nature biotechnology

Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification

Jeffrey J Quinn[1,2], Ibrahim A Ilik[3,4], Kun Qu[1,4], Plamen Georgiev[3,4], Ci Chu[1], Asifa Akhtar[3] & Howard Y Chang[1]

Little is known about the functional domain architecture of long noncoding RNAs (lncRNAs) because of a relative paucity of suitable methods to analyze RNA function at a domain level. Here we describe domain-specific chromatin isolation by RNA purification (dChIRP), a scalable technique to dissect pairwise RNA-RNA, RNA-protein and RNA-chromatin interactions at the level of individual RNA domains in living cells. dChIRP of roX1, a lncRNA essential for *Drosophila melanogaster* X-chromosome dosage compensation, reveals a 'three-fingered hand' ribonucleoprotein topology. Each RNA finger binds chromatin and the male-specific lethal (MSL) protein complex and can individually rescue male lethality in *roX*-null flies, thus defining a minimal RNA domain for chromosome-wide dosage compensation. dChIRP improves the RNA genomic localization signal by >20-fold relative to previous techniques, and these binding sites are correlated with chromosome conformation data, indicating that most roX-bound loci cluster in a nuclear territory. These results suggest dChIRP can reveal lncRNA architecture and function with high precision and sensitivity.

lncRNAs are a recently recognized class of molecules that participate in diverse biological processes. Many lncRNAs act at the interface of chromatin-modifying machinery and the genome and regulate homeotic gene expression, epigenetic imprinting and dosage compensation of entire chromosomes[1,2]. Although thousands of lncRNAs with tissue- and disease-specific expression have been discovered, the biological functions of the vast majority remain unknown or have not been mechanistically characterized[3,4]. One prevailing theory states that the functional diversity of lncRNAs is achieved through modularity of specific RNA domains that coordinate combinatorial RNA-RNA, RNA-DNA and RNA-protein interactions[5].

Many existing protein-centric technologies can detect the interaction of RNAs with other biomolecules. Cross-linking immunoprecipitation (CLIP) and related methods can identify RNAs bound to specific RNA-binding proteins (RBPs)[6,7]. However, multiple CLIP experiments are required to reveal multivalent interactions of one RNA with multiple RBPs, and separate proximity ligation assays must be conducted to map RNA-RNA interactions[8]. By contrast, ChIRP and related strategies are RNA-centric techniques for exploring chromatin-associated lncRNA function[9–11]. Specifically, ChIRP enables the genome-wide identification of RNA-chromatin binding sites and has been used to provide insights into the mechanisms of dosage compensation, cancer progression, viral pathogenesis and *FMR1* gene silencing[9,12–14]. But how do lncRNAs interact with chromatin-modifying complexes? Do these chromatin-associated lncRNAs have modular domains—much like their protein counterparts, transcription factors—that are responsible for their varied functions?

Here we describe dChIRP, a technique that dissects lncRNAs domain by domain to discover functional elements. We demonstrate the utility of the method by identifying functional domains in the roX1 lncRNA. The roX RNAs are essential for dosage compensation in male flies, wherein gene expression from the single male X chromosome is doubled to match that of females' two[15]. This X upregulation is directed by the MSL ribonucleoprotein complex, composed of roX1 and roX2 lncRNAs and five MSL proteins (MSL1–MSL3, MLE and MOF), which spread in *cis* along the X chromosome and deposit activating histone marks at defined loci[16–19]. Both RNAs are known to interact with specific sites on the X chromosome called chromatin entry sites (CESs) that are co-occupied by the MSL proteins[9,17]. CLAMP, a zinc finger protein, directly binds to the MSL recognition element (MRE, a GA repeat) within CESs and somehow links the roX-MSL complex to DNA[20]. The two roX RNAs are functionally redundant and individually dispensable despite sharing limited sequence homology, differing in size by an order of magnitude (roX1 ~3.8 kb, roX2 ~600 bp) and having different developmental expression patterns[21]. Although researchers have conducted genetic and phylogenetic studies on these RNAs, their biochemical role in dosage compensation remains poorly defined[22–24]. Here, application of dChIRP uncovers several features of roX1 RNA's architecture and function.

RESULTS

Concept of dChIRP

The goal of a dChIRP experiment is to dissect the functional domains of an RNA of interest within its native cellular context. For a target RNA, dChIRP can simultaneously map domain-level RNA-RNA,

[1]Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California, USA. [2]Department of Bioengineering, Stanford University School of Medicine and School of Engineering, Stanford, California, USA. [3]Max Planck Institute of Immunobiology and Epigenetics, Stübeweg 51, Freiburg im Breisgau, Germany. [4]These authors contributed equally to this work. Correspondence should be addressed to H.Y.C. (howchang@stanford.edu).

Received 27 November 2013; accepted 3 June 2014; published online 6 July 2014; doi:10.1038/nbt.2943

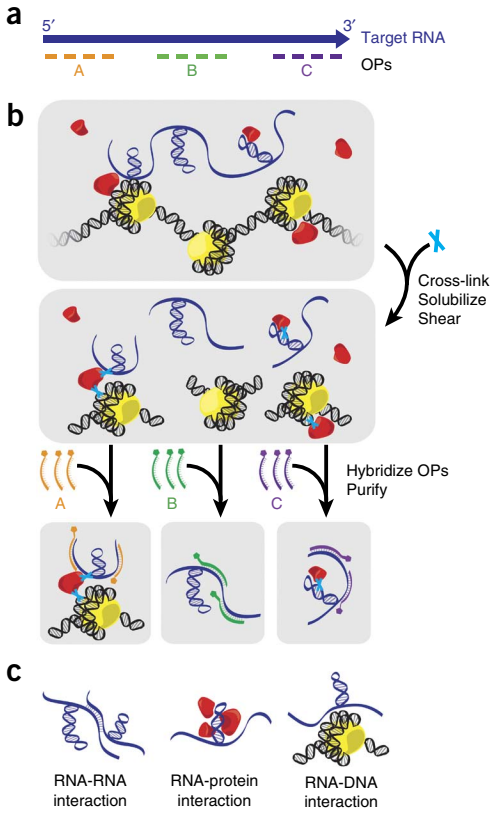

Figure 1 dChIRP uses antisense oligonucleotides to purify specific RNA domains and associated RNAs, proteins and chromatin. (**a**) dChIRP oligonucleotide design strategy. Biotinylated antisense oligonucleotide pools (OPs) are designed to tile specific regions of the target RNA. (**b**) dChIRP workflow. To prepare chromatin, whole cells are cross-linked to preserve protein–nucleic acid interactions. Sonication is used to solubilize the nuclear fraction and shear nucleic acids. Next, the chromatin is subdivided into equal samples. OPs are added to each sample, which hybridize to the targeted RNA fragments. The biotinylated oligonucleotides, RNA targets and cross-linked biomolecules are then purified on magnetic streptavidin beads, and unbound material is washed away. (**c**) RNA-, protein- and DNA-sensitive modalities of dChIRP. RNA, protein and DNA fractions are extracted from each dChIRP sample. Intra- or intermolecular RNA-RNA, RNA-protein and RNA-DNA interactions may be measured by RT-qPCR, immunoblotting, and qPCR or sequencing, respectively.

