A mutually exclusive stem–loop arrangement in roX2 RNA is essential for X-chromosome regulation in *Drosophila*

Ibrahim Avsar Ilik,[1] Daniel Maticzka,[2] Plamen Georgiev,[1] Noel Marie Gutierrez,[1] Rolf Backofen,[2,3] and Asifa Akhtar[1]

[1]Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg im Breisgau, Germany; [2]Bioinformatics Group, Department of Computer Science, University of Freiburg, 79110 Freiburg, Germany; [3]Centre for Biological Signalling Studies (BIOSS), University of Freiburg, 79104 Freiburg, Germany

The X chromosome provides an ideal model system to study the contribution of RNA–protein interactions in epigenetic regulation. In male flies, roX long noncoding RNAs (lncRNAs) harbor several redundant domains to interact with the ubiquitin ligase male-specific lethal 2 (MSL2) and the RNA helicase Maleless (MLE) for X-chromosomal regulation. However, how these interactions provide the mechanics of spreading remains unknown. By using the uvCLAP (UV cross-linking and affinity purification) methodology, which provides unprecedented information about RNA secondary structures in vivo, we identified the minimal functional unit of roX2 RNA. By using wild-type and various MLE mutant derivatives, including a catalytically inactive MLE derivative, MLE^{GET}, we show that the minimal roX RNA contains two mutually exclusive stem–loops that exist in a peculiar structural arrangement: When one stem–loop is unwound by MLE, an alternate structure can form, likely trapping MLE in this perpetually structured region. We show that this functional unit is necessary for dosage compensation, as mutations that disrupt this formation lead to male lethality. Thus, we propose that roX2 lncRNA contains an MLE-dependent affinity switch to enable reversible interactions of the MSL complex to allow dosage compensation of the X chromosome.

[*Keywords*: dosage compensation; H4K16ac; MLE; X chromosome; acetylation; roX RNA]

Supplemental material is available for this article.

Received July 11, 2017; revised version accepted September 22, 2017.

RNA helicases are a large family of proteins that have important roles in many aspects of RNA biology. Although the name "helicase" suggests that the main function of these proteins is to destabilize and remove RNA secondary structures, there are several RNA helicases, especially of the DEAD-box subfamily, that use their helicase domains to simply interact with RNA (such as eIF4A3) and not for remodeling. On the other hand, a closely related RNA helicase, eIF4A1, can unwind RNA structures but only when they are relatively weak and short (Rogers et al. 2001). At the other extreme, several helicases are proposed to act as annealers that facilitate the formation of structured RNA rather than the other way around (Jankowsky 2011).

Thus, although the helicase family is quite large, specific functions assigned to helicases that involve the actual removal of secondary structures in vivo are rather sparse. Maleless (MLE), known primarily for its role in *Drosophila* dosage compensation, is an RNA helicase of the DExH family, which is composed of relatively large multi-domain helicases that (unlike the DEAD-box family) are thought to work as processive RNA helicases. MLE's human ortholog, RNA helicase A (RHA, also known as DHX9), has been shown to remove very large secondary structures formed by inverted *Alu* elements in the transcribed part of the human genome (Aktaş et al. 2017). DHX9 was also shown to be able to unwind long stretches of dsRNA without apparent dissociation from its substrate in vitro (Koh et al. 2014). In *Drosophila*, genetic and biochemical evidence suggests that MLE incorporates roX long noncoding RNAs (lncRNAs) into the male-specific lethal (MSL) complex that is composed also of MSL1, MSL2, MSL3, and MOF (Ilik and Akhtar 2009; Ilik et al. 2013; Maenner et al. 2013). However, how MLE carries out this function and why an RNA helicase is even required to incorporate these lncRNAs into the MSL complex are not known.lncRNAs have been proposed to have many functions, ranging from the classical assembly platforms or scaffolds that bring together various enzymatic activities for regulation of transcription to decoys that work by sequestering protein factors in a

Corresponding author: akhtar@ie-freiburg.mpg.de

Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.304600.117. Freely available online through the *Genes & Development* Open Access option.

regulated way, while a few possess enzymatic activities themselves, such as RNaseP and the ribosome (Steitz and Moore 2003). Understanding how these RNAs contribute to cellular complexity is key to elucidating how ~20,000 protein-coding genes give rise to an organism as complex as a human being.

We showed previously that MLE has an unusually strong preference for two lncRNAs in *Drosophila* cells: roX1 and roX2 (Ilik et al. 2013). Binding events on these two RNAs, as determined by iCLIP (individual nucleotide-resolution cross-linking and immunoprecipitation) (König et al. 2010), comprise ~80% of all binding events. To our knowledge, this is a unique situation, as protein–RNA cross-linking studies generally report thousands of targets without any particular target being enriched to such an extent (Yang et al. 2015). Since MLE is a modular protein with distinct RNA-binding domains (RBDs) and helicase domains, we decided to explore how individual domains of MLE affect its unique interaction with roX RNAs. Importantly, to achieve this, we developed an improved method (UV cross-linking and affinity purification [uvCLAP]) that can be used to identify transcriptome-wide targets of RNA-binding proteins in vivo.

In this study, we show that roX2 lncRNA contains a hitherto unknown secondary structure that is partially buried in a previously known stem–loop (R2H5) and is only revealed when MLE unwinds R2H5 through its helicase activity, as we show that a helicase-dead mutant of MLE, MLE^{GET}, shows reduced binding to this newly identified structure. We propose that this peculiar arrangement of two mutually exclusive stem–loops creates an affinity switch for MSL2 that is continuously turned on and off through the RNA helicase activity of MLE for the purpose of modulating the interaction of the MSL complex with the chromatin, leading to the spreading of the MSL complex over the entire X chromosome in male flies.

Results

uvCLAP provides an efficient method to capture RNA–protein interactions in vivo

Classical CLIP analysis—or its variants such as iCLIP (Zarnack et al. 2013), PAR-CLIP (photoactivatable ribonucleoside-enhanced CLIP) (Hafner et al. 2010), and CRAC (cross-linking and analysis of cDNA by high-throughput sequencing) (Granneman et al. 2009)—can be used to determine binding sites of RNA-binding proteins in living cells with very high resolution, up to a single nucleotide. However, these protocols are labor-intensive, contain multiple polyacrylamide gel electrophoresis steps and radioactive labeling of cross-linked RNA, and generally take more than a week to process a small number of samples. We reasoned that in order to investigate multiple point mutants of the same protein in a quantitative manner using biological replicates, we needed to improve on existing methods. To overcome these hurdles, we developed the uvCLAP method (Fig. 1A), which does not use radioactive labeling and SDS-PAGE-mediated separation of RNA–protein complexes, thus greatly speeding up the protocol and allowing us to process many samples simultaneously. This approach also has the potential to provide more quantitative data when comparing binding profiles of point mutants with each other in comparison with other methods.

In order to overcome the frequent problems associated with the general lack of specific antibodies and be able to compare the binding profile of a protein with its mutant derivatives, we epitope-tagged our transgenes with the HBH tag (Tagwerker et al. 2006) and generated *Drosophila* stable cell lines under the control of a copper-inducible promoter (Fig. 1D). This tag combination allows very stringent washes (up to 1% SDS) to reduce and remove nonspecific interactions and therefore improves the signal to noise ratio. Briefly, cells were cross-linked with UV-C light, after which a lysate was prepared, and the samples were subjected to a quick tandem affinity purification scheme that uses buffer conditions that are more stringent than those reported to eliminate noncovalent protein–RNA interactions (0.5 M LiCl with 0.5% LiDS, as reported in Castello et al. [2012], and 0.5 M LiCl with 1% SDS). After trimming and repairing the ends of cross-linked RNA, a pair of 3′ adapters, one for each replicate, was ligated. These adapters contained random and semirandom nucleotides that reduce ligation bias (Sorefan et al. 2012) and also allowed us to mix biological replicates immediately after ligation, which could then be separated in silico (Fig. 1C). The bound RNA adapter adducts were recovered directly from beads with proteinase K digestion and reverse-transcribed with specific primers that contained unique barcodes. Following reverse transcription, all samples were mixed, and the mixed cDNA was separated on a 6 M urea-PAGE (see the Materials and Methods). This step removed excess oligonucleotides (3′ adapter and RT primer) that can degrade the quality of the final data. Finally, the cDNA was circularized, linearized, and subjected to PCR to generate the sequencing library. Beginning from cells on plates, this procedure took 4 d to complete (Fig. 1A).

We developed uvCLAP to study RNA-binding properties of MLE in vivo. For this purpose, we first stably expressed wild-type MLE (MLE^{wt}) and several point mutants (Fig. 1B–D) in S2 cells and followed the uvCLAP protocol (Fig. 1A). Comparison with the previous iCLIP method revealed two interesting observations: First, the top target of MLE^{wt} protein in S2 cells is roX2 (Fig. 1E; Supplemental Fig. S1A,B), which is consistent with our previous report and shows that the new method is indeed comparable with iCLIP (Ilik et al. 2013). Second, in the relative absence of roX1 RNA in S2 cells (Johansson et al. 2011), we also observed several mRNAs that were bound by MLE at their introns and some that were bound at their 3′ untranslated regions (UTRs; 145 peaks are located on introns of 58 genes, and 72 peaks are located on 3′ UTRs of 13 genes). These observations suggest that MLE can show a dynamic target range in different cell lines in addition to the dominating roX RNAs.

