Supporting Information

Simon et al. 10.1073/pnas.1113536108

SI Materials and Methods.

Preparation of Cross-linked Nuclei. *Drosophila* S2 cells expressing male-specific lethal complex (MSL3-TAP, where TAP is a tandem affinity purification epitope tag) (S1) were grown in shaker flasks in serum-free CCM3 media (HyClone). Cells (approximately 10^{10} cells) were harvested by centrifugation (500 × *g*, 15 min, 4 °C), rinsed once with PBS, resuspended to 200 mL with PBS, and cross-linked [1% formaldehyde, 10 min, room temperature (RT)], rinsed three times with PBS, and stored at −80 °C or carried forward directly to prepare nuclei. Nuclei were enriched essentially as described (S2). Briefly, cells (approximately 10^9 cells) were washed with PBS and nuclei were enriched by disrupting cells with a Dounce homogenizer in sucrose buffer (0.3 M sucrose, 1% Triton X-100, 10 mM Hepes pH 7.5, 100 mM KOAc, 0.1 mM EGTA, 0.5 mM spermidine, 0.15 mM spermine, 1× Roche protease inhibitor tablet, 1 mM DTT, 10 u/mL SUPERasIN), diluted with an equal volume of glycerol buffer (25% glycerol, 10 mM Hepes pH 7.5, 100 mM KOAc, 1 mM EDTA, 0.1 mM EGTA, 0.5 mM spermidine, 0.15 mM spermine, 1× Roche protease inhibitor tablet, 1 mM DTT, 10 u/mL SUPERasIN), and layered on top of glycerol buffer (4 mL). The cross-linked nuclei were collected by centrifugation (1,000 × *g*, 15 min, 4 °C). This protocol was also used to prepare cross-linked nuclei from HeLa cells, except by using tenfold fewer cells for the same volumes (i.e., preparing 10^8 nuclei).

Chromatin Extract for RNase-H Mapping. Chromatin extract for RNase-H mapping was prepared by rinsing nuclei with shearing buffer (50 mM Hepes pH 7.5, 75 mM NaCl, 0.1 mM EGTA, 0.5% *N*-lauroylsarcosine, 0.1% sodium deoxycholate, 20 u/mL SUPERasIN, 5 mM DTT) and resuspending into 4 mL buffer/10^9 nuclei for S2 cells or 4 mL buffer/10^8 HeLa nuclei. This material was sheered using a Covaris S2 instrument (30-min program, 10% duty cycle, intensity of 5, 4 °C) and then cleared by centrifugation (16,100 × *g*, 10 min, RT). The cleared extract was divided into aliquots, flash frozen with N_2, and stored at −80 °C or used directly for RNase-H mapping reactions.

RNase-H Mapping. Cross-linked extract was divided into individual 10 μL reactions supplemented with $MgCl_2$ (3 mM final), DTT (10 mM final), SUPERasIN (10 u), and RNase H (5U). To each reaction a different oligonucleotide (100 pmol) was added and the reaction was allowed to proceed for 30 min at 30 °C. The DNA was hydrolyzed by adding RQ1 DNase (1 μL, Promega) and $CaCl_2$ (500 μM final) and incubating for an additional 10 min at 30 °C. The reaction was stopped by adding quenching buffer (2 μL of 125 mM EDTA, 250 mM Tris•HCl pH 7.2, 0.5 mg/mL Proteinase K, 5% SDS), incubated for 1 h at 55 °C, and then 30 min at 65 °C. RNA was recovered using a PureLink RNA purification kit (Invitrogen) and analyzed by qPCR for RNase H sensitivity.

$$\text{RNase-H sensitivity} = \left(\frac{\text{efficiency}_{\text{TARGET PRIMERS}}^{C_{T,\text{olgio}}-C_{T,\text{no oligo}}}}{\text{efficiency}_{\text{CONTROL PRIMERS}}^{C_{T,\text{olgio}-}-C_{T,\text{no oligo}}}}\right).$$

Capture Oligonucleotides. Peaks from RNase-H mapping were identified and used to design 24–25 nt C-oligos using BLAST to avoid complementarity to other RNA sequences. The resulting C-oligos were synthesized on an Expidite DNA synthesizer with 3′-desthiobiotin (DSB-TEG) and four oligoethyleneglycol spacers. The oligonucleotides were synthesized 4,4'-dimethoxytrityl-on for purification using PolyPak II cartridges (Glen Research). C-oligos used for RNase-H–eluted capture hybridization analysis of RNA targets (CHART) were 3′-modified by a single oligoethyleneglycol spacer and biotin-TEG.