RNA-protein and RNA-chromatin interactions as well as identifying genomic binding sites with increased sensitivity. First, biotinylated antisense 20-mer oligonucleotides are designed with non-overlapping and nonredundant sequences, avoiding regions of low complexity or high occurrence in the reference genome (**Fig. 1a**). Instead of dividing the oligonucleotides into two equal groups ('even' and 'odd' pools) that tile the whole RNA, as with traditional ChIRP experiments, in dChIRP the oligonucleotides are divided into domain-specific oligonucleotide pools, such that each pool targets a distinct RNA domain. The targeted RNA regions may be devised arbitrarily (such as by even subdivision of the RNA length) or defined by biochemical-, genetic- or conservation-based functional evidence.

Next, whole cells are cross-linked to preserve protein–nucleic acid interactions (**Fig. 1b**). We have found that a combination of fixation with 1% glutaraldehyde or 1% formaldehyde followed by cross-linking with 3% formaldehyde, such as is used in capture hybridization analysis of RNA targets (CHART)[9,10], gives the best results. The nuclei from fixed cells are then extracted and lysed. Sonication is used to solubilize the chromatin fraction and shear nucleic acids. It is important to fragment DNA to ~500 bp for sequencing, and RNAs should be sheared to roughly the size of the target RNA regions (200–500 nt) such that domain-specific interactions can be independently purified. The sheared chromatin is then divided into equal samples. Oligonucleotide pools are added to each sample and allowed to hybridize under stringent conditions. After hybridization, the biotinylated oligonucleotides, hybridized RNA and associated biomolecules are purified on magnetic streptavidin beads and washed thoroughly to remove nonspecific interactions.

The recovered material from each dChIRP sample is further divided for RNA, DNA and protein extraction and then analyzed. The RNA fraction can be analyzed by RT-qPCR with primers designed to amplify the targeted RNA regions or other RNA species. This analysis is used to confirm efficient, domain-specific RNA recovery and identify potential intramolecular or intermolecular RNA-RNA interactions. The protein fraction may also be analyzed by immunoblotting against suspected RNA-associated proteins, thus identifying relevant protein-binding RNA domains. In this way, dChIRP is the reciprocal of CLIP[6,7]. Lastly, analyzing the DNA by qPCR reveals domain-level RNA-DNA or RNA-chromatin interactions. Recovered DNA may also be sequenced to identify RNA-occupied sites across the genome. Thus, in one *in situ* experiment, dChIRP can simultaneously map the RNA-, DNA- and protein-interacting domains of an RNA (**Fig. 1c**).

The roX1 D domains form topological 'fingers'

We tested and validated the dChIRP method using roX1 lncRNA. We have previously reported by individual-nucleotide-resolution CLIP (iCLIP) that MLE and MSL2 directly contact roX1 RNA at three distinct domains (denoted D1, D2 and D3), whereas the intervening domains (U1, U2 and U3) exhibit very limited binding (**Fig. 2a**)[18]. Using these empirically determined domains as a guide, we designed 6 dChIRP oligonucleotide pools, each comprised of 12 distinct biotinylated oligonucleotides that tile roughly equal lengths of roX1 (OP-U1 to OP-D3). We performed dChIRP in chromatin prepared from Clone 8 cells (a male *D. melanogaster* line) cross-linked with 1% + 3% formaldehyde using the 6 roX1 oligonucleotide pools and a negative control pool directed against the absent LacZ mRNA.

To confirm that dChIRP could recover the intended fragments of roX1 RNA, we purified the RNA fraction from the dChIRP samples and analyzed RNA recovery by RT-qPCR using primers for each of the six roX1 domains and GAPDH, a control mRNA that should not be enriched by roX1 dChIRP. RNA recovery of each domain was quantitated against input RNA. We confirmed that roX1 dChIRP specifically retrieved roX1 RNA (>1,000-fold enriched over GAPDH mRNA), whereas LacZ ChIRP did not enrich for roX1 RNA (**Fig. 2b**). For each dChIRP sample, we normalized roX1 RNA fragment recovery to total roX1 RNA recovery (percent roX1 RNA recovery) and found that each dChIRP oligonucleotide pool best enriched for its targeted RNA fragment (**Fig. 2c**, along the diagonal). D1, D2 and D3 dChIRP recovered their target fragment nearly exclusively and independently, whereas U1, U2 and U3 dChIRP each predominantly recovered all three U domains. This is unexpected because the U domains are not contiguous and are distant in one-dimensional space. For example, U3 dChIRP efficiently retrieves the U3 fragment without the neighboring D2 and D3 fragments, and yet it also retrieves the more distant U1 and U2 domains. The co-recovery suggests that the U3 domain is associated with both U1 and U2, whereas the neighboring D2 and D3 domains are sheared off during chromatin preparation.

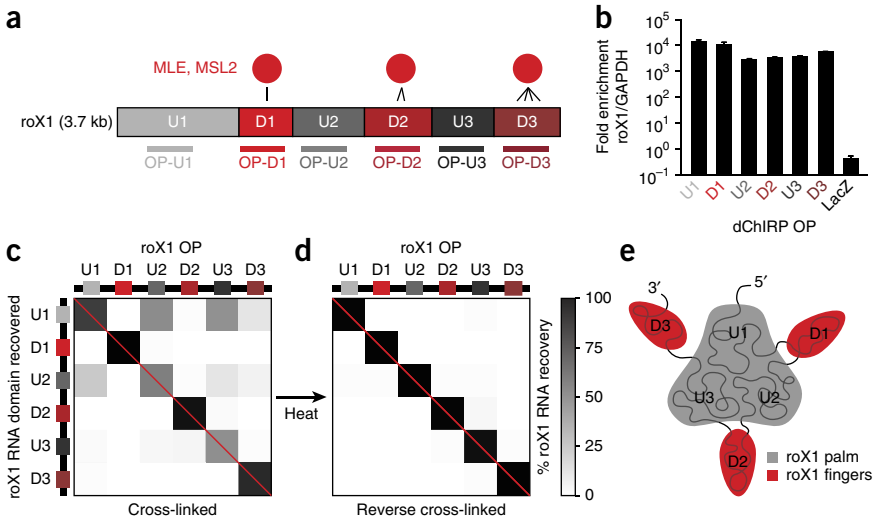

Figure 2 dChIRP RNA co-recovery reveals roX1's topological architecture. (**a**) Schematic representation of known roX1 domain interactions with MLE and MSL2 proteins and dChIRP oligonucleotide pool (OP) design strategy. MLE and MSL2 directly contact the three D domains (D1, D2 and D3), whereas the three intervening U domains (U1, U2 and U3) exhibit minimal binding. Six pools were designed to target and recover each domain. (**b**) roX1 dChIRP specifically enriches for roX1 RNA. roX1 RNA is enriched >1,000-fold over the abundant GAPDH mRNA in roX1 dChIRP samples. LacZ ChIRP does not enrich for roX1 over GAPDH. Average of technical triplicates + s.d. shown. (**c**,**d**) roX1 RNA recovery by dChIRP. Within each sample, roX1 domain recovery was quantified against input and normalized to total roX1 RNA recovery (percent roX1 RNA recovery). Each column is the OP used for dChIRP; each row is the retrieved RNA domain. As expected, each oligonucleotide pool best enriches for the target roX1 domain (**c**, red diagonal). Off-diagonal signal indicates interactions between RNA comains (e.g., U domain co-recovery). (**d**) dChIRP retrieval of RNA after thermal reverse cross-linking. Each of the domains of roX1 was then independently recovered. (**e**) Schematic representation of roX1 intramolecular topology. Domains U1, U2 and U3 are topologically proximal to one another, forming the core palm of roX1. Domains D1, D2 and D3 extend as fingers and are distant from one another and the intervening U domains.

