SUPPLEMENTARY INFORMATION

Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts

Jason C. Bell[1,†], David Jukam[2], Nicole A. Teran[1,3], Viviana I. Risca[3], Owen K. Smith[1,5], Whitney L. Johnson[1], Jan Skotheim[2], William J. Greenleaf [3,4] and Aaron F. Straight[1,2,5,†]

Departments of Biochemistry[1], Biology[2], Genetics[3], Applied Physics[4], and Chemical & Systems Biology[5]
Stanford University, Stanford, CA 94305

[†]Correspondence and requests for materials should be addressed to:

Jason C. Bell & Aaron F. Straight
jcbell@stanford.edu & astraigh@stanford.edu
Beckman Center for Molecular & Genetic Medicine B409
Stanford University
279 West Campus Drive
Stanford, CA 94305

Supplementary Figures Legends

Supplementary Figure 1: Diagram of the oligonucleotide 'bridge' and efficiency of bridge ligation and capture. The oligonucleotide 'bridge' contains a 5'-adenylated (App) six nucleotide ssDNA tail (green), a single biotin modification (purple), a DpnII site and 3'-three carbon spacer (Sp3), which prevents circularization during the adenylation of the 5'-end. The 3'-Sp3 is removed from the bridge during genomic DpnII digestion, permitting subsequent ligation to genomic DNA. *Lower panel*, Bar plot of the fraction of reads at each step of the data processing pipeline from a representative library preparation.

Supplementary Figure 2: *In vitro* optimization of RNA-to-DNA ligation conditions. *Upper panel*, Ten pmols of 17-nt adenylated ssDNA oligonucleotide (Universal App DNA) was incubated with 5 pmols of a 17nt ssRNA test probe in 1x NEB T4 RNA Ligase Buffer with the indicated ligase (Therm 5' Ligase, trT4K or trT4K9) and/or supplements (PEG, BSA, ATP, RNaseOUT). Products were then analyzed using denaturing polyacrylamide gel electrophoresis and stained with SYBR-gold. Bands were quantified and the percent product was calculated using (shifted / (total * 0.66)) to account for the molar excess of DNA over RNA. No adjustment was made to account for preferential staining of ssDNA over ssRNA. Residual signal is expected in the lower band owing to the molar excess of DNA over RNA.

Supplementary Figure 3: DpnII cut site frequency in the fly genome. DpnII sites were identified genome-wide using HOMER, which we then used to calculate the expected fragment size distribution, plotted as cumulative distribution with respect to fragment size. Dotted lines represent the 50th, 75th and 90th percentile, and are presented for visual clarity.

Supplementary Figure 4: Diagram of the ChAR-seq data processing pipeline. Data was processed using a custom pipeline, which can be accessed and is fully documented at: https://gitlab.com/charseq/flypipe. Red lines indicate reads that do not align to any transcriptome in the sense orientation, and are then permited to test alignment in the antisense orientation.

Supplementary Figure 5: ChAR-seq RNA-to-bridge ligation is sensitive to RNaseA. Bar plot of the number of reads containing bridge after PCR duplicate removal and relative quantification of the fraction of those reads that contained both RNA and DNA on each side. The RNase-treated cross-linked sample was incubated with 0.25 mg/mL RNase A and 12.5 Units of RNase H for 1 hour at 37°C between steps 4 and 5 of the extended protocol, followed by an additional wash step identical to step 4.

Supplementary Figure 6: Abundance of cis contacts. For each gene in our dataset, we functionally defined cis contacts as RNA-to-DNA contacts that lie within the gene body (+/- 2 kb) for a given RNA (i.e,, contacts that arise from nascent transcription). We then calculated a cis score, which is equivalent to the percentage of contacts that arise from this region. Upper plot is the per gene rank order analysis based on the cis score for each RNA in our dataset. Lower plot is a histogram of the frequency distribution for each cis score (percentage).

Supplementary Figure 7: Hitlist of chromatin-associated RNAs based on number of contacts and chromatin-enrichment. RNAs are sorted according to their fold-enrichment (right-most bars) relative to RNA expression levels. Left-most bars represent the number of contacts (FPKM) for each RNA. List was compiled based on analysis in **Figure 2D** and **Figure 2E**.

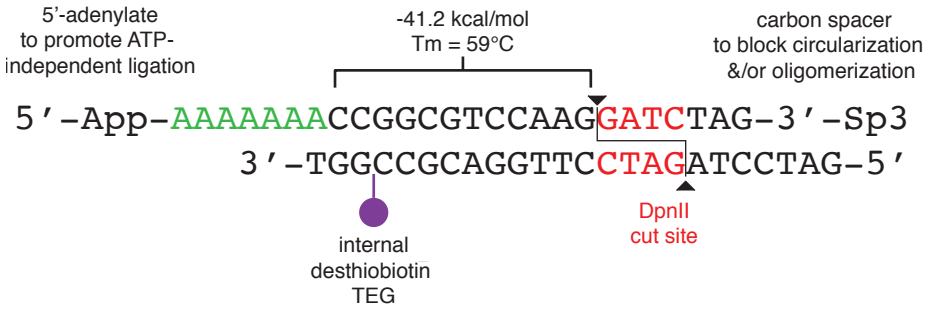
Supplementary Figure 8: Hitlist of chromatin-associated RNAs based on number of contacts and chromatin-enrichment. RNAs from Supplementary Figure 7 plotted as a bubble plot, where the x-axis represents the fold-enrichment relative to RNA expression levels (y-axis). The size of each bubble is

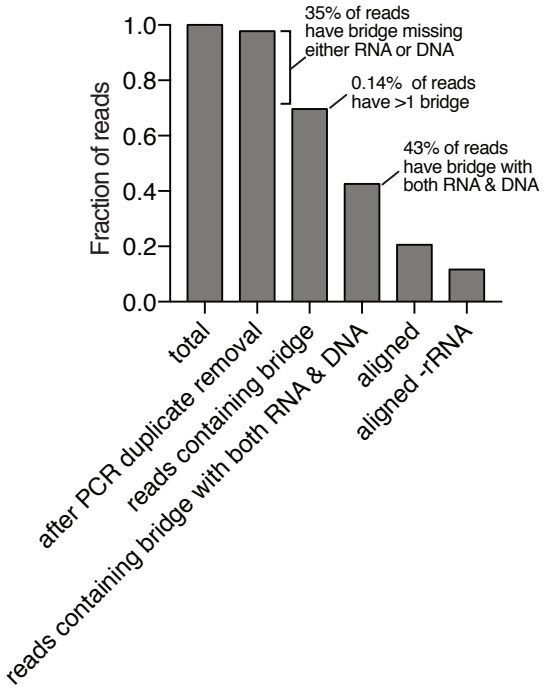
scaled according to the corresponding cis contact score (i.e., the percentage of contacts for that RNA found over its own gene body +/- 2 kbp.

