Inverted repeat structures are associated with essential and highly expressed genes on *C. elegans* autosome distal arms

DANIEL P. REICH and BRENDA L. BASS

Department of Biochemistry, University of Utah, Salt Lake City, Utah 84112, USA

ABSTRACT

Complementary sequences in cellular transcripts base-pair to form double-stranded RNA (dsRNA) structures. Because transposon-derived repeats often give rise to self-complementary sequences, dsRNA structures are prevalent in eukaryotic genomes, typically occurring in gene introns and untranslated regions (UTRs). However, the regulatory impact of double-stranded structures within genes is not fully understood. We used three independent methods to define loci in *Caenorhabditis elegans* predicted to form dsRNA and correlated these structures with patterns of gene expression, gene essentiality, and genome organization. As previously observed, dsRNA loci are enriched on distal arms of *C. elegans* autosomes, where genes typically show less conservation and lower overall expression. In contrast, we find that dsRNAs are associated with essential genes on autosome arms, and dsRNA-associated genes exhibit higher-than-expected expression and histone modification patterns associated with transcriptional elongation. Genes with significant repetitive sequence content are also highly expressed, and, thus, observed gene expression trends may relate either to dsRNA structures or to repeat content. Our results raise the possibility that as-yet-undescribed mechanisms promote expression of loci that produce dsRNAs, despite their well-characterized roles in gene silencing.

Keywords: RNA editing; dsRNAome; ADARs; gene regulation; genome organization; intron

INTRODUCTION

Interactions between long dsRNAs and dsRNA-binding proteins (dsRBPs) mediate critical cellular processes. One example is RNA interference (RNAi), which is initiated when the dsRNA-specific endoribonuclease Dicer binds and processes dsRNA into ~20–30 nucleotide (nt) small interfering RNAs (siRNAs) that silence target transcripts (Carthew and Sontheimer 2009). dsRNAs are also bound by the regulatory dsRBP Staufen (Heraud-Farlow and Kiebler 2014) and adenosine deaminases that act on RNA (ADARs), which edit adenosine to inosine in dsRNA (Nishikura 2016). Interactions of dsRBPs are largely with the phosphodiester backbone of A-form dsRNA, and thus they bind their substrates regardless of sequence (Tian et al. 2004). Complementary RNA sequences that form dsRNA often arise from transposon-derived repetitive sequences (Porath et al. 2014; Blango and Bass 2016; Reich et al. 2018), which are abundant in metazoan genomes (Feschotte and Pritham 2007; Chuong et al. 2017). Despite the ubiquity of dsRNAs and the essential roles of many dsRBPs, the functions of most cellular dsRNA structures are poorly understood.

The majority of long cellular dsRNAs are within introns and 3′ UTRs of protein-coding genes (Whipple et al. 2015; Blango and Bass 2016; Reich et al. 2018), and a growing body of evidence indicates that dsRNAs impact expression of their associated genes. In *Caenorhabditis elegans*, ADARs edit dsRNAs to prevent their processing by Dicer and silencing of associated genes by antiviral RNAi (Reich et al. 2018). Duplex structures may also impact translation efficiency, as *C. elegans* mRNAs containing dsRNA structures in 3′ UTRs are associated with lighter polysome fractions than mRNAs lacking such structures (Hundley et al. 2008). In human cells, intermolecular interactions between long noncoding RNAs and certain mRNA 3′-UTR sequences recruit Staufen and trigger mRNA decay (Gong and Maquat 2011). Together, these studies indicate that dsRNA structures regulate gene expression by multiple mechanisms.

The organization of eukaryotic genomes into distinct domains also influences gene expression. Genomic domains differ widely by transcription levels, nucleosome

---

organization, and post-translational histone modification patterns (Dixon et al. 2016; Ahringer and Gasser 2018). The *C. elegans* genome is comprised of the sex chromosome, Chromosome X, and five autosomal chromosomes divided into distinct distal arm and central domains. Autosome center domains are dense with genes that exhibit higher levels of expression and are more often essential and conserved (*C. elegans* Sequencing Consortium 1998; Cutter et al. 2009), in part because of low meiotic recombination frequency in autosome centers. In contrast, autosome distal arms are characterized by high recombination frequency, high density of genomic repeats (mostly DNA transposons), and histone modifications associated with gene repression, particularly trimethylation of histone H3 at lysine 9 (H3K9me3) and lysine 27 (H3K27me3) (Liu et al. 2011; Towbin et al. 2012; Ho et al. 2014). Chromatin immunoprecipitation (ChIP) experiments using the nuclear lamina component LEM-2 show that autosome arms are physically associated with the lamina, a characteristic of repressive heterochromatin (Ikegami et al. 2010). However, distal arms are not completely silent, as they contain many essential genes and display subdomains with marks of active chromatin that are not lamina-associated (Ikegami et al. 2010; Liu et al. 2011). Although genomic domains clearly exhibit distinct properties, the mechanisms that establish and regulate chromosomal domains are far from understood.

Using three independent methods to identify dsRNAs, we characterized the genomic distribution of *C. elegans* dsRNA loci and report their association with essential and highly expressed genes on the distal arms of autosomes. Correspondingly, histone modifications associated with active transcription are enriched over dsRNA loci. Our observations identify a property common to many highly expressed, essential genes on autosome arms, setting the stage for insight into their regulation and the roles of dsRNA.

RESULTS

Predicted dsRNA structures correlate with editing-enriched regions

We previously reported that *C. elegans* editing-enriched regions (EERs), dsRNAs containing clustered A-to-I editing, are strikingly enriched on the distal arms of autosomes (Whipple et al. 2015; Reich et al. 2018). To determine whether all *C. elegans* dsRNAs or only the edited dsRNAs of earlier analyses exhibit this pattern, we compiled experimental data sets of *C. elegans* dsRNAs by three independent methods.

Our first data set comprised the 1523 EERs previously determined from RNA-seq of *C. elegans* populations at four developmental stages: embryos, L1–L2 larval stages, L3–L4 larval stages, and young adults (Reich et al. 2018).

As detailed in our previous publication, EERs were defined by identifying transcribed regions covered by five or more reads that contained clustered A-to-G changes (Fig. 1A). As a control data set, we randomly chose regions of the transcriptome covered by greater than or equal to five reads and representing the lengths found in the EER population (length-matched, expressed random regions). Because the EER method excluded dsRNAs in lowly expressed regions, as described below, we also sought to identify dsRNAs in an unbiased manner that did not depend on gene expression.