To determine whether the co-recovery of U domains is dependent on cross-linking, we performed roX1 dChIRP in thermally reverse-cross-linked chromatin (**Fig. 2d**). Here, each roX1 RNA fragment was uniquely recovered; co-recovery of the U domains was absent. Furthermore, to confirm that the observed co-recovery was not caused by cross-hybridization of oligonucleotides within each oligonucleotide pool, we subdivided the 6 pools into 12 non-overlapping, even-odd paired pools (**Supplementary Fig. 1a**). In these subdivided pools, the pattern of roX1 U domain co-recovery is reproduced in the U domain even-odd pairs, demonstrating that co-recovery is not an artifact of oligonucleotides from one pool mis-hybridizing to other roX1 RNA fragments (**Supplementary Fig. 1b**,**c**). dChIRP of human HOTAIR lncRNA from MDA-MB-231 breast cancer cells retrieved domain-specific RNA regions (**Supplementary Fig. 2**), highlighting the generality of the dChIRP approach and the unique architecture of roX1.

The cross-linking-dependent co-recovery of roX1's U domains indicates that these three domains are topologically associated in three-dimensional space, cross-linked together possibly through accessory RBPs, RNA-RNA interactions or both. Conversely, the unique recovery of domains D1, D2 and D3 suggests that these domains are physically distant from all others. One topological model consistent with this pattern is a 'three-fingered hand' architecture, such that the U domains form a palm from which the D domains individually extend like fingers (**Fig. 2e**).

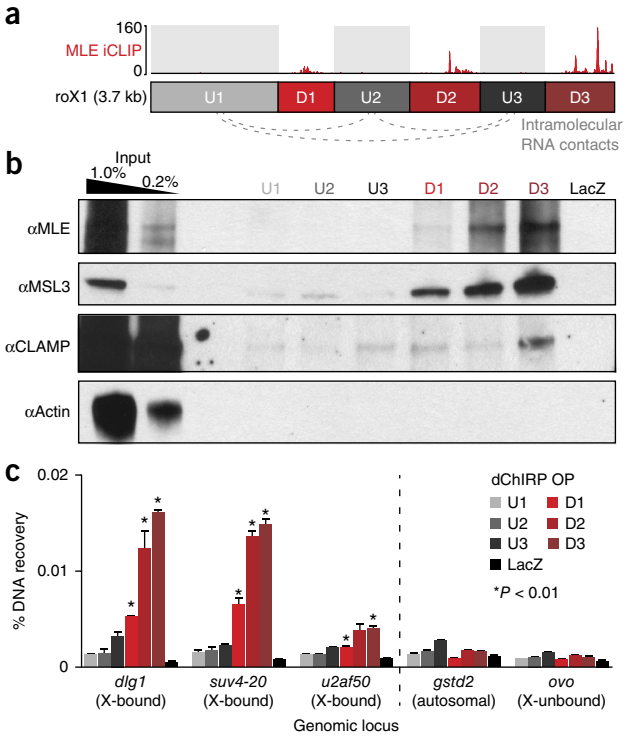

Figure 3 roX1 D domains interact with the MSL complex and chromatin on the X chromosome. (**a**) Schematic representation of roX1 interactions. The three D domains (D1, D2 and D3) directly contact MLE by iCLIP[18]. The three intervening U domains do not contact MLE but are topologically associated (gray dashed lines). (**b**) dChIRP–western blot confirms known MLE-bound domains of roX1. We analyzed the protein fraction from each roX1 dChIRP sample by immunoblotting against MLE, MSL3, CLAMP and actin. roX1 domains D1, D2 and D3 efficiently recovered MLE and MSL3 proteins. D3 recovered more protein than D2, and D2 recovered more than D1. Domains U1, U2 and U3 recovered minimal or undetectable MLE and MSL3. Only D3 recovered CLAMP appreciably, albeit very weakly. LacZ ChIRP recovered no detectable protein. Actin was not detected in any sample. (**c**) The three D domains of roX1 are associated with chromatin at dosage-compensated loci on the X chromosome. We analyzed DNA fractions from each roX1 dChIRP sample by qPCR and normalized to input. Five genomic loci were investigated: three MSL-bound X-linked loci (*dlg1*, *suv4-20* and *u2af50*), one locus from an autosome (*gstd2*) and an unbound X-linked locus (*ovo*). dChIRP of domains D1, D2 and D3 significantly enriched for X-bound loci relative to control loci (*$P < 0.01$, *t*-test). Domains D2 and D3 recovered significantly more X-bound DNA than D1 or the three U domains. LacZ ChIRP failed to recover substantial DNA from any locus. Average of technical triplicates + s.d. shown.