Next, to study the relative contribution of the various MLE domains, we first analyzed point mutants of MLE that target the dsRBDs dsRBD1, dsRBD2, or both (MLE^{K4E} [MLE^{K}], $MLE^{H196E\ R198E}$ [MLE^{HR}], and MLE $^{K4E\ H196E\ R198E}$

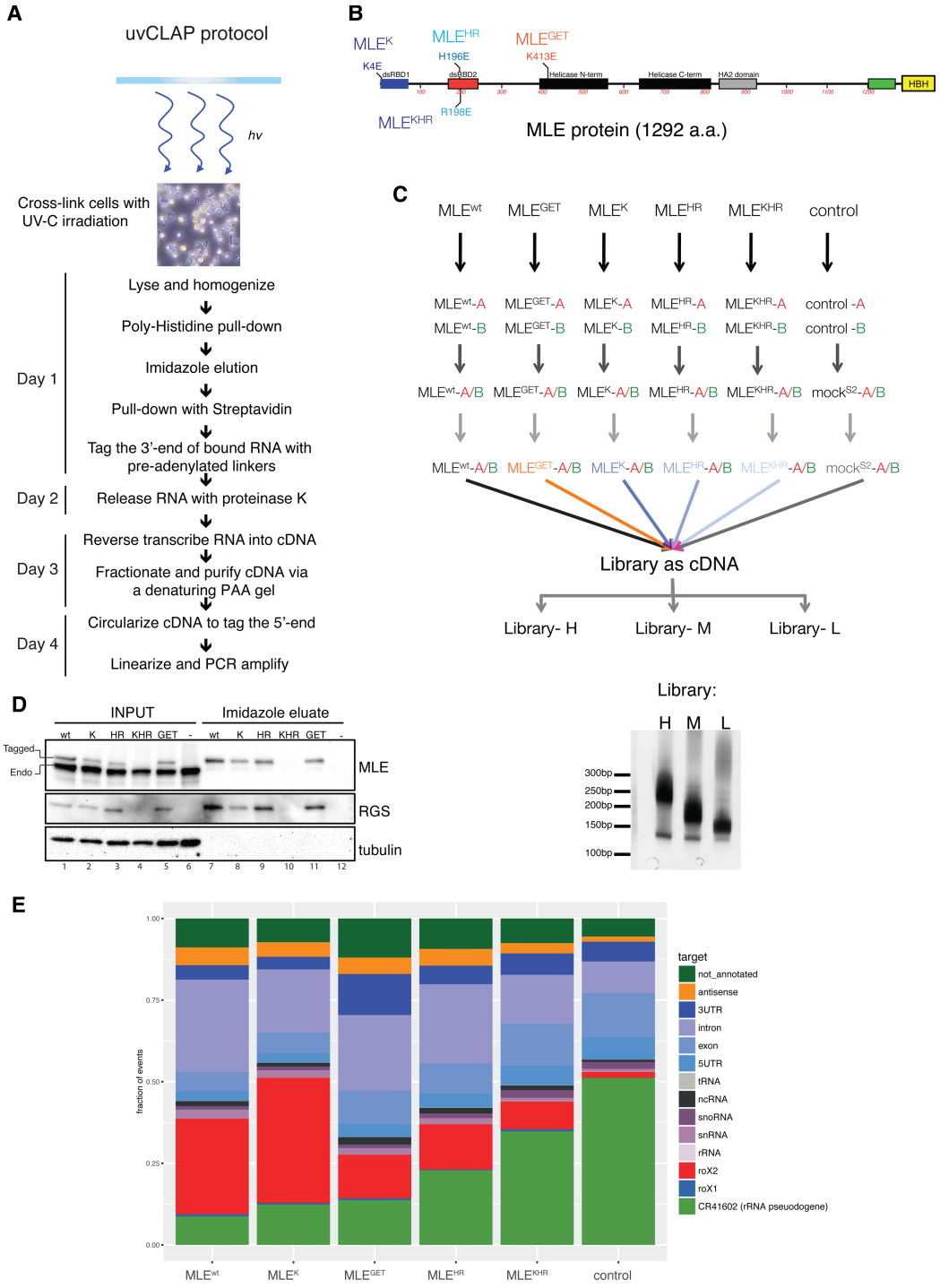

Figure 1. Experimental setup for uvCLAP methodology and analysis of MLE derivatives. (*A*) Schematic representation of the uvCLAP method. S2 cells stably expressing the transgene of interest were cross-linked with UV-C light. After lysis, the target protein was purified using a tandem affinity purification scheme. RNA that is bound to the beads was cloned and sequenced. (*B*) MLE is a 1292-amino-acid-long RNA helicase with multiple domains. The mutations used in this study were (short versions of names in parenthesis): MLE^{K4E} (MLE^{K}), $MLE^{H196E\ R198E}$ (MLE^{HR}), MLE $^{K4E\ H196E\ R198E}$ (MLE^{KHR}), and MLE^{K413E} (MLE^{GET}). (*C*) Schematic overview of wild-type MLE and its mutants that were used to construct sequencing libraries. Barcoded 3′ linkers allowed mixing of biological replicates first, which was followed by mixing all samples after reverse transcription. (*D*, *left*) Western blot analysis showing the expression of the MLE transgenes and their purification after the first elution step using imidazole. (*Right*) PAA gel showing the final sequenced libraries. (*E*) Number of cross-linking events normalized by CR41602 across several relevant RNA categories. Each event was assigned to a single category according to the order roX1, roX2, ribosomal RNA (rRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), ncRNA, transfer RNA (tRNA), 3′ untranslated region (UTR), 5′ UTR, exon, intron, and antisense.

[MLE^{KHR}], respectively). Expression of MLE^{K}, MLE^{HR}, and MLE^{wt} was comparable in S2 cells (Fig. 1D, lanes 1–3,7–9), while MLE^{KHR} was poorly expressed (Fig. 1D, lane 4). The uvCLAP profile for MLE^{K} showed a reduced number of overall binding events (mostly from introns) (Fig. 1E; Supplemental Fig. S1A), but, surprisingly, the interaction of this mutant with roX2 was largely unaffected; rather, it was slightly enhanced compared with MLE^{wt} (Supplemental Fig. S1A,C). In contrast, mutating the second dsRBD (MLE^{HR} mutant) or both dsRBDs (MLE^{KHR}) of MLE simultaneously results in a severe loss of interactions from both roX2 and elsewhere (Fig. 1E). However, the roX2 interaction of MLE^{KHR} is still significantly higher than our background control ($P < 10^{-176}$) (Supplemental Fig. S1F, DESeq2), which can mean either that our mutations are not severe enough to completely eliminate dsRBD's ability to interact with dsRNA or that MLE uses other domains in addition to its N-terminal dsRBDs to interact with roX RNAs in vivo.

In order to characterize the potential effect of MLE's helicase activity on its interaction with roX RNAs, we mutated Lys413 (K413) of MLE to an aspartic acid residue, which is an established mutation also known as MLE^{K413E} (MLE^{GET}) that severely reduces the ability of MLE to interact with ATP or other NTPs (Lee et al. 1997). This "helicase-dead" mutant is unable to localize to the X chromosome or rescue male lethality in *mle^{null}* flies, which shows that the enzymatic activity of MLE, and not just its RBDs, is essential for dosage compensation (Lee et al. 1997). At the level of total cross-linking events, MLE^{GET} interacts significantly less with roX2 compared with the wild-type protein (Fig. 1E). Intriguingly, concomitant with the decrease in roX2 binding, MLE^{GET} appears to be redistributed over the 3′ UTRs of several mRNAs, partially compensating for the loss in roX2 binding in terms of total cross-linking events (130,217 cross-linking events are lost on roX2 in the MLE^{GET} data [68% decrease compared with MLE^{wt}], while 29,313 cross-linking events are gained in the 3′ UTRs of mRNAs in the MLE^{GET} data [100% increase compared with MLE^{wt}]) (Fig. 1E; Supplemental Fig. S1A,D).

MLE has been reported to play a major role in dosage compensation in flies (with the exception of Reenan et al. [2000], where a dominant-negative allele of MLE [*mle^{napts}*], but not a null allele of MLE, has been shown to interfere with *para* splicing), and, consistently, our earlier MLE iCLIP profiles did not identify any significant RNA target of MLE other than roX1 and roX2 (Ilik et al. 2013); thus, we were quite puzzled to observe new and specific targets of MLE in vivo, especially the new 3′ UTR targets of the helicase-dead mutant MLE^{GET}. A careful examination of the raw uvCLAP data on the 3′ UTRs that had enhanced MLE^{GET} binding revealed several reads that were split while they were being mapped to the genome, as if they were being mapped to exon–exon junctions but in the absence of any annotated introns. The 3′ UTR of Rps29 is one such example: The uvCLAP profile for MLE^{GET} has a clear enrichment over this region compared with MLE^{wt} and MLE^{K} (Fig. 2A, top). In order to find whether there is an underlying logic to these split reads (i.e., whether they map to specific features of these target 3′ UTRs), we systematically looked at these gapped alignments using HMMSplicer, which was originally developed to accurately identify noncanonical splice sites (Dimon et al. 2010). At the 3′ UTR of Rps29, gapped alignments identified by HMMSplicer clustered around two regions with quite imprecise junctions, arguing against the basic hypothesis that these reads originate from unannotated noncanonical introns (Fig. 2A, "gapped alignments"). To visualize these events more clearly, we generated coverage maps of these deletions within gapped alignments, which were then converted to a heat map to see where the gaps are most likely to be found on linear DNA using a color scale that shows the deletion density in a particular region (Fig. 2A, below). When this heat map was overlaid on the minimum free energy (MFE) structure of Rps29's 3′ UTR, we observed that these gaps fit almost perfectly to the loop regions of two predicted consecutive stem–loops (Fig. 2B; see also Supplemental Fig. 2A for the same analysis on the 3′ UTR of the ifc gene). Since it is established that MLE interacts with structured regions of roX RNAs using its dsRBDs (MLE^{HR} and MLE^{KHR} mutants) (discussed above; Ilik et al. 2013), we argue that these deletions originate from stem–loops that are bound by MLE in vivo.

Even though gapped alignments fit well with the MFE structures of 3′ UTRs of Rps29 (Fig. 2A,B) and ifc (Supplemental Fig. S2A), it is possible that these are nevertheless coincidences and that gapped alignments might be originating spuriously in our data and fit MFE structures by random chance. In order to rule out the null hypothesis that gapped alignments are random phenomena and that MFE predictions fit gapped alignment data by random chance, we first looked at all of the gapped alignments in the MLE^{GET} data transcriptome-wide and observed that the increase in binding to 3′ UTRs that we scored by assessing the cross-linked nucleotides of MLE^{GET} data (Fig. 1E) was followed by an increase in the amount of gapped alignments at 3′ UTRs (75% increase in gapped alignments; 100% increase in cross-linking events compared with MLE^{wt}; Fisher's exact, *P*-value $< 2.2 \times 10^{-16}$) (Fig. 2C). This establishes that gapped alignments indeed originate from in vivo targets of MLE and are highly unlikely to be originating from random regions.