Because most EERs occur in introns of protein-coding genes (Whipple et al. 2015; Reich et al. 2018), as an independent method for determining dsRNAs, we used
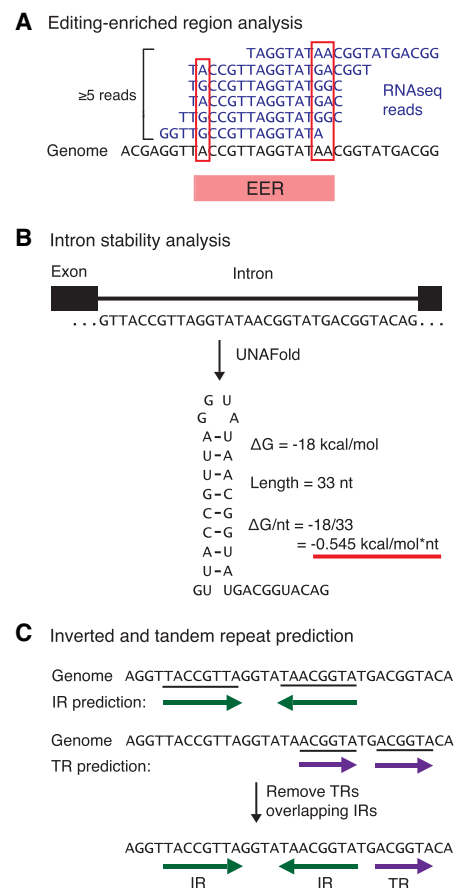

FIGURE 1. Methods to identify dsRNA loci. (*A*) Clustered A-to-G changes in RNA-seq reads defined EERs, transcribed regions that form long dsRNA in vivo and are edited by ADARs. (*B*) Intronic duplex structures were predicted using UNAFold (Markham and Zuker 2008) to determine the folding free energy of intronic sequences of >40 and <9000 nt (example structure not to scale). To compare the stability of introns with different lengths, we normalized each intron's folding free energy to the number of nucleotides (ΔG/nt). (*C*) Inverted (green arrows) and tandem repeat (TR) (purple arrows) sequences were determined computationally from the *C. elegans* genome sequence using previously published tools (Materials and Methods) (Benson 1999; Warburton et al. 2004). Each defined repeat was ≥20 nt; arrows not to scale.

UNAFold to predict folding free energies of *C. elegans* introns at 20°C (Fig. 1B; Supplemental File S1; Markham and Zuker 2008). Introns of >9000 nt, the maximum length for folding with UNAFold, were excluded from analysis, as well as introns of <40 nt, the minimum length needed to accommodate consensus splicing sequences and an intramolecular duplex long enough to bind ADAR (~15 bp) (Herbert and Rich 2001). This data set comprised 96.2% of all *C. elegans* introns annotated by the UCSC Genome Browser (https://genome.ucsc.edu/; ce10/WS220). To control for differences in intron length, we normalized the predicted folding free energy of each intron to its length, giving a total of 109,982 introns with ΔG/nt values that ranged from 0.03 to −1.01 kcal/mol × nt (where more negative values equate to more stable structures). Using ΔG/nt values, we also determined the most structured intron (i.e., most thermodynamically stable per base) for each of 20,735 intron-containing genes (Supplemental File S2).

For the vast majority of intron-containing genes (75.6%), the most structured intron had ΔG/nt values between −0.2 and −0.4 kcal/mol × nt. However, 7.3% of intron-containing genes carried at least one exceptionally structured intron (ΔG/nt < −0.5 kcal/mol × nt). The latter highly structured introns overlapped a large fraction of EERs (710/1523 EERs) suggesting many structured introns are edited and thus form dsRNA. In contrast, only 105 length-matched random regions overlapped a structured intron. Our results demonstrate that intron folding free energy is an accurate and specific predictor of dsRNA formation.

Most EERs form dsRNA through intramolecular interactions of inverted repeat (IR) sequences (Whipple et al. 2015; Blango and Bass 2016; Reich et al. 2018), so for the third data set we used the publicly available Inverted Repeats Finder tool (Warburton et al. 2004) to predict IRs of ≥20 bp in the *C. elegans* genome (Fig. 1C; Supplemental File S3). IRs were only defined where the two repeats were separated by no more than 2 kilobases (kb). Unlike our data sets for EERs and structured introns, the list of IRs only included regions predicted to form dsRNA, and predicted intervening loop sequences were excluded. As a control data set for IRs, we used the Tandem Repeats Finder (Benson 1999) to predict TRs in the *C. elegans* genome, which are less likely to fold into stable intramolecular structures (Supplemental File S3). To avoid low complexity sequences (e.g., dinucleotide tracts, telomeric repeats), we only considered TRs comprised of repeated units of a unique sequence of ≥20 nt long. We reasoned that TR sequences might also form dsRNA if they occurred nearby an inverted copy of the TR sequence, and thus we removed TR sequences predicted to overlap IRs. Emphasizing that these methods effectively predicted dsRNAs, we found that 1476 EERs (96.9%) overlapped IRs, whereas only 159 EERs (10.4%) overlapped TRs.

Gene-associated dsRNA structures cluster on autosome distal arms

To characterize the genomic distribution of dsRNAs, we first compared the relative position of EERs on *C. elegans* chromosomes to the control set of length-matched expressed random regions (Reich et al. 2018). We divided each chromosome into 10 equal segments and determined the number of EERs and control random regions in each chromosomal segment (Fig. 2A). Although 89.5% of EERs occurred outside the central four segments of any chromosome, only 59.8% of control regions were found in distal arm segments, a significantly different distribution ($P < 0.0001$, Kolmogorov–Smirnov test). EERs were also more prevalent on the five *C. elegans* autosomes, as only 50 EERs (3.3%) occurred on Chromosome X, compared to 100 random regions (6.6%). These results confirm previous findings that EERs are enriched on distal arms of *C. elegans* autosomes (Whipple et al. 2015).

To determine if EER locations reflected the distribution of dsRNA loci regardless of editing, we next plotted intron ΔG/nt values along each chromosome (Fig. 2B). Resembling EER distribution patterns, intron ΔG/nt decreased in the distal arms of all five *C. elegans* autosomes, but not Chromosome X. The differences in intronic ΔG/nt between arm and center domains were remarkably clear in some cases (e.g., on Chromosomes I and III) allowing us to refine boundaries of "arm" and "center" domains for each chromosome (dotted vertical lines in Fig. 2B; Supplemental Table S1; see Materials and Methods), which we used in all subsequent analyses, except where noted.

To determine if dsRNA clustering on autosome arms was conserved, we evaluated a related nematode species, *C. briggsae*, which also exhibits chromosome arm and center domains with respect to recombination rates, gene density, and repeat density (Hillier et al. 2007). Although neither ADAR editing sites nor EERs have been mapped for *C. briggsae*, we predicted intron ΔG/nt values for *C. briggsae* (Supplemental File S1) as we did for *C. elegans*. When we plotted intron ΔG/nt values relative to chromosome position, we found that, like in *C. elegans*, introns on *C. briggsae* distal autosome arms formed more stable dsRNA structures than those in the central autosome regions or on Chromosome X (Fig. 2B).