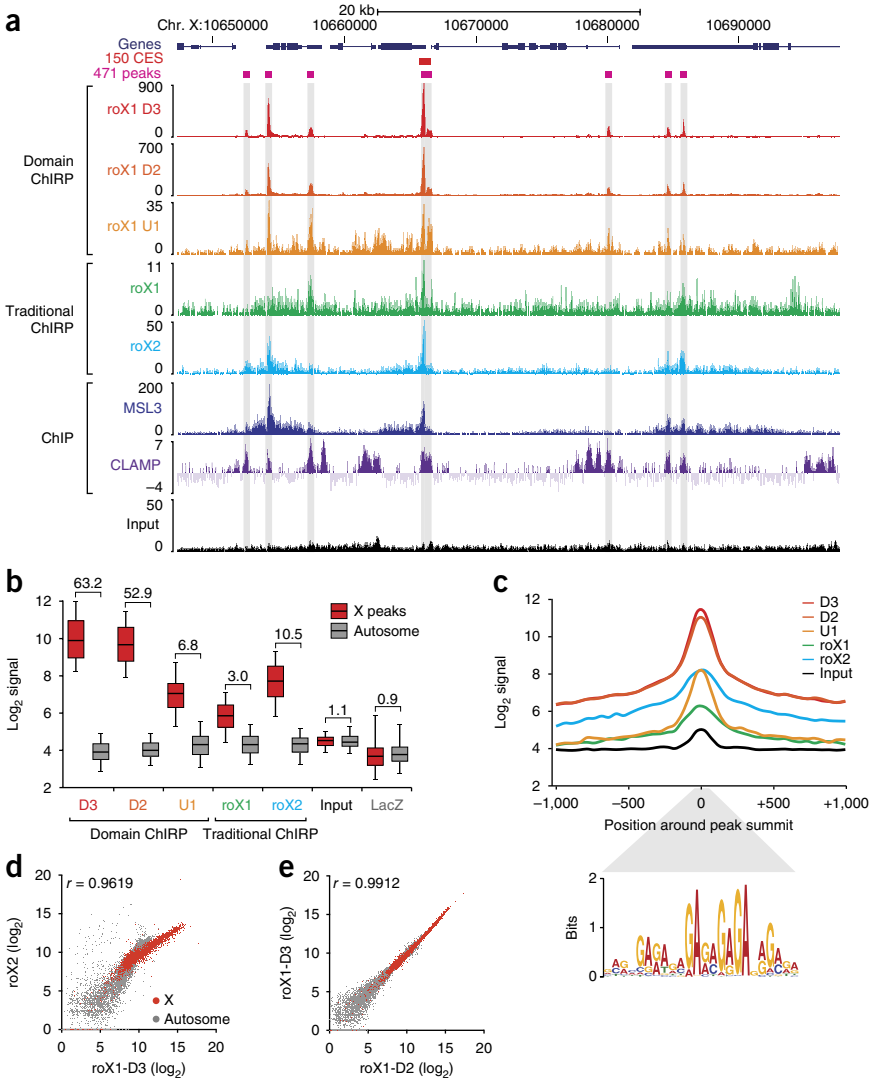

Figure 4 dChIRP boosts genomic occupancy signal relative to traditional ChIRP-seq. (**a**) Genomic tracks of dChIRP-seq results at a representative X-linked locus. Sequencing tracks from roX1 dChIRP (U1, D2, D3) and traditional ChIRP (roX1, roX2). roX1 peaks (gray highlight) aligned with peaks from roX2, MSL3 ChIP and CLAMP ChIP[17,20]. (**b**) Comparison of signal from the X chromosome and noise on autosomes. Signal was calculated at 457 peaks on the X and 457 random loci on autosomes. Box-plot represents 1st, 2nd and 3rd quartiles; whiskers denote 5th and 95th percentiles. SNR (X-peak mean to autosome mean) is indicated above each sample. (**c**) Average peak diagram of 457 peaks on X. roX1 dChIRP produces higher signal and more focal peaks than traditional ChIRP. The MRE GA-repeat motif is significantly enriched within peaks ($P = 5.2e\text{-}526$, multiple EM for motif elicitation (MEME)[32]) and is located specifically at peak summits ($P = 5.3e\text{-}182$, CentriMo[32]). (**d**,**e**) Correlation between ChIRP-seq experiments. (**d**) roX1-D3 dChIRP and roX2 ChIRP signal are highly correlated ($r = 0.9619$), especially on the X chromosome (red). (**e**) roX1 D2 and D3 dChIRP are very highly correlated ($r = 0.9912$). roX1 D2 and D3, and roX2 co-occupy the same loci on the X chromosome.

We wanted to know if these U domain interactions were mediated by base pairing, and we used Mfold to produce *in silico* models of secondary structures of roX1[25]. The structure modeling did not predict substantial secondary structures or complementary sequences between or within the U domains. This suggests that the U1-U2-U3 interaction is not likely to be caused by Watson-Crick base pairing but rather by tertiary RNA structures or interacting proteins. Mfold did, however, predict that the three D domains would adopt long, linear stem-loops (**Supplementary Fig. 3**), not unlike those observed in roX2 (refs. 18,26). These putative structures are coincident with MLE-bound residues and roX boxes, a repeated 8-nt motif in roX1 and roX2. The D3 structures were previously validated by biochemical structure mapping[18].

roX1 fingers bind the MSL complex and chromatin

Having mapped intramolecular interactions within roX1, we next used dChIRP to verify the previous iCLIP finding that MLE directly contacts roX1's D domains (**Fig. 3a**)[18]. From each dChIRP sample, we extracted the protein fraction and performed western blotting with MLE, MSL3, CLAMP and actin antibodies (**Fig. 3b**). We found that U1, U2 and U3 recovered little or no detectable MLE or MSL3, whereas the D1, D2 and D3 domains recovered substantially more, with distinct efficiencies (D3 > D2 > D1). This hierarchy is consistent with the iCLIP binding pattern[18] and demonstrates that dChIRP is sensitive to such differences in affinity. Furthermore, interactions between the roX1 D domains and MSL3 suggest that the entire core MSL complex interacts with the roX RNAs en masse, although by dChIRP alone we cannot establish whether roX1 contacts MSL3 directly or indirectly via MSL2. By contrast, only D3 could recover CLAMP, albeit substantially less successfully than it recovered MLE or MSL3. This weak recovery suggests that the interaction between roX1 and CLAMP may be less direct or have a lower affinity than the roX-MSL interaction. As negative controls, the LacZ oligonucleotide pool recovered no proteins, and actin was not detected in any sample. roX1 dChIRP from reverse cross-linked chromatin recovered substantially less protein than dChIRP from cross-linked chromatin, indicating that the protein recovery is dependent upon cross-linking (**Supplementary Fig. 4a**). To further investigate CLAMP's association with the MSL complex, we performed immunoprecipitation of CLAMP in untreated, DNase-treated or RNase-treated chromatin. MLE was co-precipitated regardless of nuclease treatment, suggesting that protein-mediated interaction with MLE may link CLAMP to the core MSL complex (**Supplementary Fig. 4b**).

Next, to discover which domains of roX1 most closely contact chromatin, we analyzed the DNA fraction from each dChIRP sample by qPCR. We used primers against three known MSL- and roX2-bound loci on the X chromosome (*dlg1*, *suv4-20* and *u2af50*), as well as against two negative control loci: one autosomal (*gstd2*) and one on the X chromosome (*ovo*). As expected, roX1 dChIRP significantly enriches for X-bound loci relative to the control loci (**Fig. 3c**). The D1, D2 and D3 domains recover more X-bound DNA than the U1, U2 and U3 domains. This suggests that the D domains of roX1, which exclusively interact with MSL proteins, are more closely associated