Next, if gapped alignments do indeed originate from MLE-bound RNA in vivo as established above but have nothing to do with in vivo RNA structures, then gapped alignments should not correlate with base-pairing probabilities of the targeted regions as obtained from MFE structures. On the contrary, not only did we observe that gapped alignments indeed frequently fall on putative stem–loop structures that have a high probability of base-pairing, but this enrichment was lost upon dinucleotide shuffling of the target regions (MLE^{wt} vs. control with mean probabilities 0.239 and 0.076, respectively, *P*-value $< 2.2 \times 10^{-16}$; MLE^{GET} vs. control with mean probabilities of 0.239 and 0.076, respectively, *P*-value $< 2.2 \times 10^{-16}$) (Fig. 2D). Moreover, MLE^{GET}, which has more 3′ UTR targets that produce more gapped alignments compared with MLE^{wt} (Figs. 1E, 2C), also shows a higher probability of

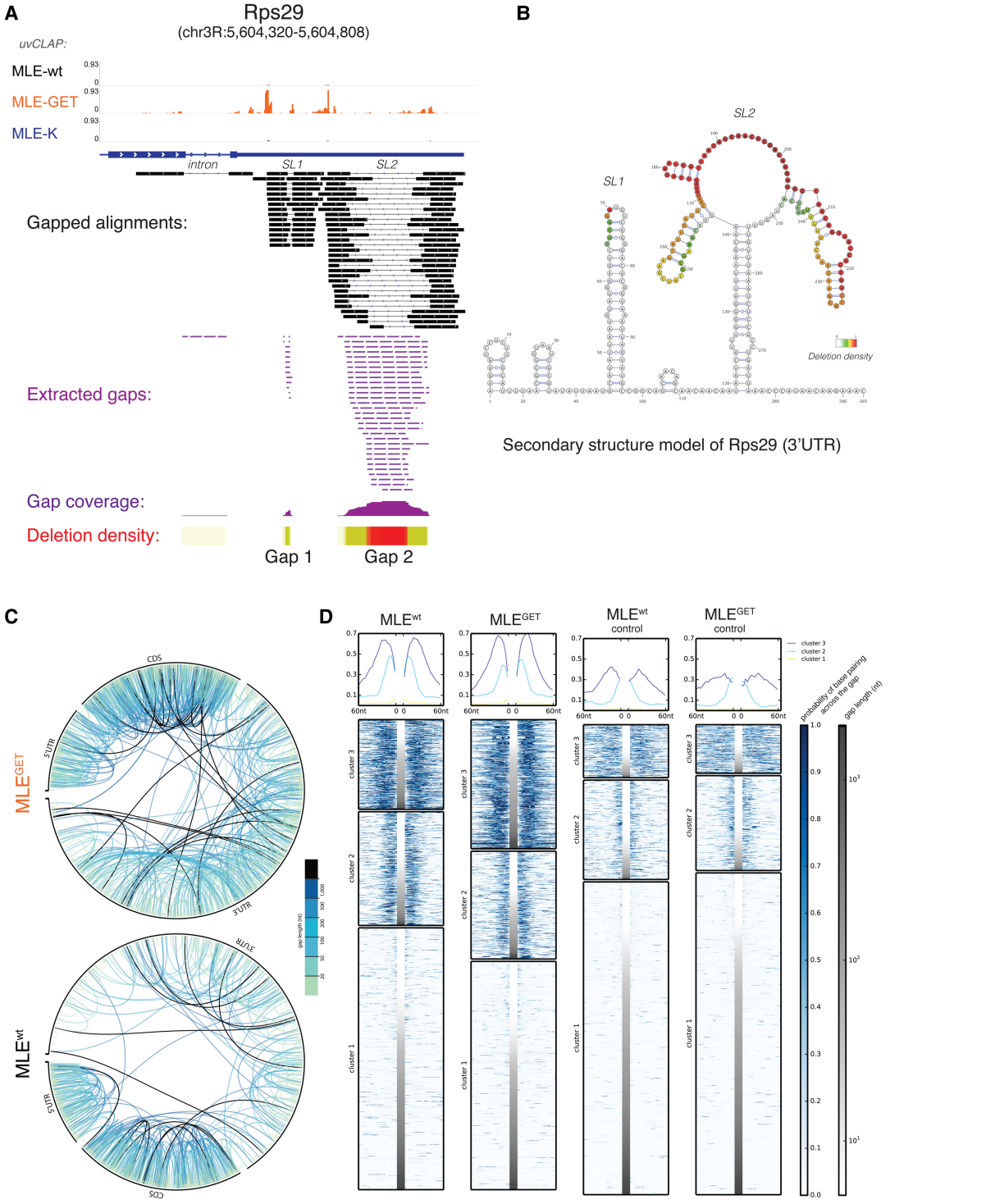

Figure 2. Gapped alignments reveal secondary structures of target RNAs. (*A*) *Rps29* is shown as a representative gene showing uvCLAP cross-linking events for MLE^{wt}, MLE^{K}, and MLE^{GET} (*top*) and gapped alignments (*bottom*). In order to focus on the gaps in gapped alignments, regions containing gaps were converted into a coverage file and finally represented as a heat map. (*B*) Overlay of gapped alignment density (deletion frequencies from *A*) on the in silico determined MFE structure of the relevant region in the Rps29 3′ UTR. (*C*) A transcriptome-wide survey of all of the gapped alignments that were detected in MLE^{wt} and MLE^{GET} data sets. The Circos (Krzywinski et al. 2009) plot shows MLE^{wt} and MLE^{GET} uvCLAP gaps on an mRNA metagene. Gap ends are connected by lines colored according to gap length and are positioned according to their relative position in the 3′ UTR, exon, and 5′ UTR. (*D*) Autosomal gaps in MLE^{wt} and MLE^{GET} uvCLAP correspond to stem–loop structures in the *Drosophila* transcriptome. The heat map shows the probabilities of nucleotides surrounding gaps forming gap-spanning base pairs. Control sequences for the selected MLE^{wt} and MLE^{GET} gaps were created by dinucleotide shuffling.

base-pairing at its targets compared with MLE^{wt} (mean probabilities of 0.239 vs. 0.206, P-value $< 2.2 \times 10^{-16}$) (Fig. 2D). These analyses show that there is a clear, reproducible, and statistical tendency for gapped alignments to originate from MLE-bound secondary structures in vivo.

The secondary structure landscape of roX RNAs as revealed by uvCLAP

Having established that we can look at secondary structures in our uvCLAP data with a certain degree of statistical certainty, we turned our attention to the prime targets of MLE: roX RNAs. Our original MLE iCLIP data were generated from clone 8 cells, which express both roX1 and roX2 lncRNAs. We were unable to generate stable cell lines using these cells and used S2 cells instead, which readily express roX2 and (in only a very small subset of cells) roX1 (<0.5% of S2 cells express roX1 according to Johansson et al. [2011]). However, we were positively surprised to see that, with uvCLAP, we could comfortably detect not only roX2 but also roX1 in S2 cells (albeit with much fewer reads compared with roX2), showing that our data have high dynamic range and sensitivity (Fig. 3A).

Reflecting the fact that roX1 is expressed in only a small subset of S2 cells, we detected a small number of gapped alignments in roX1 that were in and around the three domains that were shown to be the functional parts of roX1 (Supplemental Fig. S3A; Ilik et al. 2013; Quinn et al. 2014). Looking at D3, the ~450-nucleotide (nt) region of roX1 (~3.7 kb) that can rescue male lethality up to ~80% in the absence of any other roX RNA (highest rescue potential among any other region of roX1 except full-length roX1) (Quinn et al. 2014), we saw evidence for R1H1 and also for the presence of a long-range stem–loop structure that we and others (Park et al. 2007; Kelley et al. 2008; Quinn et al. 2014) predicted to form between the IRB element and roX1 box 1 ~120 nt away (Supplemental Fig. S3B). We also saw reads that could support the presence of a competing structure formed between IRB and roX1 box 2, which is suggestive of a dynamic switch between these two structures, possibly promoted by MLE's helicase activity (Supplemental Fig. S3B).

As expected and in contrast to roX1, we detected a much higher number of uvCLAP cross-linking events and gapped alignments that map to roX2 in S2 cells (in MLE^{wt} data: 192,600 cross-linking events in roX2 vs. 4772 in roX1; 4206 gapped alignments in roX2 vs. 75 in roX1). The overall binding pattern resembles our previous iCLIP data, with sharper peaks (Fig. 3B). As established above, gapped alignments in uvCLAP data tend to originate from structured RNA elements bound by MLE in vivo (Fig. 2). Since the secondary structure of roX2 has been determined by chemical and enzymatic RNA structure determination methods (Ilik et al. 2013; Maenner et al. 2013) and since the predicted structures have been functionally verified by fly genetics and immunofluorescence analysis of MSL complex spreading (Ilik et al. 2013), we applied the same gapped alignment analyses described above to roX2 in order to determine whether gapped alignments indeed correspond to these well-characterized stem–loops in roX2 lncRNA (for an overview, see Supplemental Fig. S4). The gapped alignments in roX2 fall predominantly on four regions (Fig. 3C, top, "deletion hot spots #1–#4"). In agreement with our transcriptome-wide analysis (Fig. 2), which indicates that gapped alignments originate from MLE-bound secondary structures in vivo, the first and most dominant deletion hot spot (#1) in roX2 falls on the loop region of R2H1 (Supplemental Fig. S4), while the second deletion hot spot (#2) corresponds to R2H2/3, confirming our previous inferences from our iCLIP, SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension), and PARS (parallel analysis of RNA structure) data (Fig. 4B; Supplemental Fig. S4; Ilik et al. 2013).

We expected the last two deletion hot spots (#3 and #4) to correspond to R2H5 and R2H6; however, we were surprised to see that, while region #4 did correspond to the loop region of R2H5, region #3 did not correspond to a stem–loop that we or others had characterized before. Looking at the raw mapped reads (Fig. 3C, bottom; Supplemental Fig. S4), we realized that this stem–loop shares a stem with R2H5; i.e., the 5′ stem of R2H5 is the 3′ stem of this new structure. As these stem–loops cannot exist at the same time, we refer to this newly identified stem–loop as the "alternate stem–loop" (ASL).