Preparation of Chromatin Extract for CHART. Rinsed, cross-linked nuclei were further cross-linked with formaldehyde (10^9 S2 nuclei in 50 mL of PBS supplemented with 3% formaldehyde, 30 min, RT). The nuclei were rinsed with PBS and then resuspended in WB100 (100 mM NaCl, 10 mM Hepes pH 7.5, 2 mM EDTA, 1 mM EGTA, 0.2% SDS, 0.1% *N*-lauroylsarcosine). This material was sheared using a Bransen sonicator to 2–3 kb average DNA fragment sizes and then cleared by centrifugation (16,100 × *g*, 10 min, RT). The cleared extract was divided into aliquots, flash frozen with N_2, and stored at −80 °C, or used directly for CHART. HeLa and MCF7 (a breast adenocarcinoma cell line) extracts were made following the same protocol, except using 10^8 nuclei and shearing with a Covaris S2 instrument (15-min program, 10% duty cycle, intensity of 5, 4 °C).

Nucleic Acid Analysis. CHART-enriched samples were deproteinized with proteinase K and cross-links were reversed with Proteinase K (1 mg/mL), SDS (0.5%), and Tris pH 7.4 (100 mM) at 55 °C for 1 h and then 65 °C for 30 min.

qPCR Analysis. Nucleic acids were purified with QIAGEN columns according to the manufacturer's directions. CHART-enriched material was assayed in comparison with supernatant from a no-oligo control (to control for handling loss, hereafter referred to as input). The yields are reported relative to input signal without further normalization:

$$\text{Yield} = \left(\frac{\text{Input dilution factor}}{\text{efficiency}_{\text{PRIMERS}}^{C_{T,\text{CHART}}-C_{T,\text{INPUT}}}}\right).$$

In cases where the C_T was not reached within 40 cycles, a value of 40 was assigned for purposes of analysis, thereby conservatively underestimating enrichment.

Protein Analysis. CHART samples were treated with SDS (1.0%), Tris pH 8.8 (100 mM), and β-mercaptoethanol (1 M) for 1 h at 95 °C. These samples were resolved by SDS PAGE, transferred to PVDF, and analyzed using peroxidase antiperoxidase (to detect MSL3-TAP, Sigma), anti-DSP1 antisera (S3), anti-PSPC1 antisera (sc-84577), anti-p54/nrb (sc-67016), or anti-histone H3 (ab1791).

Sequence Analysis. DNA fragments were isolated, further sheared (S4), sequenced (Illumina GAIIx or HiSeq) and mapped to the *Drosophila* genome (dm3, Bowtie aligner, ref. S5), recording positions of uniquely mappable reads. The enrichment of the biotin-CHART signal was determined relative to the sense-oligo controls and the RNase-H–eluted CHART signal was determined relative to input. Conservative enrichment profiles were determined using the SPP (Solexa Processing Pipeline) package (S6) (lower bound of enrichment was determined based on a Poisson model, with a confidence interval of $p = 0.001$). Positions of top CHART sites were determined as peaks of the conservative enrichment profiles (with minimum separation of 3 kb). The top peaks were selected for subsequent analysis based on 90% specificity to chrX (Fig. S4 *B* and *F*). To determine sequence motifs corresponding to the top CHART peaks (Fig. 4*D*), 200-bp sequences flanking the peaks were analyzed using MEME (Multiple EM for Motif Elicitation) (S7).

1. Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI (2006) High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev* 20:848–857.
2. Dennis JH, et al. (2007) Independent and complementary methods for large-scale structural analysis of mammalian chromatin. *Genome Res* 17:928–939.
3. Mosrin-Huaman C, Canaple L, Locker D, Decoville M (1998) DSP1 gene of Drosophila melanogaster encodes an HMG-domain protein that plays multiple roles in development. *Dev Genet* 23:324–334.
4. Lieberman-Aiden E, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–293.
5. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
6. Kharchenko PV, Tolstorukov MY, Park PJ (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26:1351–1359.
7. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34:W369–373.