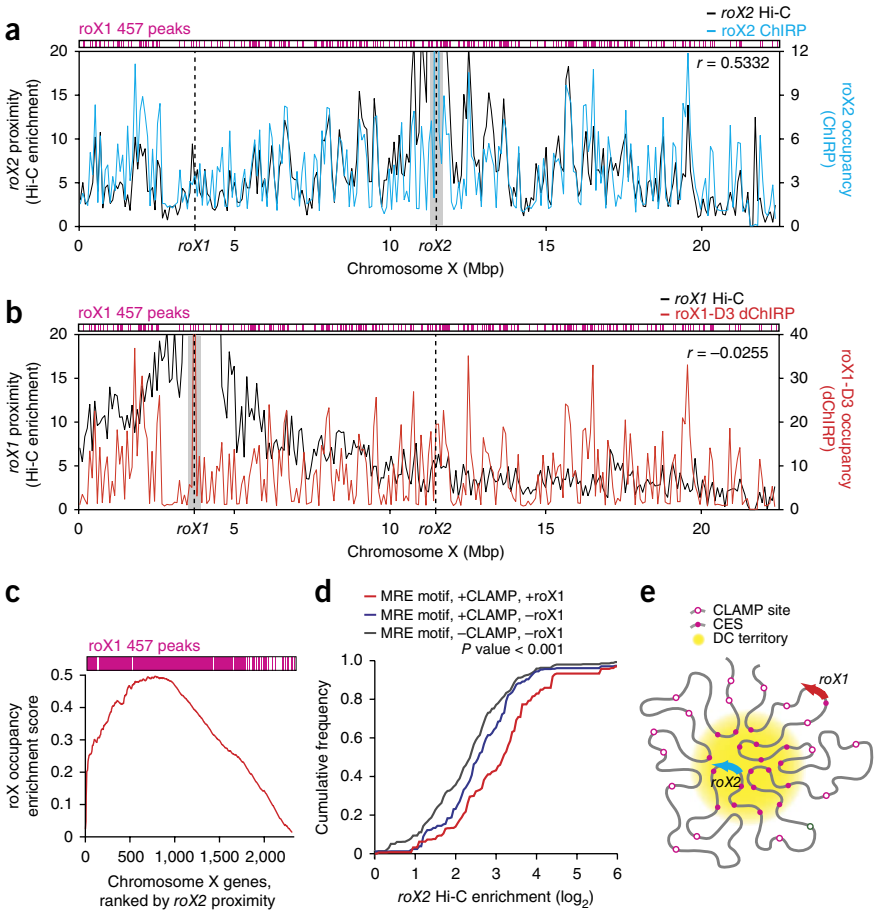

Figure 5 CESs cluster together in a dosage-compensation territory of the nucleus. (**a**,**b**) Correlation between (**a**) roX2 occupancy by ChIRP and *roX2* proximity as determined by Hi-C and (**b**) roX1 D3 occupancy by ChIRP and *roX1* proximity by Hi-C[27]. roX2 RNA occupancy is correlated with *roX2* proximity ($r = 0.5332$); roX1 RNA occupancy is not correlated with *roX1* proximity ($r = -0.0255$). roX1 457 peaks (magenta) are clustered at sites of high roX RNA occupancy. 400 kb around the *roX* gene loci were excluded (gray mask) for correlation calculation (Pearson's *r*), so as to exclude signal from direct ChIRP oligonucleotide-DNA recovery and one-dimensionally proximal chromosome sites. (**c**) GSEA of roX-occupied genes. Genes that are occupied by roX RNAs are significantly more likely to be proximal to the *roX2* locus (FDR < 0.001). (**d**) Instances of the MRE motif that are more proximal to the *roX2* locus are significantly more likely to be bound by CLAMP and co-occupied by roX RNAs ($P < 0.001$, Kolmogorov-Smirnov test). (**e**) Model of X-chromosome conformation. The *roX2* locus and CESs are clustered in a dosage-compensation (DC) territory. The *roX1* locus lies outside of the DC territory.

with chromatin than are the U domains. D2 and D3 also significantly enrich for each X-bound locus relative to D1, recapitulating the protein-binding affinity hierarchy. The negative control LacZ oligonucleotide pool does not enrich for X-bound loci.

Combining these results with the roX1 RNA architecture, we concluded that each of the roX1 D domain fingers can independently bind to the MSL proteins to form a ribonucleoprotein complex, which together grasp chromatin at hyper-expressed loci on the X chromosome.

dChIRP maps the genome-wide binding sites of roX1

As dChIRP of roX1 D domains recovered comparatively large amounts of DNA (**Fig. 3c**), we suspected that this domain-specific strategy could improve the signal from ChIRP-sequencing (ChIRP-seq) experiments and thus facilitate better identification of genome-wide, lncRNA-associated loci. To directly compare the two techniques, we performed both traditional ChIRP and dChIRP in Clone 8 cells. We used 11 different oligonucleotide pools: even and odd pools for full-length roX2 (12 oligonucleotides each), full-length roX1 (75 each), roX1 domain U1 (9 each), roX1 domain D2 (9 each) and roX1 domain D3 (9 each), as well as a single pool for LacZ (12 oligonucleotides) as a control (**Supplementary Fig. 1a**). We prepared sequencing libraries from the DNA fractions and sequenced each one. Even and odd lanes were aligned separately and normalized to mappable reads. The even-odd pairs were then merged and plotted as previously described[9].

We observed that the roX1 dChIRP and ChIRP tracks showed clear peaks that aligned precisely with known roX2-, MSL3- and CLAMP-binding sites[9,17,20] (**Fig. 4a**). Most prominently, the peaks from roX1 D3 and D2 were much higher in magnitude and focally tighter than those from roX1 U1 dChIRP or traditional roX1 ChIRP (note the *y* axes of dChIRP vs. traditional ChIRP signals require different scales in the figure because of the greater signal of the former).

We next sought to quantify dChIRP's improvement in signal-to-noise ratio (SNR) in the sequencing data relative to ChIRP. First, we used the MACS and ZINBA software suites to identify peaks and locate summits from the sequencing data. We then filtered the peaks based on signal magnitude, even-odd pair correlation and enrichment score, yielding 471 peaks, 457 of which (>97%) were on the X chromosome . By contrast, MSL3 ChIP and roX2 ChIRP identified 150 and 308 CESs, respectively. To represent background noise, we randomly selected an equal number of autosomal sites with nonzero signal. We then calculated the average signal over each site from the peaks on the X chromosome and the autosomes (**Fig. 4b**). The SNR was calculated by dividing the X peak signal by autosomal background. Traditional ChIRP of roX1 produced especially noisy results (SNR = 3.0), whereas dChIRP of roX1 U1—the domain with the weakest chromatin association—increased SNR compared with traditional ChIRP (SNR = 6.8) more than twofold. dChIRP of D2 and D3 further increased the SNR compared with traditional ChIRP (SNR = 52.9 and 63.2, respectively).

We believe that this improvement in SNR is the result of two factors. First, using fewer oligonucleotides per oligonucleotide pool (e.g., 9 for dChIRP versus 75 for traditional ChIRP) decreases the likelihood of recovering DNA in an RNA-independent manner through direct oligonucleotide-DNA hybridization. In addition, using fewer oligonucleotides decreases the risk of having two oligonucleotides in opposite pools with homologous sequences, which may produce false-positive peaks. Second, traditional roX1 ChIRP further dilutes signal by targeting domains of the RNA that are not involved in chromatin interaction, such as the U domains; this is observed in the SNR boost between U1 and D3 dChIRP (from 6.8 to 63.2).

