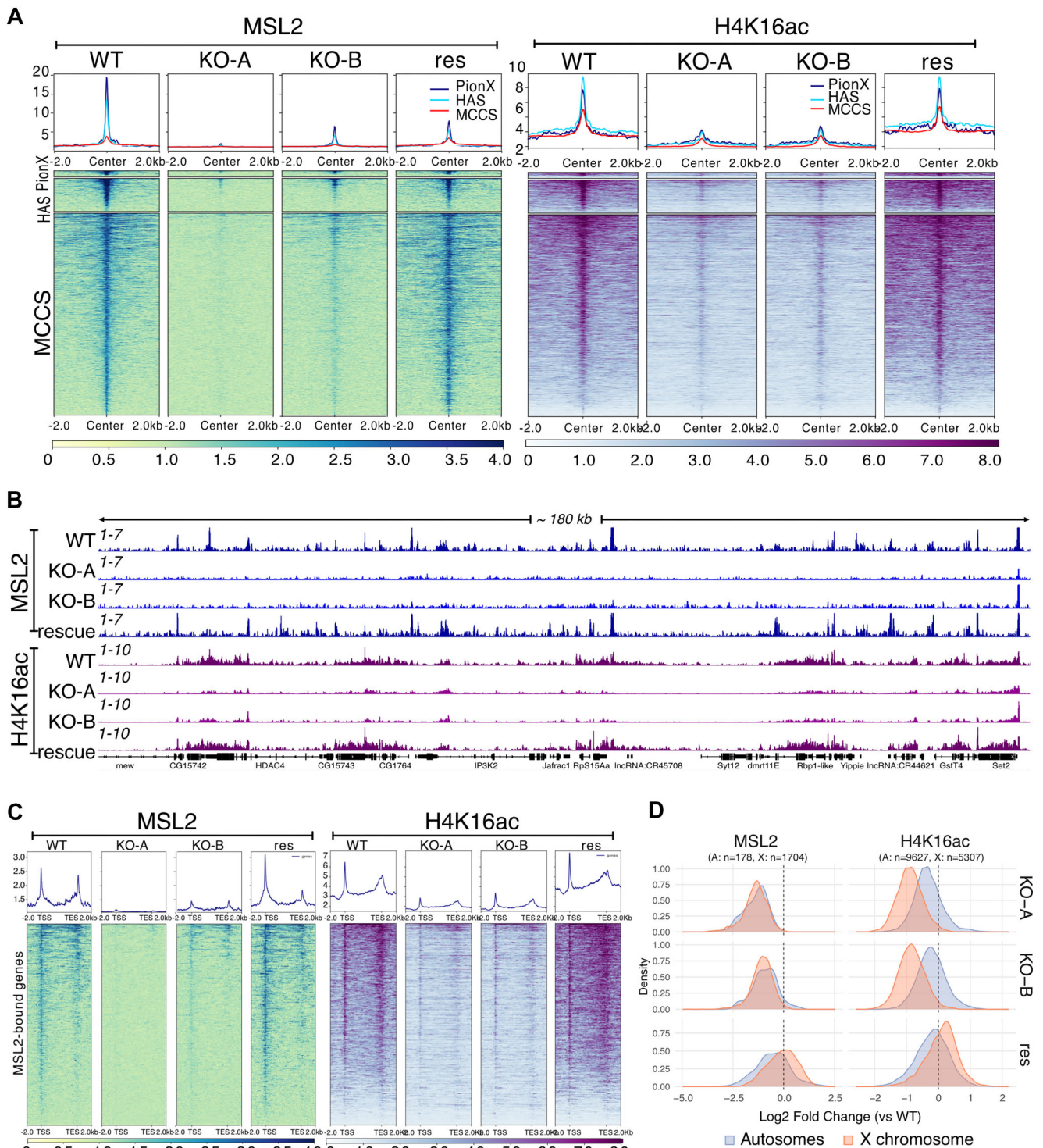
Dosage compensation is maintained in *roX2* knockout cells

In view of the global loss of DCC binding to the X chromosome in *roX2* KO cells, we wished to confirm that dosage compensation was impaired at the level of gene expression and identify *roX*-dependent gene expression changes. Towards this end, we established the transcriptomes of WT, KO, and rescue cells through RNA-seq analyses. Principal component analysis (PCA) of all independent biological replicates confirmed the clear separation of the datasets according to genotype, and the absence of *roX2* in the KO clones (Supplementary Fig. S4A).

Despite the loss of *roX2*, all samples correlate highly, with near perfect correlation for replicates of the same genotype (Supplementary Fig. S4B). Regardless, when comparing the KO and rescue conditions with the WT, most statistically significant gene expression changes differ according to cell line and not perturbation of the DCC *per se* (Supplementary Fig. S4C). To our surprise, we found only minor global changes of transcription relative to the WT, with log₂FC distributions for autosomal and X-linked genes centered around zero in KO-A, KO-B, and rescue cells (Fig. 5A). Apparently, dosage compensation was largely maintained despite loss of a functional DCC from the X chromosome.

To understand this unexpected finding, we performed a small-scale karyotype analysis of ~20 individual cells from each condition (see Supplementary Table S2 for a comprehensive summary of karyotypes). S2 cells are known to be male and mostly tetraploid, with a chromosome complement of four of each autosome and two X chromosomes (typical WT = 2 × X, 4 × second, 4 × third, 2 × fourth, and 1 × 2L = 13 chromosomes). We found the expected karyotype in all WT cells. Strikingly, about half of KO-A nuclei and about a quarter of KO-B nuclei contained a third X chromosome, which is clearly identified because of its acrocentric nature (representative images in Fig. 5B). We also found many cases in which an autosome appeared to be missing, but these presumed aneuploidies were obscured by the presence of broken chromosomes in our karyotypes (Supplementary Table S2). The variability of karyotype between individual cells across all lines was best illustrated by the variable number of the fourth chromosome, ranging from two to seven copies. Remarkably, the WT karyotype with exactly two X chromosomes was recovered in all rescue cells (Fig. 5B). Despite the low sample numbers and some uncertainty about the identification of all autosomes, the results of this qualitative analysis provided an important clue to explain dosage compensation in the absence of DCC interactions with the X chromosome.