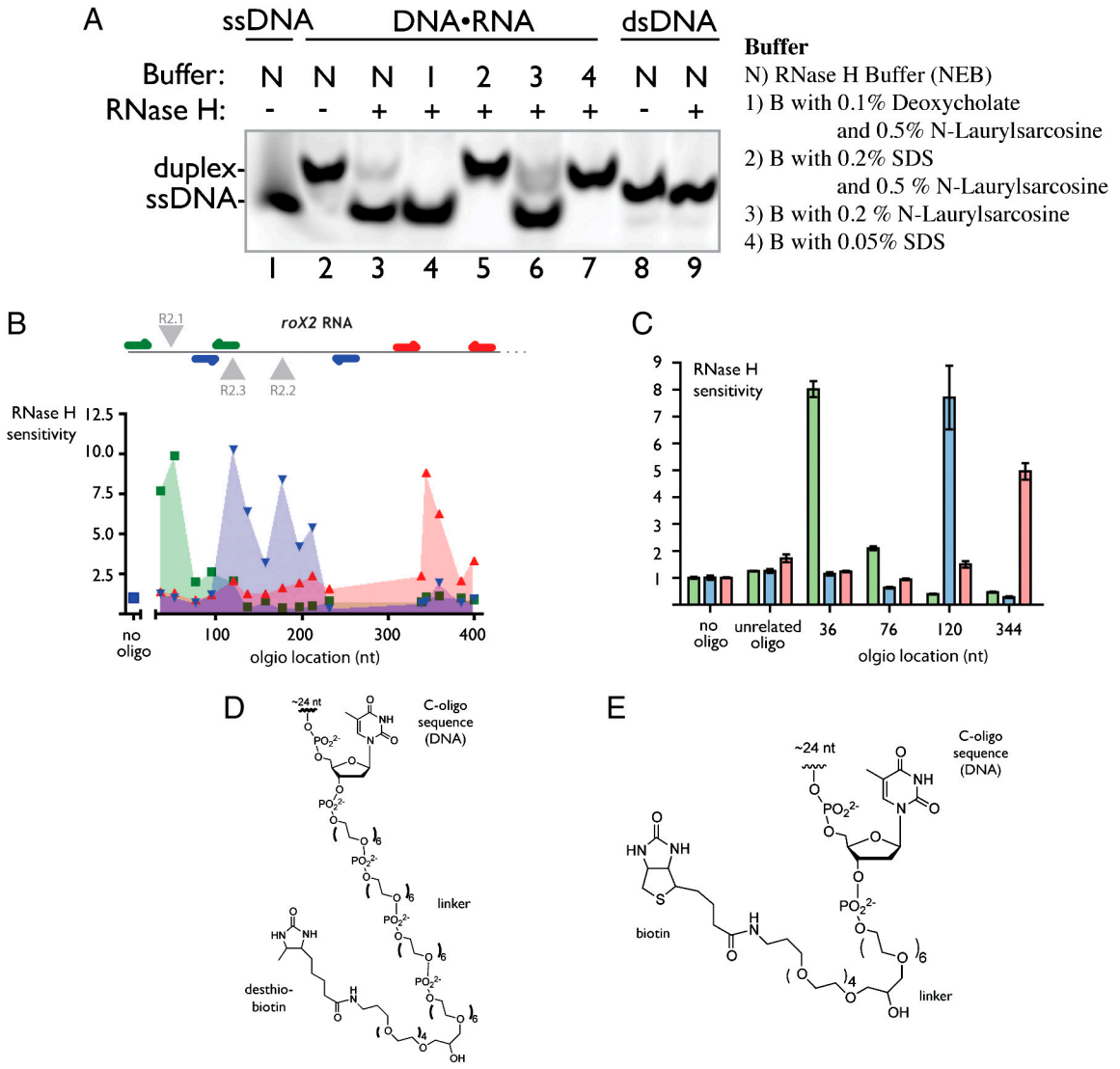

Fig. S1. The development of C-oligos for CHART. (*A*) Analysis of RNase-H activity using synthetic nucleotides analyzed by native PAGE. A Cy5-fluorescently labeled DNA oligonucleotide was hybridized to either a complementary RNA (lanes 2–7) or DNA (lanes 8 and 9) or run without hybridization as a control (lane 1). These oligonucleotides were incubated with RNase H (5U) under the indicated buffer conditions. Buffer B is 50 mM Hepes pH 7.5, 75 mM KCl, 3 mM $MgCl_2$, 0.1 mM EGTA, 20 u/mL SUPERasIN, 5 mM DTT, 7.5% glycerol to which the indicated detergents were added. The reaction was incubated at 30 °C for 30 min and quenched with EDTA and Proteinase K. The products of the reaction were resolved by 10% native polyacrylamide gel and the gel scanned for Cy5-fluorescence on a Typhoon imager. From this analysis, buffer 1 conditions were chosen because these conditions were found to be compatible with RNase-H activity. (*B*) Top, the 5′ region of *roX2* RNA examined using RNase-H mapping. Three primer sets (indicated here in green, blue, and red) were used to assay cleavage by RT-qPCR. Below, each point on this plot depicts the RNase-H sensitivity induced by a single oligonucleotide, and cleavage measured using the primer sets shown above. RNase-H sensitivity represents the ratio of cleaved to uncleaved RNA (e.g., a value of 9 corresponds