Supplementary Figure 9: Comparison of RNA-to-DNA contacts between cell type. Scatter plot of the number of contacts for chromatin-associated RNAs identified in CME-W1-cl8+ (x-axis) and Kc167 (y-axis) cell lines.

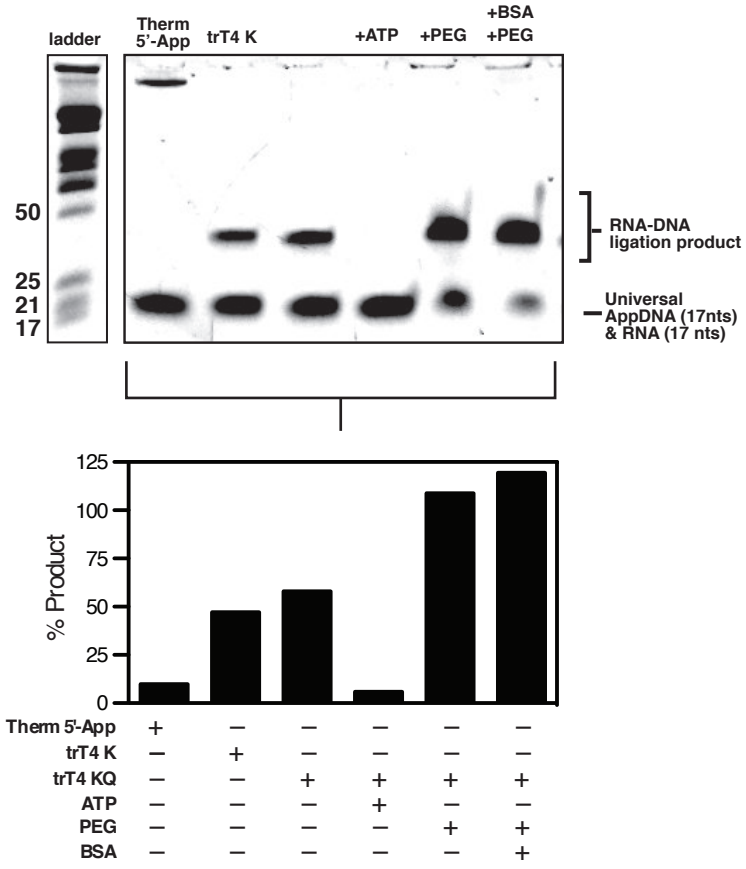
Supplementary Figure 10: Sequence similarity of snRNAs. Pairwise alignment of snRNA gene variants annotated in FlyBase was performed using Genious R7 using a global alignment with free-end gaps and a 51% similarity cost matrix and the Tamura-Nei distance model. Distance scores were then clustered using the heatmap module in R. Low scores (blue) represent a high similarity, while a higher score (red) represents higher levels of dissimilarity.