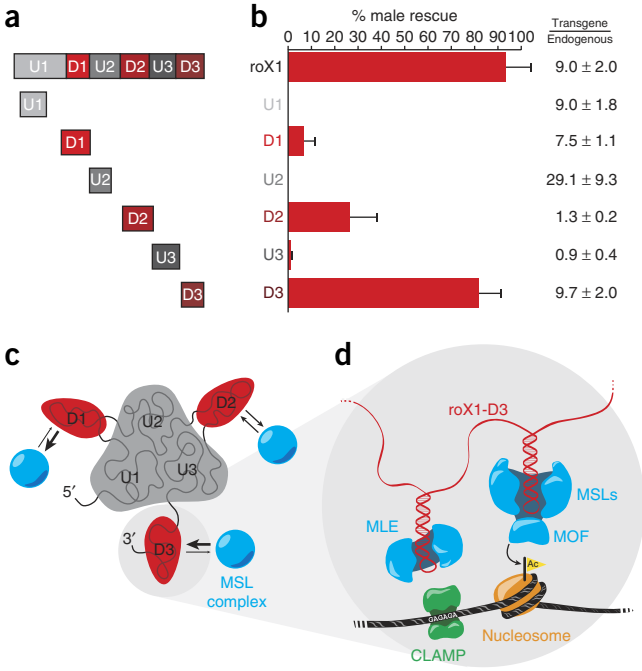

Figure 6 roX1's D domains are independent, functional RNA subunits. (**a**) Transgene designs. We cloned transgenic constructs of full-length roX1 and the six individual domains and then chromosomally integrated and expressed them under the tubulin-GAL4 promoter in *roX*-null flies. (**b**) Rescue of male lethality by roX1 transgenes. Transgenic males surviving to adulthood were counted and normalized to females. Only the D domains rescued males appreciably. Rescue by D3 is not significantly different from that of full-length roX1 (P value = 0.20, t-test). Average of three separate crosses + s.d. shown (on average, n = 800). roX transgene expression was quantified and normalized to endogenous roX1 expression in wild-type males, represented as relative fold (transgene/endogenous) ± s.d. (**c**,**d**) Integrated interaction map of the dosage-compensation complex with chromatin. (**c**) roX1 RNA is topologically organized such that the three U domains form a core palm and each of the D domains extends independently as a finger. Each D domain finger directly binds to proteins of the MSL complex, with domain D3 having the highest affinity and D1 the weakest. (**d**) CLAMP binds the GAGA motif at X-linked CESs and associates with MLE. MLE binds to stem-loops on roX1, which tethers MLE to the core MSL complex. MOF of the MSL complex recognizes and acetylates H4K16 in adjacent nucleosomes.

By minimizing oligonucleotide pools and targeting domains with strong chromatin associations, dChIRP can improve SNR >20-fold over traditional ChIRP.

To further demonstrate the increase in SNR achievable by dChIRP, we plotted the average signal around X chromosome peaks in 50-bp bins (**Fig. 4c**). Notably, roX1 dChIRP peaks have higher amplitude and are more focal than peaks from traditional roX1 or roX2 ChIRP. All 457 roX1 peaks on the X chromosome contain the MRE motif, which is significantly enriched directly at the peak summit. This motif is virtually indistinguishable from the motifs identified by roX2 ChIRP, MSL3 ChIP and CLAMP ChIP[9,17,20].

dChIRP of roX1 also reveals that roX1 occupies both its own genomic locus and the *roX2* locus; similarly, roX2 occupies the *roX1* locus (**Supplementary Fig. 5**). roX1 dChIRP also identified 11 autosomal sites that are weakly occupied by roX1, predominantly at the transcriptional start site of genes (**Supplementary Fig. 6**). The bound sites contain the MRE motif and are co-occupied by CLAMP but not MSL3. These sites may not be related to canonical dosage compensation and may represent misguided roX RNAs.

roX1 dChIRP-seq also allowed us to resolve similarities in chromatin occupancy between roX1 and roX2. Signal from roX1 D3 dChIRP and roX2 ChIRP are strongly correlated, especially on the X chromosome (**Fig. 4d**), indicating that these two RNAs bind the same loci with equivalent relative affinities. roX1 D2 and D3 dChIRP are also highly correlated and therefore bind the same loci (**Fig. 4e**). These findings support the observation that roX1 and roX2 are genetically redundant and that roX1 exhibits internal redundancy[21,23].

CESs cluster in a dosage-compensation territory

We next wanted to determine whether roX RNA occupancy is related to three-dimensional chromosome conformation to better understand how the roX RNAs spread along the X chromosome. Overlaying Hi-C enrichment data of *roX1* and *roX2* gene loci[27] with the roX ChIRP data revealed two notable patterns (**Fig. 5a**,**b**). Regions proximal to the *roX2* locus and roX2 RNA occupancy are correlated (R = 0.53), indicating that the *roX2* gene locus and CESs of dosage-compensated genes are spatially proximal and reside within the same chromosome territory (**Fig. 5a**). This is consistent with previous DNA fluorescence *in situ* hybridization (FISH) experiments that show three CESs cluster in an MSL2-occupied nuclear territory in a male-specific, MSL-dependent manner[28]. By contrast, the *roX1* locus makes few long-range contacts with distant chromosomal regions, and the correlation between roX1 RNA occupancy and *roX1* proximity is poor (R = −0.03; **Fig. 5b**). Gene set enrichment analysis (GSEA) showed that authentic CESs are significantly enriched for spatial proximity to *roX2* locus (**Fig. 5c**; false discovery rate (FDR) < 0.001). Furthermore, CLAMP-bound sites that are proximal to the *roX2* locus are significantly more likely to be co-occupied by roX2 RNA and MSL3 than other CLAMP-bound sites not in *roX2* proximity (**Fig. 5d**, P < 0.001). Thus, the *roX2* locus and CESs (but not the *roX1* locus) cluster into a dosage-compensation territory formed by large-scale chromosomal conformation (**Fig. 5e**).

roX1's D domains are independent RNA modules

The three D domains of roX1 are topologically independent and interact with MSL proteins and chromatin with distinct affinities (D3 > D2 > D1). These findings suggest that the D domains are independently functional RNA modules and that each may suffice for dosage compensation. To test this hypothesis, we overexpressed single U or D domains of roX1 as tubulin-GAL4–driven transgenes inserted in position 65B2 of chr3L in *roX*-null flies and tested their ability to rescue male lethality (**Fig. 6a**). None of the U domain constructs appreciably rescued males, but all three of the D domain constructs rescued roX deficiency, albeit with different efficiencies (**Fig. 6b**). The D domain constructs' rescue efficiency echoes the previously observed affinity hierarchy (D3 > D2 > D1; **Fig. 3**). The D3 construct alone was able to rescue male lethality as efficiently as full-length roX1. We did not observe a direct correlation between rescue and transgene expression relative to endogenous roX1 in wild-type males.

To test whether multiple D domains can enhance rescue of male lethality, we expressed a D1-D3 fusion, but we did not observe a notable change in rescue efficiency over D3 alone (**Supplementary Fig. 7a**). Only when the transgenes are driven by a weaker promoter (daughterless-GAL4) at near-endogenous levels does the two-domain fusion increase male rescue (**Supplementary Fig. 7b**). This suggests that the multiple D domains act cooperatively, increasing the RNA's functional output at lower concentrations. When one of the putative stem-loops in D2 is disrupted by truncation (D2ΔSL and D2ΔSL-D3),

male rescue is greatly decreased compared to that with wild-type D2 domain, indicating that this sequence is essential to transgene function (**Supplementary Fig. 7**).

DISCUSSION

dChIRP is an RNA-centric technology for dissecting RNA functional domains involved in RNA-RNA, RNA-protein and RNA-chromatin interactions. dChIRP is a broadly applicable method for dissecting lncRNAs of sufficient length. As traditional ChIRP has been successfully applied to diverse RNAs with a wide range of abundances, the improved sensitivity and information content of dChIRP should expand the use of this technology[9,12–14]. dChIRP traps endogenous RNA-chromatin interactions in living cells and then breaks the RNA apart to decipher which section of the RNA is doing what job. dChIRP investigates domain-level interactions with protein, RNA and chromatin. The identified domains are then appropriate subjects for further dissection by additional methods, such as secondary structure probing by selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE). The lower limit of dChIRP resolution is ~200–500 bp as determined by RNA shearing, and different regions may be targeted and iteratively refined. Targeted regions may be determined arbitrarily or based on existing knowledge of the RNA's biology.