to 90% cleavage). Note that the sites of high sensitivity are only observed with the appropriate primer sets. The targets of the C-oligos based on this mapping are indicated with gray arrowheads. (*C*) The same as *B* but repeated in three independent experiments for each oligo shown. Error bars represent ±SEM. (*D*) The design of C-oligos used in biotin-eluted CHART and (*E*) the design of C-oligos used for RNase-H–eluted CHART.

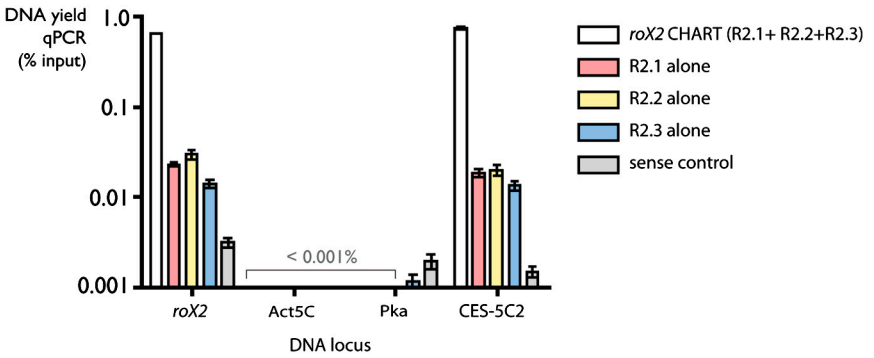

Fig. S2. The C-oligos used in *roX2* CHART each independently enrich *roX2* binding sites and act synergistically. *roX2* CHART was performed either with the standard mixture of three C-oligo nucleotides (white) or with the individual C-oligos (red, yellow, and blue). As a control, a mixture of three sense oligos corresponding to each of the *roX2* C-oligo cocktail was used (gray). The results are plotted on a $\log_{10}$ scale relative to input. The individual C-oligos each have yields greater than threefold lower (40-, 37-, and 56-fold lower, respectively) than the combined cocktail, demonstrating that the C-oligos act synergistically. Where indicated, the two negative-control loci (Pka and Act5C) amplified but did not achieve 0.001% yield (corresponding to a qPCR C_T value of >35 with input C_T values of approximately 20 for all four loci). Error bars represent ±SEM of three qPCR replicates.