Interestingly, the free energy of formation for these structures is not equal; R2H5 is a significantly more stable stem–loop than ASL (free energy of formation of R2H5 and ASL is −23.69 kcal/mol and −15.73 kcal/mol, respectively), meaning that under equilibrium conditions and in the absence of external factors that affect the formation or unwinding of either structure, R2H5 will be the stem–loop that will form most of the time, in accordance with the associated partition function. Moreover, once formed, the energy barrier between the R2H5 state and the ASL state is calculated to be ~20 kcal/mol, which would not only strongly favor the R2H5 state as indicated above but also imply the existence of a regulated switch that requires the net input of energy to shift from the R2H5 state to the ASL state.

If we assume that the relative numbers of gapped alignments correlate positively with the relative stabilities of R2H5 and ASL, it would thus be expected that since R2H5 is the more stable stem–loop, it should be the source of the higher numbers of gapped alignments as compared with ASL. However, our data show that this is not the case for MLE^{wt}. In contrast, ASL and R2H5 are represented by similar numbers of gapped alignments (Fig. 3C, top, deletion hot spots #3 and #4, respectively). This result can mean that in silico calculations are either inaccurate (and ASL and R2H5 can, in fact, form with equal probability) or accurate. In the latter case, since uvCLAP captures the in vivo binding events of a large number of MLE–roX interactions, we actually scored MLE's initial binding to R2H5 as well as the end product of its unwinding activity: ASL. If the former is true, then, similar to MLE^{wt}, one would expect MLE^{GET} to interact with ASL and R2H5 equally well, as it possesses all of the domains known to interact with dsRNA or ssRNA and is mutated at only a single amino acid that affects its ability to bind ATP/NTP. If the latter is true, we would

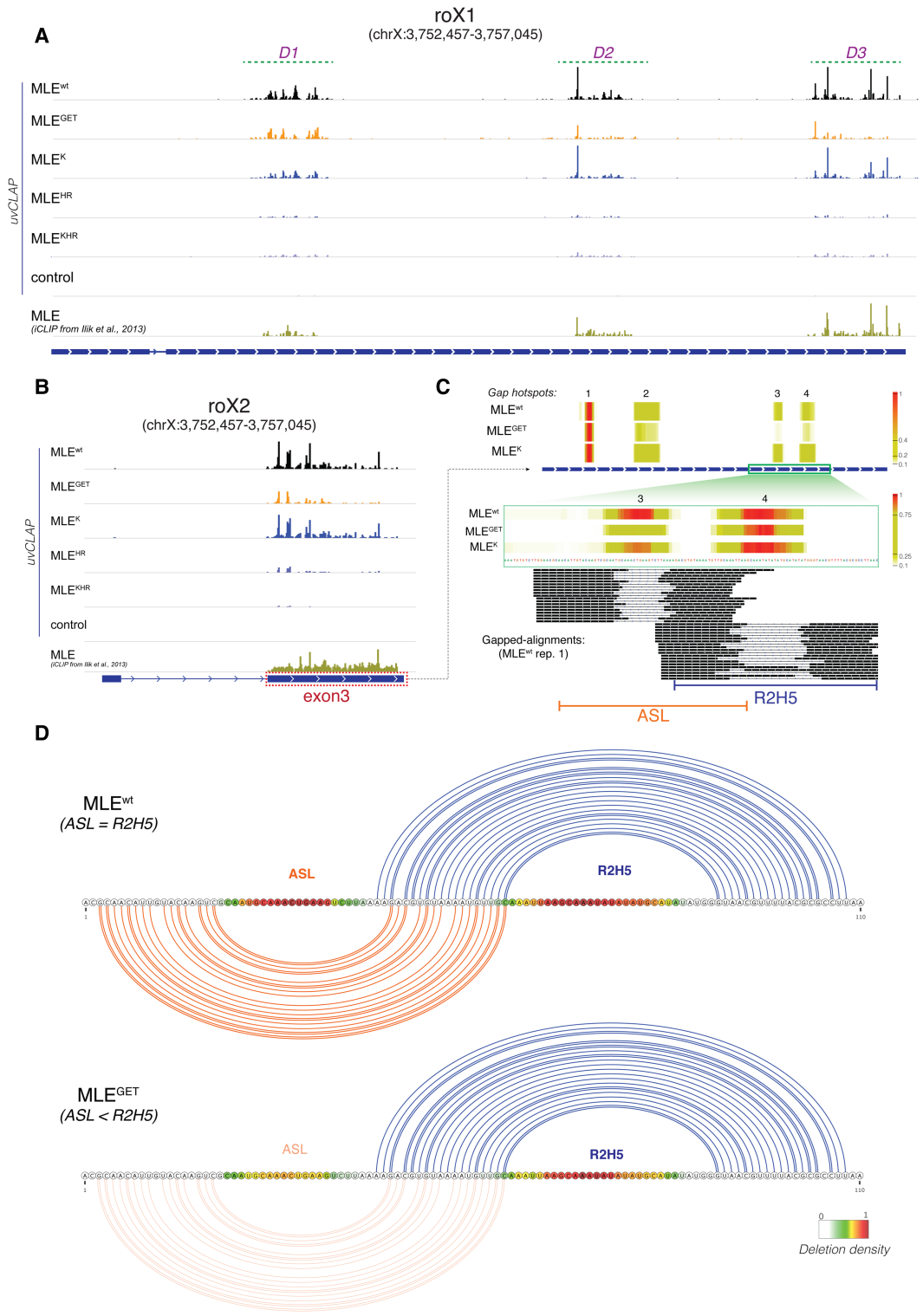

Figure 3. uvCLAP reveals a mutually exclusive stem–loop in roX2 exon 3. (*A*) uvCLAP data of MLE^{wt} and mutant constructs on the roX1 gene. Previously determined roX1 functional domains are depicted at the *top* as D1, D2, and D3. (*B*) uvCLAP data on the roX2 gene. (*C*, *top*) Depiction of gapped alignments that fall on roX2 exon 3 (red box from *B*). Gapped alignments falling to the intron were removed for clarity. (*Bottom*) Zoomed-in image of deletion hot spots #3 and #4 (green box) with a subsample of MLE^{wt} raw HMMSplicer alignments showing that the stem–loops that result in deletion hot spots #3 and #4 are mutually exclusive and cannot be found on the same piece of RNA at the same time. Heat maps represent deletion hot spots as in Figure 2A. (*D*) A close-up of the 110-nucleotide (nt) region of roX2 that can form either ASL (alternate stem–loop) or R2H5. Base-pairing regions are connected with lines. (Blue) R2H5; (orange) ASL. Deletion densities for MLE^{wt} and MLE^{GET} from *C* are overlaid over the sequence.

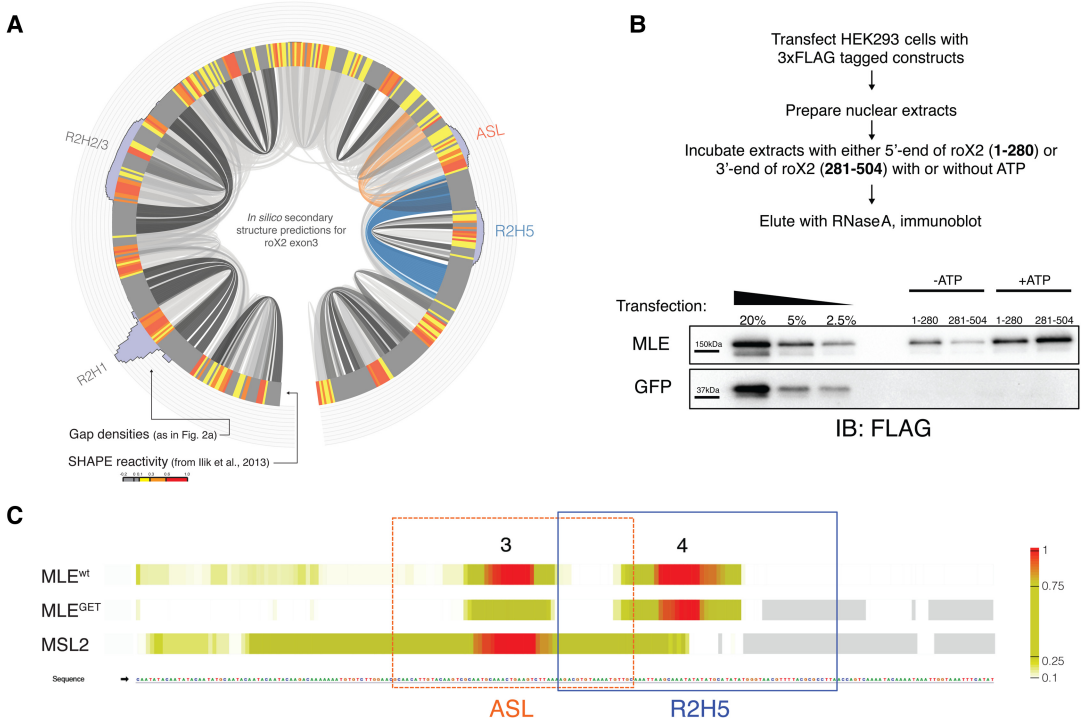

Figure 4. MSL2 shows a preference for ASL versus R2H5 in roX2 in vivo. (*A*) A Circos plot showing deletion hot spots (*outside*), SHAPE reactivity (*middle*; from Ilik et al. [2013]), and in silico prediction of secondary structures (*inside*) shaded according to their average base-pairing probability (RNAplfold version 2.1.8 [Lorenz et al. 2011], folded at 25°C using parameters "-W 150 -L 100" as recommended by Lange et al. [2012]). R2H5 and ASL are in shades of blue and orange. (*B*) GST-RNA (GRNA) affinity chromatography shows that MLE interacts with the 3′ end of roX2 RNA (281–504) in an ATP-augmentable manner. This ATP-dependent trap is shown in Supplemental Movie 1. (*C*) A closer look at the 3′ end of roX2 shows that the mutually exclusive structures ASL and R2H5 are equally present in MLE^{wt}, whereas there is a preference for R2H5 in MLE^{GET}. In the MSL2 data, gapped alignments support the presence of ASL but not R2H5.

expect to see higher numbers of gapped alignments supporting the presence of R2H5, as MLE^{GET} cannot unwind R2H5. Analysis of the gapped alignments mapping to the ASL–R2H5 region shows that, unlike MLE^{wt} data, MLE^{GET} data are enriched for R2H5, supporting the hypothesis that MLE unwinds R2H5 to promote the formation of ASL, a process that requires MLE's helicase activity (Fig. 3C [bottom], D).

