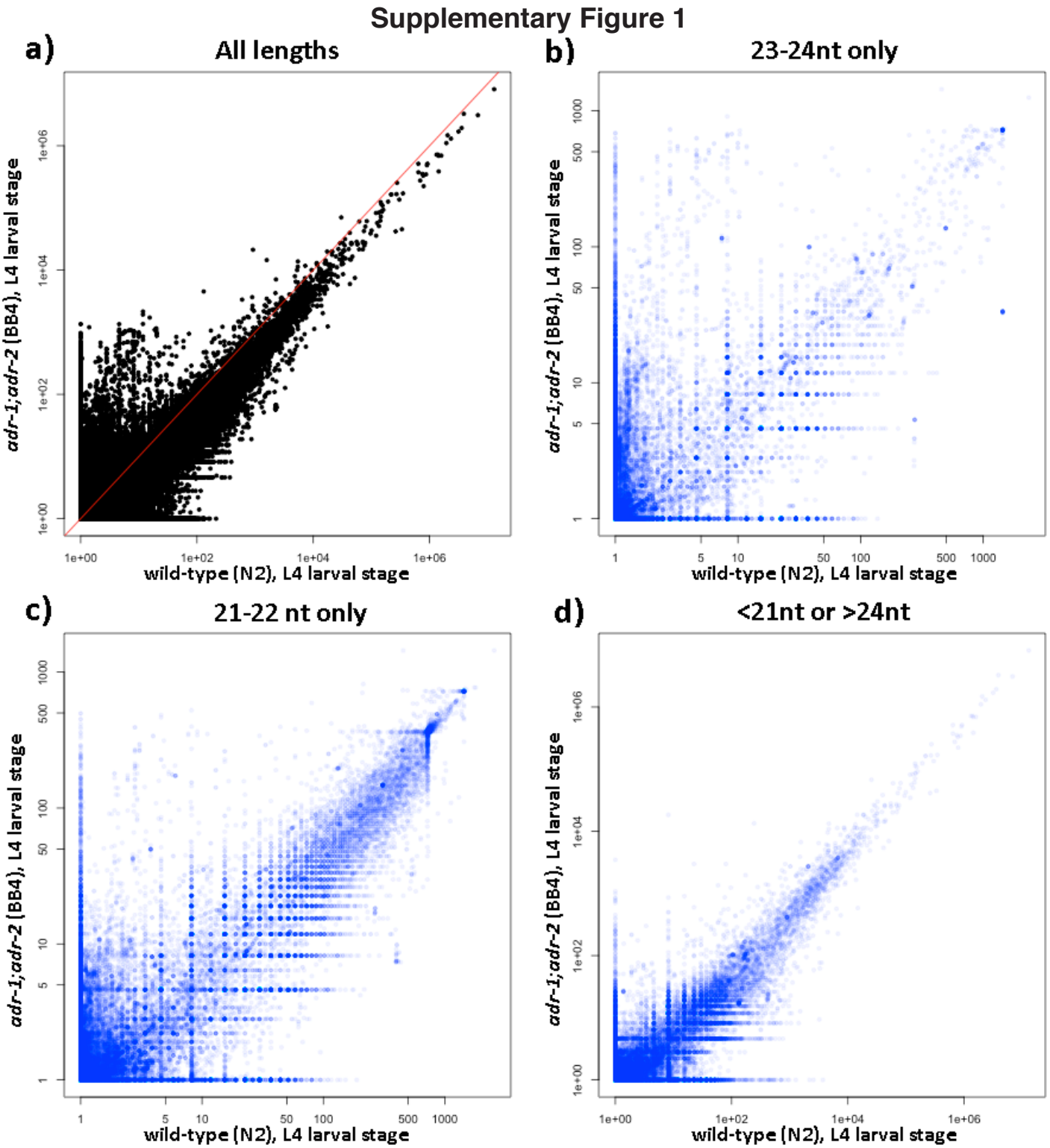
Competition between ADAR and RNAi pathways for an extensive class of RNA targets

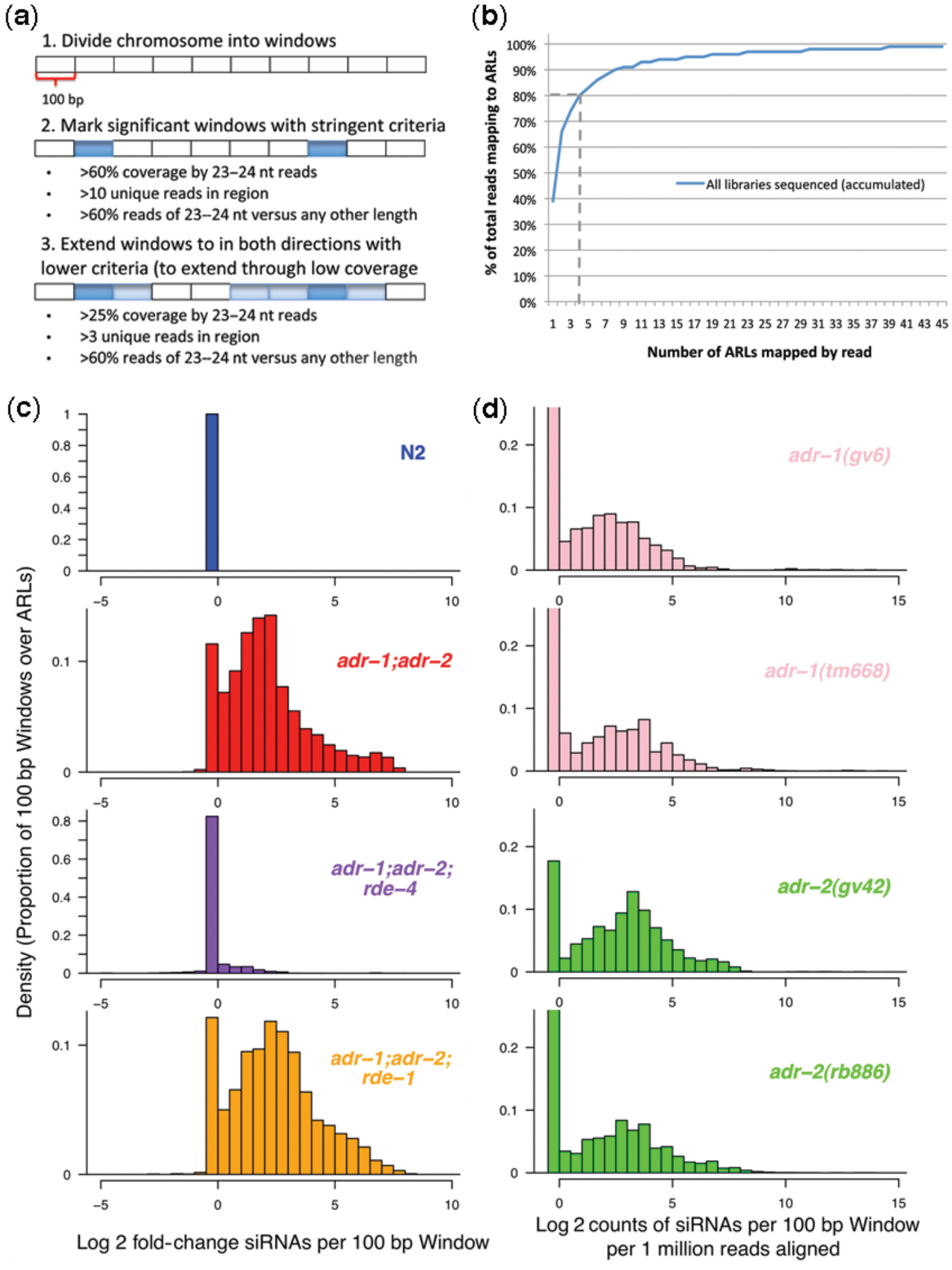
Diane Wu, Ayelet T. Lamm , and Andrew Z. Fire