When we mapped predicted IR and TR regions to chromosomal segments shown in Figure 2A, we found that 77.6% of IRs and 75.3% of TRs occurred in autosome distal arm regions (Fig. 2C), which comprise only 49.4% of total genomic sequence. To determine if IRs and TRs were associated with protein-coding genes, we determined the fraction of IR and TR sequences that mapped to annotated exons, introns, UTRs, and noncoding RNAs (ncRNAs), as well as the fraction that mapped to intergenic sequences either proximal (<1 kb) or distal (>1 kb) to genes (Fig. 2D). Similar to EERs (Whipple et al. 2015; Reich et al.
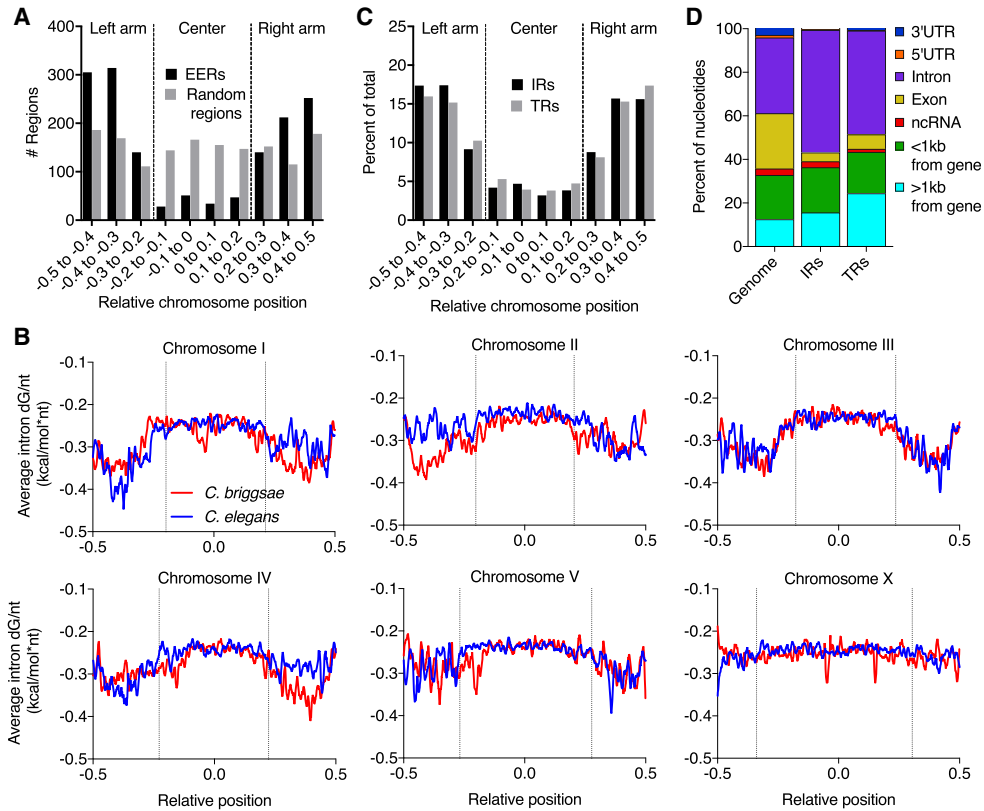
FIGURE 2. dsRNAs cluster on the distal arms of two *Caenorhabditis* species. (*A*) The distribution of EERs (black) and random expressed regions (gray) over the six major *C. elegans* chromosomes divided into 10 equal segments. The center of each chromosome is defined as position 0, the end of the left arm is −0.5 and the end of the right arm is position 0.5. (*B*) Smoothed distributions of average intron ΔG/nt across *C. elegans* (blue) and *Caenorhabditis briggsae* (red) chromosomes. More structured introns have lower (more negative) ΔG/nt values. Vertical dotted lines depict chromosome arm and center domains, determined by intron ΔG/nt (Materials and Methods). (Adapted from Reich and Bass 2019, with permission from Cold Spring Harbor Laboratory Press.) (*C*) The distribution of inverted repeats (IRs; black) and tandem repeats (TRs; gray) across relative chromosomal segments. (*D*) The percentages of the entire *C. elegans* genome, IR regions, and TR regions that overlap annotated gene features.

2018), the majority of IR (56.3%) sequences mapped to intronic sequences, which comprise just 34.7% of total genomic sequence. Intronic sequences also made up the largest fraction of TRs (47.6%), although a higher fraction of TR sequences occurred outside annotated genes compared to IRs. Our observations suggest that IR and TR sequences are enriched on distal autosome arms and largely occur in noncoding regions of protein-coding genes, similar to EERs. Thus, from our three different analyses, we conclude that gene-associated dsRNAs cluster on distal autosome arms in two *Caenorhabditis* species.

RNA structures are enriched in essential genes

In *adr-1;adr-2* mutant worms, EERs associated with protein-coding genes are cleaved by Dicer to generate siRNAs, which induce gene silencing by RNAi (Reich et al. 2018). However, the function of EERs in wild-type animals is unknown, and we wondered whether dsRNAs show signatures of functional importance. Mammalian EER sequences are not conserved between mouse and hu-

man, but mouse EER-associated genes (EAGs) have orthologous human genes containing EERs at a greater fraction than expected by chance (Blango and Bass 2016). To assess if similar sets of orthologous *Caenorhabditis* genes were associated with dsRNAs, we compared *C. elegans* genes containing structured introns (ΔG/nt < −0.5 kcal/ mol × nt) to genes whose *C. briggsae* ortholog had a structured intron (Supplemental File S4). We found that 147 of 1521 *C. elegans* genes (9.6%) containing structured introns had an orthologous gene in *C. briggsae* that contained a structured intron, more than double the number expected by chance (expected 62 genes; $P < 0.0001$, χ^2 test). This suggested that dsRNA-associated genes maintain association with duplex structures over many millions of years. However, as observed for mouse and human EAGs (Blango and Bass 2016), structured introns did not always occur in the same intron in orthologous genes (Supplemental Fig. S1). Additionally, when we compared minimum intron ΔG/nt values of orthologous *C. elegans* and *C. briggsae* genes containing structured introns, we found no correlation (Pearson correlation coefficient =

−0.03), consistent with a lack of direct sequence conservation. Thus, the exact structures do not appear conserved between species, but common sets of genes associate with dsRNA, possibly for some functional purpose.

If dsRNAs serve a function in their associated genes, we hypothesized that they would be enriched among physiologically important, essential genes. To test this, we mined WormBase (www.wormbase.org) for genes that induced lethal or sterile phenotypes when disrupted by RNAi or genetic mutation, a total of 1906 genes that we defined as essential (Supplemental File S2). When we overlapped this list of essential genes with the complete list of 955 EAGs defined previously (Reich et al. 2018), we observed 213 essential EAGs, more than double the number expected by chance (Fig. 3A). Further, essential genes were significantly enriched among EAGs expressed and edited in individual developmental stages, except young adults, where only 128 EERs were defined. Our results were surprising because EERs and EAGs cluster on autosome distal arms, whereas conserved and essential genes are more abundant in chromosome centers (Cutter et al. 2009; Liu et al. 2011). Strikingly, we found that a large subset of essential genes on autosome arms contained EERs (Fig. 3B). More than 30% of the essential genes on the arms of Chromosomes I and V were EER-associated. In contrast, essential genes in autosome centers or on Chromosome X did not show enriched association with EERs. Further, a much lower fraction of nonessential (NE) genes on autosome arms contained EERs than essential genes. These results indicate that EERs occur within functionally important genes on distal autosome arms.