We reported previously that MLE interacts with the 5′ end of roX2 equally well with or without ATP, while MLE's interaction with the 3′ end of roX2 (that we now show contains the ASL–R2H5 metastructure) is augmented with ATP in vitro (Ilik et al. 2013). Furthermore, we show that this binding was abolished when MLE^{GET} was used instead of MLE^{wt}, suggesting that, paradoxically, enhanced binding of MLE to the 3′ end of roX2 is boosted by its helicase activity. In order to show that this enhanced binding in the presence of ATP is a property of MLE and not of auxiliary factors that could be found in a *Drosophila* extract, we transfected human HEK293 cells with a construct that expresses MLE and carried out GST-RNA (GRNA) affinity chromatography experiments using the 5′ and 3′ ends of roX2 with or without ATP, verifying that, within the context of a human extract, MLE can interact better with the 3′ end of roX2 in the presence of ATP (Fig. 4C). Intriguingly, uvCLAP data show that, in vivo, MLE^{GET} loses a significant portion of its roX2 binding; however, this loss is partially accounted for by an increased interaction with 3′ UTRs of target mRNAs that tend to harbor RNA secondary structures (Fig. 1E). This result shows that, unlike the $MLE^{HR/KHR}$ mutants, MLE^{GET} has not lost critical domains that mediate RNA interactions. Moreover, uvCLAP cross-linking data (Fig. 3B) and gapped alignments (Fig. 3C) show that MLE^{GET} binding is lost preferentially from the 3′ end of roX2, suggesting that a peculiar RNA element in this region is able to "trap" MLE as long as it can unwind RNA structures, while the helicase-dead mutant MLE^{GET} cannot be trapped and is free to explore the dsRNA content of the nucleus. This model explains the equal distribution of ASL and R2H5, supporting gapped alignments in MLE^{wt} data as well as the disappearance of MLE^{GET} from the 3′ end of roX2 in vivo (especially from ASL) and its redistribution to other structured RNA elements at the 3′ UTRs of new mRNA targets.

MLE helicase activity is important for MSL2 to bind roX2 in vivo

One critical aspect of the ASL–R2H5 arrangement that is essential to understanding how it can act as an MLE trap is that, unlike traditional stem–loops, unwinding of one

stem–loop within this metastructure will very likely lead to the spontaneous formation of the other stem–loop due to the short distance (~20 nt) between competing structures. It is conceivable that when an RNA helicase (MLE) has completed unwinding one stem–loop (e.g., R2H5), the other stem–loop could already have formed in the vicinity (e.g., ASL), as they are physically linked to each other. Thus, an MLE^{wt} molecule can continue unwinding ASL and R2H5 into each other as long as there is available energy (Supplemental Movie 1). Importantly, the structural and sequence requirements for such a "trap," such as sharing a base-pairing region between two stem–loops, is strikingly conserved from *Drosophila melanogaster* to *Chymomyza amoena* (see Supplemental Fig. 5; Quinn et al. 2016).

RNA switches, such as riboswitches, are on either an "on" state or an "off" state, which is generally determined by ligand binding (Mandal et al. 2003). What can be the purpose of a switch if the switch is continuously turned on and off? Since it has been proposed that MLE unwinds roX RNA for MSL2 interactions based on in vitro experiments (Maenner et al. 2013), we reanalyzed our MSL2 iCLIP data for the presence of gapped alignments, expecting that if MSL2 interacts with a single-stranded element within the ASL–R2H5 region of roX2, it could stabilize the formation of one stem–loop at the expense of the other. Intriguingly, MSL2 data contain gapped alignments that come from only ASL but not R2H5 (Fig. 4A). This result suggests that MLE unwinds R2H5, which leads to the formation of ASL, simultaneously exposing roX2 box 2 for MSL2 interactions. However, our data and published work suggest that this model is incomplete. For example, from the core MSL complex perspective, recent work on MSL2 has shown that MSL2 can recognize high-affinity sites (HASs) in the absence of other factors in vitro, but how would binding to HASs by MSL2 lead to the spreading of the complex to the rest of the X chromosome in a way that depends on roX RNAs and MLE (Villa et al. 2016)? Moreover, from the roX lncRNA perspective, Park et al. (2007) have shown that concatenating six copies of R2H5 without the sequence that is required to also form ASL can rescue only ~17% male lethality. Therefore, we decided to evaluate the relative importance of the ASL–R2H5 structure in vivo using fly genetics by expressing several different roX RNA derivatives in the absence of endogenous roX expression.

A minimal roX RNA rescues male lethality

Integrating the wealth of previous recent data (Ilik et al. 2013; Quinn et al. 2014, 2016) with our updated understanding as a result of our uvCLAP analysis described above, we constructed a "minimal" roX RNA that is comprised of two parts: (1) R2H1 that serves as the ATP-independent contact point for MLE and (2) the ASL–R2H5 metastructure (Fig. 5A,B). Interestingly, when expressed in a roX1/roX2 double-mutant background, this transgene is able to rescue male lethality with an average rescue level of 72% (Fig. 5C) and fully support spreading of the MSL complex on the male X chromosome as the sole source of roX RNA (Fig. 5G). To our knowledge, this construct, containing 240 nt of roX2 RNA, is the smallest roX RNA that that has been used to rescue male lethality in the absence of any endogenous roX RNA expression.

Next, we decided to probe this significantly smaller RNA for even smaller functional units, whether sequences or secondary structures. First, we inverted the sequence between nucleotides 174 and 240 that encodes R2H5 and the right stem of ASL (Fig. 5C,D; Supplemental Fig. S6). This mutant (mut1) preserves the overall structure of the wild-type minimal roX RNA (Fig. 5, wild type [wt]) but prevents the formation of ASL. This mutant was completely nonfunctional, without any escaper males (Fig. 5E). Importantly, immunofluorescence analysis on polytene chromosome isolated from third instar larvae revealed that the expression of the mut1 construct led to severely compromised binding of the MSL complex to the X chromosome (Fig. 5G).

While mut1 results show that the overall structure of the roX cannot by itself account for its functionality, they do not prove that ASL is functionally relevant, as the orientation of one or more sequence elements within R2H5 could also be critical for function. To address this issue, we constructed another mutant (mut2) in which we inverted 17 nt of the minimal roX (nucleotides 135–151) that forms the left stem of ASL, thereby making it impossible for ASL to form, while R2H1, R2H5, and, overall, >95% of the minimal roX sequence was left untouched (Fig. 5D). This construct could rescue only 2.4% of male flies, strongly suggesting that ASL is necessary for a functional roX RNA. Moreover, polytene chromosomal staining of MOF and MSL3 again showed a severely compromised binding pattern with just a few sites showing a residual signal of MSL3, reminiscent of loss of spreading. In order to test whether ASL rather than R2H5 could be sufficient for roX function, we constructed another mutant (mut3) in which we mutated 10 nt at the right stem of R2H5 (predicted to prevent the formation of R2H5 and lead to constitutive ASL formation) while leaving four consecutive thymidines that form the core of roX2 box 2 untouched (Fig. 5D). This construct, similar to mut1, was completely nonfunctional, without any escaper males (Fig. 5E). These three mutants clearly show that the presence of both structures is necessary (but neither ASL nor R2H5 alone is sufficient) for a fully functional roX2 RNA.

Intrigued by the sensitivity of minimal roX to small mutations, we then tested whether separating ASL and R2H5 would have an effect on the functionality of our transgene (mut4) and found that, by all of the metrics that we checked (viability, stability, MSL-complex spreading) (Fig. 5E–G), this construct performed very similar to the wild-type construct. In contrast, completely inverting minimal roX RNA (mut5) or restoring the orientation of roX2 box 2 within the inverted R2H5 sequence together with the complementary parts within R2H5 as well as within ASL (mut6) led to two completely nonfunctional RNAs (Fig. 5E). In addition, using antaRNA (Kleinkauf et al. 2015), we designed two further constructs to mimic the overall structure of minimal roX (mut7 and mut8) (Fig.