To substantiate these findings quantitatively, we determined the chromosome copy numbers at the population level by counting DNA reads in input samples of our ChIP analysis. Consistent with the single-cell karyotype data, both *roX2* KO clones showed an average X-chromosome copy number of ~2.7 when scaled to the WT expected autosomal baseline, while the rescue line value was similar to that of the WT (Fig. 5C). The inferred increase in copy number of the X chromosome in the two independent KO clones and across biological replicates is chromosome wide rather than segmen-



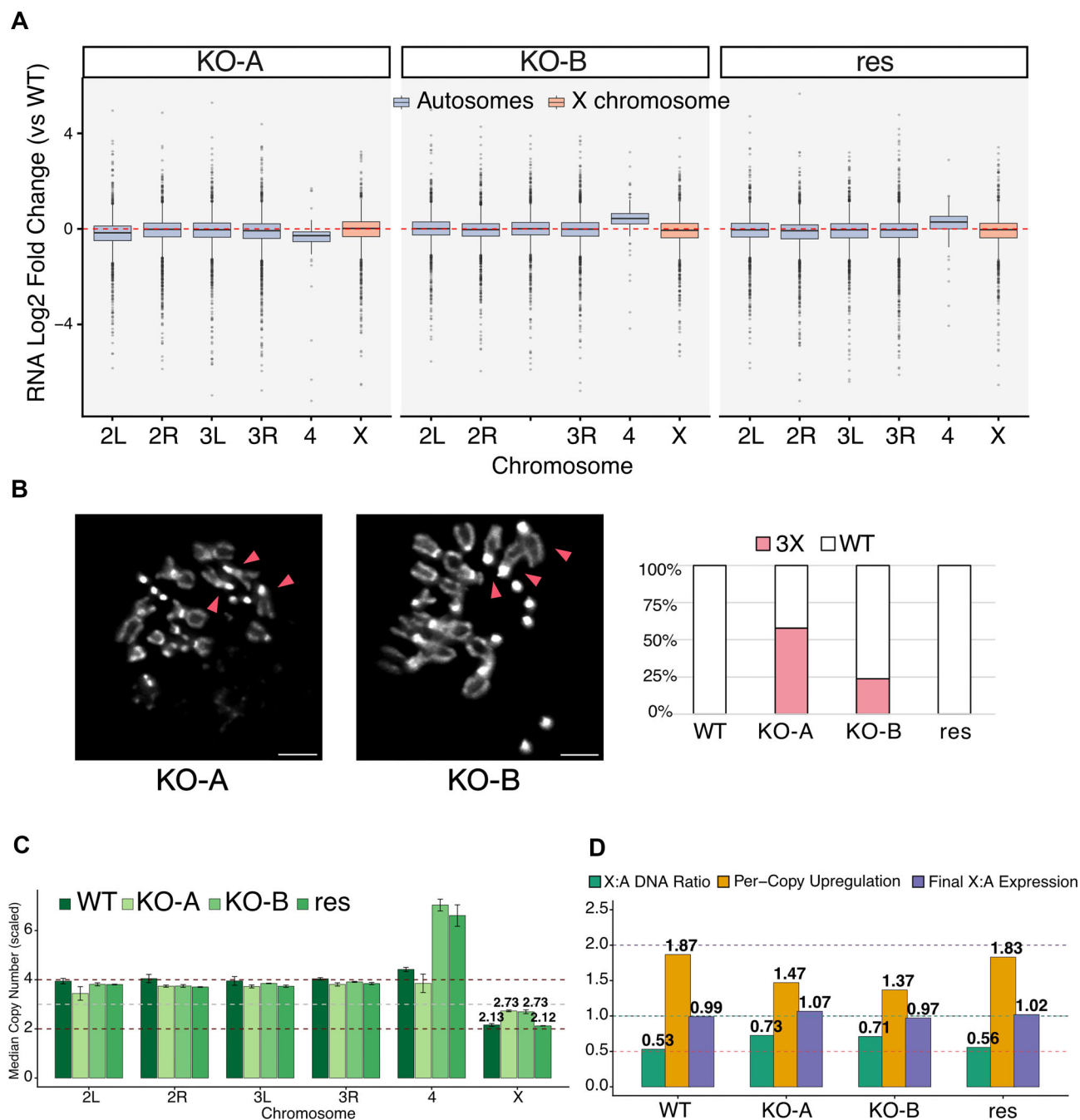


Figure 5. Dosage compensation is maintained in the *roX2* KO cells. **(A)** RNA-seq analysis of gene expression changes for each chromosome. Boxplots show per-gene log₂FCs in RNA levels transcribed from protein-coding genes in KO-A, KO-B, and rescue cells relative to the WT, grouped by chromosomal arm (2L, 2R, 3L, 3R, 4, and X). Distributions for all chromosomes except the fourth are centered near zero, indicating only very minor global expression changes. Boxes represent the interquartile range (IQR) with median lines; whiskers extend to 1.5× the IQR. The dashed red line marks no change (log₂FC = 0). **(B)** Heterogeneity of karyotypes in KO cells. Left, exemplary images illustrating the existence of three X chromosomes (marked by pink arrows) in KO-A and KO-B. Left, barplot showing the percentage of cells having either three X chromosomes or the WT karyotype (see also [Supplementary Table S2](#)). **(C)** Chromosome copy number estimates derived from ChIP-MNase input sequencing. Median inferred copy numbers for the indicated cell lines are scaled to an autosomal baseline defined by the tetraploidy of WT cells (excluding the fourth chromosome). Dashed lines denote the theoretical values. Both KO lines show elevated average X chromosome copy number (values above the bars), whereas the rescue line restores X chromosome copy number toward ~2. Note the amplification of the fourth chromosome in KO-B and its descendant rescue line. Error bars indicate the SD across three independent biological replicates. **(D)** Dosage compensation by adjustment of X/autosome ratios. Bar plots show X/autosome ratios: X:A DNA ratios from median inferred copy numbers as in (C) (green), copy number-corrected X:A expression ratios (per-copy up-regulation, yellow) computed by dividing per gene expression by the inferred copy number and raw X:A expression ratios (purple) computed as mean per gene expression of protein-coding genes on the X divided by the corresponding mean for autosomal protein-coding genes. The dashed lines indicate the theoretical ratios expected for full dosage compensation (X:A DNA 0.5, X:A copy number-corrected 2, X:A expression 1). The actual values are indicated above the bars.

tal (Supplementary Fig. S5). This observation, in combination with karyotyping that directly identified cells with an additional X chromosome (Fig. 5B; Supplementary Table S2), supports the conclusion that the elevated X/autosome ratios reflect a robust and genuine gain of X chromosome rather than regional MNase sampling bias. This suggests that KO clones maintained balanced expression by altering the X/autosome copy number ratios upon impairment of DCC-mediated regulation of transcription (Fig. 5C). When expression ratios were corrected for DNA copy number (per-copy up-regulation), the KO clones showed reduced per-copy expression from the X chromosome, consistent with impaired DCC function. However, the overall X:A (X:autosome) expression ratio remained near 1.0 (Fig. 5D), due to a combination of increases in X chromosome sequences and the generic feedback mechanism described by Zhang and Oliver [28]. The sequencing also confirmed the amplification of chromosome 4 in KO-B and its derived rescue line. Evidently, chromosome segregation in S2 cells is error prone to an extent that allows reversible adjustment of chromosome content under selective pressure.