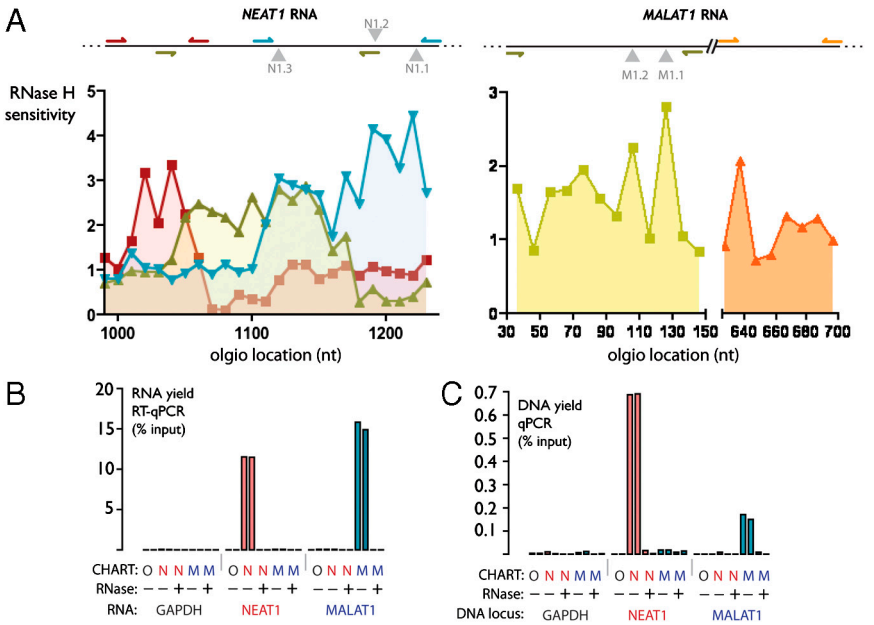

Fig. S3. (*A*) Using HeLa cell extract, peaks of RNase-H sensitivity were used to design C-oligos to a mammalian lncRNA. Similar to analysis depicted in Fig. 1*B*, RNase-H mapping of a region of *NEAT1* (980–1240 nt of NR_028272.1) and *MALAT1* revealed sites of high RNase-H sensitivity that were used to design C-oligos (sites indicated by gray arrowheads). (*C*) *NEAT1* CHART, but not *MALAT1* CHART, enriches *NEAT1* RNA in MCF7 cells, similar to analysis depicted in Fig. 3*A* demonstrating enrichment of specific RNAs by RT-qPCR with different CHART experiments (O, *Mock*; N, *NEAT1*; M, *MALAT1*) and RNase refers to pretreatment of the extract with RNase prior to CHART analysis. (*C*) *NEAT1* CHART, but not *MALAT1* CHART, enriches the *NEAT1* endogenous locus in MCF7 cells, similar to analysis depicted in Fig. 3*B* demonstrating enrichment of specific DNA loci by qPCR with different CHART experiments as in *B*.