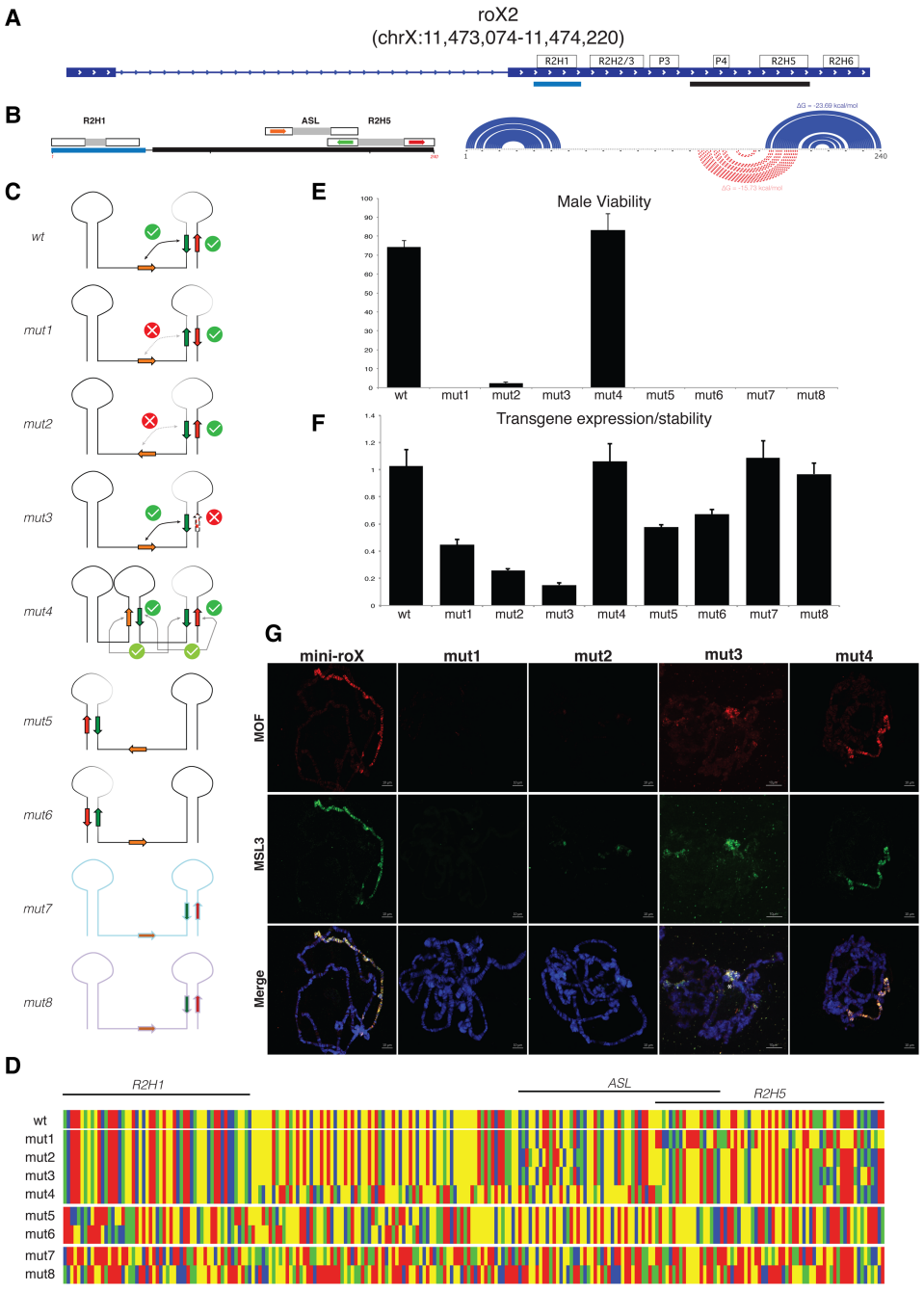

Figure 5. Minimal roX RNA with a functional ASL ↔ R2H5 switch is sufficient to rescue male lethality. (*A*) Minimal roX is constructed by taking R2H1 as the high-affinity handle for MLE (blue box) and fusing it to the ~100-nt stretch at the 3′ end that contains ASL and R2H5 (black box). (*B*) Minimal roX RNA is 240 nt long (*left*) and contains R2H1 and either R2H5 or ASL stem–loops (*right*). (*C*) Schematic representation of roX derivatives. Minimal roX is able to rescue male lethality in the absence of endogenous roX expression (wild type). Mutations that prevent the formation of ASL (mut1 and mut2) or R2H5 (mut3) lead to nonfunctional roX RNAs. Separating ASL from R2H5 leads to a functional minimal roX RNA indistinguishable from the wild-type construct (mut4). Inverting the minimal roX sequence (mut5), reinstating the correct orientation of the important sequences (mut6), or designing constructs that mimic minimal roX RNA structure (mut7,8) all lead to nonfunctional RNAs that cannot rescue male lethality. All transgenes were expressed with a tubGAL4 driver from the same genomic locus (see *E* and *F* and the Materials and Methods). (*D*) Sequences of all constructs described in *C*, presented as colored boxes. (Green) Guanine; (blue) cytosine; (red) thymine; (yellow) adenine. (*E*) Rescue frequencies of the constructs described in *C*. (*F*) Expression/stability of the constructs described in *E* as determined by RT-qPCR from third instar larvae. All values were normalized to pfk and the wild-type minimal roX RNA expression. (*G*) Polytene stainings show the proper spreading and colocalization of MOF and MSL3 on the male X chromosome in wild-type and mut4-expressing transgenes that also rescue male lethality. In contrast, mut1, mut2, and mut3 do not rescue male lethality or support MOF or MSL3 binding on the X chromosome. (Red) MOF; (green) MSL3; (blue) DNA; (*) mislocalization to chromocenter. Bar, 10 µm.

5C,D). While they expressed as well as the wild-type construct, neither of these RNAs could rescue male lethality (Fig. 5E,F).

In summary, these results collectively show that minimal roX RNA made up of R2H1 and the ASL–R2H5 metastructure is the shortest functional roX RNA reported to date. Mutants of minimal roX that cannot form ASL or R2H5 are nonfunctional, underscoring the importance of the ASL–R2H5 metastructure while showing that neither structure by itself is sufficient for dosage compensation.

Discussion

In this study, using MLE as an example, we showed that uvCLAP can be used to study RNA-binding interactions of an RNA helicase and its mutants in vivo. Our results show that, as expected from our previous knowledge of MLE–roX interactions, the top target of MLE in S2 cells is roX2. Through the analysis of point mutants, we verified that the N-terminal dsRBDs (and especially dsRBD2) are crucial for MLE's interactions with RNA, in agreement with a recent crystal structure of MLE (Rajan Prabu et al. 2015). While analyzing the enzymatically inactive MLE mutant MLE^{GET}, we serendipitously discovered that uvCLAP data, in addition to providing us with nucleotide-resolution binding sites, also contain crucial secondary structure information about MLE's in vivo RNA targets. Using this knowledge, we took a closer look at MLE's interactions with roX RNAs and were surprised to discover a previously unknown stem–loop at the 3′ end of roX2: ASL. Our data show that this stem–loop is adjacent to R2H5 and can be formed only if and when R2H5 is unwound, since the 3′ stem of ASL is exactly the same stretch of RNA that forms the 5′ stem of R2H5, making these two stem–loops mutually exclusive.

There are several interesting properties and implications of this mutually exclusive stem–loop arrangement between ASL and R2H5. First, despite the poor sequence conservation of roX RNAs across *Drosophila* species, RNA elements that are able to form this arrangement are conserved from *D. melanogaster* to *C. amoena* (see Supplemental Fig. S5; Quinn et al. 2016). Second, the free energy of formation for these structures is not equal, R2H5 is a more stable stem–loop than ASL (free energy of formation of R2H5 and ASL is −23.69 kcal/mol and −15.73 kcal/mol, respectively). Third, once formed, the energy barrier between the R2H5 state and the ASL state was calculated to be ~20 kcal/mol, which not only strongly favors the R2H5 state but also implies the existence of a regulated switch that requires net input of energy to shift from the R2H5 state to the ASL state. Based on the plethora of genetic and biochemical evidence and in the absence of any other candidate, we propose that MLE is the RNA helicase that can carry out this transition from the R2H5 state to the ASL state. In this study, we provide several lines of evidence in support of this hypothesis and show how, by viewing ASL and R2H5 as a single metastructure rather than two independent stem–loops, we can explain MLE's perplexing ability to interact with certain RNAs better in vitro in the presence of ATP.

Importantly, we observed that uvCLAP data from MLE^{GET}, which cannot unwind secondary structures, contained more gapped alignments, supporting R2H5's presence, whereas, in the MLE^{wt} data, gapped alignments support ASL and R2H5 equally, supporting the idea that MLE is using its helicase activity to unwind R2H5 to allow the formation of ASL, and this process is impaired by the K413E mutation. Finally, we constructed a minimal roX RNA to understand whether ASL has any physiological role in flies. We found that mutations that disrupt the formation of the ASL while keeping R2H5 intact are detrimental to the function of our roX transgene, attesting to the fact that R2H5 together with ASL forms the long-sought-after functional unit in roX2 that is necessary for dosage compensation.

In summary, based on (1) the evolutionary conservation of the ASL–R2H5 metastructure, (2) ASL and R2H5 gapped alignments supporting equal formation of ASL and R2H5 in MLE^{wt} uvCLAP data despite a large energy difference between the structures, (3) reduction of gapped alignments that map to ASL in MLE^{GET} uvCLAP data, (4) MSL2 gapped alignment data supporting the presence of only ASL and not R2H5, (5) MLE's ATP-augmented interaction with the 3′ end of roX2 that contains the ASL–R2H5 metastructure, and (6) genetic data showing that the minimal roX RNA that we designed is functional only when both ASL and R2H5 are intact, we suggest the following model to explain how the MSL complex can spread from a small number of HASs to the rest of the X chromosome: roX2 RNA is transcribed in a conformation that does not allow it to interact with MSL2 (e.g., R2H5 state, roX2 box 2 buried). In the absence of roX interaction, MSL2 together with MSL1 binds to a few sites on the X chromosome, the HASs. As roX2 is remodeled by MLE, it can now interact with MSL2. This interaction can disrupt MSL2's interaction with the chromatin as suggested by Li et al. (2008). As MSL2 diffuses away (with the MSL complex), MLE unwinds ASL, favoring the reformation R2H5, which competes with the MSL2–roX interaction and thus makes the MSL complex once again competent to bind chromatin at a nearby locus. This MLE-powered cycle can lead to spreading of the MSL complex from a small number of HASs locally to the rest of the X chromosome by transiently breaking MSL2's contact with HASs using the ASL–R2H5 switch (Fig. 6). Further experiments that target the MSL2–roX RNA surface will be essential to test this model of MSL complex spreading on the male X chromosome.