We wondered if the observed enrichment of EERs in essential genes was due to differences in expression. If essential genes had higher read coverage than NE genes in RNA-seq data sets used to define EERs, EERs might appear as enriched in essential genes because editing in NE genes was missed as a result of low coverage. In intersecting EAGs and essential genes (Fig. 3A), we limited comparisons to genes with the read coverage depth we used to define an EER (greater than or equal to five reads); however, we still worried that gene expression differences might create bias. To avoid complications from differences in gene expression, we analyzed ΔG/nt of the most structured intron for all intron-containing genes, essential and NE, on chromosome arms and centers (Fig. 3C). Compared to genes in chromosome centers, both essential and NE genes on distal arms had introns with more stable predicted structures (i.e., lower ΔG/nt). However, on autosome arms, but not Chromosome X, introns of essential genes had more stable predicted structures than NE genes. Thus, independent of gene expression, we observed an association between essential genes and the predicted thermodynamic stability of dsRNA structures.

We next determined if IRs or TRs were enriched within essential genes. We determined the number of IR and TR nucleotides in noncoding regions (i.e., introns and UTRs) of each gene on autosome arms (Supplemental File S2). Then, we separated genes into four categories of increasing IR or TR content and determined the fraction of essential and NE genes in each category (Fig. 3D). Consistent with earlier observations, we found that IRs were associated with essential genes, because 18.2% of essential genes on autosome arms contained >1000 nt of IR sequence, compared to only 6.2% of NE genes. TRs were
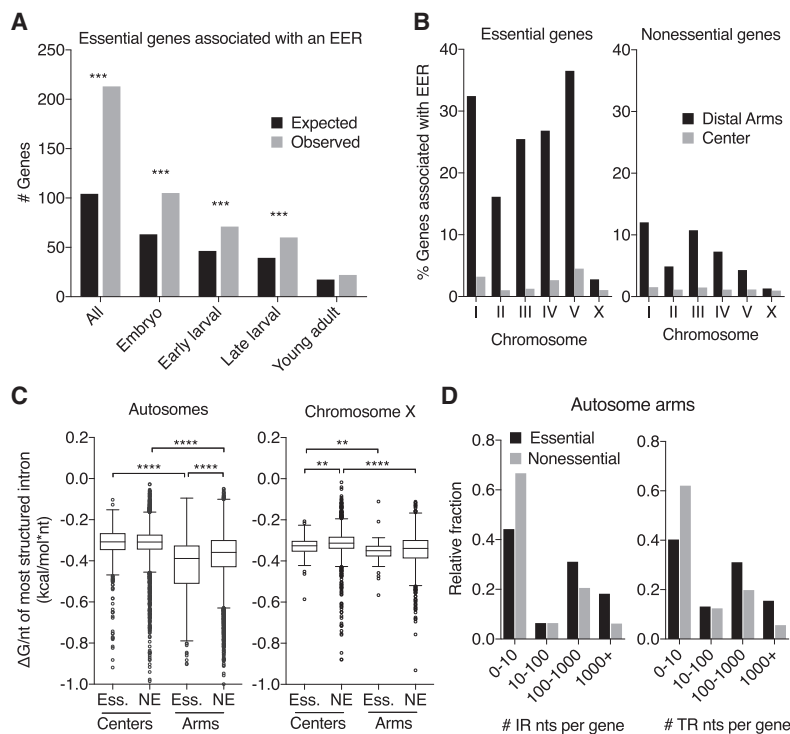

FIGURE 3. dsRNA structures are enriched in essential genes. (A) The number of essential genes expected by chance to overlap genes associated with EERs from all stages combined (All) or individual development stages compared to the actual observed number of essential genes that overlapped EAGs. (***) $P<0.001$; χ^2 test. (B) The percentage of essential (left) and nonessential (NE; right) genes associated with an EER on the distal arm and center domains of each chromosome. (C) Tukey box plots of ΔG/nt of the most structured intron in essential (Ess.) and NE genes in center and arm domains of C. elegans autosomes and Chromosome X. (**) $P<0.01$; (****) $P<0.0001$; Mann–Whitney test. (D) Relative fractions of essential and NE genes on autosome arms, separated by the amount of IR (left) or TR (right) sequence in each gene's noncoding elements (i.e., introns, UTRs).

also associated with essential genes, although to a lesser extent than IRs, as 15.5% of essential genes carried >1000 nt of TR sequence, compared to 5.6% of NE genes.

Genes with dsRNA structures are highly expressed

Genes on distal autosome arms typically exhibit lower gene expression and elevated levels of H3K9me3 and H3K27me3 relative to genes in chromosome centers (Liu et al. 2011). Given their enrichment on distal autosome arms, we wondered if EAGs would follow these trends. We separated genes by their position on chromosome arms or centers, and plotted embryo-stage expression of EAGs and all expressed genes (Fig. 4A). As expected, when considering all expressed intron-containing genes, those on autosome arms exhibited lower overall expression than genes in autosome centers. However, EAGs showed the opposite trend, with those on autosome arms showing higher expression than those in autosome centers. Furthermore, EAGs on autosome arms displayed significantly higher expression than the set of all autosome arm genes. On Chromosome X, we observed no significant relationship between gene expression and EER association. Using gene expression data from four different developmental stages, we found that arm EAGs had higher expression than all autosome arm genes at each stage (Fig. 4B).