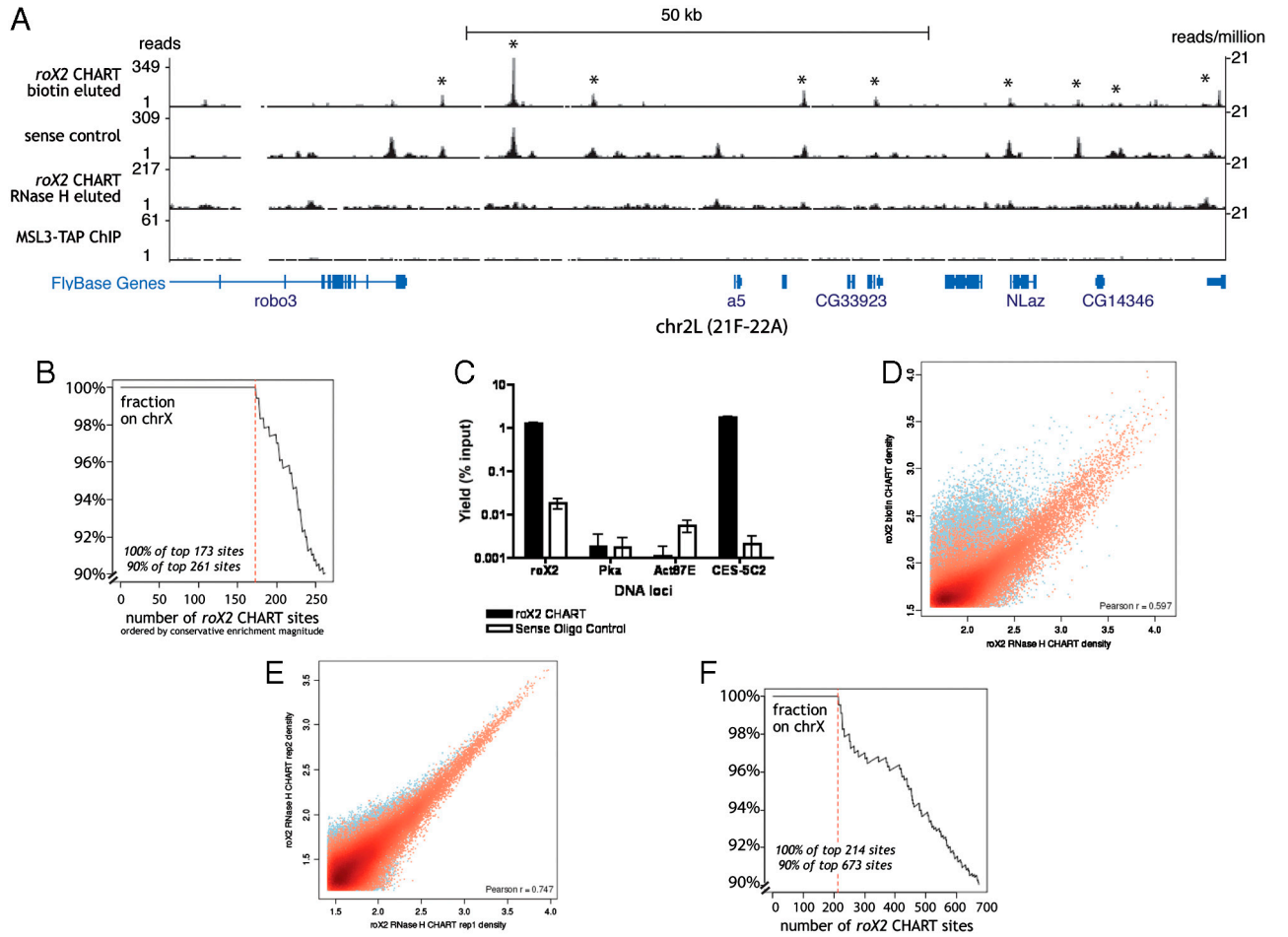

Fig. S4. (*A*) Whereas the top *roX2* CHART peaks are found on chrX, some of the lower-significance peaks from biotin-eluted *roX2* CHART-seq correspond to sites that are caused by direct binding of the C-oligos to DNA. Comparison of normalized sequencing reads across a region of chr2L demonstrating several *roX2* CHART peaks (marked by asterisks) that correspond to peaks also observed in the sense control, suggesting they are caused by direct binding of the C-oligos to DNA and are not *roX2* binding sites. Supporting this conclusion, these peaks were greatly reduced in an RNase-H–eluted *roX2* CHART and did not correspond to peaks in the MSL3-TAP ChIP experiment (1). (*B*) Peaks from biotin-eluted *roX2* CHART data were ordered by the enrichment magnitude relative to the sense-oligo control and plotted for their cumulative fraction found on the chrX. The red dashed line shows the cutoff at 173 peaks where 100% of peaks are found on chrX. (*C*) Analysis of DNA enrichment by RNase-H–eluted *roX2* CHART based on qPCR and similar to Fig. 2*B*. Results are plotted on a $\log_{10}$ scale. Error bars are ±SEM from three qPCR replicates. (*D*) Genome-wide correlation of biotin-eluted and RNase-H–eluted CHART read density (200-bp bandwidth) plotted on a $\log_{10}$ scale. chrX peaks are plotted in red, the autosomal peaks in blue. The correlation is for all data (red and blue together). (*E*) Similar to *D* but demonstrating the correlation between two RNase-H–eluted CHART replicates. (*F*) Similar to *B*, peaks from RNase-H–eluted *roX2* CHART data were ordered by the enrichment magnitude relative to the input and plotted for their cumulative fraction found on the chrX. The red dashed line shows the cutoff at 214 peaks where 100% of peaks are found on chrX.