Supplementary Figure 1

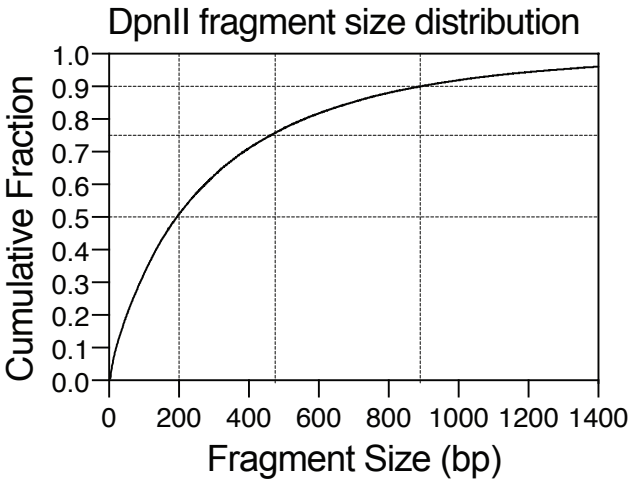




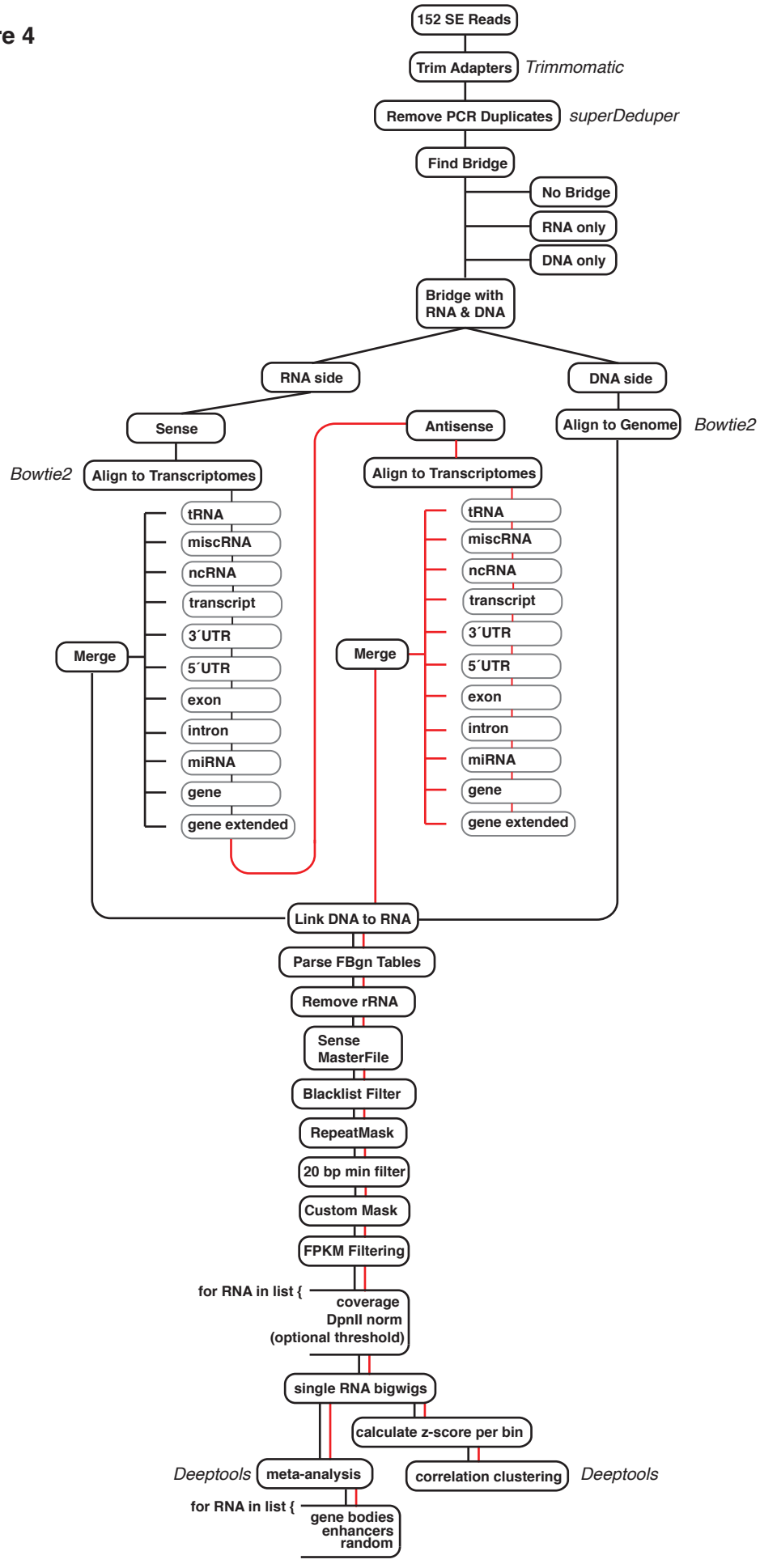
Supplementary Figure 2



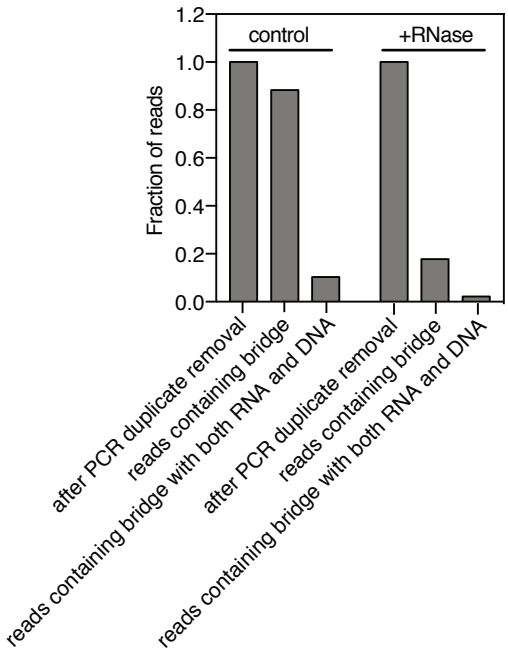
Supplementary Figure 3



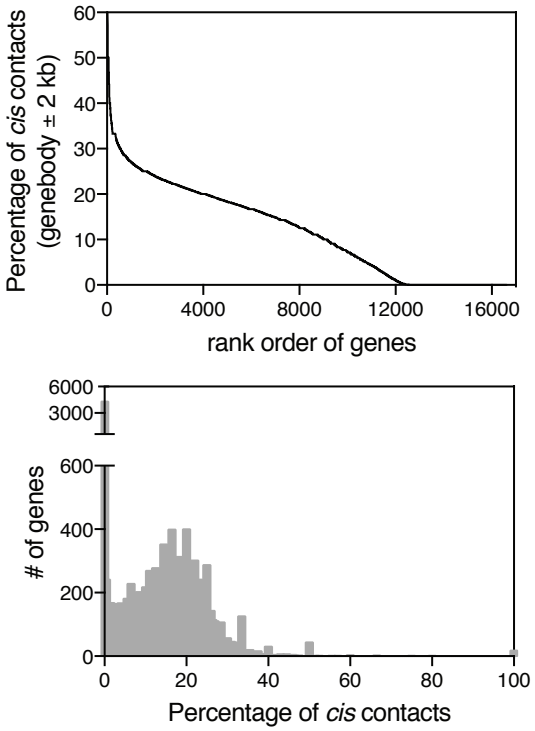
Supplementary Figure 4



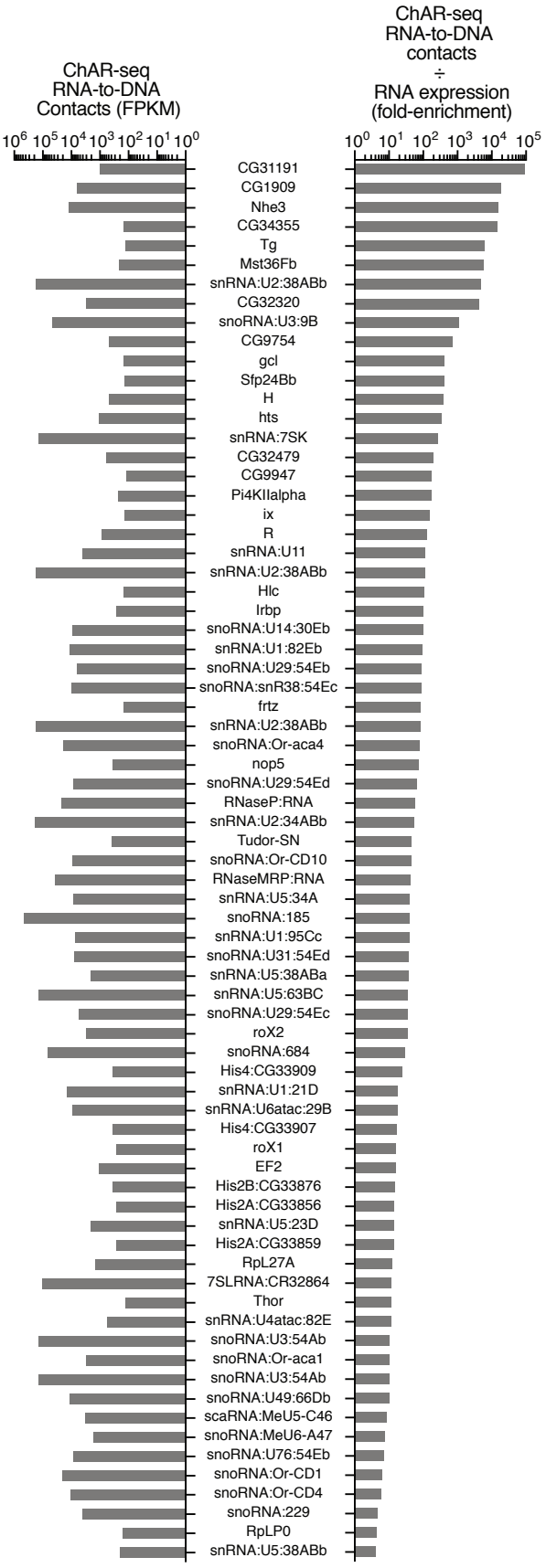
Supplementary Figure 5



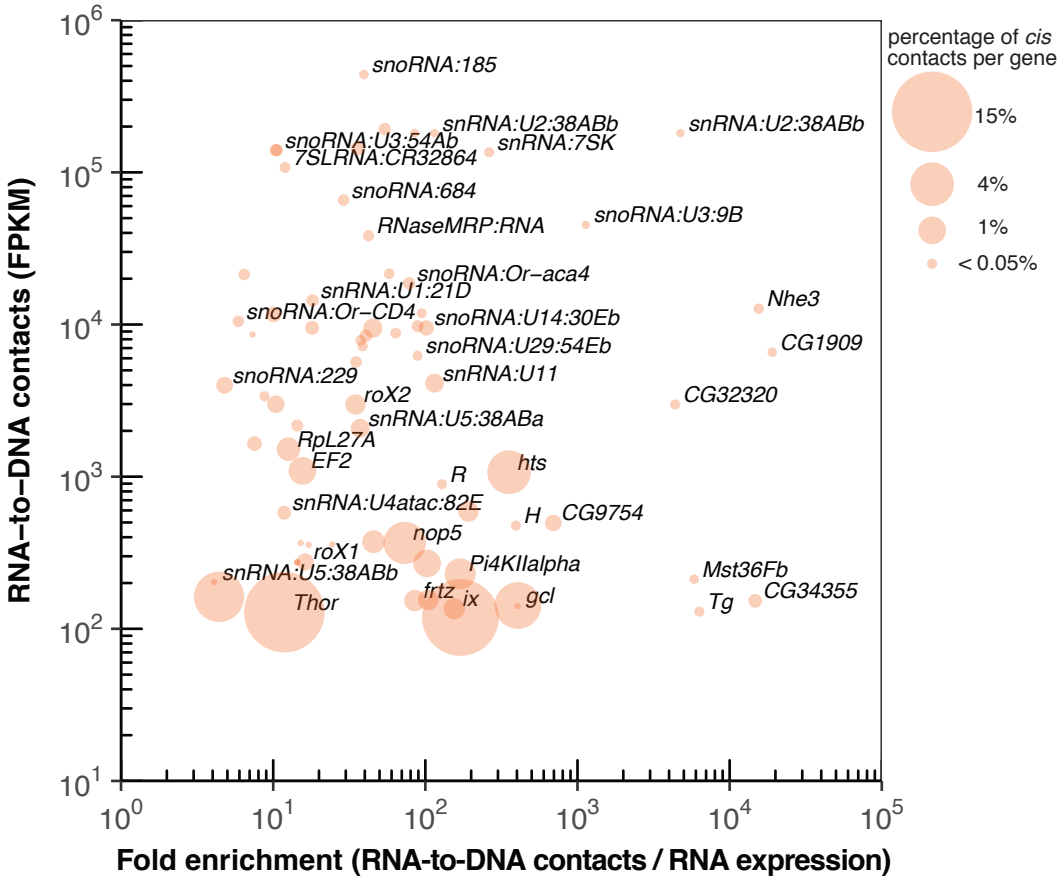
Supplementary Figure 6



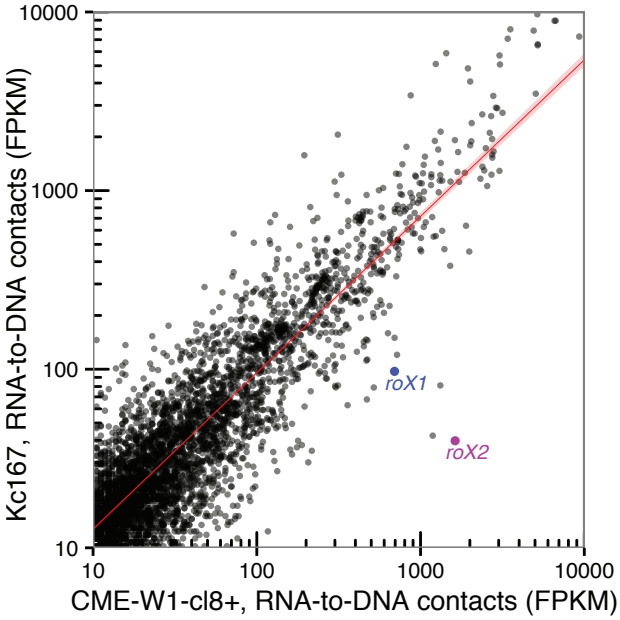
Supplementary Figure 7



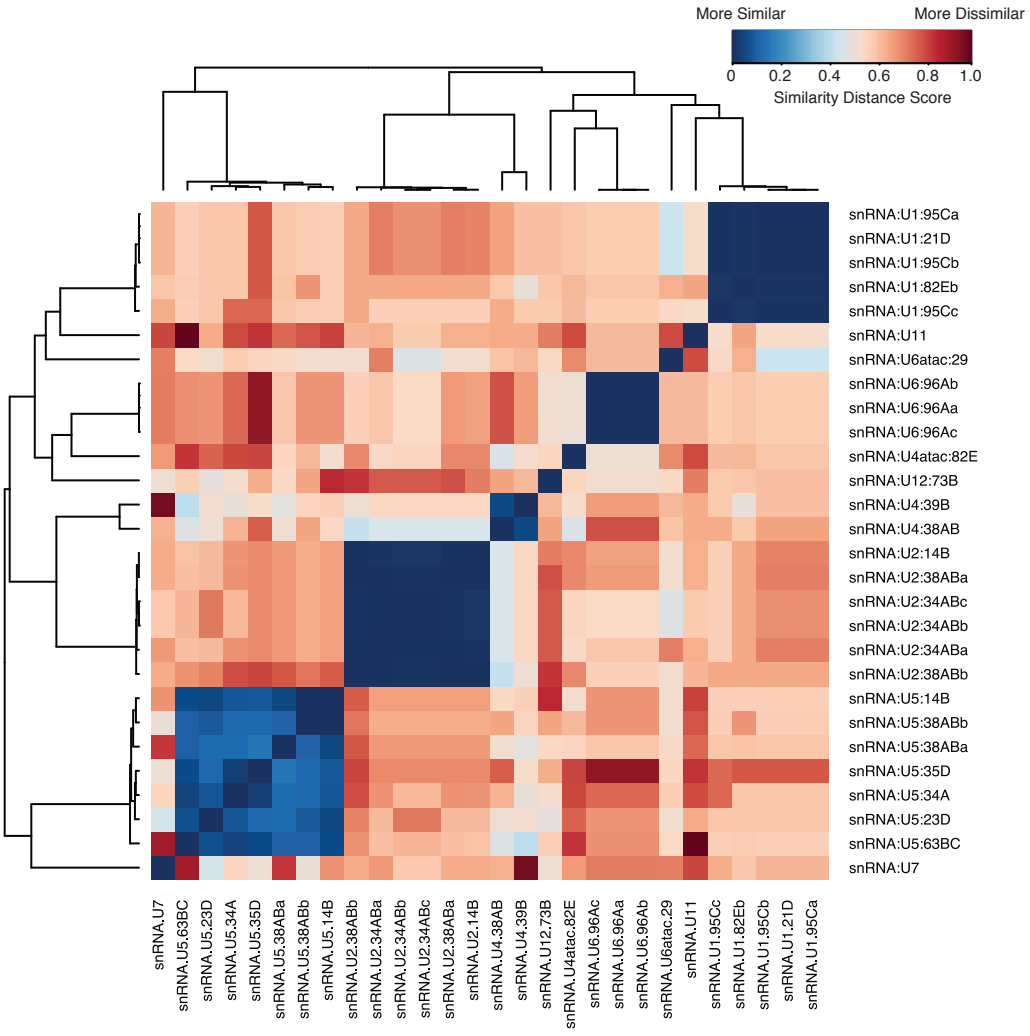
Supplementary Figure 8



Supplementary Figure 9



Supplementary Figure 10



Extended Methods and Detailed ChAR-seq Protocol

Shields and Sang M3 Media, for 0.5 L
- 19.6 g S&S M3 powder (sigma)
- 0.25 g bicarb
- 5 ug/mL insulin
- 2% FBS
- 2% fly extract
- 1x Pen/Strep

Sterile filter media using 0.2 um filter. Store media at 4C. Warm to room temperature before splitting cells. Culture cells at 27C in 75 flasks (1-2 flasks per library), splitting or harvesting approximately every 3-4 days. Cells are detached from the flask surface by vigorously pipetting up and down with a serological pipette. Confirm that >95% of cells are healthy and viable using trypan blue assay at each harvest.

All buffers and solutions prepared with DEPC-treated water. All buffers are made with fresh, unopened chemicals to minimize RNase contamination. All enzyme stocks and reagents are dedicated for RNA work and kept RNase-free.

Step 1: Harvesting cells & crosslinking (can be done in advance)
- Pipette cells up and down to detach them from the surface. The cells grow in small strings or clumps that are very weakly attached to the surface.
- Count the cell using a hemocytometer. Typically, each T75 flask should yield between 100-150 M cells.
- Spin (~2000xg) the cells in either one or two 50 µL conical tubes in a swinging bucket centrifuge for 5 minutes.