To confirm that our observations were not biased by the minimum read coverage required to identify EERs, we also parsed genes into those with and without a structured intron (ΔG/nt < −0.5 kcal/mol × nt), which would be unrelated to read coverage. We plotted expression of these structured genes and all genes in each chromosomal domain (Fig. 4C). As observed with EAGs, genes with structured introns exhibited significantly higher expression than the set of all genes on autosome distal arms, but not in autosome centers or on Chromosome X. We conclude that dsRNA structures are associated with highly expressed genes on autosome arms.
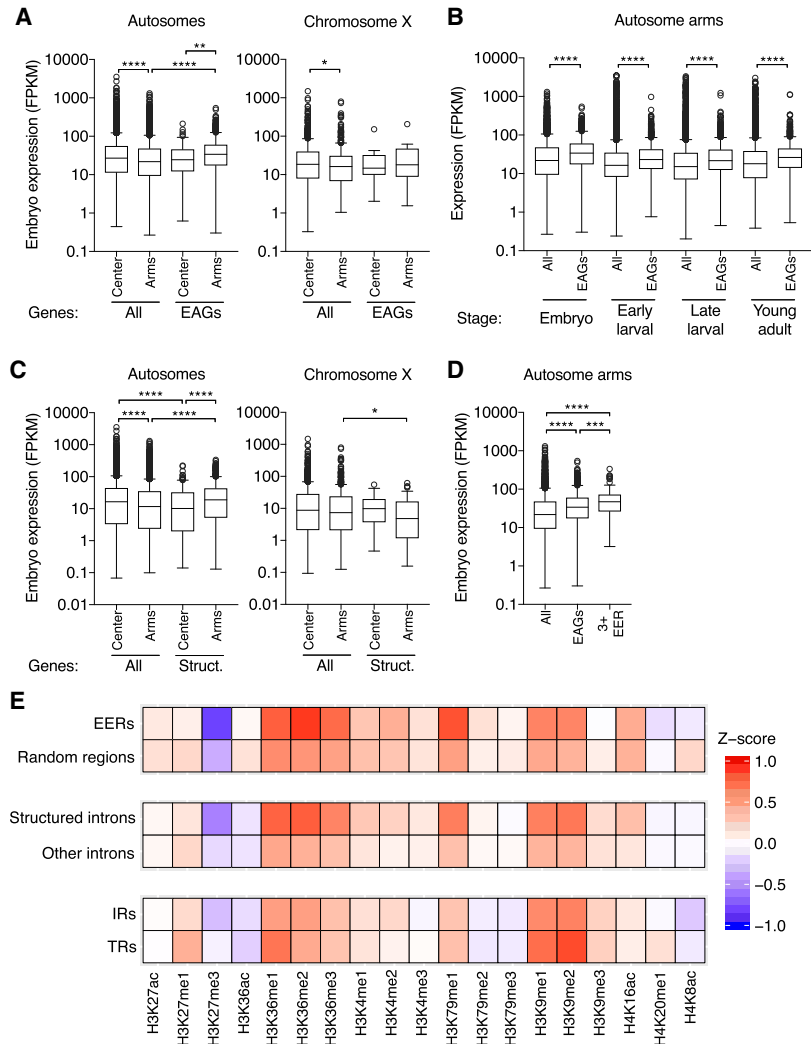

FIGURE 4. dsRNAs are associated with highly expressed genes and marks of transcriptional elongation. (*A*) Tukey box plots show the distributions of embryo-stage gene expression (FPKM: fragments per kilobase-million reads) for all genes or EAGs on distal arm or center domains of autosomes or Chromosome X. (*) $P < 0.05$; (**) $P < 0.01$; (****) $P < 0.0001$; Mann–Whitney test. (*B*) Tukey box plots as in *A* showing expression in each developmental stage of all genes or EAGs on autosome distal arms. (****) $P < 0.0001$; Mann–Whitney test. (*C*) Tukey box plots as in *A* showing embryo-stage expression of all genes or genes containing a structured intron (Struct.; intron ΔG/nt < −0.5 kcal/mol × nt), separated by chromosomal domain. (*) $P < 0.05$; (****) $P < 0.0001$; Mann–Whitney test. (*D*) Tukey box plot as in *A* showing embryo-stage expression of all genes, EAGs, or genes containing three or more distinct EER-associated introns and/or UTRs (3+ EER) present on autosome arms. (***) $P < 0.001$; (****) $P < 0.0001$; Mann–Whitney test. (*E*) Heatmap displays the relative ChIP-chip signal for 19 histone modifications over EERs, random regions, structured (ΔG/nt < −0.5 kcal/mol × nt) or unstructured introns, IR loci, and TR loci present on autosome distal arms.

In considering the correlation between dsRNA structure and elevated expression, we noted that many genes contained multiple dsRNA structures. Specifically, we identified 125 genes, all located on autosome distal arms, where EERs overlapped three or more distinct introns, UTRs (5′ or 3′), or proximal (<1 kb) upstream or downstream regions, which often include unannotated UTRs (Whipple et al. 2015). Included among these highly

structured genes were examples such as *daf-2* and *ssl-1*, which contained seven and 12 edited introns/UTRs, respectively. Of these 125 genes, 43 were essential, more than three times the number expected by chance (expected 14 genes, $P < 0.0001$, χ^2 test). Strikingly, the 125 EAGs with three or more EER-associated introns/UTRs displayed significantly higher gene expression than the set of all autosome arm genes (Fig. 4D). Indeed, when we ranked autosome arm genes by their expression in embryos, 83 of the 125 highly structured genes (66.4%) were in the top expression quartile, suggesting that these genes are among the most highly expressed on autosome arms. Together, our observations provide a compelling correlation between gene-associated dsRNA structures and elevated gene expression on distal autosome arms.

Histone modifications linked to transcriptional elongation are enriched over dsRNA loci

Gene expression correlates with histone modification patterns, which have been mapped in detail for the *C. elegans* genome (Liu et al. 2011; Towbin et al. 2012; Ho et al. 2014; Evans et al. 2016). Modifications like H3K4me3 and acetylated H3K27 (H3K27ac) typically mark active promoters, whereas H3K36me marks (mono-, di-, and trimethylation) often correlate with transcriptional elongation (Venkatesh and Workman 2013; Evans et al. 2016; Ahringer and Gasser 2018). In contrast, H3K9me3 and H3K27me3 are enriched over transcriptionally silent regions (Liu et al. 2011; Towbin et al. 2012; Ho et al. 2014) and deposited on targets of nuclear RNAi (Gu et al. 2012; Mao et al. 2015). Because cellular dsRNAs can be targets of RNAi (Reich et al. 2018), but are also associated with higher-than-normal expression on autosome arms, we wondered if they would exhibit chromatin modification patterns associated with either silencing or transcription.

We analyzed modENCODE ChIP and microarray (ChIP-chip) data to determine the frequency of 19 histone modifications at dsRNA or control loci on autosome arms (Fig. 4E). Compared to the control set of random regions, EERs showed enrichment for H3K36me1, me2, and me3, as well as H3K9me1 and me2, but not H3K9me3. Additionally, EERs exhibited very low levels of H3K27me3, a mark of transcriptional repression. When we compared histone modification patterns over structured introns ($\Delta G/nt < -0.5$ kcal/mol × nt) and all other autosome arm introns, we observed that chromatin marks over structured introns closely resembled those over EERs. IRs displayed more subtle enrichment of H3K36me marks

and H3K9me1 and me2, but, like EERs and structured introns, did not exhibit high levels of the silencing marks H3K9me3 and H3K27me3. Overall, modification patterns over IRs were similar to those at TR loci. A large fraction of IRs and TRs are not in genes (Fig. 2D), and we observed histone modification patterns more similar in magnitude to those at EERs and structured introns if we restricted our IR/TR analyses to regions associated with genes (Supplemental Fig. S2). Our analyses indicate that gene-associated dsRNAs are associated with marks of transcriptional elongation rather than those of gene silencing.