Supplementary Material



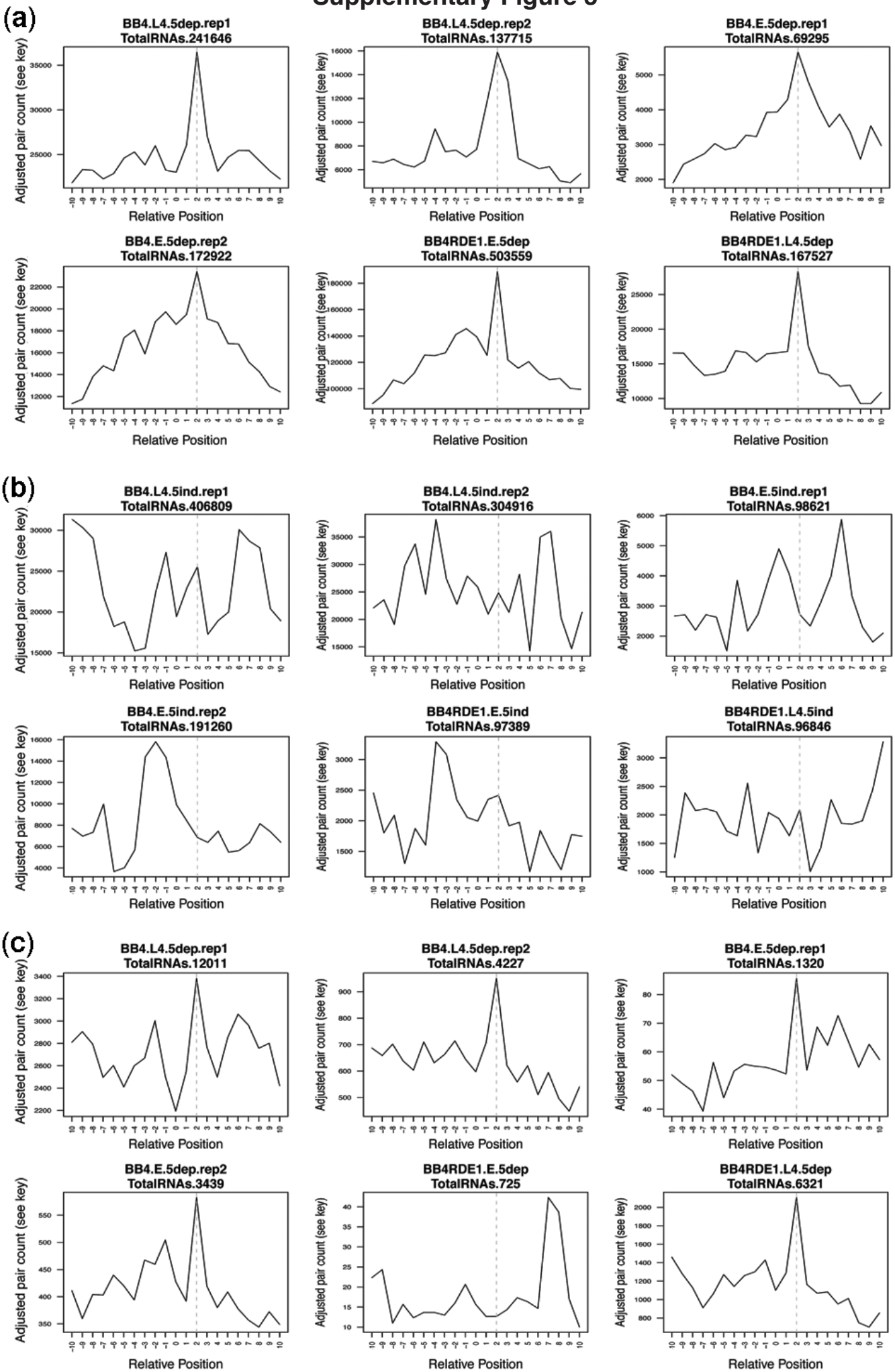
Supplementary Figure 1. Small RNA accumulation across the *C. elegans* genome. (**a**) Each point represents a 100 bp locus in the *C. elegans* genome. Small RNA sequences from 5' monophosphate dependent capture were aligned to the genome and tallied for each locus. Counts per locus were then normalized to counts per 10 million total counts from aligned reads, using the scaling normalization method[61] against a reference set (**Supplementary Methods**), and data from a *adr-1;adr-2* sample was plotted against a wild-type sample (both at L4 stage). The *adr-1;adr-2* editing deficient mutant sample shows an enrichment of small RNA counts for many loci in the genome. (**b**) Small RNA reads in the 23–24 nt size range from normalized counts in panel **a** were tallied and plotted for each 100 bp region in the genome. Each point is plotted in light blue and represents a 100 bp locus in the genome. Dark blue regions indicate increased density of points. Extreme enrichment for 23–24 nt reads can be seen for many 100 bp windows (light blue cloud on top left of graph; dark blue streak on bottom left). (**c**) Similar plot as in panel **b**, except for the 21–22 nt size range. There is still an observable enrichment of small RNAs in the *adr-1;adr-2* mutant sample, but it is much less drastic than that seen in the 23–24 nt size range (many more points are along the diagonal). (**d**) Similar plot as in panel **b**, except for small RNAs >24 nt or <21 nt in length. No drastic enrichment of small RNAs over a large class of windows is seen in either direction.

Supplementary Figure 2



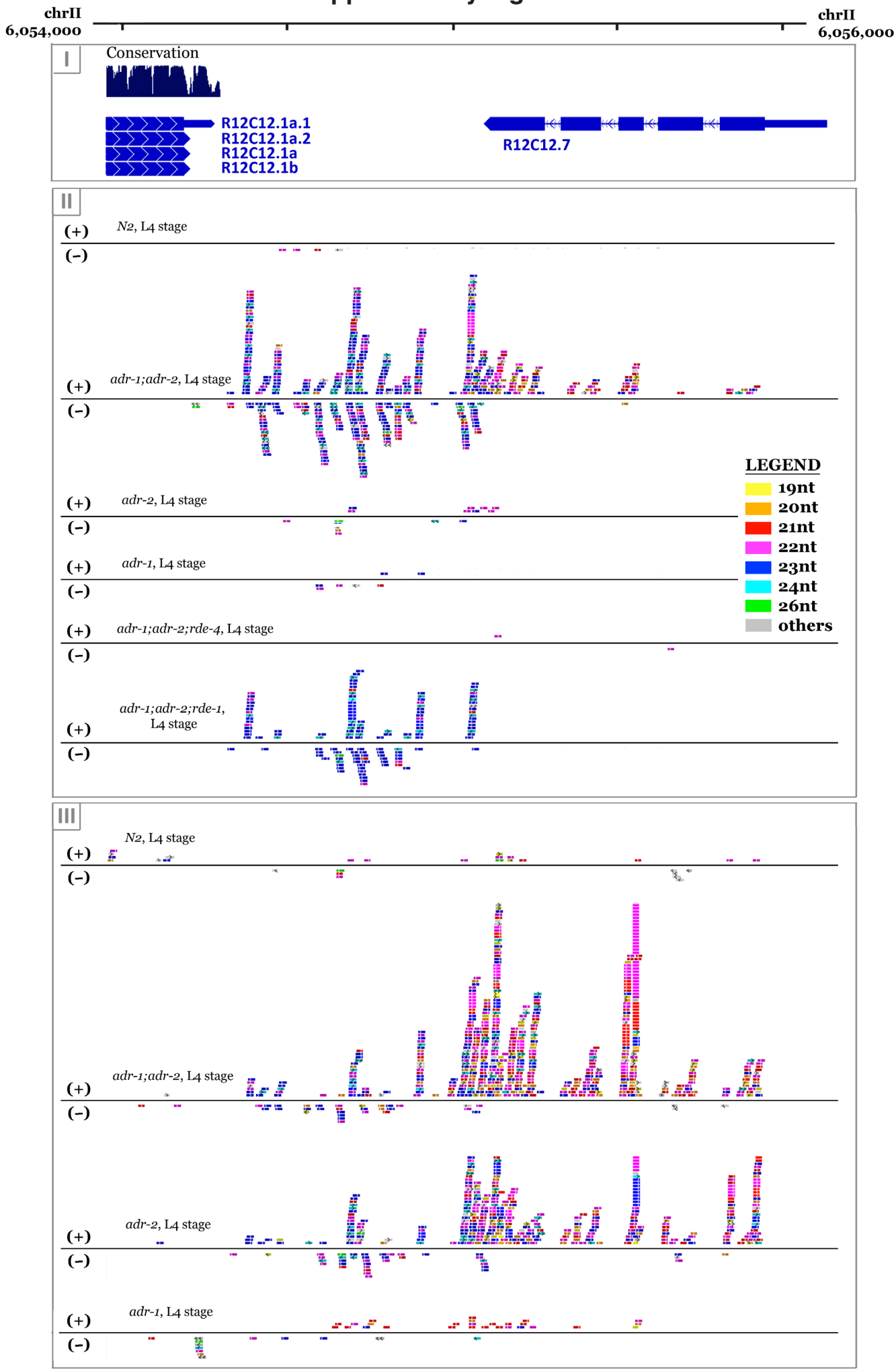
Supplementary Figure 2. Identification of ARLs. (**a**) Flowchart schematic of the algorithm used to identify ARLs, as described in **Supplementary Methods**. (1) The genomic sequence was divided into windows of 100 bp or 200 bp. (2) Significant windows exhibiting extreme enrichment for 23–24 nt sRNAs were marked. (3) Boundaries of significant windows were extended iteratively if neighboring windows exhibited a lower coverage but still showed size enrichment for 23–24 nt sRNAs. After applying this unsupervised discovery technique to each sample, we removed regions that were not significantly enriched in *adr-1;adr-2* samples over wild-type levels. (**b**) Most small RNAs align with low redundancy across ARLs. For every small RNA sequence that maps within an ARL, the number of different mapped positions of that small RNA sequence within all ARLs was recorded. Shown is the cumulative distribution of the mapping frequency within ARLs for ARL-associated small RNAs. Over 90% of the ARL-associated small RNA reads map to less than 10 locations, with a majority mapping to one or two locations. (**c**) Relative small RNA abundances for each 100 bp genomic segment was calculated as described in **Figure 4** and further normalized to the wild-type levels of the corresponding segment. Shown above are the fold changes of small RNA abundances (23–24 nt only) for all 100 bp windows within ARLs. (**d**) Small RNA counts over ARLs (calculated as described in **Fig. 4**, not normalized to wild-type) in animals at embryo and L4 stage (merged), for each of: *adr-1(gv6)* (pink) and *adr-1(tm668)* (pink), *adr-2(gv42)* (green) and *adr-2(rb886)* (green). Values are shown as counts per million total reads aligned.