- Resuspend the cells in ~36 mL of fresh media
- Add formaldehyde to 1% (1 mL of 37% form)
    - incubate at RT for 10 minutes mixing end over end
- Quench by adding 3 mL of 2.5 M glycine (final concentration = 0.2 M), mix for 5 minutes end over end
- Spin the cells at 2000xg for 5 min
- Resuspend in ~1 mL of sterile filtered, ice cold PBS + DEPC
    - count cells and split into an appropriate number of tubes to aliquot 100M cells per tube
- Spin again at 2000xg for 5 minutes
- Aspirate off supernatant and flash freeze in liquid nitrogen

Step 2: Lysis (Day 1, start in early afternoon)

Lysis buffer:
- 10 mM TrisHCl pH 8
- 10 mM NaCl
- 0.2% igepal
- 1 mM DTT
- 0.2 mM EDTA

- Mix 750 µL lysis buffer + 150 µL Sigma protease inhibitor (P8340, comes as a DMSO stock) + 45 µL RNaseOut
- Add lysis buffer plus inhibitors to STILL FROZEN crosslinked cells and thaw with gentle mixing. Incubate for 2 minutes on ice after thawing is complete. *DO NOT let cells thaw on ice before adding lysis buffer!*
- Centrifuge at 2.5k x g for 2 minutes, discard supernatant.
- Wash pellet with 500 µL of lysis buffer + inhibitors, spin again for 1 min.
- Gently resuspend the pellet in 100 µL 0.5% SDS and incubate at 62C for 5 minutes
- Pre-mix 290 µL water + 50 µL 10% Triton-X 100
- Quench SDS with diluted Triton-X 100. Mix well by gentle pipetting; incubate at 37C for 15 minutes.

Step 3: RNA Fragmentation (Day 1)
- Add 0.25X (~7 µL) T4 Rnl2tr KQ (trKQ) RNA ligase buffer (final concentration ~2.5 mM Mg^{+2})
- Incubate for **exactly 4 min**, 70'C

- Snap cool on ice

Step 4: Wash cells to remove SDS and non-crosslinked RNA fragments (Day 1)

- Centrifuge at 2.5k x g for 2 minutes, discard supernatant
- Add 1000 µL DEPC-treated PBS, mix gently
- Centrifuge at 2.5k x g for 2 minutes, discard supernatant