Currently the standard approach to dissect RNA function involves generating many deletion mutants; each mutant is individually tested for physical interaction or function. This is laborious and has many potential caveats involving unintended changes in expression, folding, stability or cytotopic localization. By contrast, dChIRP stabilizes the endogenous interactions by cross-linking and then dissects the RNA domains involved *in situ*. No mutant constructs are initially required, the number of configurations tested is readily scaled to the number of oligonucleotide pools desired and multiple types of RNA-mediated interactions can be tested simultaneously.

We applied dChIRP to roX1 RNA to dissect the nature of the interactions between roX1, MSL proteins, chromatin and CLAMP, elucidating an integrated interaction model. First, roX1 is topologically organized such that the three U domains form a core, or palm, from which each of the D domains extends independently as fingers (**Fig. 6c**). The U domains exhibit neither chromatin nor MSL binding and are genetically dispensable, implying that these domains and their association are not essential to dosage compensation. The D domains contain double stem-loops and the roX-box motif that we previously found to be the target of MLE and MSL2 interaction[18]. We found that each D domain finger independently binds the MSL proteins and chromatin, for which D1 has the weakest affinity and D3 has the strongest. MLE and the core MSL complex bind to double-stranded regions within roX1's D domains at or near roX-boxes (**Fig. 6d**). CLAMP binds to the MRE motif (GA repeat) at X-linked CESs and is associated with MLE. MLE is tethered to the core MSL complex via roX1 RNA binding. The MSL proteins bind chromatin via MOF, which acetylates H4K16 in adjacent nucleosomes[19]. Finally, these findings allowed us to design hypothesis-driven genetic mutants that proved the D domains are each minimally sufficient for dosage compensation. Despite being approximately one-tenth the size of wild-type roX1, D3 can rescue *roX*-null male flies as efficiently as the wild-type gene, defining—to our knowledge—the smallest RNA unit sufficient for chromosome-wide dosage compensation. Additional D domains may enhance D3 function, as suggested by prior genetic studies[23], and our data also support the idea that roX1 contains multiple D domains that act cooperatively and are functionally equivalent.

In addition to studying individual RNA domains, dChIRP improves the signal-to-noise ratio of sequencing experiments by more than an order of magnitude, enabling unbiased genome-wide mapping of RNA occupancy with greater precision and confidence. We used dChIRP-seq to map the genomic binding sites of roX1, which are nearly identical to roX2- and MSL3-binding sites, providing molecular evidence of redundant function between roX1 and roX2. The signal improvement is most relevant for longer RNAs, for which the use of many oligonucleotides to tile the target RNA increases false positives and background noise while sacrificing true signal by unproductively targeting nonfunctional RNA regions. This strategy is an example of RNA partitioning, wherein functional interactions are partitioned from the nonfunctional; therefore, sequencing depth need not be wasted on nonfunctional elements. Just as genome partitioning technologies such as exome sequencing have revolutionized human genetics, this RNA partitioning technology may catalyze advances in RNA genetics and genomics.

We found that many roX1 and roX2 target sites (including the *roX2* locus) cluster in a dosage compensation territory, extending an idea suggested by previous DNA FISH experiments[28]. Because autosomally integrated roX transgenes can still target the X chromosome and rescue male lethality[18], fly dosage compensation represents a strikingly different strategy of sex-chromosome targeting than that seen in mammals. Recent studies in mammals suggest that Xist targeting depends on the chromosomal location of the *Xist* gene locus[11,29]. One important consequence of CES clustering is an increase in the local density of target sites, which may increase the avidity of the dosage-compensation complex for CESs and distinguish the X chromosome from autosomes. Our findings with the roX RNAs are reminiscent of mammalian lncRNAs such as HOTTIP and some enhancer-like RNAs that connect chromosome conformation to three-dimensionally proximal gene activation[12,30,31]. Nonetheless, the existing Hi-C data is from mixed-sex embryos and reflects a gender-averaged map of the X chromosome. The relationship between chromosome conformation, the dynamics of roX1 versus roX2 spreading, and dosage compensation merits further investigation.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Raw sequencing reads, merged ChIRP lanes and called peaks can be accessed at GEO: GSE53020.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of the Chang and Akhtar labs, P. Sharp and X. Wu (Massachusetts Institute of Technology) for meaningful discussion and E. Larschan (Brown University) for CLAMP antibody. Supported by US National Institutes of Health R01-CA118750 and R01-ES023168 (H.Y.C.), Max Planck Society (A.A.) and Bio-X Fellowship (J.J.Q.). This work was supported by DFG-funded SFB992 and SFB746 and EU-funded EpiGeneSys awarded to A.A. H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute; A.A. is part of the BIOSS excellence initiative.

AUTHOR CONTRIBUTIONS

J.J.Q., I.A.I., P.G., C.C., H.Y.C. and A.A. designed the research. J.J.Q., I.A.I. and P.G. performed the research. K.Q. and J.J.Q. performed bioinformatic analyses. J.J.Q. and H.Y.C. wrote the manuscript. All authors discussed the results and reviewed the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Batista, P.J. & Chang, H.Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298–1307 (2013).
2. Lee, J.T. & Bartolomei, M.S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308–1323 (2013).
3. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
4. Mercer, T.R., Dinger, M.E. & Mattick, J.S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155–159 (2009).
5. Guttman, M. & Rinn, J.L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
6. Ule, J., Jensen, K., Mele, A. & Darnell, R.B. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).
7. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
8. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent non-canonical binding. *Cell* **153**, 654–665 (2013).
9. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveals principles of RNA-chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
10. Simon, M.D. *et al.* The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. USA* **108**, 20497–20502 (2011).
11. Engreitz, J.M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
12. Yang, L. *et al.* lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* **500**, 598–602 (2013).
13. Rossetto, C.C., Tarrant-Elorza, M., Verma, S., Purushothaman, P. & Pari, G.S. Regulation of viral and cellular gene expression by Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA. *J. Virol.* **87**, 5540–5553 (2013).
14. Colak, D. *et al.* Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* **343**, 1002–1005 (2014).
15. Conrad, T. & Akhtar, A. Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nat. Rev. Genet.* **13**, 123–134 (2012).
16. Straub, T., Gilfillan, G., Maier, V.K. & Becker, P.B. The *Drosophila* MSL complex activates the transcription of target genes. *Genes Dev.* **19**, 2284–2288 (2005).
17. Alekseyenko, A.A. *et al.* A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* **134**, 599–609 (2008).
18. Ilik, I.A. *et al.* Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell* **51**, 156–173 (2013).
19. Kind, J. *et al.* Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell* **133**, 813–828 (2008).
20. Soruco, M.M.L. *et al.* The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes Dev.* **27**, 1551–1556 (2013).
21. Meller, V.H. & Rattner, B.P. The roX genes encode redundant Male-Specific-Lethal transcripts required for targeting of the MSL complex. *EMBO J.* **21**, 1084–1091 (2002).
22. Park, S.W. *et al.* An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys19 acetylation on the *Drosophila* X chromosome. *Genetics* **177**, 1429–1437 (2007).
23. Stuckenholz, C., Meller, V.H. & Kuroda, M.I. Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in *Drosophila melanogaster*. *Genetics* **164**, 1003–1014 (2003).
24. Kelley, R.L., Lee, O.K. & Shim, Y.K. Transcription rate of noncoding roX1 RNA controls local spreading of the *Drosophila* MSL chromatin remodeling complex. *Mech. Dev.* **125**, 1009–1019 (2008).
25. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
26. Maenner, S., Müller, M., Fröhlich, J., Langer, D. & Becker, P.B. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol. Cell* **51**, 174–184 (2013).
27. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
28. Grimaud, C. & Becker, P.B. The dosage compensation complex shapes the conformation of the X chromosome in *Drosophila*. *Genes Dev.* **23**, 2490–2495 (2009).
29. Simon, M.D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465–469 (2013).
30. Wang, K.C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
31. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
32. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).