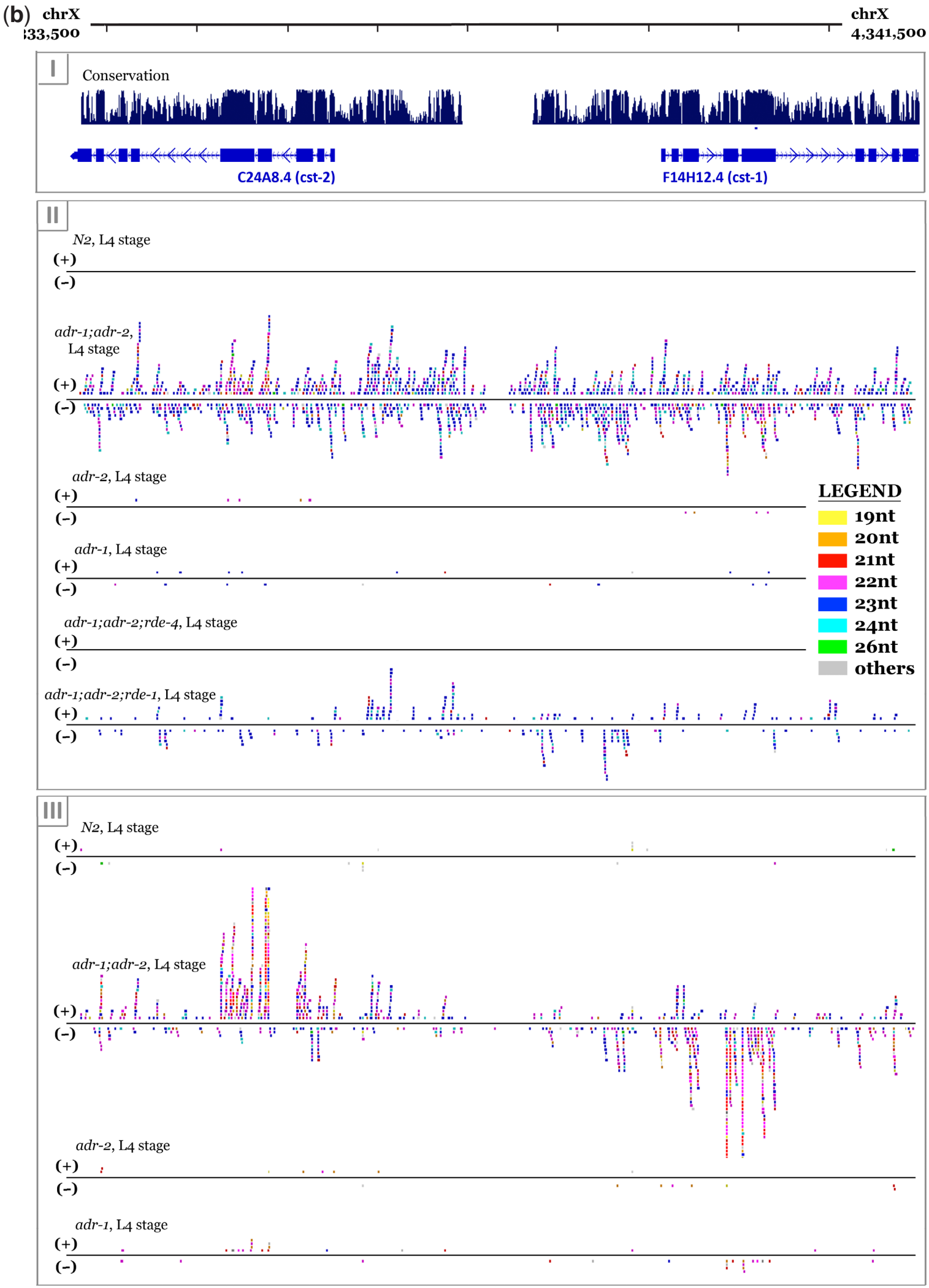
Supplementary Figure 3

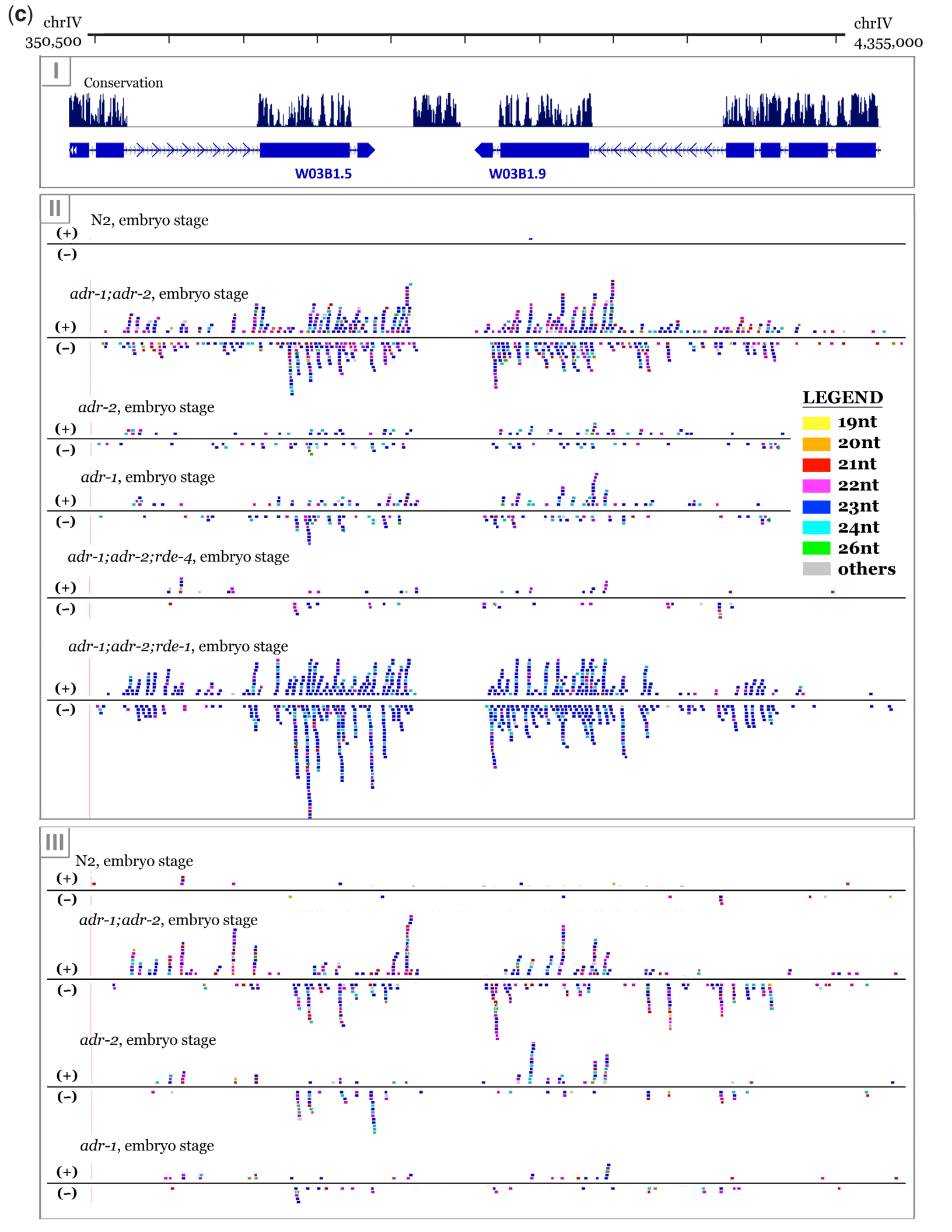


Supplementary Figure 3. Small RNA alignment offset distribution. (**a**) Alignment offsets for each pair of 23–24 nt reads aligning in opposite orientations within ARLs are aggregated and calculated as described in **Supplementary Methods**. Panels depict aggregate statistics over ARLs for each of the 5′ monophosphate dependent libraries from *adr-1;adr-2* mutant animals. Samples are named in each panel as Strain.Stage.Sample-prep.replicate, where strains are described in Table S1, stage = L4 (L4 larval stage) or E (embryo stage), and sample-prep = 5dep (5′ monophosphate dependent) or 5ind (5′ monophosphate independent). The x-axis on each panel depicts, for any pair of reads aligning to opposite strands, the position of the leftmost alignment position of the (+) strand read with respect to left most alignment position of the (–) strand read. The y-axis depicts the total sum of normalized pair-counts for each alignment offset in the exemplary region. Pair-counts for each pair of reads were adjusted for amplification bias by (**i**) taking the cube-root of all pair-counts, and (**ii**) restricting any pair of sequences to take on at most 1,000 counts (such that the maximum value for any pair-count at each position is 10). A distance of +2 indicates a 3′ 2 nt offset in a hypothetical duplex. The total number of RNAs of specified size aligning over ARLs from each sample and that are used in this calculation are indicated in each panel. (**b**) To generate a negative control, equivalent statistics as those described in **a** were calculated for 21–22 nt RNAs over ARLs for 5′ monophosphate independent libraries from *adr-1*;*adr-2* animals. The majority of 21–22 nt RNAs from 5′ monophosphate independent libraries are believed to be RdRP products, generated from unprimed synthesis[51]. Aggregate data over all 5′ phosphate independent samples fail to show a consistent pattern as those observed from 5′ dependent samples. (**c**) Calculations were performed as described in panel **a** using only RNAs aligning to the exemplary region shown in **Figure 2** (chromosome V 8848500–8851500). Peaks at +2 nt can be seen for all samples that have sufficient coverage of RNAs (i.e. BB4RDE1 at embryo stage contains few RNAs over this 3 kb region).