High repeat content is associated with highly expressed distal arm genes

Most ADAR editing in *C. elegans* occurs in repetitive sequences, primarily those derived from DNA transposons (Zhao et al. 2015; Reich et al. 2018). Thus, we wondered if the observed association between dsRNAs and highly expressed genes could be explained by a correlation between repeat content and gene expression. To test if this was the case, we examined embryo-stage expression of autosome arm genes stratified by the amount of IR and TR sequences in introns and UTRs (Fig. 5A,B; Supplemental File S2). We hypothesized that if gene expression correlated with dsRNA structure, rather than repeat content, we would observe higher expression of genes with greater IR content, but not TR content. However, although we observed higher gene expression of genes with greater IR content, we observed the same trend in genes stratified by TR content. To test if expression correlated with repeat content generally, we determined the overlap between RepeatMasker-annotated DNA transposons and introns/UTRs of each gene. When we stratified genes by DNA transposon content and plotted expression (Fig. 5C), again we observed that higher repeat content correlated with higher expression. We cannot rule out that repeat content correlates with
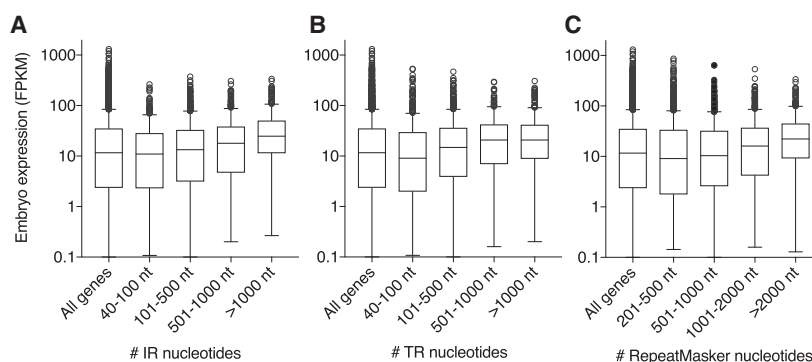
FIGURE 5. Expression of autosome arm genes by repeat content. Tukey box plots show embryo-stage expression of all autosome arm genes (*leftmost* box in each plot) and genes segregated by the amount of noncoding sequence in each gene that overlaps (A) IRs, (B) TRs, or (C) annotated RepeatMasker transposon repeats.

expression because of increased prevalence of dsRNA structures. However, because of the close relationship between duplex structure and repetitive sequences, we cannot definitively conclude that associations we describe relate explicitly to RNA structure.

To understand if specific types of transposons were associated with elevated gene expression, we determined if essential genes and highly expressed genes (>50 FPKM) on autosome arms had specific transposons that were significantly enriched or depleted ($P < 0.05$, binomial test; Supplemental File S5; Supplemental Methods; Kapusta et al. 2013). Interestingly, four out of five significantly enriched transposons in essential autosome arm genes (PAL5A_CE, IR3_CE, CELE1, and PALTTAA3_CE) were annotated in Repbase as "palindromic" repeats, suggesting that they comprise long IRs and likely form dsRNA; indeed all four were commonly found within EERs (Reich et al. 2018). These same four palindromic repeats, as well as CELE46B, another common element in EERs, were among the eight transposons significantly enriched in highly expressed genes. In contrast, when considering transposons significantly depleted from essential and highly expressed (>50 FPKM) autosome arm genes (Supplemental File S5), palindromic repeats comprised only 12% (3/25) and 14% (4/29), respectively. Although we cannot rule out that transposon-derived sequences might contribute to elevated expression of dsRNA-associated genes, the disproportionate enrichment of transposons that form dsRNA in highly expressed autosome arm genes hints at a functional role for dsRNA structure.

If repeat-associated dsRNA structures promote gene expression, in theory, expression differences should be observed in identical genes with and without dsRNA elements. Two well-studied and genetically diverse *C. elegans* strains, Bristol N2 and the Hawaiian strain CB4856, exhibit distinct transposon insertion and deletion patterns (Vergara et al. 2014; Laricchia et al. 2017) that result in some genes containing dsRNA elements in one strain but not the other. We used mRNA expression data from each strain (Kamkina et al. 2016) to compare strain-specific expression of genes in which a palindromic transposon in N2 was deleted in CB4856 (Supplemental Fig. S3A; Supplemental Methods). Of 10 such genes, three exhibited significantly reduced expression in CB4856, whereas one gene was significantly up-regulated and six were unchanged. When we looked at genes where a palindromic transposon had inserted in CB4856, but not N2, most genes (17/22) were not significantly different, whereas two genes were higher in N2 and three significantly lower in CB4856 (Supplemental Fig. S3B). These observations do not provide obvious support for our hypothesis, although the small number of genes containing strain-specific palindromic transposons may have been insufficient to observe modest trends that would be more apparent in a larger gene set. Further, additional genetic changes might

underlie the expression differences of the genes studied in this analysis. We hope to more conclusively test the impact of dsRNA loci on gene expression in future studies.

DISCUSSION

Here we used three methods to identify *C. elegans* loci expressing long dsRNA and evaluated these loci for essential genes and expression. We find that dsRNAs are enriched on autosome distal arms, in essential and highly expressed genes. Although we cannot determine if these associations are driven by a functional effect of dsRNA or a related property, like transposable element content, our results suggest dsRNAs positively correlate with gene expression in *C. elegans*.

How well do the three methods identify dsRNAs?

Previous work established *C. elegans* EERs as long dsRNAs that are edited by ADARs (Whipple et al. 2015; Reich et al. 2018), and these EERs served as one data set. Structured introns and predicted IRs comprised two additional data sets that, importantly, predicted dsRNAs from genomic sequence alone, regardless of expression; both data sets overlapped with EERs, lending confidence that they included dsRNAs. Overlap with other data sets of presumptive dsRNAs reiterates that EERs include regions of dsRNA structure. For example, prior analyses of *C. elegans* dsRNAs showed 18% of EERs overlapped transcribed regions resistant to ssRNA-specific nucleases (Li et al. 2012; Whipple et al. 2015). In the current analysis, we found that 57% of EAGs overlapped genes enriched by immunoprecipitation with the J2 dsRNA-specific antibody (545/955 EAGs, expected 154, $P < 0.0001$, χ^2 test) (Saldi et al. 2014).

However, 63.1% (1248/1979) of structured introns (ΔG/nt < −0.5 kcal/mol × nt) and 89.5% (28762/32128) of IRs did not overlap EERs (Supplemental Files S1, S3). Possibly these transcripts escape editing by ADARs, for example, because of competitive binding by other dsRBPs. However, the read coverage parameters used to define EERs better explain why many predicted dsRNAs do not overlap EERs. Specifically, 55.9% of structured introns and 78.1% of IRs that did not overlap EERs also failed to overlap any region covered by five or more reads in our original RNA-seq data sets (Supplemental Files S1, S3, S6). Of the five *C. elegans* introns with the most stable predicted structures, only two overlapped EERs. The remaining three structured introns failed to meet the five-read coverage depth threshold needed to define EERs, but all exhibited aligned reads containing A-to-G mismatches indicative of ADAR editing (Supplemental Fig. S4). Additionally, 53.4% of EERs did not overlap a structured intron, perhaps because our stability cutoff of −0.5 kcal/mol × nt was too stringent. Because EERs sometimes comprise

only a portion of an intron, introns containing duplex structures may also have additional unstructured sequences that cause ΔG/nt values to exceed the −0.5 kcal/mol × nt threshold.