ONLINE METHODS

dChIRP oligonucleotide design. Biotinylated 20-mer antisense oligonucleotides were designed according to ChIRP using Stellaris single-molecule FISH probe designer (http://biosearch.com/) following Chu *et al.*[33]. ChIRP oligonucleotides are listed in **Supplementary Table 1**.

Tissue culture. Clone 8 cells (*Drosophila* Genomics Resource Center, https://dgrc.cgb.indiana.edu/Home) were grown in M3 insect medium (Sigma) supplemented with 5 µg/mL human insulin (Sigma), 1× penicillin/streptomycin (Gibco), 2% heat-inactivated FBS (HyClone) and 2.5% fly extract and maintained at 27 °C. Cultures were split every 5 d to a concentration of 5×10^6 cells/mL. MDA-MB-231 breast cancer cell line overexpressing HOTAIR was cultured according to Gupta *et al.*[34].

Chromatin preparation. Two-step 1% + 3% formaldehyde cross-linking was performed as previously described[10]. Chromatin was prepared and sonicated according to ChIRP[33] with the exception that chromatin was sheared by sonication until the bulk of nucleic acids was between 200–400 bp. MDA-MB-231 HOTAIR cells were cross-linked in 1% glutaraldehyde according to Chu *et al.*[33]. For thermally reverse cross-linked samples, chromatin was heated at 65 °C for 4 h.

ChIRP. ChIRP was performed as previously described[33] using the oligonucleotide pools listed in **Supplementary Table 1**. RNA and DNA were extracted and quantified as described, using RT-qPCR and qPCR primers listed in **Supplementary Table 1**. Western blots were performed following Ilik *et al.*[18] using MLE and MSL3, CLAMP (1:1,000; courtesy of Erica Larschan[20]) and actin (ab1801, Abcam) primary antibodies.

Sequencing and data analysis. High-throughput sequencing libraries were constructed as previously described[9] and sequenced on Genome Analyzer IIx or HiSeq 2000 (Illumina) with read lengths of 36 or 50, respectively.

dChIRP-seq bioinformatics analysis was performed as previously described[9]. dChIRP peaks were defined by MACS[35]; peak summits were identified by ZINBA[36]. We filtered the resulting raw peaks by maximum signal (>10,000), median peak signal (>0) and *P* value (>130), yielding 471 total peaks. For the correlation analysis, the *D. melanogaster* genome (dm3 assembly) was divided into 33,739 5-kb windows. For each sample, we calculated $\log_2$-transformed values and plotted pairwise as scatter plots the normalized number of reads that fall into each window (roX1 D2 vs. D3 dChIRPs; roX1 D3 dChIRP vs. roX2 ChIRP). We performed motif analysis of the peaks using MEME[32]. Hi-C data showing chromosome conformation in mixed male/female embryos was obtained from Sexton *et al.*[27] (GSM849422) at a resolution of 80 kb. The proximities between all 80-kb chromosome bins on X chromosomes with *roX1* and *roX2* loci were defined as the observed counts divided by the expected counts in Hi-C experiment. roX2 ChIRP and roX1 dChIRP occupancy ($\log_2$-transformed number of reads within each 80-kb window) versus *roX2* and *roX1* Hi-C proximities, respectively, were shown, and we calculated Pearson correlations between each pair. Signals within a 400-kb window around the *roX1* and *roX2* loci were masked before calculating the Pearson correlation to avoid bias caused by strong correlations of ChIRP/dChIRP signals with Hi-C signals at these loci. The dosage-compensated gene set was defined by X-linked genes containing a roX1-D3 dChIRP peak, and enrichment of this gene set of roX2 chromosomal proximity was estimated by GSEA analysis[37] (FDR < 0.001). We obtained 15,051 loci with GAGA motif by searching for a perfect match of GAGAGA sequences on the X chromosome. These loci were then ranked by their occupancy of CLAMP signals. The top 2,000 loci were defined as GAGA motif with CLAMP signal, and the bottom 1,000 loci were defined as GAGA motif without CLAMP signal. Loci with CLAMP signal were further segregated into two groups: those with or without roX1 D3 dChIRP peaks. Cumulative frequencies of *roX2* proximity for these three subgroups (GAGA motif without CLAMP signal, GAGA motif with CLAMP signal and without roX1 peak, and GAGA motif with CLAMP signal and roX1 peak) were obtained, and significance of the difference was estimated using a Kolmogorov-Smirnov test.

Genetic mutants. Fly work has been done essentially as described in Ilik *et al.*[18]. Briefly, all roX1 constructs were cloned into pUASattB vector, and we created transgenic flies using phiC31 integrase–mediated germ-line transformation as previously described[38]. To score male viability, $roX1^{SMC17A}$, $roX2^{\Delta}$; *tubGal4/TM6Tb* or *daGAL4* virgin females were crossed to *UAS-roX1*U1, D1, U2, D2, U3, D3, D1-D3* or *D2-D3* males. We counted male and female adult flies from at least 3 independent crosses daily for a period of 10 d from the start of eclosion, without blinding. The total number of non-Tubby males was divided by the total number of non-Tubby females that eclosed during the 10-d period, which was used as an internal control for 100% viability.

For gene-expression analysis, wandering 3rd instar larvae of the correct genotype were homogenized in TRIzol (Qiagen), and we extracted total RNA from these lysates using the Direct-zol kit (Zymo) as per manufacturer's instructions. Total RNA was then reverse transcribed using Superscript III (Life Tech.) and random hexamers, after which transcript abundances were calculated using qPCR and the $2^{-\Delta\Delta Ct}$ method.

33. Chu, C., Quinn, J.J. & Chang, H.Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* **61**, e3912 (2012).
34. Gupta, R.A. *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
35. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
36. Rashid, N.U., Giresi, P.G., Ibrahim, J.G., Sun, W. & Lieb, J.D. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. *Genome Biol.* **12**, R67 (2011).
37. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
38. Groth, A.C., Fish, M., Nusse, R. & Calos, M.P. Construction of transgenic *Drosophila* by using the site-specific integrase from phage PhiC31. *Genetics* **166**, 1775–1782 (2004).