1 Alekseyenko AA, et al. (2008) A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* 134:599–609.

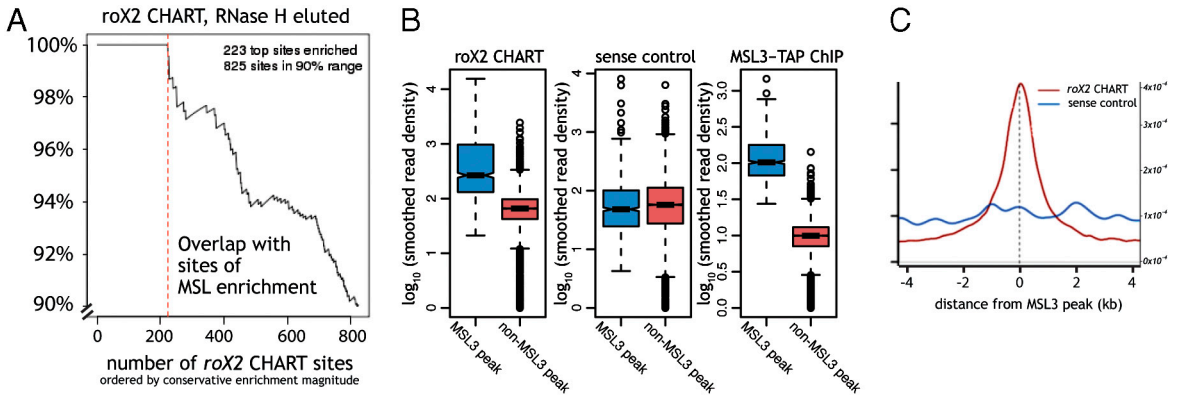

Fig. S5. (*A*) The top sites of *roX2* CHART enrichment are all sites of MSL enrichment. Similar to the analysis in Fig. 4*F*, *roX2* CHART sites were ordered by significance. The plot shows the cumulative fraction of the top peaks that have at least twofold enrichment of MSL3-TAP ChIP (1). The red dashed line represents the cutoff at 223 peaks above which 100% of top peaks are found to have MSL enrichment. (*B*) Box plot comparing the distribution of read densities for each dataset for either the top MSL3-enriched sites (blue, based on ref. 1; red, 1,000 non-MSL3 enriched sites chosen at random). (*C*) Average *roX2* CHART data (red, biotin-eluted) and sense-oligo CHART data (blue) aligned around sites of MSL3 enrichment (top 625 peaks used based on Fig. S4*B*).

1 Alekseyenko AA, et al. (2008) A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* 134:599–609.

Table S1. Watson-Crick complementarity between *roX2* RNA and genomic sequence of *roX2* CHART targets

	5-mers	7-mers	10-mers
Full *roX2*-RA	190.0/196.9	30.3/34.4	0.48/0.87
roX2-RA 72nt loop	38.8/43.1	4.21/4.90	0.169/0.160

To test whether *roX2* CHART targets show prevalence of sequences complementary to *roX2* RNA we have extracted all *n*-mers from *roX2* RNA and compared the number of direct and reverse-complement occurrences between a set of *roX2* CHART target regions and a set of randomly selected control regions. Each cell in the table shows the average number of *n*-mer matches for *roX2* CHART targets, followed by the average number of matches observed in control sequences (separated by */*). Around each *roX2* CHART target site, 300-bp regions were analyzed. A tenfold set of control regions was chosen randomly from the X chromosome. The results suggest that the *roX2* CHART target regions do not show increased frequency of *n*-mers complementary to either full *roX2* RNA product (first row) or the 72-nt step loop (second row) critical for the *roX2* function (1). In fact the overall match frequency appears to be slightly below that of randomly selected controls. Comparison with control regions selected from entire genome yields analogous results.

1 Park SW, et al. (2007) An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the Drosophila X chromosome. *Genetics* 177:1429–1437.