Unlike structured introns, nearly every EER (96.9%) overlapped a predicted IR. Because IRs form dsRNA through intramolecular base-pairing, this finding suggests nearly all EERs can form intramolecular duplexes. The prevalence of IRs within EERs further indicates that edited intermolecular dsRNAs may be rare in *C. elegans*. If complementary RNAs interact to form intermolecular dsRNAs, perhaps this occurs primarily in the cytoplasm where they cannot be readily edited by ADR-2, which is mainly nuclear (Ohta et al. 2008). Alternatively, intermolecular dsRNA formation may be limited by mechanisms that prevent both sense and antisense transcription in the same cell, as has been observed at loci in yeast and *Arabidopsis* (Castelnuovo et al. 2013; Rosa et al. 2016).

What is the important property of dsRNA loci?

We observed a correlation between dsRNA-producing loci and highly expressed and essential genes, but as yet we do not know if this correlation depends on transcription of loci into dsRNAs. In fact, dsRNA loci have other characteristics that could underlie the associations we describe.

For one, most dsRNA loci are repetitive, with 62.9% of EER sequences (Reich et al. 2018), 27.9% of structured introns, and 25.9% of IRs mapping to RepeatMasker-annotated DNA transposons, which comprise ~12% of the total genome. If particular transposon sequences in dsRNA loci promote gene expression at the DNA or RNA level, this might explain the observed association between dsRNA loci and highly expressed genes. Transposons are known to provide gene regulatory sequences that act as transcription factor binding sites, splicing regulatory elements, and noncoding RNAs (Lev-Maor et al. 2007; Chuong et al. 2017), so it is conceivable that transposon sequences might promote gene expression (Supplemental Fig. S5A). Indeed, we observed a positive correlation between the amount of noncoding RepeatMasker transposon sequence in a gene and its relative expression (Fig. 5C). If transposons provide sequences that promote gene expression, we would expect to find specific transposons or transposon families enriched within highly expressed genes. Although we did observe several significantly enriched transposons, most of these derive from different families and have long, IR (i.e., palindromic) sequences (Supplemental File S5), suggesting that the propensity to form dsRNA may be a relevant property of transposons associated with elevated gene expression.

Alternatively, the association between highly expressed genes and transposon-derived dsRNAs could be explained by a propensity for transposons to insert into actively transcribed genes or genes with open chromatin. Observations

in mouse embryonic stem cells provide mixed support for this model, as murine leukemia virus and PiggyBac transposons preferentially insert within highly expressed genes, whereas insertion profiles of mouse mammary tumor virus and Tol2 and Sleeping Beauty transposons exhibit a weak preference or no preference at all for highly expressed genes (de Jong et al. 2014; Yoshida et al. 2017). If transposon insertion preferences underlie the association between *C. elegans* dsRNA loci and highly expressed genes, heritable dsRNA loci should specifically associate with genes highly expressed in germline tissues, but not necessarily somatic tissues. To assess this, we examined gene expression data from oogenic, spermatogenic, neuronal, intestinal, and pharyngeal tissues (Ortiz et al. 2014; Spencer et al. 2014; Blazie et al. 2015), binning genes into categories of low to high expression in each tissue and measuring the fraction of genes with EERs or structured introns (Supplemental Fig S6). In oogenic and spermatogenic tissues, EERs and structured introns were more prevalent in moderately-to-highly expressed genes, and this trend was absent in intestine and pharyngeal tissues. However, dsRNA loci were also prevalent in genes highly expressed in NSM neurons, a somatic tissue where transposon insertions would not be heritable. Further, in no tissue were dsRNA loci most prevalent in the most highly expressed genes (>100 FPKM), an observation inconsistent with the model that dsRNA-forming transposons preferentially insert in highly expressed regions. Thus, transposon insertion preferences may only partly explain the association between dsRNA loci and highly expressed genes.

Differences in chromatin modifications at dsRNA loci, like enrichment of transcription-associated H3K36me marks (Fig. 4E), may shed light on our observations. However, the relevance of enrichment of certain modified histones (e.g., H3K9me1/2 compared to H3K9me3) is not yet clear. H3K9me1/2 and H3K9me3 exhibit distinct genome-wide patterns in *C. elegans*, with greater H3K9me3 enrichment over retroelements and transcriptionally silent genes (Zeller et al. 2016; Ahringer and Gasser 2018), consistent with a role in nuclear RNAi (Gu et al. 2012; Billi et al. 2014). H3K9me1/2 deposition also occurs over repetitive DNA, but these marks are not excluded from expressed regions (Liu et al. 2011; Garrigues et al. 2015; Zeller et al. 2016). In *Schizosaccharomyces pombe*, RNAi-dependent deposition of H3K9me2 or H3K9me3 leads to different transcriptional states, with H3K9me3 domains silent and H3K9me2 domains remaining transcriptionally active (Jih et al. 2017). We speculate that enriched H3K9me1/2 over dsRNA loci indicates a similar transcriptionally permissive state, although perhaps one poised for silencing in particular tissues or conditions.

Structured and edited sequences share other properties that might relate to their association with highly expressed genes. For instance, we observed that intron structure was moderately correlated with both intron length and

with the occurrence of periodic A_n/T_n clusters (PATCs) (Supplemental Fig. S7A,B; Supplemental Methods), repeated intronic DNA elements that promote germline gene expression in *C. elegans* (Frøkjær-Jensen et al. 2016). Because structured introns also tend to be long and have more PATC sequences, these properties could underlie the association observed between intron structure and elevated gene expression. However, among genes with long introns (>500 nt) and high PATC content, genes with structured introns (ΔG/nt < −0.5 kcal/mol × nt) still exhibited higher gene expression than those without intronic structures (Supplemental Fig. S7C,D), suggesting intron length and PATC content may not entirely explain the association between dsRNA structure and elevated expression. Evolutionary pressure might contribute to the association, as essential genes silenced because of transposon insertions could accumulate transcription-promoting elements over time to counteract silencing, leading to heightened expression of genes with transposon-associated dsRNA loci. Finally, we cannot rule out the possibility that structures formed by DNA sequences, rather than RNA, underlie the association between dsRNAs and highly expressed genes.

How might dsRNA promote gene expression?