Supplementary Figure 4

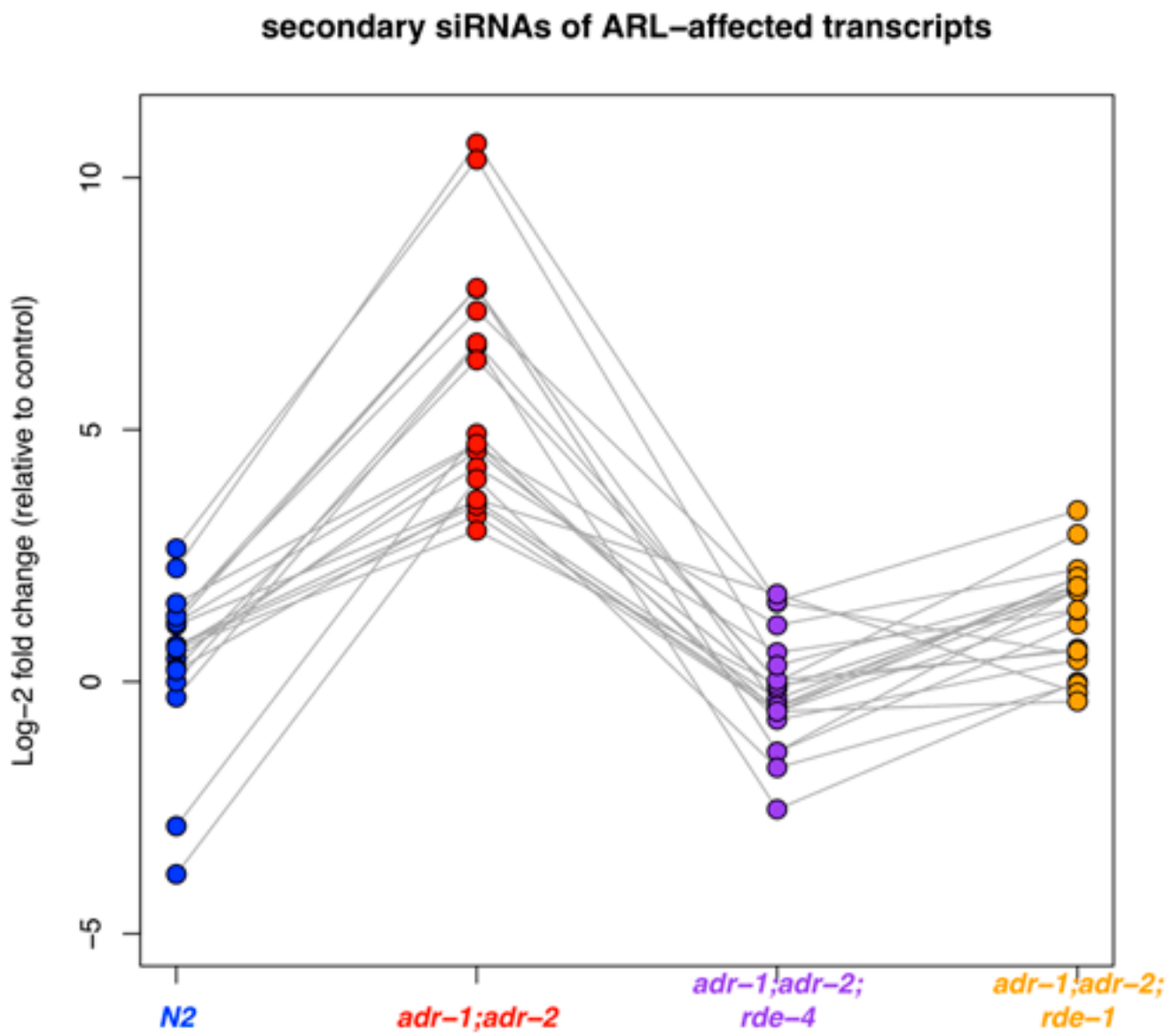



(b)
chrX
:33,500
chrX
4,341,500
I
Conservation
C24A8.4 (cst-2)
F14H12.4 (cst-1)
II
N2, L4 stage
(+)
(–)
adr-1;adr-2, L4 stage
adr-2, L4 stage
adr-1, L4 stage
adr-1;adr-2;rde-4, L4 stage
adr-1;adr-2;rde-1, L4 stage
LEGEND
19nt
20nt
21nt
22nt
23nt
24nt
26nt
others
III
N2, L4 stage
adr-1;adr-2, L4 stage
adr-2, L4 stage
adr-1, L4 stage


(c)
chrIV
350,500
chrIV
4,355,000
I
Conservation
W03B1.5
W03B1.9
II
N2, embryo stage
(+)
(–)
adr-1;adr-2, embryo stage
adr-2, embryo stage
adr-1, embryo stage
adr-1;adr-2;rde-4, embryo stage
adr-1;adr-2;rde-1, embryo stage
LEGEND
19nt
20nt
21nt
22nt
23nt
24nt
26nt
others
III
N2, embryo stage
adr-1;adr-2, embryo stage
adr-2, embryo stage
adr-1, embryo stage

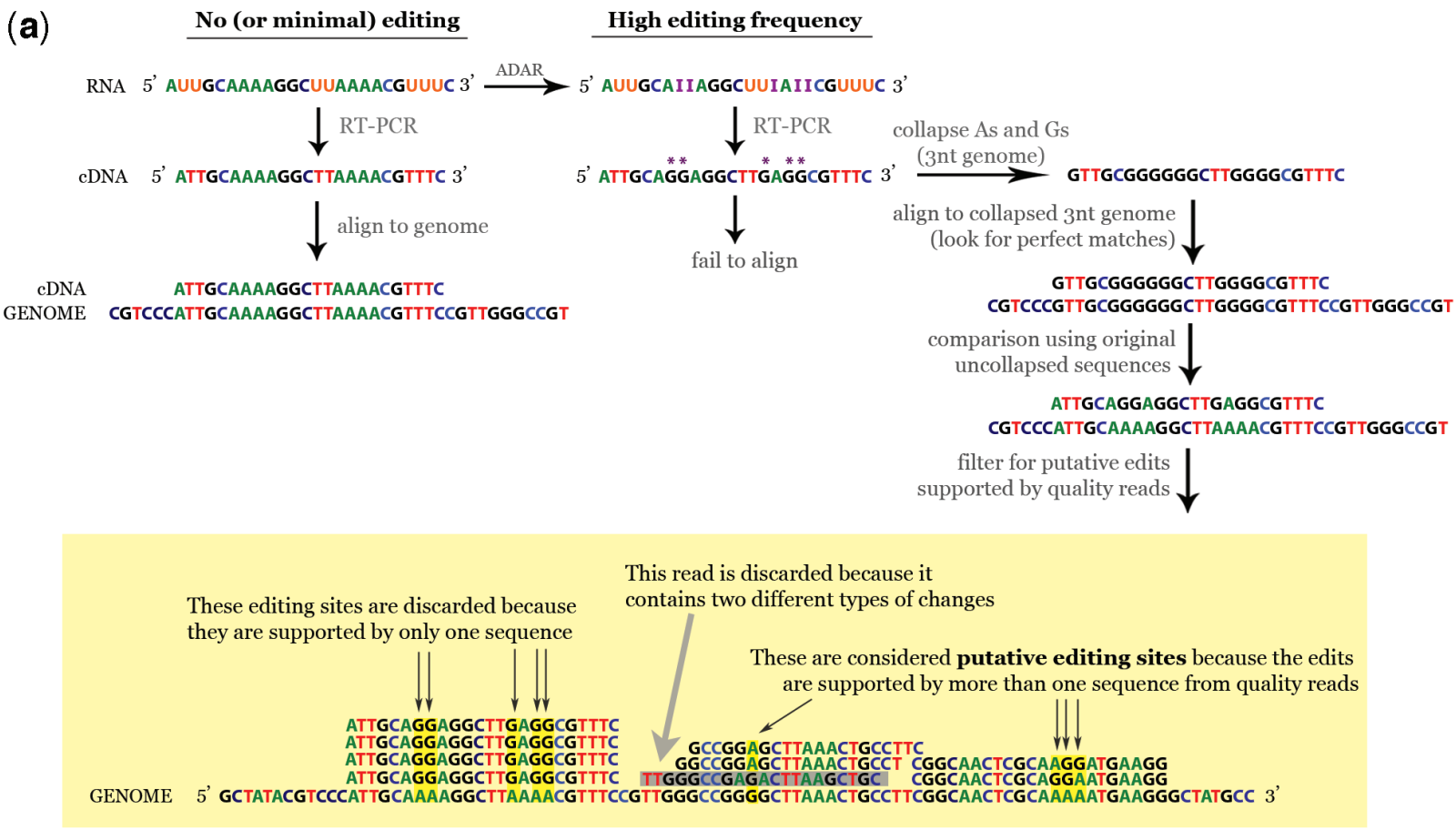
Supplementary Figure 4. Examples of additional ARL regions. Figures use identical schema as that used in **Figures 2** and **5**. (**a**) A 1 kb region on chromosome II that exhibits siRNA accumulation in the intergenic region between two convergent transcripts. (**i**) Identified coding region and conservation are diagrammed as UCSC Genome Browser tracks (*C. elegans* genome version WS190; ref. 59). Direction of transcription is shown as arrows. (**ii**) Populations of 5′ monophosphorylated siRNAs at L4 stage in wild type (N2), *adr-1;adr-2* double mutants, *adr-2* mutants, *adr-1* mutants, *adr-1;adr-2;rde-4* triple mutants, and *adr-1;adr-2;rde-1* triple mutants. (**iii**) Populations of siRNAs treated by phosphatase-kinase method at 5′ end, at L4 stage in wild-type, *adr-1;adr-2* double mutants, *adr-2* mutants, and *adr-1* mutants. Colors indicate sizes (as on figure legend): yellow=19 nt, orange=20 nt, red=21 nt, magenta=22 nt, blue=23 nt, cyan=24 nt, green=26 nt, grey=all other lengths. (**b**) A 9.2 kb ARL on chromosome X, overlapping two paralogous genes. Shown as in panel **a**: (**i**) coding region and conservation, (**ii**) 5′ monophosphate small RNA population at L4 stage from various mutant backgrounds, (**iii**) 5′ treated small RNA population at L4 stage from various mutant backgrounds. (**c**) Two adjacent 2 kb ARLs on chromosome IV. Shown as in panel **a**: (**i**) coding region and conservation, (**ii**) 5′ monophosphate small RNA population at embryo stage from various mutant backgrounds, (**iii**) 5′ treated small RNA population at embryo stage from various mutant backgrounds. As shown in this example on chromosome IV, the accumulation of ADAR-modulated small RNAs at the embryonic stage is often equivalent to or greater than that at the L4 larval stage. However, unlike what is observed at the L4 larval stage, these ARLs show little evidence of secondary siRNA accumulation over adjacent transcripts at the embryonic stage.