Even with such a well-studied RNA as roX2, discovering this functionally essential, mutually exclusive stem–loop arrangement has been difficult; thus, new tools might be necessary to uncover more of them, if they exist, in other RNAs. Using reversible AMT cross-linking, a recent study identified many previously unknown RNA–RNA contacts in human and mouse cells (Lu et al. 2016), mostly in the form of duplexes. Interestingly, this study also uncovered several putative alternating structures, similar to what we observed in our study of

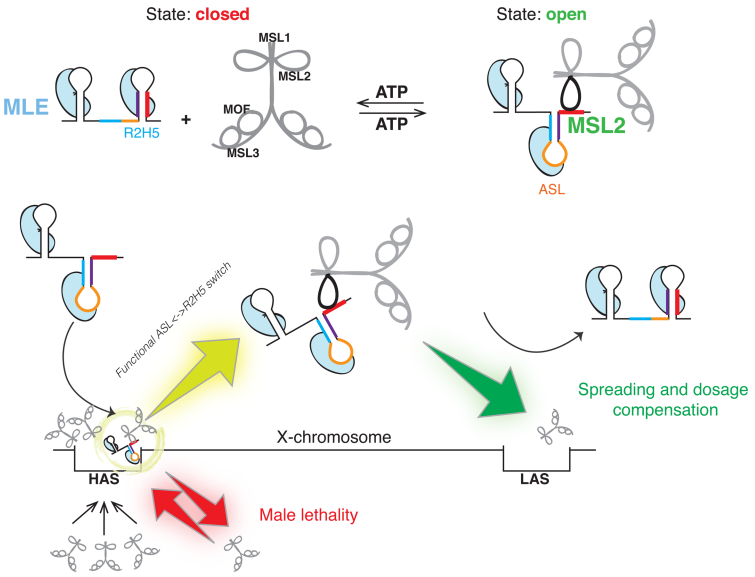

Figure 6. Unwinding of the mutually exclusive stem–loops in roX RNAs and MSL complex spreading on the male X chromosome. A model explaining how the ASL ↔ R2H5 switch can be used to spread the MSL complex from a few HASs to the entire male X chromosome. HASs on the X chromosome are able to recruit MSL complexes in an RNA-independent manner and act as a sink for the MSL complex. roX RNA, by virtue of its highly structured 5′ end, interacts with MLE, which, in the presence of ATP, constantly unwinds R2H5 and then ASL in an endless loop. This loop is interrupted when the roX–MLE complex hits MSL2 while at the "ASL" conformation, which can move the MSL complex away from the HASs. The locally diffusing MSL complex can interact with the chromatin again when MLE unwinds ASL, which leads to the formation of R2H5 and the release of MSL2 together with MSL complex. This allows the MSL complex to first be recruited to a small number of HASs and then explore nearby chromatin through reversibly interacting with roX RNAs.

roX RNAs. Whether these alternative structures in human RNAs serve any physiological function, as we demonstrated for roX2 in flies, and whether specific RNA helicases play a role in their biogenesis and maintenance remain important open questions.

Materials and methods

Drosophila *cell culture*

Drosophila S2 cells were maintained under standard conditions at 25°C. Stable cell lines were generated by transfecting S2 cells with pIBU1 plasmids that contained either MLE^{wt} or point mutants of it and selecting them with 1 µg/mL neomycin. The cells were induced with 0.2 mM $CuSO_4$ for 16 h before harvesting.

uvCLAP

For uvCLAP, copper-induced cells were washed once with ice-cold PBS and cross-linked with 0.15 mJ/cm^2 UV-C light. After cross-linking, cells were flash-frozen in $N_{2(l)}$ and kept at −80°C until use. The cells were lysed with 1× NLB (1× PBS, 300 mM NaCl [total monovalent salt concentration 450 mM], 1% Triton X-100, 0.1% Tween-20), sonicated with a Bioruptor water bath sonicator, and clarified with centrifugation. His pull-downs were done using paramagnetic His beads for ∼5 min. The imidazole eluate was then used immediately for a streptavidin pull-down (MyONE C1 beads). The SA beads were washed and treated with RNase I and T4 PNK (see also Supplemental Fig. S2; Das and Shuman 2013). Two 3′ linkers (L3A and L3B) were ligated to biological replicates using T4 RNA ligase 1. Excess linkers are washed away, biological replicates were pooled, and bound RNA was isolated by proteinase K digestion and column purification. Barcoded RT primers were then used for reverse transcription, after which all of the samples were mixed, column-purified, and loaded onto a denaturing 6 M urea/6% PAA gel. Three size fractions (H, M, and L) were isolated and purified with phenol-chloroform extraction and ethanol precipitation. The precipitates were then circularized using Circ LigaseII and linearized with BamH1 digestion. The linearized cDNA was then column-purified and subjected to PCR. The final libraries were quantitated, normalized, and sequenced on a MiSeq sequencer.

uvCLAP data analysis

uvCLAP reads were processed as described in Aktaş et al. (2017). Peaks were called using JAMM (joint analysis of next-generation sequencing replicates via mixture model clustering) (Ibrahim et al. 2015) and PEAKachu (Holmqvist et al. 2016). JAMM was designed to report a large number of peaks and relies on external post-processing for peak filtering. We used JAMM to determine unfiltered candidate peak regions for all experimental conditions using the two replicates for the respective pull-down conditions as the foreground and the eGFP pull-down condition as the background. The resulting candidate peaks for MLE^{wt} and its four mutants were used to calculate pairwise Spearman correlations between all replicates (Supplemental Fig. S1G), showing increased pairwise correlations between biological replicates within each condition over replicates of differing experimental conditions (deeptools version 2.3.5) (Ramírez et al. 2016). JAMM combined the MLE-binding sites located on roX1 and roX2 into single large regions. For that reason, we decided to perform additional detection of high-confidence peaks using PEAKachu, a peak caller that allows the identification of individual peaks within highly expressed RNAs and uses DESeq2 for rigorous statistical filtering.

Quantitative investigation of MLE point mutants

Biological replicates of the five MLE pull-down conditions and the background control were subjected to joint amplification and sequencing. This setup preserved the relative quantities of detected cross-linking events within a multiplexed library and thus allowed us to directly compare different pull-down conditions without further normalization (Maticzka et al. 2017). The preservation of relative quantities of cross-linking events within the multiplexed library for MLE was further highlighted by the similar amounts of detected events for CR41602, an 18S ribosomal RNA pseudogene. Events for CR41602 constituted 47%

of total events in the uvCLAP background control and were also detected for MLE and MSL2 iCLIP (3% and 31% of total cross-linking events, respectively) and iCLIP experiments of eight different SR proteins (SC35, SF2, SRp54, XL6, Rbp1, B52, Rsf1, and Rbp1L) (Bradley et al. 2015). The large fraction of CR41602 events in the uvCLAP background control and its independent detection by iCLIP experiments for at least 10 different proteins indicate that CR41602 constitutes an unspecific background. In line with the notion that uvCLAP preserves RNA quantities of experiments within a multiplexed library and the categorization of CR41602 as an unspecific background invariant of the pull-down condition, we detected very similar amounts of CR41602 events for the five MLE experiments and the background control (mean: 107,297 events; standard deviation: 12,979.48; relative standard deviation: 12.10%). In contrast, we observed large differences in the total number of detected cross-linking events and the composition of the target classes of the MLE^{wt} and the four MLE mutants, highlighting the ability of uvCLAP to quantitatively investigate multiple point mutants of the same protein.

Detection of gapped alignments

Gapped alignments were identified from uvCLAP data with HMMSplicer (version 0.9.5) with the command runHMM.py -o [OUTPUT] -i [INPUT.fastq] -g dm3 -c FALSE -w 5 -k 2000 -d TRUE -r FALSE and processed further with custom Python scripts to remove duplicate reads using Unique Molecular Identifiers (UMIs).

Transcriptome-wide analysis of gapped alignments

The heat map shown in Figure 2D shows the probabilities of nucleotides surrounding gaps forming gap-spanning base pairs. The regions of the gaps are not shown; gap lengths are annotated by the scale located in the center of each heat map. For all autosomal gaps spanning at least 2 nt, we selected the transcriptomic contexts resulting in minimal gap lengths (MLE^{wt}: 4311 gaps; MLE^{GET}: 3840 gaps). The resulting gaps were extended by 250 nt upstream and downstream in the corresponding transcriptomic contexts. Control sequences for the selected MLE^{wt} and MLE^{GET} gaps were created by dinucleotide shuffling. RNAfold (ViennaRNA package version 2.1.8) (Lorenz et al. 2011) was used to predict base-pairing probabilities at 25°C (parameter -p -T 25). Based on these predictions, we calculated the probability of each nucleotide forming a base pair spanning the gap. The resulting probabilities for the nucleotides 60 nt upstream of and downstream from the gaps were visualized using deeptools (version 1.5.11) (Ramirez et al. 2014) and applying the *k*-means clustering integrated with deeptools. Average probabilities of nucleotides within 30 nt of a gap forming a base pair with nucleotides on the other side of the gap were significantly increased for MLE^{wt} compared with its control (mean probabilities 0.206 vs. 0.083; Welch two-sample *t*-test $P\text{-value} < 2.2 \times 10^{-16}$), for MLE^{GET} compared with its control (mean probabilities 0.239 vs. 0.076; Welch two-sample *t*-test $P\text{-value} < 2.2 \times 10^{-16}$), and for MLE^{GET} compared with MLE^{wt} (mean probabilities 0.239 0.206; Welch two-sample *t*-test $P\text{-value} < 2.2 \times 10^{-16}$).

GRNA affinity chromatography

GRNA affinity chromatography was performed essentially as described in Ilik et al. (2013), with the exception that MLE or GFP expression plasmids were transfected into HEK293 cells, and nuclear extracts prepared from these cells were used in GRNA affinity experiments.