These findings demonstrate that lack of dosage compensation leads to a severely reduced fitness of aneuploid S2 cells in culture. The recovery of cells with additional X chromosomes in both KO clones, and the subsequent loss of the extra X upon *roX* restoration, strongly suggest that balanced genome expression provides a selective advantage during cell line establishment.

The error-prone chromosome segregation inherent to S2 cells, evidenced also by variable chromosome 4 copy numbers across all genotypes, apparently allows adjustment of chromosome content when cells experience prolonged selection pressure from impaired dosage compensation.

Together, these results establish *roX2* as essential for DCC function in cultured cells and demonstrate that the strong selective pressure to maintain dosage balance can drive rapid, reversible karyotypic changes, providing a sensitive experimental system for quantitative evaluation of *roX* RNA variants in chromosome targeting.

Discussion

Non-coding *roX* RNA is required for all chromosomal interactions of the dosage compensation complex

Our CUT&RUN/csw analysis identified >1800 MCCSs, vastly expanding the catalog of DCC occupancy compared with previous studies that identified only a few hundred of the strongest binding sites [25, 29, 56]. This improved detection is due to refinement of mapping by CUT&RUN and the implementation of csw. CUT&RUN yields better signal-to-noise ratios and reduced background at active promoter regions, and avoids fixation artifacts inherent to ChIP [79], with some degree of accessibility bias remaining [35, 80, 81]. Compared with conventional peak callers, csw's window-based approach is more flexible and sensitive, allowing it to detect a broader range of enrichment events while delivering robust, statistical inference suitable for complex and replicated experimental designs [52, 73].

The expanded MCCS catalog includes most of the previously characterized HASs and PionX sites, but in addition numerous MREs of presumed lower affinity. The enrichment of MCCSs on the X chromosome demonstrates the robustness and specificity of our methodology in delineating the

DCC binding landscape. Approximately half of these sites map within 1 kb of TSSs, while a significant fraction localize to introns and 3'-untranslated regions (UTRs) of genes. This resonates with earlier lower-resolution ChIP studies that reported some of the HASs mapping to promoter-proximal regions, but a more substantial fraction at introns and 3'UTRs [29, 56]. The gene body binding on the X, which is mirrored by the H4K16ac distribution, is in line with a proposed mechanism of dosage compensation through enhanced processivity of transcription elongation [1]. The observation that most MCCSs are located within genes implies that the presumed "spreading" of DCC activity from primary DNA elements to active chromatin marks (H3K36me2/3) does not require bridging vast distances, but can occur rather locally. Tian *et al.* proposed that *roX* dynamically regulates DCC distribution by forming TES-TSS loops that facilitate local RNA polymerase II (Pol II) access rather than long-range spreading [82]. These observations are compatible with facilitated reinitiation of the RNA Pol II machinery that may contribute to the transcription boost.

The binding of the DCC close to TSSs may reflect interactions of MSL2 with CLAMP, a ubiquitous transcription factor that directly interacts and cooperates with MSL2 to bind to MREs *in vivo* [36, 83, 84] and *in vitro* [25, 76]. Interestingly, sequence motif searches among the expanded set of MCCSs not only identified the known MRE motif, but also found the motif (YTATCGATAR) significantly enriched at MCCSs, which matches consensus binding sites for BEAF-32 and DREF. BEAF-32 [77, 85, 86] and DREF [87–90, 91] are known to bind many promoters and insulators, and act as boundary elements. BEAF-32 has previously been shown to be enriched close to TSSs of MSL-bound genes [77]. Consistent with this, BEAF-32-like motif prevalence was specifically elevated at promoter-proximal X-chromosomal MCCSs, while MRE prevalence remained stable, and MCCSs lacking active gene promoter overlap showed high MRE but minimal BEAF-32 occupancy, suggesting that these two motifs mark compositionally distinct binding site subsets. Whether BEAF-32 and/or DREF contribute directly to DCC recruitment or are confounding factors associated with accessible chromatin remains to be determined.

Implications for X-chromosome targeting by the dosage compensation complex

Our findings challenge "two-step" models for DCC targeting on the X chromosome, which propose that initial binding occurs independently of *roX* RNA, followed by *roX*-dependent spreading over considerable distances [1]. Key to these models is that binding sites differ qualitatively in that they are either *roX* dependent or *roX* independent. In contrast, the "affinity hierarchy" model argues for a continuum of DCC binding sites with varying affinities. Spreading may then involve the local enrichment of DCC around sites of higher affinity, followed by the dynamic association with MREs or H3K36-methylated nucleosomes along the concentration gradient [12, 23, 24]. It has been suggested that the dynamic association of *roX* RNA, mediated by the helicase MLE, may define cycles of chromatin release and "reloading", which would facilitate the "spreading" of the DCC to all potential binding sites in the neighborhood [18, 92].

Our genome-wide mapping in *roX2* KO S2 cells demonstrates that loss of *roX* RNA results in a dramatic reduction

of MSL2 and H4K16ac occupancy not only at lower-affinity MREs, but also at HASs and PionX sites, which are known to be selectively bound by recombinant MSL2 in the absence of *roX* *in vitro* [56]. Although H4K16ac enrichment from the X chromosome is lost, the global levels of the mark remain largely unchanged because MOF maintains global H4K16ac levels on autosomes independently of *roX* RNA in the context of the non-specific lethal (NSL) complex [90, 93].