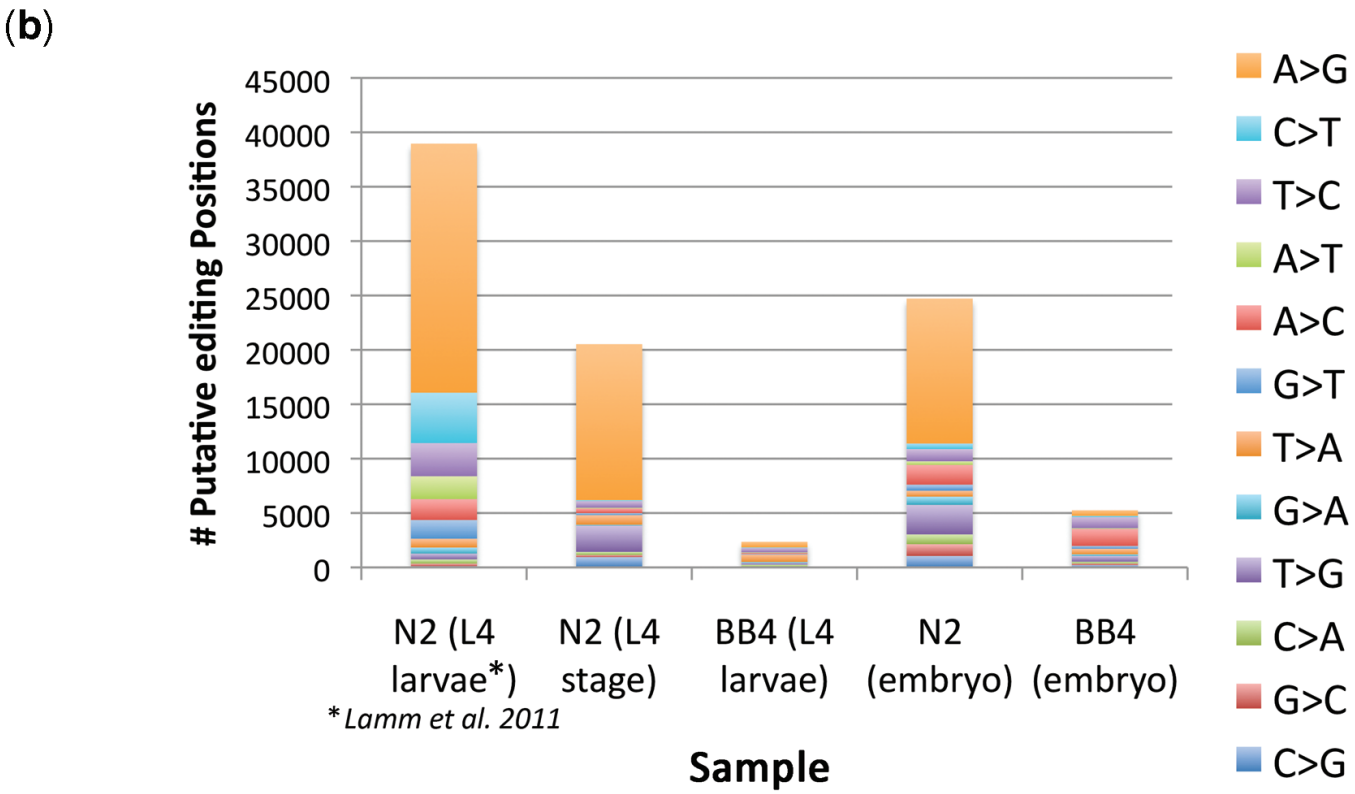
Supplementary Figure 5

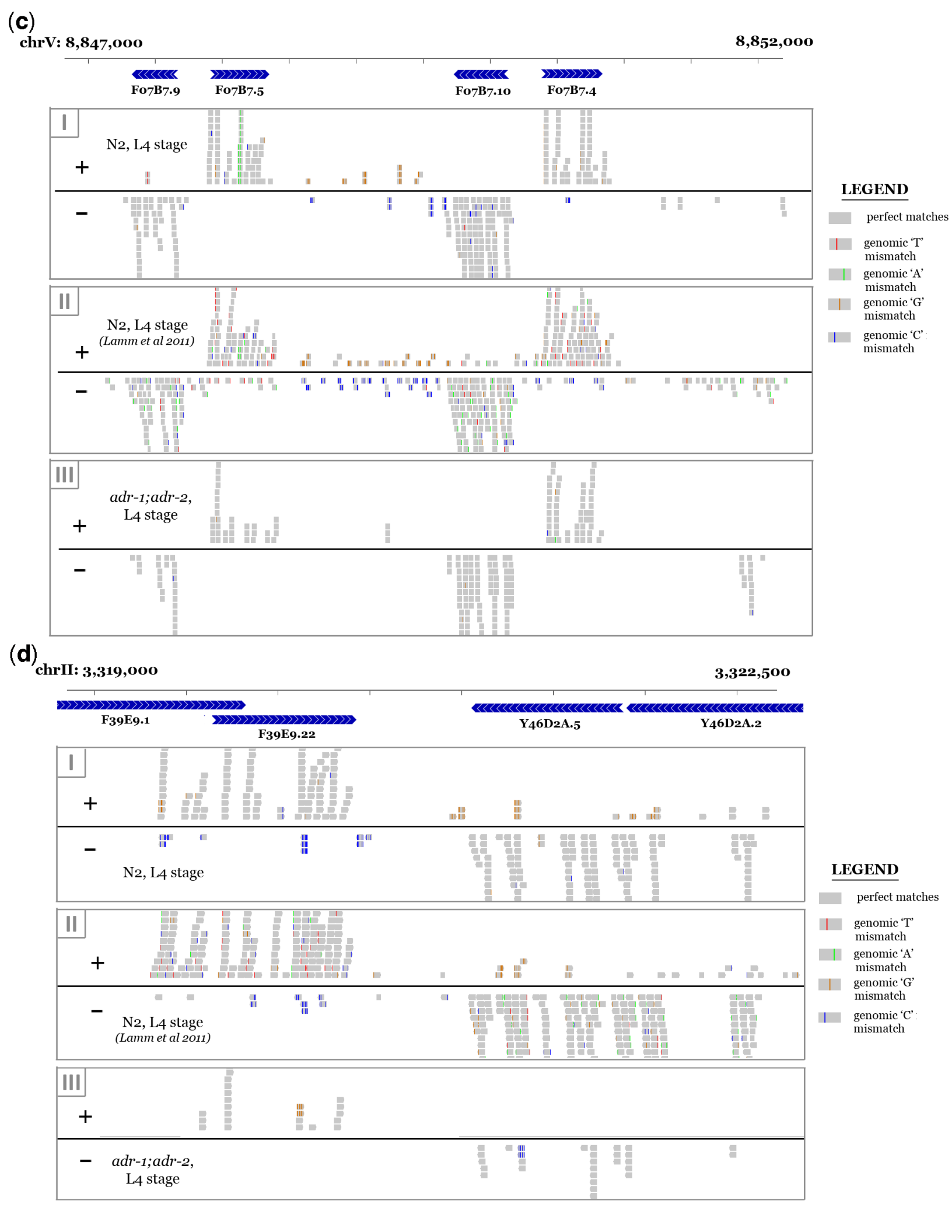


Supplementary Figure 5. Effects of RNAi mutants on the incidence of Secondary ARL sRNA levels. Antisense siRNA counts for ARL-affected transcripts (as described in **Fig. 6a**) were normalized to their respective wild-type levels. The y-axis denotes log-2 fold change over that of wild-type levels. The first column (blue) denotes a separate biological preparation of wild-type animals at an earlier L4 stage (2 h earlier at 20°C). The second, third, and fourth columns denote expression of corresponding transcripts in *adr-1;adr-2* (red), *adr-1;adr-2;rde-4* (purple), and *adr-1;adr-2;rde-1* (orange) mutants, respectively. To indicate corresponding transcripts, grey lines connect expression levels in various mutants. Although outliers of the expression distribution overlap slightly between mutants (as seen in **Fig. 6b**), the change in expression level for each individual gene is significant, with genes upregulated in *adr-1;adr-2* animals showing drastic downregulation to near wild-type levels in *adr-1;adr-2;rde-4* and *adr-1;adr-2;rde-1* animals.