Drosophila *husbandry and genetics*

Drosophila husbandry and genetics were essentially as described in Ilik et al. (2013). Following are the mini-roX sequences that were used to rescue male lethality in roX-double-null flies: mini-roX (wild type) (GCTTTAGAGATCGTTTCGAATCACA
TTGATAATCGTTCGAAACGTTCTCCGAAGCAAAAAATG
AATGATATACAATATACAATATACAATATGCAATACAAT
ACAATACAAGACAAAAAAATGTGTCTTGGAACGCAACA
TTGTACAAGTCGCAATGCAAACTGAAGTCTTAAAAGAC
GTGTAAAATGTTGCAAATTAAGCAAATATATATGCATA
TATGGGTAACGTTTTACGCGCCTTAACCAC), mut1 (GC
TTTAGAGATCGTTTCGAATCACATTGATAATCGTTCGA
AACGTTCTCCGAAGCAAAAAATGAATGATATACAATAT
ACAATATACAATATGCAATACAATACAATACAAGACAA
AAAAATGTGTCTTGGAACGCAACATTGTACAAGTCGCAA
TGCAAACTGAAGTCTTAATTCCGCGCATTTTGCAATGGG
TATATACGTATATATAAACGAATTAAACGTTGTAAAATG
TGCAGAA), mut2 (GCTTTAGAGATCGTTTCGAATCACATT
GATAATCGTTCGAAACGTTCTCCGAAGCAAAAAATGAA
TGATATACAATATACAATATACAATATGCAATACAATAC
AATACAAGACAAAAAAATGTGTCTTGGAACCTGAACAT
GTTACAACGGCAATGCAAACTGAAGTCTTAAAAGACGT
GTAAAATGTTGCAAATTAAGCAAATATATATGCATATA
TGGGTAACGTTTTACGCGCCTT), mut3 (GCTTTAGAGA
TCGTTTCGAATCACATTGATAATCGTTCGAAACGTTCT
CCGAAGCAAAAAATGAATGATATACAATATACAATATAC
AATATGCAATACAATACAATACAAGACAAAAAAATGTGT
CTTGGAACGCAACATTGTACAAGTCGCAATGCAAACTG
AAGTCTTAAAAGACGTGTAAAATGTTGCAAATTAAGC
AAATATATATGCATATATGGCGCGCATTTTGCAATGC
TT), mut4 (GCTTTAGAGATCGTTTCGAATCACATTGAT
AATCGTTCGAAACGTTCTCCGAAGCAAGCAACATTGTA
CAAGTCGCAATGCAAACTGAAGTCTTAAAAGACGTGT
AAAATGTTGCAAAATGAATGATATACAATATGCAATAC
AATACAATACAAGACAAAAAAATGTGTCTTGGAACGTG
TAAAATGTTGCAAATTAAGCAAATATATATGCATATATG
GGTAACGTTTTACGCGCCTT), mut5 (TTCCGCGCATTTT
GCAATGGGTATATACGTATAACTTAACGAATTAGACGT
TGTAAAATGTGCAGAAAATTCTGAAGTCAAACGTAACG
CTGAACATGTTACAACGCAAGGTTCTGTGTAAAAAAAC
AGAACATAACATAACATAACGTATAACATATAACATATA
ACATATAGTAAGTAAAAAACGAAGCCTCTTGCAAAGC
TTGCTAATAGTTACACTAAGCTTTGCTAGAGATTTCG),
mut6 (TTTAACGTTTTACGCGCCGGGTATATACGTATAA
CTTAACGAATTAGACGACGTGTAAAATGTTGAAAATTC
TGAAGTCAAACGTAACGCAACATTGTACAAGTCGCAA
GGTTCTGTGTAAAAAAACAGAACATAACATAACATAA
CGTATAACATATAACATATAACATATAGTAAGTAAAAAA
CGAAGCCTCTTGCAAAGCTTGCTAATAGTTACACTAAG
CTTTGCTAGAGATTTCG), mut7 (CTTATTACATTTTATTA
TTAGAACTATCTCACTTCTGGTAATAACATGTCGTAAGA
GGCGAGAGATGATATTGGAAGATGTGATCTAGTGGTAT
CATATTTTGGACATCGTGAATTTGTGGATATTCACGTCG
TGCAACATTGTACAAGCCATAATGCTACTTATTGTTGGA
ATTGGCTTGTAAAATGTTGTTTAGATTATATTAGAGATT
ATGGTCTTCAATAATGTTTTACGAGCCAT), and mut8 (GC
ATACGTTAATGTTTTCAGGCTTCTATCAGAACCCTGGA
AATAATGATTTATGCACTCCAGAAATAGTGATTCTTTT
TTTGAGAGGTGATTATGATTTCTAATCATAAGGTACAA
CTGTCTATGTACCTTTTCTTAATATTGTACACGTTGCG
ATATGTCCCATTATTGGTGGTGATGTGTAAAATATTGA

TGTCTTTATAATCCTTTTCGTAAAGGAATTTAGTGTTTT ACGCAGACT).

Transgenes were expressed with the tub-GAL4 driver at 25°C. Polytene squashes were prepared from wandering third instar male larvae as described in Ilik et al. (2013).

Data accession

The MLE uvCLAP data in this study have been deposited in the Gene Expression Omnibus database under accession number GSE87792.

Acknowledgments

We thank current and former members of the Akhtar laboratory for helpful discussions, especially Tugce Aktas for her insights into the physiological potential of gapped alignments; Thomas Conrad and Giuseppe Semplicio for their ideas on the nature of gapped alignments and Supplemental Figure 2D; Ken Lam for his input regarding minimal roX in flies; Ulrike Bönisch; and the Deep Sequencing Facility. We are grateful to Aline Gaub for help with polytene squashes. This work was supported by Collaborative Research Centre (CRC) 992 (A.A. and R.B.), CRC 746 (A.A.), Transregio 167/1 (R.B.), Graduiertenkollegs 2344/1 (R.B.), Deutsche Forschungsgemeinschaft grant 2168/14-1 (R.B.), and CRC 1140 (A.A.).

References

Aktas T, Ilik IA, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, Manke T, Backofen R, Akhtar A. 2017. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* **544:** 115–119.

Bradley T, Cook ME, Blanchette M. 2015. SR proteins control a complex network of RNA-processing events. *RNA* **21:** 75–92.

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149:** 1393–1406.

Das U, Shuman S. 2013. Mechanism of RNA 2′,3′-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Res* **41:** 355–365.

Dimon MT, Sorber K, DeRisi JL. 2010. HMMSplicer: a tool for efficient and sensitive discovery of known and novel splice junctions in RNA-seq data. *PLoS One* **5:** e13875.

Granneman S, Kudla G, Petfalski E, Tollervey D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci* **106:** 9613–9618.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp A-C, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141:** 129–141.

Holmqvist E, Wright PR, Li L, Bischler T, Barquist L, Reinhardt R, Backofen R, Vogel J. 2016. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J* **35:** 991–1011.

Ibrahim MM, Lacadie SA, Ohler U. 2015. JAMM: a peak finder for joint analysis of NGS replicates. *Bioinformatics* **31:** 48–55.

Ilik I, Akhtar A. 2009. roX RNAs: non-coding regulators of the male X chromosome in flies. *RNA Biol* **6:** 113–121.

Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R, et al. 2013. Tandem stem–loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol Cell* **51:** 156–173.

Jankowsky E. 2011. RNA helicases at work: binding and rearranging. *Trends Biochem Sci* **36:** 19–29.

Johansson A-M, Allgardsson A, Stenberg P, Larsson J. 2011. msl2 mRNA is bound by free nuclear MSL complex in *Drosophila melanogaster*. *Nucleic Acids Res* **39:** 6428–6439.

Kelley RL, Lee O-K, Shim Y-K. 2008. Transcription rate of noncoding roX1 RNA controls local spreading of the *Drosophila* MSL chromatin remodeling complex. *Mech Dev* **125:** 1009–1019.

Kleinkauf R, Mann M, Backofen R. 2015. antaRNA—ant colony based RNA sequence design. *Bioinformatics* **31:** 3114–3121.

Koh HR, Xing L, Kleiman L, Myong S. 2014. Repetitive RNA unwinding by RNA helicase A facilitates RNA annealing. *Nucleic Acids Res* **42:** 8556–8564.

König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17:** 909–915.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645.

Lange SJ, Maticzka D, Möhl M, Gagnon JN, Brown CM, Backofen R. 2012. Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res* **40:** 5215–5226.

Lee CG, Chang KA, Kuroda MI, Hurwitz J. 1997. The NTPase/helicase activities of *Drosophila* maleless, an essential factor in dosage compensation. *EMBO J* **16:** 2671–2681.

Li F, Schiemann AH, Scott MJ. 2008. Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the *Drosophila* MSL1/MSL2 complex. *Mol Cell Biol* **28:** 1252–1264.

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6:** 26.

Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, Davidovich C, Gooding AR, Goodrich KJ, Mattick JS, et al. 2016. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165:** 1267–1279.

Maenner S, Müller M, Fröhlich J, Langer D, Becker PB. 2013. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol Cell* **51:** 174–184.

Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR. 2003. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113:** 577–586.

Maticzka D, Ilik IA, Aktas T, Backofen R, Akhtar A. 2017. uvCLAP: a fast, non-radioactive method to identify in vivo targets of RNA-binding proteins. bioRxiv doi: 10.1101/158410.

Park S-W, Kang YI, Sypula JG, Choi J, Oh H, Park Y. 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.

Quinn JJ, Ilik IA, Qu K, Georgiev P, Chu C, Akhtar A, Chang HY. 2014. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat Biotechnol* **32:** 933–940.

Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY. 2016. Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes Dev* **30:** 191–207.

Rajan Prabu J, Müller M, Thomae AW, Schüssler S, Bonneau F, Becker PB, Conti E. 2015. Structure of the RNA helicase MLE reveals the molecular mechanisms for uridine specificity and RNA-ATP coupling. *Mol Cell* **60:** 487–499.

Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42:** W187–W191.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44:** W160–W165.

Reenan RA, Hanrahan CJ, Ganetzky B. 2000. The mle(napts) RNA helicase mutation in *Drosophila* results in a splicing catastrophe of the para Na^+ channel transcript in a region of RNA editing. *Neuron* **25:** 139–149.

Rogers GW Jr, Lima WF, Merrick WC. 2001. Further characterization of the helicase activity of eIF4A. Substrate specificity. *J Biol Chem* **276:** 12598–12608.

Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, Dalmay T. 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3:** 4.

Steitz TA, Moore PB. 2003. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* **28:** 411–418.

Tagwerker C, Flick K, Cui M, Guerrero C, Dou Y, Auer B, Baldi P, Huang L, Kaiser P. 2006. A tandem affinity tag for two-step purification under fully denaturing conditions application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. *Mol Cell Proteomics* **5:** 737–748.

Villa R, Schauer T, Smialowski P, Straub T, Becker PB. 2016. PionX sites mark the X chromosome for dosage compensation. *Nature* **537:** 244–248.

Yang Y-CT, Di C, Hu B, Zhou M, Liu Y, Song N, Li Y, Umetsu J, Lu ZJ. 2015. CLIPdb: a CLIP-seq database for protein–RNA interactions. *BMC Genomics* **16:** 1–8.

Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe NM, Ule J. 2013. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152:** 453–466.