- Immediately proceed to the next step, with pre-mixed reaction buffer already prepared

Step 5: RNA to linker ligation, RNA first, Day 1

- Pre-mix the following and then add to cells
    - 20 µL 10X T4 trKQ RNA ligase buffer
    - 100 µL 50% PEG
    - 20 µL RNaseOUT
    - 20 µL 50 uM annealed bridge (2.5 nmols)
    - 10 µL T4 trKQ RNA ligase (4000 units)
- Incubate at 23C overnight, shaking 900 rpm in ThermoMixer

END DAY 1

START DAY 2

Step 6: Wash to remove PEG (Day 2, morning)

- Centrifuge at 2.5 k x g for 4 minutes at RT, remove and discard supernatant
- Resuspend pellet with 1 mL DEPC-treated PBS

NOTE, this applies for all subsequent steps using PBS

- Centrifuge at 2.5 k x g for 5 minutes at room temperature
- While spinning, make 200 µL 1x T4 trKQ RNA ligase buffer
    - add 0.2 µL 1 M DTT
    - add 10 µL RNaseOUT
- Resuspend pellet in 200 µL 1x T4 trKQ RNA ligase buffer + DTT and RNaseOUT

Step 7: First strand synthesis, Bst 3.0 (Day 2, morning)

- Add 20 µL 10 mM (each) dNTP mix
- Add 15 µL Bst 3.0 (120 U/ µL) for 1800 Units
- Incubate for 10 minutes at RT
- Increase temp to 37C for 10 minutes (shaking 900 rpm in ThermoMixer)
- Increase temp again to 50C for 20 minutes
- Add 1mL ice cold PBS, mix, and cool on ice 1 minute

Step 8: Inactivate and Wash (Day 2)

- Centrifuge at 2.5k x g for 2 minutes, discard supernatant
- Repeat wash with another 1 mL of PBS
- Centrifuge at 2.5k x g for 2 minutes, discard supernatant
- Resuspend pellet in 100 µL 0.5% SDS + 0.1 mM EDTA
- Heat inactivate for 10 min at 55C
- Quench SDS with 290 µL water + 50 µL 10% Triton-X 100 (**pre-mixed**). Mix well, incubate at 37C for 15 minutes.
- Centrifuge at 2.5 k x g for 2 minutes, remove and discard supernatant.
- Wash nuclear pellet by adding 1 mL PBS, mix, and spin at 2.5k x g for 2 minutes
- Remove and discard supernatant, and proceed IMMEDIATELY to digestion.