Table S2. Capture oligonucleotides used in this study

R2.1:	TAA	CAC	CAA	TTT	ACC	CTT	TCG	ATG	LLL	L-DSB
R2.2:	TCT	CAC	TGT	CCG	TAA	GAC	AAT	TCA	ALL	LL-DSB
R2.3:	CTC	TTG	CTT	GAT	TTT	GCT	TCG	GAG	ALL	LL-DSB
CNTL:	TAA	TGG	CTC	CTA	CAT	ACT	ACA	TCT	LLL	L-DSB
R2.so1:	CAT	CGA	AAG	GGT	AAA	TTG	GTG	TTA	LLL	L-DSB
R2.so2:	TTG	AAT	TGT	CTT	ACG	GAC	AGT	GAG	ALL	LL-DSB
R2.so3:	TCT	CCG	AAG	CAA	AAT	CAA	GCA	AGA	GLL	LL-DSB
N1.1:	GCT	AGG	ACT	CAC	ACT	GGC	CAG	GGA	CLL	LL-DSB
N1.2:	TCC	ATG	TCT	CCC	GGT	TCC	ATC	TGC	TLL	LL-DSB
N1.3:	CAT	GAA	GCA	TTT	TTG	TAA	CTT	TCA	GLL	LL-DSB
M1.1:	GGA	CTC	TGG	GAA	ACC	TGG	GCT	CCC	GLL	LL-DSB
M1.2:	GAG	GCG	TCA	GAG	GGG	ACC	TGC	CTT	CLL	LL-DSB

All sequences are listed 5′ to 3′. "L" stands for the spacer C18 residue. "DSB" stands for desthiobiotin-TEG; "so" stands for sense oligo.

Table S3. Primer sequences used in this study

RNASE H MAPPING	
R2.GREEN.F	AGCTCGGATGGCCATCGA
R2.GREEN.R	CGTTACTCTTGCTTGATTTTGC
R2.BLUE.F	CATTGATAATCGTTCGAAACGTTC
R2.BLUE.R	GACAAGCGCGTCAACC
R2.RED.F	TGTCTTGGAACGCAACATT
R2.RED.R	GCATATATATTTGCTTAATTTGCAACAT
N1.RED.F	GTGGGCCTGCAGCCATCCAG
N1.RED.R	GCGGGCTCTCTCCTCCAGGG
N1.YELLOW.F	GGGGCGGATCGGTGTTGCTT
N1.YELLOW.R	CCCGGTTCCATCTGCTCGCC
N1.BLUE.F	AGCCCGGGACAGTAAGCCGA
N1.BLUE.R	TCCCCACCCTCTCTGCAGGC
QPCR/RT-QPCR	
ACT5C.F	CAGCTCCTCGTTGGAGAAGT
ACT5C.R	AAGCCTCCATTCCCAAGAAC
CES-11B16.F	TCGCCGAACCCCAACACCAA
CES-11B16.R	GCGCGGTGTTCATCGGCCAT
CES-3A1.F	GTTGGCGGAGTGCTTGCCCT
CES-3A1.R	CGGACGCAGAAGTCCTCGCC
CES-3F3.F	CCGCTTGCGATGCAAACGCC
CES-3F3.R	ATGTGGCGGTACGCGGATGC
CES-5C2.F	AGAGCGAGATAGTTGGAAG
CES-5C2.R	TCAAGTTGAGATCGCTTCG
CG14438.F	GACCGGATTACTGGGTTTCGC
CG14438.R	CATATGGCCGATCAAGTGCTC
PEAK-5A1.F	AACGGCGTAGTGGGAGGCCA
PEAK-5A1.R	CCGCCCACCACAGCTGTCTG
PKA.F	CAATCAGCAGATTCTCCGGCT
PKA.R	AGCCGCACTCGCGCTTCTAC
ROX2.F	AGCTCGGATGGCCATCGA
ROX2.R	CGTTACTCTTGCTTGATTTTGC
RPL17.F	TCAGTAGTTGTCACCGGCTTG
RPL17.R	CCCGCCAAGAAGAAGCTCTC
GAPDH.F	AAGGTGAAGGTCGGAGTCAA
GAPDH.R	GGAAGATGGTGATGGGATTT
MALAT1.F	CGCAACTGGCCTCTCCTGCC
MALAT1.R	CTCGTCGCTGCGTCCCAAGG
NEAT1.F	GGGGCGGATCGGTGTTGCTT
NEAT1.R	CCCGGTTCCATCTGCTCGCC

All sequences are listed 5′ to 3′.