Supplementary Figure 6

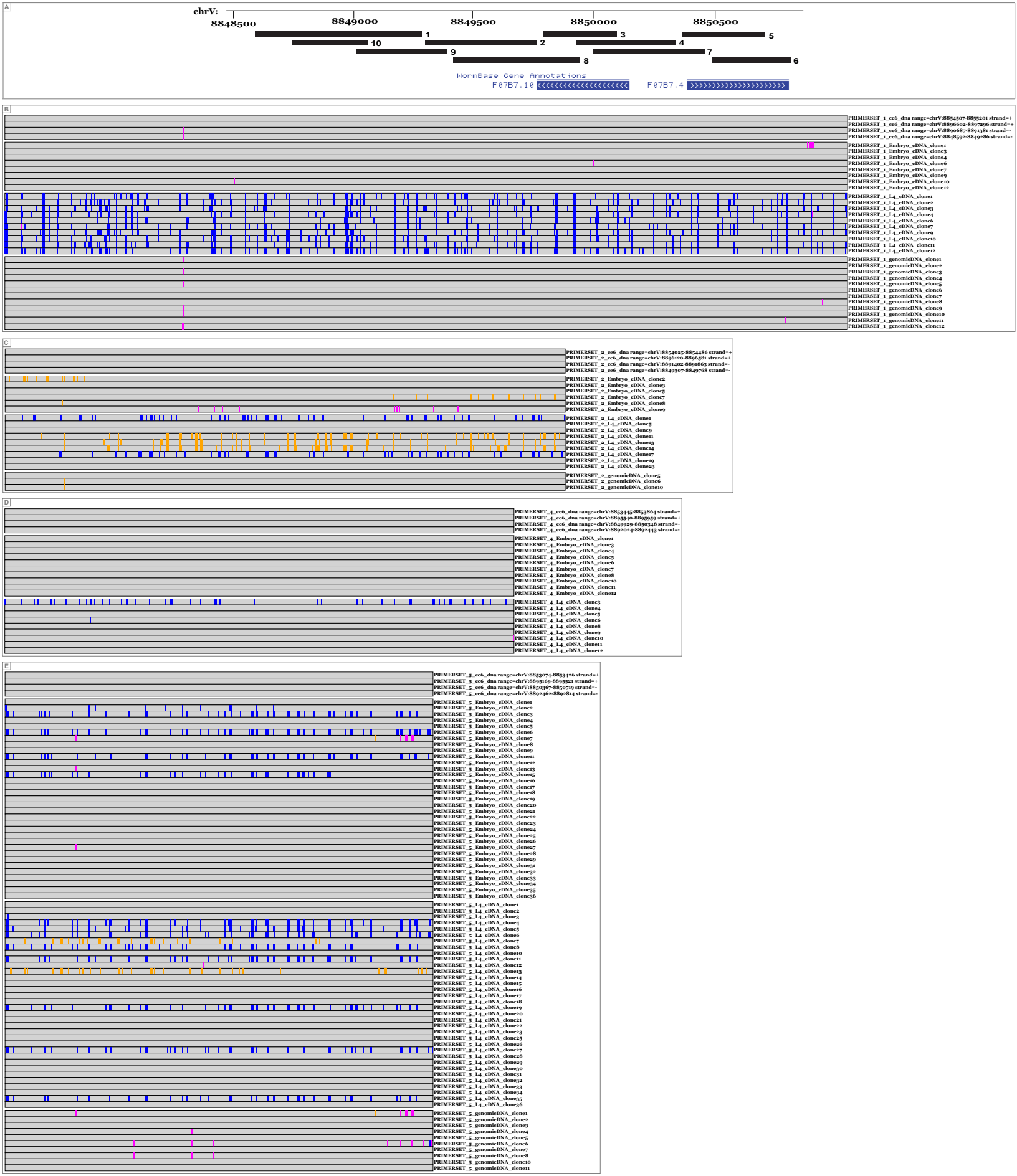


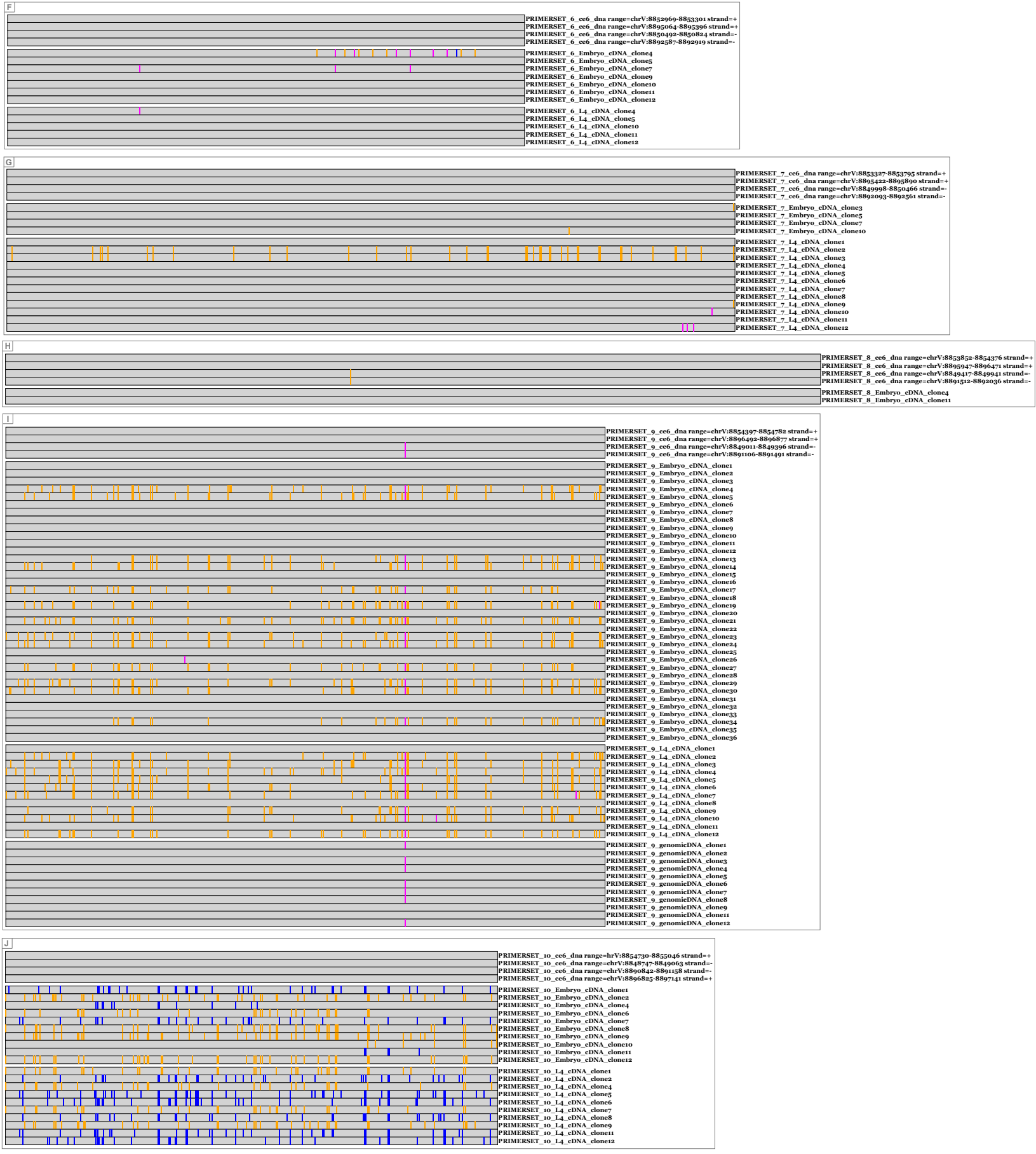



(c)
chrV: 8,847,000
8,852,000
F07B7.9
F07B7.5
F07B7.10
F07B7.4
I
N2, L4 stage
+
−
II
N2, L4 stage
(Lamm et al 2011)
III
adr-1;adr-2,
L4 stage
LEGEND
perfect matches
genomic 'T' mismatch
genomic 'A' mismatch
genomic 'G' mismatch
genomic 'C' mismatch
(d)
chrII: 3,319,000
3,322,500
F39E9.1
F39E9.22
Y46D2A.5
Y46D2A.2
N2, L4 stage
N2, L4 stage
(Lamm et al 2011)
adr-1;adr-2,
L4 stage
LEGEND
perfect matches
genomic 'T' mismatch
genomic 'A' mismatch
genomic 'G' mismatch
genomic 'C' mismatch