In the absence of *roX*, MSL2 protein levels are moderately reduced, which may contribute to the observed global loss of binding. Under such limiting MSL2 conditions, one might predict preferential occupancy of the highest-affinity sites. Instead, we find that MSL2 binding is abrogated at all sites, including those previously classified as high affinity. Evidently, *roX* RNA is required for MSL2 association at all genomic locations, independent of type and affinity, consistent with direct MSL2–*roX*2 interactions that promote stable chromosome association [10, 20, 22]. While we cannot exclude that the intrinsic affinity of MSL2 for ideal MRE sequences leads to *roX*-independent binding under certain circumstances, our data support a model in which the DCC, through MSL2–DNA recognition and cooperation with accessory factors such as CLAMP, engages its binding sites with a range of intrinsic affinities, and *roX* RNA stabilizes these interactions on the X chromosome. We observed earlier through fluorescence recovery after photobleaching (FRAP) studies that *roX*2 endows the X chromosome-bound DCC with “low-diffusion” properties that effectively trap the DCC on the X [10, 21]. In combination with the known homeostasis mechanisms that ensure that the numbers of DCCs do not exceed the binding capacity of sites on the X chromosome [1, 15, 69, 94], such “trapping” could be the basis of X chromosome-selective binding.

Taken together, our findings support models that posit that all chromosomal binding sites for the DCC require *roX* for robust interaction, independent of their DNA sequence signature or composition.

A strategy for systematic structure–function analysis of *roX* RNA requirements for dosage compensation complex loading

CRISPR-engineered *roX*2 KO S2 cells can form the basis of an experimental strategy to dissect the structural and functional elements that allow *roX* RNA to interact with MSL proteins and to target the complex to the X-chromosomal territory. S2 cells naturally depend on *roX*2 for dosage compensation, allowing us to avoid the tedious generation of *roX*1 *roX*2 double mutants that are not straightforward in the organism [14]. Similar structure–function analyses in *Drosophila* commonly score “male viability” as a readout for DCC functionality [11, 17, 18] and are complicated by developmental dynamics and potential cell-specific effects [95]. We sought to test the cellular system by analyzing *roX*2 derivatives which had previously been shown to support some male viability in corresponding rescue experiments in flies.

In our rescue experiments, the robust DCC localization to the X chromosome, assessed by immunostaining and single-cell imaging analyses, was only observed with full-length *roX*2 RNA. Despite overexpression, several truncated *roX*2 variants as well as *roX*1-D3, constructs previously reported to support organismal viability when overexpressed [17, 18], failed to establish faithful DCC targeting in our cellular sys-

tem. This discrepancy and the karyotypic adaptation that occurred during the selection of *roX*2-deleted cells highlight a tight dosage sensitivity of cultured cells that could not have been anticipated. It is possible that in flies even cryptic DCC function may be compatible with some viability due to compensatory buffering mechanisms, whereas the systematic image-based profiling employed here detects more nuanced deficiencies in DCC targeting. On the other hand, our rescue strategy involved a population of cells that derived from a clonal event. The apparent discrepancy between our data and the rescue of flies can only be resolved by more systematic studies.

Nevertheless, the texture-based analysis of MSL protein staining in individual cells revealed subtle differences in nuclear staining for the different *roX* constructs, suggesting that the approach may be useful to classify intermediate phenotypes. The inability of certain *roX* derivatives to support DCC targeting may reflect sequestration of MSL proteins to repetitive, heterochromatic regions [71, 72] or other non-functional nuclear foci, patterns that might distinctly contribute to the classification texture-based image analysis.

Dosage compensation is essential even in aneuploid cultured cells

Dosage compensation in *Drosophila* occurs throughout the organism, though its extent varies across tissues and cell types [95]. Male S2 cells, an immortalized hematopoietic line, have served as a valuable model for studying dosage compensation mechanisms [28, 29]. These cells are predominantly tetraploid but exhibit substantial aneuploidy [28, 44, 96]. While aneuploidy is typically detrimental at the organismal level due to cumulative gene dosage effects and regulatory network collapse [97, 98], cultured cells display greater tolerance to chromosomal imbalances [28]. This resilience initially led us to hypothesize that S2 cells might tolerate perturbation of the DCC. Indeed, transient RNAi-mediated depletion of MSL proteins does not impair S2 cell proliferation over experimental time scales of 6–10 days [27, 28, 29], raising the question of whether dosage compensation is truly essential in this context.

We therefore assumed that generating *roX*-deficient S2 cells would not pose any problem. However, generating a stable *roX*2 gene deletion line, the first of its kind, turned out to be very challenging, in stark contrast to the ease of generating equivalent KOs in female Kc cells (unpublished observations). The fact that the two *roX*-deficient lines we finally obtained during the lengthy selection process had increased X chromosome copy numbers demonstrates the critical requirement for maintaining a balanced X/autosome expression ratio even in aneuploid cultured cells. This is remarkable, given that the DCC compensates the 2-fold reduced X chromosome copy number by only a 1.4-fold boost in transcription [28]. This dosage compensation is completely defective in the absence of *roX*, since MSL2 levels are diminished and DCC binding across all sites is dramatically lost.

Mechanisms of karyotypic adaptation in *roX*-deficient cells

The gain of an additional X chromosome in *roX*2 KO cells represents a remarkable example of rapid evolution under selection. S2 cells are known to exhibit high rates of chromosome segregation defects during mitosis and variable chromo-

some copy numbers, particularly of chromosome 4 [28, 44, 96, 99]. Mitotic mistakes that lead to abnormal karyotypes may be as frequent as 18% [98]. Such defects arise from multiple mechanisms, including weakened spindle checkpoints and aberrant chromosome segregation [100]. Like many tumor cells, S2 cells overexpress the protooncogene *ras* [96], which is known to trigger chromosome segregation defects and have multiple centromeres [101, 102]. Individual cells from Drosophila Genomics Resource Center (DGRC) lines, including S2 cells, display substantial karyotypic heterogeneity even within clonal populations [96]. Therefore, S2 cells are constantly confronted with changes in chromosome number through segregation defects. Aneuploidy leads to reduced fitness and impairs proliferation [103]. Under normal growth conditions, cells that deviate from the optimal tetraploid-like genotype lack fitness and are outgrown quickly. We propose that the perturbation of the balanced genome expression by abrogating dosage compensation changes the selection paradigm. During the extended selection process following CRISPR/Cas9 editing (~6 weeks), cells that stochastically acquired a third X chromosome or lost autosomal DNA would have gained a selective growth advantage by restoring X/autosome expression balance without requiring a functional DCC. This hypothesis is strongly supported by our single-cell karyotyping and DNA copy number analyses, which showed an average of ~2.7 X chromosomes in both independent KO populations, suggesting that this copy number represents an adaptive equilibrium that partially compensates for impaired DCC function while avoiding excessive aneuploidy-associated fitness costs. Heterogeneous losses of autosomal DNA probably also contribute to rebalancing the X/autosome expression ratio [28]. We assume that, collectively, these karyotypic changes compensate for the loss of DCC-dependent dosage compensation.