NOTE: we find it useful to remove the supernatant in this and all subsequent steps using a mechanical micropipette rather than an aspirator, which risks loss of sample.

Step 9: Genomic DNA digestions (Day 2, afternoon)

- Resuspend in 200 µL 1x T4 RNA ligase + 1 mM DTT + 10µl RNaseOUT (pre-mixed).
- Add 15 µL DpnII (750 Units)
- Incubate at 37C for at least 3 hours, up to overnight if desired.
- Terminate digestion by adding EDTA to a final concentration of 15 mM

Step 10: Pellet & wash cells

- Spin the cells at 2500 x g for 2 min
- Remove and discard the supernatant
- Re-suspend the X-linked cells with 1 mL PBS (DEPC),
- Spin again at 2500 x g for 2 min
- Remove and discard the supernatant using a P200

Step 11: Inactivate DpnII

- Re-suspend the washed cells in 200 µL 0.5% SDS + 0.1 mM EDTA
- Incubate at 55C for 10 minutes
- Quench SDS by adding 240 µL 4.2% TritonX-100
- Incubate at 37C for 5 minutes
- Add 1 mL of PBS and incubate on ice for 5-10 minutes

Step 12: Pellet & wash cells three times

- Spin the cells at 2500 x g for 2 min
- Remove and discard the supernatant
- Re-suspend the pellet with 1 mL PBS
- Spin again at 2500 x g for 2 min
- Remove and discard the supernatant
- Re-suspend the pellet with 1 mL PBS
- Spin again at 2500 x g for 2 min
- Remove and discard the supernatant

Step 13: DNA-linker Ligation (Day 2 overnight to Day 3)

- Resuspend pellet in 200 µL 1x T4 DNA ligase buffer (pre-mixed)
- Add 2000 U T4 ligase (5 µL of 400 U/ µL)
- Incubate at 16C overnight

END DAY 2

START DAY 3

Step 14: Pellet and wash cells, to remove T4 ligase

- Add 7.5 µL 0.5 M EDTA (~15 mM final)
- Spin the cells at 2500 x g for 2 min
- Remove and discard the supernatant
- Re-suspend the pellet with 1 mL PBS
- Spin the cells at 2500 x g for 2 min
- Remove and discard the supernatant
- Re-suspend the pellet with 1 mL PBS
- Re-suspend each pellet in 100 µL 1x trKQ RNA ligase buffer + 1 mM DTT + 1/20 volume RNaseOUT

Step 15: Second strand synthesis

- 2x cDNA bufffer
    - 10 mM TrisHCl (8.0)
    - 180 mM KCl
    - 100 mM $(NH_4)_2SO_4$
- Add 100 µL 2x cDNA buffer

- Add 20 µL 10 mM dNTP (each) mix
- Add 100 Units E. coli DNA Polymerase I (NEB)
- Add 5 Units RNaseH (NEB)
- Mix and incubate at 16C for at least two hours

Step 16: Crosslink reversal (end of Day 3)

- Centrifuge at 2.5k x g for 5 minutes
- Remove and discard supernatant
- Resuspend pellet in 400 µL PBS
- To ~ 450 µL of sample, add
    - 55 µL 10% SDS
    - 55 µL 5 M NaCl
    - 3 µL 20 mg/mL ProteinaseK
    - incubate overnight at 70C

END DAY 3

START DAY 4

Step 17: Precipitate DNA

- Cool sample to room temperature
- Add 50 µL 3 M NaOAc
- Add 2 µL (5mg/ml) glycogen
- Add 850 µL ice cold 100% EtOH.

NOTE: Flocculent should be visible after inverting the tube gently 3-4 times. If no flocculent is visible, add 200 µL of additional ice cold 100% EtOH.

- Incubate on ice or at -20C for 30 minutes
- Centrifuge at ~21,000 x g for 5-10 minutes
- Wash pellet with 70% ice cold EtOH
- Gently dry pellet and resuspend in 130 µL TE

Step 19: Covaris shear DNA

- Shear DNA to ~200 bp fragments using the following settings: 10% duty factor, 175 peak incident power, 200 cycles per burst, for 180 seconds

Step 19: Isolation of biotinylated DNA fragments

1X Tween Wash Buffer (TWB)
5 mM TrisHCl (7.5)
0.5 mM EDTA
1 M NaCl
0.05% Tween 20

2X Bead Binding Buffer (BBB)
10 mM TrisHCl (7.5)
1 mM EDTA
2 M NaCl

- Wash 150 µL Dynabeads MyOne SA T1 beads (Life Tech) with 400 µL TWB. Pull down beads with a magnetic tube rack and discard supernatant.
- Resuspend the beads in 130 µL 2X BBB
- Mix the bead slurry with the DNA (final volume of 260 µL)
- Incubate at RT for 15 minutes with rotation to bind biotinylated DNA
- Pull down beads with a magnetic tube rack and discard supernatant.

NOTE: Do not allow the beads to dry out.

- Wash the beads by adding 750 µL 1x TWB and mixing

- Warm the tubes to 50C for 2 minutes
- Pull down beads with a magnetic tube rack and discard supernatant
- Wash the beads by adding 750 µL 1x TWB and mixing
- Pull down beads with a magnetic tube rack and discard supernatant

Step 20: End Repair, and dA tailing

- Resuspend the beads in 40 µL TE
- add 7 µL NEB Next End Prep Buffer
- add 3 µL NEB Next End Prep Enzyme Mix
- Incubate at RT for 20 minutes with agitation to keep beads suspended
- Increase temp to 65C for 30 minutes
- Cool to room temperature

Step 21: Adapter ligation (on bead)

NOTE: The following order of addition is important.