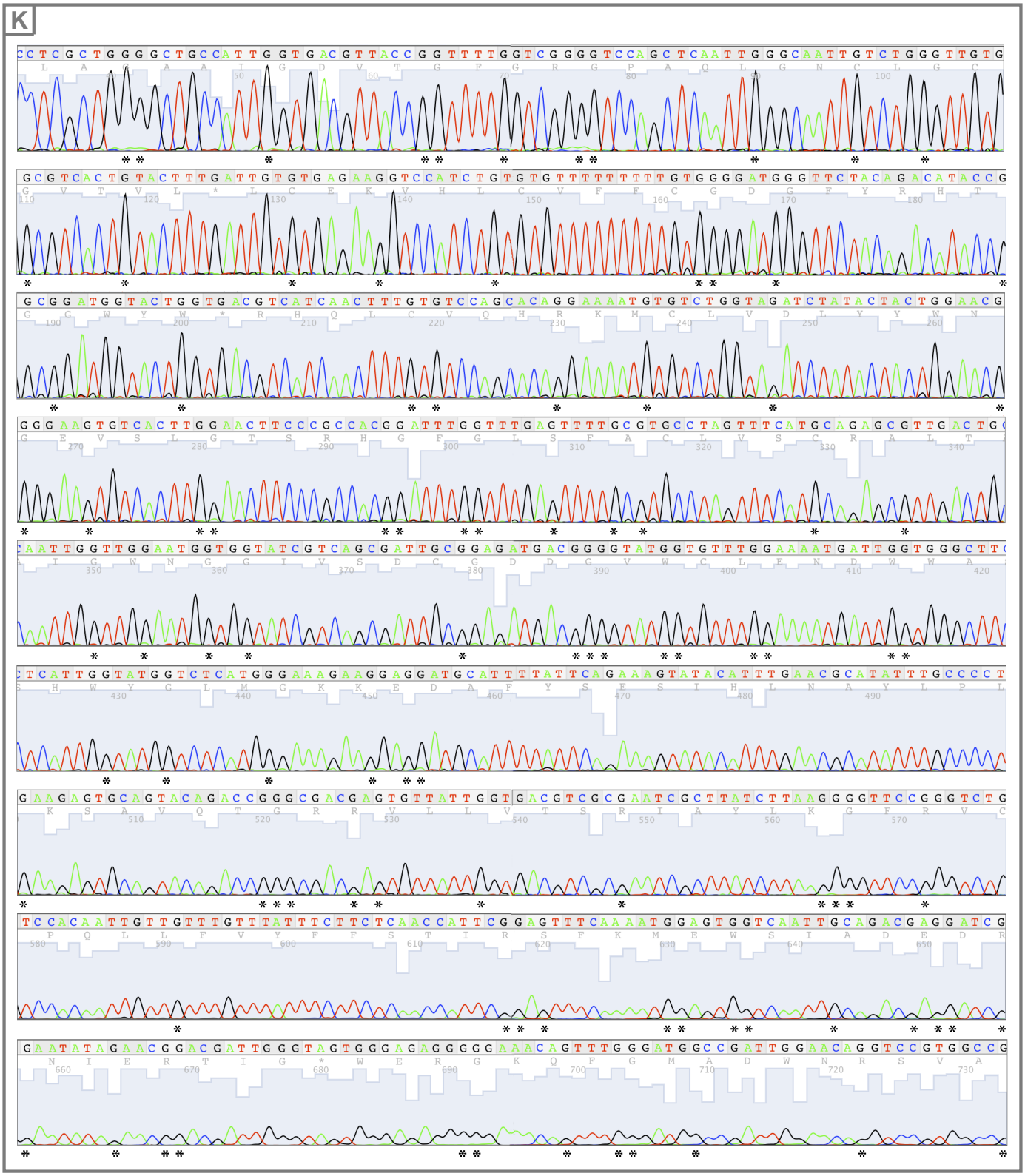
Supplementary Figure 6. Identification of putative editing sites over highly edited regions using a 3-base collapsed-genome strategy. (**a**) Algorithm for putative editing detection, as described in **Supplementary Methods**. To detect putative A-to-G edits, all As in read sequences and genomic sequences were "pre-edited" into Gs. Alignment using collapsed sequences followed by reversal to original read and genomic sequences gives a list of positions with A-to-G changes as well as G-to-A changes from genomic to read sequence. These positions were further filtered for stringency as described in **Methods**. The algorithm was repeated for all 6 possible nucleotide pairs, for detection of all 12 possible nucleotide edits. (**b**) Distribution of classes of putative editing events. Procedures described in panel **a** were applied to each sample: wild-type (N2) at L4 larval stage, *adr-1;adr-2* (BB4) at L4 larval stage, wild-type (N2) at embryo stage, *adr-1;adr-2* (BB4) at embryo stage, and an independent RNA-Seq dataset obtained from wild-type (N2) animals at L4 larval stage using the Circligase sequence capture protocol [56]. A-to-G changes are the most abundant class amongst all nucleotide-changes detected in the wild-type sample at L4 stage from both capture protocols and at the embryo stage (this study). A-to-G are significantly depleted in the *adr-1;adr-2* double-mutant at both embryo and L4 stage, and enrichment for any other class of nucleotide change is absent. (**c** and **d**) Examples of frequent mRNA editing at ARLs. (**c**) Exemplary region from **Figure 5** (inverted repeat from paralogous duplication at Y46D2A.5/F39E9.22) is shown, for (**i**) N2 at L4 larval stage, (**ii**) N2 at L4 larval stage from Lamm *et al* [56], and (**iii**) *adr-1;adr-2* at L4 larval stage. Results from the first-pass 4-base alignment were collated with those from the 3-base alignment procedure, and visualized using the Integrative Genomics Viewer[60]. Aligned mRNA reads across this region are shown in grey boxes. Reads aligning to the (+) strand are drawn above the line and those aligning to the (–) strand are drawn below the line. Mismatches are shown as colored lines at the position of the mismatch. Colors indicate the genomic nucleotide (Watson strand) of the mismatch (e.g. RNA with 'G' mismatch aligning to (–) strand is presented as a 'C' mismatch). Red = 'T', green= 'A', orange= 'G', blue = 'C'. Hence, G mismatches on the (+) strand are colored in orange while those on the (–) strand are colored in blue. Over 90% of the G mismatches in this region were verified to be A-to-G changes. (**d**) Exemplary region from **Figure 2** (F07B7 histone locus) is shown, for (**i**) N2, (**ii**) N2 from Lamm *et al* [56], and (**iii**) *adr-1;adr-2*, all at L4 larval stage. Aligned reads with mismatches are depicted as described in (**c**). Over 90% of the G mismatches in the duplicated region were verified to be A-to-G changes.

Supplementary Figure 7





Supplementary Figure 7. Sanger sequencing validation of edited regions.
cDNA from wild-type animals was amplified (as described in **Supplementary Methods**) using sequence-specific primers designed to target several subregions of the exemplary ARL shown in **Figure 2**, which overlaps the F07B7 histone locus. (**a**) Each primer set targets a specific region (up to 800 bp) in the F07B7 locus, as depicted by black bars (which are numbered by primer set). (**b–j**) Each panel depicts clones from a separate primer set, as indicated. Each clone was sequenced using Sanger sequencing; primer sequences were removed and the insert sequence



was aligned to the genome. Sequences are depicted as a grey slices with single-nucleotide mismatches from the first (reference) genomic sequence depicted by vertical lines: genomic A-to-G mismatches are shown in orange, genomic T-to-C mismatches are shown in blue, and all other genomic mismatches are shown in magenta (including Ns in the query sequence). The first four slices for each primer pair depict the four nearly identical copies of each segment in the published genomic sequence (there are 4 copies of the F07B7 histone locus). Subsequent slices each depict a single cDNA sequence. Strikingly, each edited cDNA clone predominantly exhibited only one of the two types of changes (A-to-G or T-to-C). Editing patterns exhibited by these clones are extremely diverse, and very few clones are identical in sequence. Clones are named as PRIMERSET_<#>_<Stage>_cDNA_clone<#>. To rule out DNA diversity as the source for the sequence variation at the RNA level, we also used these primers for highly edited regions to amplify genomic DNA. No such sequence diversity was observed in these sequences (although a limited number of differences from the reference genome were observed). Sequences of clones and aligning regions shown in this figure are listed in **Supplementary Table 6**. (**k**) Sanger trace of one of the clones with high levels of A-to-G mismatches (panel **b**, PRIMERSET_1_L4_cDNA_clone3). Asterisks below the trace denote the positions of the mismatches; mismatch locations show no correlation with sequence quality.

Supplementary Methods

C. elegans strains

The following strains were used in this study: N2 'Bristol' wild type [62], BB2:*adr-1*(gv6), BB3: *adr-2*(gv42), BB4: *adr-1*(gv6);*adr-2*(gv42), TM668: *adr-1*(tm668); RB886: *adr-2*(ok735), *adr-1*(gv6);*adr-2*(gv42); *rde-4*(ne299), and *adr-1*(gv6);*adr-2*(gv42);*rde-1*(ne219). All strains were grown on OP50 on enriched agar plates.

Sample Preparation

L4 stage animals were harvested as synchronized populations during early vulva development. Synchronization was performed as follows (i) sodium hypochlorite treatment to remove all stages except the embryos, (ii) allowing embryos to hatch at 20^{o}C on agar plates containing no bacteria (24 hrs) (iii) growth on enriched agar plates with OP50 (ref. 62) for 40 h at 20^{o}C. Animals were then washed, centrifuged, and pellets were frozen in EN50 (100mM NaCl, 0.25M EDTA) using liquid nitrogen. Embryos were harvested using the same sodium hypochlorite treatment used for staging, followed immediately by washing and freezing in liquid nitrogen.