The remaining compensation is likely to be achieved through a combination of gene-specific feedback mechanisms that operate whenever large parts of chromosomes are deleted [28, 104, 105] and post-transcriptional buffering mechanisms, including regulation of mRNA stability, translation, and proteostasis [69, 106, 107]. The gains of X chromosome in combination with such buffering mechanisms probably enabled our KO cells to stay competitive during selection despite complete loss of DCC function.

An important observation made through karyotyping of individual cells is that the established *roX*-depleted cells are no longer clonal but represent a mixed population of cells with either three X chromosomes or autosomal aneuploidies that arose from stochastic mitotic mis-segregation events in the wake of *roX2* gene deletion. This karyotypic heterogeneity also explains the reversibility of the phenotype upon re-expression of *roX2* in our rescue experiment, when the levels of MSL2 were partly restored and the binding of DCC to the X chromosomal territory was re-established. Under these circumstances, an extra X chromosome is disadvantageous due to overcompensation and is selected against during the establishment of the rescue lines, which again takes several weeks. This has been demonstrated in CHO cell lines, where cells under stringent phenotypic selection (e.g. expression of a transgene) are part of a more homogeneous population with fewer chromosome variations, while subcloning does not increase said homogeneity or chromosomal stability [108].

The relevance of our findings for male flies is unclear at this point. However, *roX*-deficient and *msl*-deficient male flies like-

wise fail to up-regulate X chromosomal transcription leading to male lethality, yet 5% of escapees are observed for *roX1 roX2* mutants [13, 14], whereas escapees are not found for *msl* mutants. This shows that, in principle, a *roX* deficiency can be overcome *in vivo*. The overexpression of MSL proteins is one documented mechanism [15], which clearly is not employed in our system.

In summary, our findings establish *roX*-deficient S2 cells as a versatile and tractable platform for future structure–function studies of lncRNA-guided chromatin regulation in the context of dosage compensation. Reversible switching of dosage compensation provides a unique experimental opportunity for studying chromosomal evolution under controlled selective pressure. The unexpected karyotypic adaptation observed in these aneuploid cells upon *roX2* gene deletion underscores the existential importance of maintaining a balanced genome expression and the vital contribution of the DCC to this end. Our findings also reveal a surprising cellular plasticity in response to (epi)genetic perturbation and selection, which could be subject to further studies. Refinement of the experimental strategy by rapid and reversible switching of *roX2* expression may avoid adaptive effects that obscure the true functions of *roX* RNAs and will pave the way for high-resolution dissection of the elements and mechanisms by which *roX* RNA mediates DCC assembly, X-chromosome targeting, and epigenetic regulation.

Acknowledgements

The authors gratefully acknowledge A. Thomae from the BMC Bioimaging Core Facility for training and invaluable advice on microscopy techniques and analyses. We thank D. Hörl from the Human Biology Division (LMU Biology) for early pilot classifications and analysis of microscopy images, T. Straub from the BMC Bioinformatics Unit for access to the computational cluster and advice on bioinformatics and helpful scripts, and S. Krebs of LMU LAFUGA for Illumina sequencing. Also, we thank the IMB Microscopy and Histology CF for the image deconvolution in this work. Finally, we thank all current lab members for fruitful discussions. The graphical abstract was created in BioRender. Jayakrishnan, M. (2026) <https://BioRender.com/osclqya>.

Author contributions: FG: Conceptualization, Investigation, Data curation, Formal analysis, Validation, Writing—original draft, Writing—review & editing; GY: Investigation, Formal analysis (karyotype experiment), Writing—review & editing; MJ: Methodology (developed initial widefield microscopy analysis), Writing—review & editing; ACS: Investigation; MM: Writing—review & editing; PH: Supervision of GY, Formal analysis (karyotype experiment), Writing—review & editing; PPB: Conceptualization, Funding acquisition, Supervision, Validation, Writing—original draft, Writing—review & editing.

Supplementary data

Supplementary data are available at NAR online.

Conflict of interest

The authors declare no competing interest.

Funding

German Research Council (DFG) grant BE1140/11-1 [to P.B.B.]; the International Max-Planck Research School “Molecules of Life” (IMPRS-ML) and the Integrated Research Training Group of the DFG-funded Collaborative Research Center “Chromatin Dynamics” [to F.G.]. Funding to pay the Open Access publication charges for this article was provided by the Ludwig-Maximilians-University, Munich.

Data availability

Sequencing data generated in this study can be found in the Gene Expression Omnibus (GEO) database. Raw sequencing files in fastq format and processed coverage tracks in bigwig format for CUT&RUN are available under the accession number GSE320392. Similarly, ChIP data are available under accession number GSE320376. Transcriptomic data generated in this study were deposited under the accession number GSE320378. Reference tracks can be viewed interactively on the UCSC genome browser at https://genome-euro.ucsc.edu/s/FotiosGk/dm6_gkountromichos_2026. Custom analysis code is available at Github (https://github.com/FotiosGk/Gkountromichos_et_al_2026/) and Zenodo (doi:10.5281/zenodo.20119753).