- add 2.5 µL NEBNext Adaptor **FIRST**
- add 1 µL ligation enhancer
- add 30 µL NEBNext UltraII Ligation Master Mix **LAST**
- mix vigorously, incubate at **Room Temperature** for 15-20 minutes with agitation to keep beads suspended
- add 3 µL USER enzyme
- incubate at 37C for 15 minutes

Step 22: Wash beads

- Pull down beads with a magnetic tube rack and discard supernatant
- Wash the beads by adding 750 µL 1x TWB and mixing
- Warm the tubes to 50C for 2 minutes
- Pull down beads with a magnetic tube rack and discard supernatant
- Wash the beads by adding 750 µL 1x TWB and mixing
- Pull down beads with a magnetic tube rack and discard supernatant

Step 23: On bead amplification (library PCR 1)

- For each index, pre-mix the following
    - 30 µL 2x NEB Next High Fidelity master mix
    - 3 µL 10 uM Universal Primer (NEBNext Multiplex Oligos for Illumina)
    - 3 µL 10 uM Indexing Primer (NEBNext Multiplex Oligos for Illumina)
    - 24 µL water
- Resuspend the beads in 50 µL amplification mix (reserve the extra 10 µL as you will need it in **step 25**)
- Run the following PCR cycle
    - 98C for 3 min
    - then 8 cycles of
        - 98C for 30s
        - 63C for 30s
        - 72C for 40s
- Stop the PCR, pull down beads in a magnetic rack

Step 24: First Ampure XP size selection to remove adaptor dimers (target size 200 to 500 bp)

NOTE: The NEBNext adaptor is a 65 nt hairpin… adaptor dimers would be 65 bp, because each hairpin is ~ 32 bp long… however, the indexing primers are also 65 nts, so amplification of adaptor dimer results in a 130 bp fragment that must be removed at this step, and adds 130 bp to the ideal target length of 150 bp, our molecules of interest are now 280-300 bp

- Allow Ampure bead slurry to warm to room temperature
- Add an equal volume of Ampure beads to the PCR supernatant, mix by pipetting and incubate for 5 minutes at room temperature
- Pull down beads with magnet, give 1-2 minutes to clear solution. Remove supernatant.

NOTE: This is, in theory, to be discarded, but save until you've validated the completed library prep by qPCR and

bioanalyzer in case needed for troubleshooting

- Wash the beads with 70% EtOH (e.g., 700 µL), pull down, remove sup, and
- elute with 30 µL of 10 mM Tris pH 8

Step 25: qPCR to determine how many more cycles to amplify

- Mix 5 µL of each library eluted from the Ampure beads with 10 µL of the reserved PCR mix from **step 23**
- Dilute 100x Syber Green 1:3 in water for a 33x working solution
- Add 0.5 µL to each 15 µL reaction for ~1x
- Run on qPCR to determine the number of cycles to amplify… target should be the number of cycles required to reach ~1/3 of the saturation

Step 25: <u>Off</u> bead amplification (library PCR 2)

- For each index, pre-mix the following
- 30 µL 2x NEB Next High Fidelity master mix
- 25 uL eluted library from **Step 24**
- 2.5 µL 10 uM Universal Primer (NEBNext Multiplex Oligos for Illumina)
- 2.5 µL 10 uM Indexing Primer (NEBNext Multiplex Oligos for Illumina)
- Run the following PCR cycle
    - 98C for 3 min
    - then N cycles of
        - 98C for 30s
        - 63C for 30s
        - 72C for 40s

*NOTE: the number of additional cycles, N, is ascertained by estimating the number of additional cycles required to reach 1/3 saturation from the qPCR analysis in **Step 24.** The number of cycles will vary between each sample library.*

Step 26: High and low size selection using Ampure XP Beads (final polishing)

- Allow Ampure bead slurry to warm to room temperature
- Add 0.5 volumes of Ampure bead slurry (30 µL) to PCR mix (60 µL). Mix by pipetting and incubate at RT for 5 minutes.
- Pull down beads with magnet, give 1-2 minutes to clear solution, then recover the supernatant. This should contain fragments from 0-500 bp. Larger fragments, if you want them, are bound to the beads. Add 70% EtOH to the beads and set aside.
- To the supernatant (90 µL), add 0.5 volumes of original volume (30 µL) for a final bead to input ratio of 1:1 (total volume should be 120 µL). Mix and incubate at RT for 5 minutes.
- Pull down the beads and remove the supernatant, which should contain small, unwanted fragments (adaptor dimers, etc). Nonetheless, save it for now just in case and set aside.
- Wash the beads with 700 µL 70% EtOH, pull down the beads in a magnetic tube rack until clear (1-2 min), discard supernatant, and repeat for a total of two washes. Do not disrupt the beads during the second wash.
- Remove the EtOH completely and let the beads passively dry for ~5 minutes
- Add 50 µL of elution buffer (**20 mM Tris pH 8**). Incubate for 2-3 minutes. Pull down the beads until clear and remove the supernatant. This should contain fragments from ~100-500 bp.

Step 27: Quality control

- QC library by Bioanalyzer (1:10 dilution, High Sensitivity)
- Quantify library by phiX qPCR

Sequencing. ChAR-seq libraries were sequenced with single-end 152-bp reads on the Illumina MiSeq (v3_150 cycle kit) and Next-seq according to manufacturer's instructions.

Data processing & software. Reads were processed using a custom pipeline, which can be accessed at: https://gitlab.com/charseq/flypipe. PCR duplicates were removed using Super Deduper[39] using the entire read length. Adapters were removed and reads trimmed for low quality with Trimmomatic[40] using a composite set of Illumina adapters, leading/trailing cut length of 3, sliding window of 4:15, and a minimum length of 36. RNA and DNA portions of each read are identified and split by a custom python script that finds the bridge, uses the bridge polarity to identify which side of the read is RNA or DNA, and verifies that there is only a copy of the bridge. All reads lacking a full bridge sequence or that contained 2 or more bridges were removed. Only reads that contained sequence on both sides of the bridge (RNA and DNA) were processed further. Unique read IDs from the initial read containing the full molecule remained linked to both the RNA and DNA sequences after splitting from the bridge sequence. Additional software used by FlyPipe or during preparation of figures include: bowtie2[41], SAMtools[42], BEDtools[43], MACS2[44], Circos[45], R, GraphPad Prism v7.0, deepTools2[46], HOMER[47], and Genieous[48].