ARL detection and assessment of ARL-associated siRNA expression

We divided the *C. elegans* genome into regions of set size (100bp) and counted the number of reads that map to each region **(Supplementary Fig. 1)**. We identified regions that satisfied the following criteria as exemplary cases of ARL-associated siRNA enrichment for further analysis: a) at least 60% of the region is covered by reads sized 23–24nt, b) reads of length 23–24nt make up at least 60% of reads of all lengths mapping in the region, and c) there are at least 10 unique reads of length 23–24nt in the region. After identifying regions that satisfy the above criteria **(Supplementary Fig. 2a)**, we then tried to extend these regions in both directions in 100bp intervals, requiring extended regions to satisfy the following criteria: 1) at least 25% of the region is covered by reads sized 23–24nt, 2) reads of length 23–24nt make up at least 60% of reads of all lengths mapping in the region, and 3) there are at least 3 unique reads of length 23–24nt in the region. Extension of regions ceases when a candidate extension region fails to satisfy the above criteria, and the end coordinates of the extended regions were recorded.

To quantify the number of small RNAs mapping to ARLs for each sample, we divided all ARLs into segments of 100bp. For each segment, we counted the number of 23–24nt reads mapping to that segment, normalized to the total number of reads aligned to the genome. We further normalized the count contribution for each read to the number of total positions that read aligned to. This normalization step corrects for any amplification of signals due to repeated mappings to repeat regions in the genome, such that the sum of aligned read counts for all reads aligning to the genome is equal to the number of sequenced reads.

To generate a final list of ARL loci, we repeated the discovery algorithm with window size set to 200bp, and merged the resulting regions discovered using the two window sizes (100bp and 200bp). Significant regions detected from both embryo and L4 larval stages were further merged into one candidate ARL list. Candidate ARLs were filtered for false positives by removing regions that do not show enrichment ($p<0.05$) for counts of 23–24nt small RNAs in *adr-1;adr-2* samples over that of wild-type.

To ensure that the enrichment for small RNAs over ARLs in *adr*-deficient samples was not due to an artificial depletion of small RNAs over ARLs in wild-type samples as a result of the assumptions used for alignment, we examined the pool of unaligned reads for edited small RNAs. We employed the approach of "pre-editing" all adenosines over ARLs *in silico (*similar to that performed on mRNAs; see "mRNA editing detection" in **Methods**), and asked if such pre-editing restores alignment to large numbers of small RNAs. Using this approach, we recovered very few additional alignments within ARLs (<0.1% of aligned unique sequences).

Normalization Procedures

Since the editing-deficient animals exhibit a drastic change in the small RNA population and misregulation of many loci, we employed the scaling normalization method as described by Robinson and Oshlack[61].

Statistical Tests for Significance of RNA enrichment

Enrichment of ADAR-affected RNAs over ARLs (23–24nt RNAs from 5′ monophosphate dependent capture)

To test for differential expression between two genes or genomic loci in our datasets, we first normalized the data as described above. After normalizing all read counts to reads per 10 million reads, we applied Laplacian smoothing by adding 1 to all read counts. We used Fisher's Exact Test and the DEGSeq package[63] in R[64] (version 2.12.0) to determine p-values for siRNA enrichments using the variation calculated from the population of transcripts with similar expression levels.

For detection of ARLs, we performed FET on all ARLs and removed those that were not significantly upregulated. For any loci in a given sample, we compared its expression in that sample set with each of the wild-type samples at the equivalent stage, and required that it be significantly upregulated in each case. We used a total of 6 wild-type sample controls for the L4 stage (**Supplementary Table 1**, samples #3–7,53), and 4 wild-type controls at the embryo stage (**Supplementary Table 1**, samples #1–2,51–52), for detection of upregulation at the respective stages. This stringent requirement is directed toward ensuring that the small RNA signals we are detecting are not the result of variations in staging, growth conditions, or genetic factors that might affect development.

Enrichment of siRNAs Originating From Annotated Transcripts (21–22 nt RNAs from 5′ independent capture, antisense to transcripts)

To test for differential incidence in 5′ independent libraries, we employed the same normalization method and statistical package, with the exception of using 5 counts for the Laplacian smoothing parameter for the counts of siRNAs antisense to transcripts (per 5 million read aligned).

Enrichment of mRNA levels

For comparisons of mRNA transcript levels, we used Laplacian smoothing parameter of 1 for mRNA counts per 10 million reads aligned. As controls, we used wild-type animals at equivalent stages (L4 and embryo), in addition to a previously published[56] wild-type sample at L4 stage (grown at 16°C for 63 h after hatching)

Genome Annotations

To annotate genomic regions containing inverted repeats, we used previously submitted annotations downloaded from WormBase (version WS215), which includes predicted long inverted repeats spanning <2 kb and with >85% homology between the repeats[65]. To annotate transcribed regions on the genome, we used the cDNA annotation coordinates as downloaded from WormBase (version WS215). To annotate transposons, we used the "Repeat Masker" track as downloaded from the UCSC Genome Browser[59] We considered "transposons" as any annotated repeat in this library; unannotated DNA segments were not recorded as transposons. Since the UCSC Genome browser only hosts the *C. elegans* genome annotation up to version WS190, we converted these coordinates to those corresponding to version WS215 using a in-house software.

Calculation of annotation enrichment using Monte Carlo simulation

Enrichment statistics for annotations overlapped by ARLs were determined by simulating ARLs with identical size distribution as those detected in this paper. ARLs were given equal likelihood of falling anywhere in the genome, and statistics for overlap annotations were calculated over 1000 simulations.

Start-to-Start Positional Correlation Analysis for siRNAs on different strands

To determine if transient signals for 2 nt 3′ overhangs are present in small RNAs aligning to ARLs, we conducted a start-to-start positional correlation analysis for all siRNAs aligning to ARLs. We calculated frequencies of the distances between all pairs of ARL-aligning siRNAs on opposite strands as follows:

1. Separate siRNAs mapping in ARL region into two pools, those mapping to the Watson strand (pool W) and those mapping to the Crick strand (pool C).
2. Record start positions for all reads in each pool. The start position is annotated as the genomic

coordinate of the start of the alignment, relative to the Watson strand of the reference genome.

3. For every possible pair of reads where one is from pool W and one is from pool C, find the distance between the start position of the read from pool W to that of the read from pool C.
4. Add one count for the relative distance calculated
5. Aggregate counts for all ARL-aligning read-pairs at each relative distance

To control for bias caused by specific highly abundant read-pairs, we conducted the following normalization procedures:

1. Normalize all read-pair counts to the number of positions it aligns to (i.e. a read pair that aligns to 5 other positions is counted 0.2 times at each position)
2. Take the cube root of all distance counts
3. Restrict the maximum value of each relative distance at each position to 100 (i.e. a highly abundant sequence-pair at a particular position on the chromosome gets a maximum value of $100^{1/3} = 4.64$

Validation of mRNA editing using Sanger Sequencing

We treated total mRNA (large RNA fraction from RNA extraction using the miRVana kit) with DNase I and generated cDNA using random hexamers. We designed primers targeting regions of up to 800bps within the exemplary ARL depicted in **Figure 2**. In particular, the following set of primers were used:

Primer Set	Forward primer	Reverse primer	Expected product size (bp)
1	GGTTCT(A/G)CTGTCCTC(A/G)CTG	CACAACCACCCCCACCCACA	734
2	TGTGGGTGGGGGTGGTTGTG	CCAAAGAAGAC(T/C)GGAGGAGACAAGG	507
3	CCTTGTCTCCTCC(A/G)GTCTTCTTTGG	CCAAGACCGGAGGAAAAGCC	351
4	TTCTGGTCTTCTTGTTGTC	CGG(T/C)GGGAAGGAGGAGCC	457
5	GGCTCCTCCTTCCC(A/G)CCG	GGAAGAA(T/C)CAAACGGACGGCAG	393
6	GCCGTTGGGTTCGGTG(A/G)TTTGCTTTC	GGAGCAAAGAAAGCCGCCAAGACCG	387
7	CGGTCTTGGCGGCTTTCTTTGCTCC	C(T/C)CAAGAACAGCAGCCAGG(T/C)AAAC	519
8	CGCTGCTG(A/G)GGTTCTCG(A/G)GTTGGCTGG	GAGAGAAC(T/C)CCAACAAGGCG	575
9	CGCCTTGTTGG(A/G)GTTCTCTC	GAGAAAGAAAGAAAA(T/C)GCA	425
10	C(A/G)TTTGTCGTTTGGTCTGT	ACAGACA(T/C)ACCAGCAGA(T/C)GG	356

Semi-degenerate primers (indicated in brackets) were created for those that contained an "A" in the primer sequence, to allow for capture of transcripts that might be edited at that location. Primer pairs 1 to 5 were designed for capture of possibly edited transcripts in the Watson direction; primers 6 to 10 were designed for capture of possibly edited transcripts in the Crick direction. If transcripts are not edited in the primer regions, the primer pair would allow for capture of transcripts originating from both directions.

Of those pairs that generated a PCR product of the expected size, we cut the corresponding band out of the gel and isolated the DNA using the QIAquick Gel Extraction kit (QIAGEN). We cloned the PCR fragments and sequenced individual clones using Sanger sequencing. For each primer pair, we sequenced between 12 and 36 clones in order to assess the complexity of the editing pool. Sequences were aligned to the genome using BLAT and only those sequences aligning with highest similarity to the target region were considered "edited" transcripts originating from that region.

Software

Figures were generated using R[64], Adobe Photoshop and Adobe Illustrator, and Microsoft Excel

61. Robinson, M.D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).
62. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94 (1974).
63. Wang, L., Feng, Z., Wang, X., Wang, X. & Zhang, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138 (2010).
64. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing: Vienna, Austria,).at <http://www.R-project.org/>
65. Wang, Y. & Leung, F.C.C. Long inverted repeats in eukaryotic genomes: recombinogenic motifs determine genomic plasticity. *FEBS Lett* **580**, 1277–84 (2006).