References

- Lucchesi JC, Kuroda MI. Dosage compensation in *Drosophila*. *Cold Spring Harb Perspect Biol* 2015;7:a019398. <https://doi.org/10.1101/cshperspect.a019398>
- Makki R, Meller VH. When down is up: heterochromatin, nuclear organization and X upregulation. *Cells* 2021;10:3416. <https://doi.org/10.3390/cells10123416>
- Kuroda MI, Hilfiker A, Lucchesi JC. Dosage compensation in *Drosophila*—a model for the coordinate regulation of transcription. *Genetics* 2016;204:435–50. <https://doi.org/10.1534/genetics.115.185108>
- Franke A, Baker BS. The roX1 and roX2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol Cell* 1999;4:117–22. [https://doi.org/10.1016/S1097-2765\(00\)80193-8](https://doi.org/10.1016/S1097-2765(00)80193-8)
- Kelley RL, Meller VH, Gordadze PR *et al.* Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* 1999;98:513–22. [https://doi.org/10.1016/S0092-8674\(00\)81979-0](https://doi.org/10.1016/S0092-8674(00)81979-0)
- Meller VH. Initiation of dosage compensation in *Drosophila* embryos depends on expression of the roX RNAs. *Mech Dev* 2003;120:759–67. [https://doi.org/10.1016/S0925-4773\(03\)00157-6](https://doi.org/10.1016/S0925-4773(03)00157-6)
- Deng X, Meller VH. roX RNAs are required for increased expression of X-linked genes in *Drosophila melanogaster* males. *Genetics* 2006;174:1859–66. <https://doi.org/10.1534/genetics.106.064568>
- Meller VH, Wu KH, Roman G *et al.* roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* 1997;88:445–57. [https://doi.org/10.1016/S0092-8674\(00\)81885-1](https://doi.org/10.1016/S0092-8674(00)81885-1)
- Meller VH, Joshi SS, Deshpande N. Modulation of chromatin by noncoding RNA. *Annu Rev Genet* 2015;49:673–95. <https://doi.org/10.1146/annurev-genet-112414-055205>
- Villa R, Jagtap PKA, Thomae AW *et al.* Divergent evolution toward sex chromosome-specific gene regulation in *Drosophila*. *Genes Dev* 2021;35:1055–70. <https://doi.org/10.1101/gad.348411.121>
- Quinn JJ, Zhang QC, Georgiev P *et al.* Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes Dev* 2016;30:191–207. <https://doi.org/10.1101/gad.272187.115>
- Prayitno K, Schauer T, Regnard C *et al.* Progressive dosage compensation during *Drosophila* embryogenesis is reflected by gene arrangement. *EMBO Rep* 2019;20:e48138. <https://doi.org/10.15252/embr.201948138>
- Meller VH, Rattner BP. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J* 2002;21:1084–91. <https://doi.org/10.1093/emboj/21.5.1084>
- Kim M, Fauchillon ML, Larsson J. RNA-on-X 1 and 2 in *Drosophila melanogaster* fulfill separate functions in dosage compensation. *PLoS Genet* 2018;14:e1007842. <https://doi.org/10.1371/journal.pgen.1007842>
- Oh H, Park Y, Kuroda MI. Local spreading of MSL complexes from roX genes on the *Drosophila* X chromosome. *Genes Dev* 2003;17:1334–9. <https://doi.org/10.1101/gad.1082003>
- Kuroda MI, Kernan MJ, Kreber R *et al.* The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila*. *Cell* 1991;66:935–47. [https://doi.org/10.1016/0092-8674\(91\)90439-6](https://doi.org/10.1016/0092-8674(91)90439-6)
- Ilik IA, Quinn JJ, Georgiev P *et al.* Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol Cell* 2013;51:156–73. <https://doi.org/10.1016/j.molcel.2013.07.001>
- Ilik IA, Maticzka D, Georgiev P *et al.* A mutually exclusive stem-loop arrangement in roX2 RNA is essential for X-chromosome regulation in *Drosophila*. *Genes Dev* 2017;31:1973–87. <https://doi.org/10.1101/gad.304600.117>
- Maenner S, Muller M, Fröhlich J *et al.* ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol Cell* 2013;51:174–84. <https://doi.org/10.1016/j.molcel.2013.06.011>
- Muller M, Schauer T, Krause S *et al.* Two-step mechanism for selective incorporation of lncRNA into a chromatin modifier. *Nucleic Acids Res* 2020;48:7483–501.
- Straub T, Neumann MF, Prestel M *et al.* Stable chromosomal association of MSL2 defines a dosage-compensated nuclear compartment. *Chromosoma* 2005;114:352–64. <https://doi.org/10.1007/s00412-005-0020-x>
- Valsecchi CIK, Basilicata MF, Georgiev P *et al.* RNA nucleation by MSL2 induces selective X chromosome compartmentalization. *Nature* 2021;589:137–42. <https://doi.org/10.1038/s41586-020-2935-z>
- Demakova OV, Kotlikova IV, Gordadze PR *et al.* The MSL complex levels are critical for its correct targeting to the chromosomes in *Drosophila melanogaster*. *Chromosoma* 2003;112:103–15. <https://doi.org/10.1007/s00412-003-0249-1>
- Dahlsveen IK, Gilfillan GD, Shelest VI *et al.* Targeting determinants of dosage compensation in *Drosophila*. *PLoS Genet* 2006;2:e5. <https://doi.org/10.1371/journal.pgen.0020005>
- Eggers N, Gkountromichos F, Krause S *et al.* Physical interaction between MSL2 and CLAMP assures direct cooperativity and prevents competition at composite binding sites. *Nucleic Acids Res* 2023;51:9039–54. <https://doi.org/10.1093/nar/gkad680>
- Baum B, Cherbas L. *Drosophila* cell lines as model systems and as an experimental tool. *Methods Mol Biol* 2008;420:391–424.
- Hamada FN, Park PJ, Gordadze PR *et al.* Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev* 2005;19:2289–94. <https://doi.org/10.1101/gad.1343705>
- Zhang Y, Malone JH, Powell SK *et al.* Expression in aneuploid *Drosophila* S2 cells. *PLoS Biol* 2010;8:e1000320. <https://doi.org/10.1371/journal.pbio.1000320>
- Straub T, Zabel A, Gilfillan GD *et al.* Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. *Genome Res* 2013;23:473–85. <https://doi.org/10.1101/gr.146407.112>