ChAR-seq DNA and RNA alignment. Four replicates were processed and merged for analysis in the following way. Reads were aligned to *Drosophila melanogaster* genome (dm3, release r5.57 using bowtie2[41] with the --very-sensitive option and allowing one mismatch. RNA was aligned separately to all dm3 transcriptomes from FlyBase (downloaded January 2016, versions r5.57): with the same bowtie2 parameters, but also first forcing forward strandedness as bridge orientation should preserve the RNA strand information, and then for reads that did not align in the sense direction, permitting antisense alignment. Reads that aligned with equal alignment score to more than one transcriptome were filtered based on a priority rank according to the following order: tRNA, miscRNA, ncRNA, transcript, three_prime_UTR, five_prime_UTR, exon, intron, miRNA, gene, gene_extended2000. For a given read, this rank preserves the annotation information from the top ranked transcripome. For example, a read aligning to the "ncRNA", "transcript", and "exon" transcriptomes would receive ncRNA annotations, and be considered a ncRNA in further analysis. Reads aligning to tRNA and miscRNA groups were removed from further analysis. Reads lacking valid sense alignments to any transcriptome, but that aligned in the antisense orientation were ranked and filtered in the same manner. Previously linked RNA and DNA sequences with sense alignments had their information re-associated using the original read ID. These 1-to-1 links are deemed "RNA-DNA contacts". Any RNA-DNA contact that contained Ribosomal RNAs (rRNA) were removed from further analysis. Additional processing was then performed to remove mapped DNA contacts that overlapped regions of poor mappabilty using the modEncode Drosophila blacklist[49] (https://sites.google.com/site/anshulkundaje/projects/blacklists), repetitive regions downloaded from the UCSC Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables) using the following selection parameters: clade: *Insect*, genome: *D. melanogaster*, assembly: *Apr. 2006 (BDGP R5/dm3)*, group: *Variation and Repeats*, track: *Repeatmasker*. Several RNAs that were re-annotated as rRNA in later *Drosophila* genome annotation versions were also removed.

RNA coverage tracks: Individual RNA tracks were generated by extracting contact information from the BED-formatted table generated by FlyPipe (described above). Coverage was calculated using BEDtools for 200 bp windows, and the number of contacts was then normalized by dividing by the number of DpnII sites for each window throughout the genome. Coverage tracks for the metagene profile analysis were similarly normalized, but were extracted and normalized using a sliding window method with 200 bp bins and 20 bp steps. For the correlation clustering, we normalized the X chromosome signal to the autosomes by doubling the raw contacts on chrX and chrXHet before DpnII normalization to account for the presence of only a single X chromosome in the CME-W1-cl8+ cell line. Where indicated, coverage tracks were calculated for 2 kb bins, then converted to the log2 ratio of the contacts and DpnII frequency, which was subsequently used to calculate a z-score for each bin based on the whole genome mean and standard deviation for each RNA signal track.

Correlation clustering and metaplots: We used deepTools (https://github.com/fidelram/deepTools) to calculate the correlation between coverage tracks using the multiBigWigSummary tool excluding the following chromosome regions: chrU, chrUextra, ChrYHet and chrM. Coverage was calculated using 100 kb bins unless indicated. The Pearson correlation coefficients were then clustered and plotted using the either the PlotCorrelation function in deepTools (snRNA clustering) or the heatplot function in R using the ward.D2 method and Euclidian distance function (modENCODE vs ChAR-seq correlation clustering). modENCODE datasets (M-values, wig-formated) for CME-W1-cl8+ were downloaded from data.modencode.org, and then re-binned into 2kb windows, and filtered to remove modENCODE blacklist sites and repeat-regions, chrYHet, chrM, chrU and any bins lacking DpnII sites. modENCODE tracks were then transformed into z-scores by dividing the mean shifted (log2 M-values) by the genome-wide standard deviation of the bin depth. Metaplots were generated using the computeMatrix and plotProfile tools with a 2kb window up and downstream of each region. Signal was then normalized to fold-change relative to mean of the random signal and re-ploted in GraphPad Prism.

ChIRP-seq and ChAR-seq correlation analysis.
ChIRP-seq binding profiles for *roX1* and *roX2* were obtained from published data (GSE53020_roX1_merge.bw, GSE53020_roX2_merge.bw)[23]. The *Drosophila melanogaster* genome (dm3) was divided into equally sized windows for a range of distinct window sizes (50 bp to 1 MB). Read counts per window were obtained using BEDtools. ChIRP-seq and ChAR-seq data were filtered identically to remove modENCODE blacklist sites and repeat-regions, chrYHet, chrM, and any bins lacking DpnII sites. *roX1* and *roX2* origins were excluded due to extreme read pileups in ChIRP-seq data[23]. Spearman's rank correlations of read-counts-per-bin across the genome between ChAR-seq and ChIRP-seq were performed in R and plots graphed and fitted with Prism.

RNA-seq library preparation. For total RNA-seq analysis, total RNA from 10-20 million CME-W1-cl8+ cells was purified using TriPure, then treated with 10 units of TURBO DNase (Life Technologies) at 37°C for 30 min according to the manufacturer's instructions. RNA was re-purified with TriPure, resuspended in DEPC-treated water, and quality checked by Bioanalyzer (Agilent). Ribosomal RNAs were depleted using the Ribo-Zero rRNA removal kit (Illumina), RNA was purified using Agencourt Ampure XP beads (Beckman Coulter), then cDNAs were generated, amplified, and indexed with the ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre) according to manufacturer's instructions. Indexed libraries were quantified by Bioanalyzer and qPCR, pooled, and sequenced on a NextSeq 500 (Illumina).

ATAC-seq library preparation. ATAC-seq using 250K-500K CME-W1-cl8+ cells per reaction was performed with the Nextera DNA Library Prep Kit (Illumina, FC-121-1030). The reaction protocol was as previously described[50], but without detergent and lysis steps. Transposition reactions were performed at 37ºC for 30 min shaking at 400 r.p.m. Libraries were purified with the QIAGEN MinElute Reaction Cleanup Kit and PCR amplified with barcoded primers. Amplification cycle number for each sample was monitored by qPCR to minimize PCR bias. PCR amplified libraries were purified with the QIAquick PCR Purification kit and excess primers removed by AMPure XP bead selection (Beckman Coulter). Final library concentrations were determined by qPCR using custom primers and PhiX sequence (Illumina) as a standard.

ATAC-seq data processing. Six technical replicates were sequenced with 75-bp paired-end reads on the Illumina Next-seq. Illumina Nextera Adapters were removed using a custom Python script. Reads were aligned to the *Drosophila melanogaster* genome (dm3) using bowtie2 (version 2.2.5) with the parameter -X 2000. Duplicates were removed with Picard; mitochondrial reads or reads with bowtie2 MAPQ score <30 were removed using SAMtools. All replicates had high similarity so their alignment files were merged to increase library complexity before analysis with ChAR-seq data.