30. Scacchetti A, Schauer T, Reim A *et al.* Drosophila SWR1 and NuA4 complexes are defined by DOMINO isoforms. *eLife* 2020;9:e56325. <https://doi.org/10.7554/eLife.56325>
31. Benchling. [Biology Software]. (2021–2026). Retrieved from <https://benchling.com> (accessed May 20, 2026).
32. Doench JG, Fusi N, Sullender M *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat Biotechnol* 2016;34:184–91. <https://doi.org/10.1038/nbt.3437>
33. Gratz SJ, Ukken FP, Rubinstein CD *et al.* Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* 2014;196:961–71. <https://doi.org/10.1534/genetics.113.160713>
34. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria. <https://www.R-project.org/>. (accessed May 20, 2026)
35. Meers MP, Bryson TD, Henikoff JG *et al.* Improved CUT&RUN chromatin profiling tools. *eLife* 2019;8:e46314. <https://doi.org/10.7554/eLife.46314>
36. Albig C, Tikhonova E, Krause S *et al.* Factor cooperation for chromosome discrimination in Drosophila. *Nucleic Acids Res* 2019;47:1706–24. <https://doi.org/10.1093/nar/gky1238>
37. Jayakrishnan M, Havlova M, Veverka V *et al.* Genomic context-dependent histone H3K36 methylation by three Drosophila methyltransferases and implications for dedicated chromatin readers. *Nucleic Acids Res* 2025;53:gkaf202. <https://doi.org/10.1093/nar/gkaf202>
38. Liu N, Hargreaves VV, Zhu Q *et al.* Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. *Cell* 2018;173:430–42. <https://doi.org/10.1016/j.cell.2018.03.016>
39. Schindelin J, Arganda-Carreras I, Frise E *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012;9:676–82. <https://doi.org/10.1038/nmeth.2019>
40. Pau G, Fuchs F, Sklyar O *et al.* EBIImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* 2010;26:979–81. <https://doi.org/10.1093/bioinformatics/btq046>
41. Wickham H, Averick M, Bryan J *et al.* Welcome to the Tidyverse. *J Open Source Softw* 2019;4:1686. <https://doi.org/10.21105/joss.01686>
42. McInnes L, Healy J, Saul N *et al.* UMAP: Uniform Manifold Approximation and Projection. *J Open Source Softw* 2018;3:861. <https://doi.org/10.21105/joss.00861>
43. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;SMC-3:610–21. <https://doi.org/10.1109/TSMC.1973.4309314>
44. Olszak AM, van Essen D, Pereira AJ *et al.* Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nat Cell Biol* 2011;13:799–808. <https://doi.org/10.1038/ncb2272>
45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9. <https://doi.org/10.1038/nmeth.1923>
46. Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>
47. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2. <https://doi.org/10.1093/bioinformatics/btq033>
48. Ramirez F, Ryan DP, Gruning B *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;44:W160–5. <https://doi.org/10.1093/nar/gkw257>
49. Kent WJ, Zweig AS, Barber G *et al.* BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 2010;26:2204–7. <https://doi.org/10.1093/bioinformatics/btq351>
50. Zerbino DR, Johnson N, Juettemann T *et al.* WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics* 2014;30:1008–9. <https://doi.org/10.1093/bioinformatics/btt737>
51. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinform* 2013;14:178–92. <https://doi.org/10.1093/bib/bbs017>
52. Lun AT, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* 2016;44:e45. <https://doi.org/10.1093/nar/gkv1191>
53. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40. <https://doi.org/10.1093/bioinformatics/btp616>
54. Lawrence M, Huber W, Pages H *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;9:e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>
55. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382–3. <https://doi.org/10.1093/bioinformatics/btv145>
56. Villa R, Schauer T, Smialowski P *et al.* PionX sites mark the X chromosome for dosage compensation. *Nature* 2016;537:244–8. <https://doi.org/10.1038/nature19338>
57. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017;33:2938–40. <https://doi.org/10.1093/bioinformatics/btx364>
58. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 2009;25:1841–2. <https://doi.org/10.1093/bioinformatics/btp328>
59. R Core Team. 2025; R: A language and environment for statistical computing, R version 2.76.70, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>
60. Bailey TL. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* 2021;37:2834–40. <https://doi.org/10.1093/bioinformatics/btab203>
61. Tanaka E, Bailey T, Grant CE *et al.* Improved similarity scores for comparing motifs. *Bioinformatics* 2011;27:1603–9. <https://doi.org/10.1093/bioinformatics/btr257>
62. Morgan M, Pagès H, Obenchain V *et al.* 2025; Rsamtools: binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 2.24.1, <https://bioconductor.org/packages/Rsamtools>. (accessed May 20, 2026)
63. Bhardwaj V, Heyne S, Sikora K *et al.* snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics* 2019;35:4757–9. <https://doi.org/10.1093/bioinformatics/btz436>
64. Andrews S. 2010; FastQC: a quality control tool for high throughput sequence data. [Online]. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed May 20, 2026)
65. Ewels P, Magnusson M, Lundin S *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;32:3047–8. <https://doi.org/10.1093/bioinformatics/btw354>
66. Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
67. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550. <https://doi.org/10.1186/s13059-014-0550-8>
68. Pedersen TL. 2025; patchwork: the Composer of Plots. R package version 1.3.2. <https://CRAN.R-project.org/package=patchwork> (accessed May 20, 2026)
69. Villa R, Forne I, Muller M *et al.* MSL2 combines sensor and effector functions in homeostatic control of the Drosophila

- dosage compensation machinery. *Mol Cell* 2012;48:647–54. <https://doi.org/10.1016/j.molcel.2012.09.012>
70. Gilfillan GD, Straub T, de Wit E *et al.* Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev* 2006;20:858–70. <https://doi.org/10.1101/gad.1399406>
 71. Li F, Schiemann AH, Scott MJ. Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the *Drosophila* MSL1/MSL2 complex. *Mol Cell Biol* 2008;28:1252–64. <https://doi.org/10.1128/MCB.00910-07>
 72. Figueiredo MLA, Kim M, Philip P *et al.* Non-coding roX RNAs prevent the binding of the MSL-complex to heterochromatic regions. *PLoS Genet* 2014;10:e1004865. <https://doi.org/10.1371/journal.pgen.1004865>
 73. Nooranikhojasteh A, Tavallaei G, Orouji E. Benchmarking peak calling methods for CUT&RUN. *Bioinformatics* 2025;41:btaf375. <https://doi.org/10.1093/bioinformatics/btaf375>
 74. Alekseyenko AA, Peng S, Larschan E *et al.* A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* 2008;134:599–609. <https://doi.org/10.1016/j.cell.2008.06.033>
 75. Straub T, Grimaud C, Gilfillan GD *et al.* The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS Genet* 2008;4:e1000302. <https://doi.org/10.1371/journal.pgen.1000302>
 76. Eggers N, Becker PB. Cell-free genomics reveal intrinsic, cooperative and competitive determinants of chromatin interactions. *Nucleic Acids Res* 2021;49:7602–17. <https://doi.org/10.1093/nar/gkab558>
 77. Philip P, Pettersson F, Stenberg P. Sequence signatures involved in targeting the male-specific lethal complex to X-chromosomal genes in *Drosophila melanogaster*. *BMC Genomics* 2012;13:97. <https://doi.org/10.1186/1471-2164-13-97>
 78. Gu W, Wei X, Pannuti A *et al.* Targeting the chromatin-remodeling MSL complex of *Drosophila* to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities. *EMBO J* 2000;19:5202–11. <https://doi.org/10.1093/emboj/19.19.5202>
 79. Jain D, Baldi S, Zabel A *et al.* Active promoters give rise to false positive ‘Phantom Peaks’ in ChIP-seq experiments. *Nucleic Acids Res* 2015;43:6959–68. <https://doi.org/10.1093/nar/gkv637>
 80. Skene PJ, Henikoff S. A simple method for generating high-resolution maps of genome-wide protein binding. *eLife* 2015;4:e09225. <https://doi.org/10.7554/eLife.09225>
 81. Henikoff S, Henikoff JG, Kaya-Okur HS *et al.* Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *eLife* 2020;9:e63274. <https://doi.org/10.7554/eLife.63274>
 82. Tian SZ, Yang Y, Ning D *et al.* 3D chromatin structures associated with ncRNA roX2 for hyperactivation and coactivation across the entire X chromosome. *Sci Adv* 2024;10:eado5716. <https://doi.org/10.1126/sciadv.ado5716>
 83. Soruco MM, Chery J, Bishop EP *et al.* The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes Dev* 2013;27:1551–6. <https://doi.org/10.1101/gad.214585.113>
 84. Tikhonova E, Fedotova A, Bonchuk A *et al.* The simultaneous interaction of MSL2 with CLAMP and DNA provides redundancy in the initiation of dosage compensation in *Drosophila* males. *Development* 2019;146:dev179663. <https://doi.org/10.1242/dev.179663r1>
 85. Jiang N, Emberly E, Cuvier O *et al.* Genome-wide mapping of Boundary Element-Associated Factor (BEAF) binding sites in *Drosophila melanogaster* links BEAF to transcription. *Mol Cell Biol* 2009;29:3556–68. <https://doi.org/10.1128/MCB.01748-08>
 86. Ngedu S, Vidrine EJ, Hart CM. Characterization of core promoter activation by the *Drosophila* insulator-binding protein BEAF. *Sci Rep* 2025;15:37129. <https://doi.org/10.1038/s41598-025-21036-z>
 87. Matsukage A, Hirose F, Yoo M-A *et al.* The DRE/DREF transcriptional regulatory system: a master key for cell proliferation. *Biochim Biophys Acta* 2008;1779:81–9. <https://doi.org/10.1016/j.bbagr.2007.11.011>
 88. Heseding C, Saumweber H, Rathke C *et al.* Widespread colocalization of the *Drosophila* histone acetyltransferase homolog MYST5 with DREF and insulator proteins at active genes. *Chromosoma* 2017;126:165–78. <https://doi.org/10.1007/s00412-016-0582-9>
 89. Tue NT, Yoshioka Y, Mizoguchi M *et al.* DREF plays multiple roles during *Drosophila* development. *Biochim Biophys Acta* 2017;1860:705–12. <https://doi.org/10.1016/j.bbagr.2017.03.004>
 90. Feller C, Prestel M, Hartmann H *et al.* The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. *Nucleic Acids Res* 2012;40:1509–22. <https://doi.org/10.1093/nar/gkr869>
 91. Lam KC, Muhlpfordt F, Vaquerizas JM *et al.* The NSL complex regulates housekeeping genes in *Drosophila*. *PLoS Genet* 2012;8:e1002736. <https://doi.org/10.1371/journal.pgen.1002736>
 92. Kiss AE, Venkatasubramani AV, Pathirana D *et al.* Processivity and specificity of histone acetylation by the male-specific lethal complex. *Nucleic Acids Res* 2024;52:4889–905. <https://doi.org/10.1093/nar/gkac123>
 93. Raja SJ, Charapitsa I, Conrad T *et al.* The nonspecific lethal complex is a transcriptional regulator in *Drosophila*. *Mol Cell* 2010;38:827–41. <https://doi.org/10.1016/j.molcel.2010.05.021>
 94. Park Y, Kelley RL, Oh H *et al.* Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. *Science* 2002;298:1620–3. <https://doi.org/10.1126/science.1076686>
 95. Pal S, Oliver B, Przytycka TM. Cell-type specific variation in X-chromosome dosage compensation in *Drosophila*. *MicroPubl Biol* 2025;2025. <https://doi.org/10.17912/micropub.biology.001501>
 96. Lee H, McManus CJ, Cho D-Y *et al.* DNA copy number evolution in *Drosophila* cell lines. *Genome Biol* 2014;15:R70. <https://doi.org/10.1186/gb-2014-15-8-r70>
 97. Cook RK, Christensen SJ, Deal JA *et al.* The generation of chromosomal deletions to provide extensive coverage and subdivision of the *Drosophila melanogaster* genome. *Genome Biol* 2012;13, R21.
 98. Lee H, Cho DY, Whitworth C *et al.* Effects of gene dose, chromatin, and network topology on expression in *Drosophila melanogaster*. *PLoS Genet* 2016;12:e1006295. <https://doi.org/10.1371/journal.pgen.1006295>
 99. Heun P, Erhardt S, Blower MD *et al.* Mislocalization of the *Drosophila* centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell* 2006;10:303–15. <https://doi.org/10.1016/j.devcel.2006.01.014>
 100. Holland AJ, Cleveland DW. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* 2009;10:478–87. <https://doi.org/10.1038/nrm2718>
 101. Chunduri NK, Storchova Z. The diverse consequences of aneuploidy. *Nat Cell Biol* 2019;21:54–62. <https://doi.org/10.1038/s41556-018-0243-8>
 102. Storchova Z, Kuffer C. The consequences of tetraploidy and aneuploidy. *J Cell Sci* 2008;121:3859–66. <https://doi.org/10.1242/jcs.039537>
 103. Dodgson SE, Kim S, Costanzo M *et al.* Chromosome-specific and global effects of aneuploidy in *Saccharomyces cerevisiae*. *Genetics* 2016;202:1395–409. <https://doi.org/10.1534/genetics.115.185660>
 104. McAnally AA, Yampolsky LY. Widespread transcriptional autosomal dosage compensation in *Drosophila* correlates with gene expression level. *Genome Biol Evol* 2009;2:44–52. <https://doi.org/10.1093/gbe/evp054>

105. Malone JH, Cho D-Y, Mattiuzzo NR *et al.* Mediation of *Drosophila* autosomal dosage effects and compensation by network interactions. *Genome Biol* 2012;13:R28. <https://doi.org/10.1186/gb-2012-13-4-r28>
106. Chen Z-X, Golovkina K, Sultana H *et al.* Transcriptional effects of gene dose reduction. *Biol Sex Differ* 2014;5:5. <https://doi.org/10.1186/2042-6410-5-5>
107. Zhang Z, Presgraves DC. Translational compensation of gene copy number alterations by aneuploidy in *Drosophila melanogaster*. *Nucleic Acids Res* 2017;45:2986–93. <https://doi.org/10.1093/nar/gkx106>
108. Vcelar S, Melcher M, Auer N *et al.* Changes in chromosome counts and patterns in CHO cell lines upon generation of recombinant cell lines and subcloning. *Biotechnol J* 2018;13:1700495. <https://doi.org/10.1002/biot.201700495>