Although we cannot rule out repetitive sequences, transposon insertion preferences, or DNA structures as the reason for the correlation between dsRNA and highly expressed genes, we are intrigued by the possibility that dsRNA might promote gene expression. However, we can only speculate on the mechanism at this stage. Although it is possible that dsRNA structures confer slower rates of mRNA decay, we do not favor this model because most dsRNAs are intronic (Whipple et al. 2015; Reich et al. 2018) and would be unlikely to affect mRNA stability after removal by splicing. Possibly intronic dsRNA structures promote efficient splicing by folding to bring 5′ and 3′ splice sites in close proximity, thus promoting mRNA expression (Supplemental Fig. S5B). Alternatively, nascent dsRNA structures might bind nuclear dsRBPs that in turn recruit transcription machinery to promote expression (Supplemental Fig. S5C).

If dsRNAs promote gene expression through interactions with dsRBPs, which proteins might mediate this phenomenon? Although Dicer primarily acts on dsRNA to silence genes (Carthew and Sontheimer 2009; Billi et al. 2014), the dsRBPs ADAR, Staufen, and NF90 are candidate factors because of well-established roles in gene regulation (Bass et al. 1994; Heraud-Farlow and Kiebler 2014; Castella et al. 2015; Nishikura 2016). Expression of ADAR-edited dsRNAs positively correlates with ADAR expression in human brain tissue (Liscovitch et al. 2014), suggesting that ADARs could promote dsRNA expression in humans. In *C. elegans*, we showed that ADARs promote EAG ex-

pression by antagonizing RNAi silencing (Reich et al. 2018). However, analysis of RNA-seq data sets of the latter studies showed that EAGs exhibit significantly higher expression than the set of all autosome arm genes even in ADAR null strains, with or without additional mutations in RNAi factors (Supplemental Fig. S8). Thus, neither ADARs nor the RNAi machinery are required to maintain the association between dsRNAs and highly expressed genes.

Besides ADARs, other dsRBPs might promote expression of their substrates. In addition to characterized functions in RNA transport (Heraud-Farlow and Kiebler 2014) and mRNA decay (Park and Maquat 2013), the dsRBP Staufen has pleiotropic effects that are not well understood. There is significant overlap between EAGs and transcripts bound by the *C. elegans* Staufen ortholog, STAU-1 (72 of 415 STAU-1 target genes, expected 29, $P < 0.0001$, χ^2 test) (LeGendre et al. 2013), raising the possibility that STAU-1 regulates some of the highly expressed, dsRNA-associated genes we identified. In mammals, the dsRBP NF90 regulates gene expression through effects on transcription, mRNA stability, and translation (Castella et al. 2015). Although *C. elegans* lacks an ortholog of NF90, it encodes an NF90-related gene orthologous to mammalian ZFR, which contains zinc finger domains predicted to bind dsRNA (Finerty and Bass 1999; Wolkowicz and Cook 2012). Thus, the *C. elegans* ZFR ortholog, *Y95B8A.8*, could also underlie the correlation between dsRNAs and highly expressed genes on autosome arms.

MATERIALS AND METHODS

Published data sets

The coordinates of 1523 *C. elegans* EERs are reported in Supplemental File S1 of Reich et al. (2018). RNA-seq data sets used to define these EERs and subsequently used for expression analyses in this report are available from Gene Expression Omnibus (GEO), accession GSE79375. RNA-seq data sets used in Supplemental Figure S8 are deposited at GEO, accession GSE106647. Alignment and read filtering parameters are described in Reich et al. (2018).

We downloaded modENCODE ChIP-chip data from data.mod encode.org. Data were downloaded as $\log_2$ ratios of signal over input, which we normalized, converted to Z-scores, and averaged over replicates.

All tissue-specific expression data sets were downloaded as average FPKM values. Oogenic and spermatogenic data were obtained from Ortiz et al. (2014), neuronal data were from Spencer et al. (2014), and intestinal and pharyngeal muscle data were from Blazie et al. (2015).

Intron stability analysis

A .bed file of annotated *C. elegans* introns (ce10/WS220) was downloaded from the UCSC Genome Browser (https://genome. ucsc. edu/). To compile a list of unique, nonoverlapping introns,

we first removed entries containing the same chromosome, start, and stop coordinates as another. Then, if remaining introns overlapped, we removed the largest intron in each overlap until no overlapping introns remained. For *C. briggsae* introns, we downloaded a file of annotated *C. briggsae* genomic features (cb4/WS248) from WormBase (ftp://ftp.wormbase.org), extracted intron coordinates, and converted them to .bed format. We determined unique *C. briggsae* introns as for *C. elegans*. For both sets of introns, we removed those of <40 nt and >9000 nt.

To predict folding free energies of intronic sequences, we used the bedtools2 (https://github.com/arq5x/bedtools2) application fastaFromBed to extract genomic sequences of each intron. The ΔG of each sequence was determined with UNAFold (Markham and Zuker 2008), using the parameters "-X 1 --mode bases -t 20". For each intron, ΔG was divided by intron length to calculate ΔG/nt.

IR and TR prediction

Inverted repeats were predicted using the Inverted Repeats Finder (http://tandem.bu.edu/irf/irf.download.html) (Warburton et al. 2004) using the following parameters: Match = 2, Mismatch = 5, Delta = 5, PM = 80, PI = 10, Minscore = 40, MaxLength = 20000, MaxLoop = 1000. Repeats comprised of 100% A-T or G-C base pairs were discarded. Overlapping IRs were merged using the bedtools application mergeBed. Loop sequences were not included.

Tandem repeats were predicted with the Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.html) (Benson 1999) using the following parameters: Match = 2, Mismatch = 5, Delta = 5, PM = 80, PI = 10, Minscore = 40, MaxPeriod = 2000. Repeats with a period of <20 or sequence entropy of ≤1 were discarded. Overlapping TRs were merged using the bedtools application mergeBed and sequences overlapping IRs were removed with the bedtools application subtractBed.

To calculate the number of IR and TR nucleotides in each gene's introns and UTRs, we first determined the coordinates for noncoding regions of each gene. We downloaded a table of annotated gene transcription start and stop sites and a table of coding exons from UCSC. We used the bedtools2 application subtractBed to subtract coding regions from annotated transcribed regions to determine each gene's noncoding regions. We then used the bedtools2 application to calculate the number of IR and TR nucleotides that overlapped the table of gene noncoding regions. Finally, for each gene, we summed IR/TR nucleotides in all noncoding regions to determine the total number of IR and TR nucleotides in each gene.

Defining chromosome arm and center boundaries

To define chromosome domain boundaries by intron ΔG/nt, we first determined the ΔG/nt of each gene's most structured intron. Then, we separated each chromosome into 100-kb segments, counted the number of genes in each segment, and determined the fraction of genes with an intron ΔG/nt < −0.5 kcal/mol × nt (% structured genes). Starting at the centermost segment of each chromosome, we moved outward toward each arm until we encountered three consecutive 100-kb segments, each with ≥20% structured genes. Of the three consecutive segments, we took

the margin closest to the chromosome center as the boundary of the distal arm.

Essential gene analysis

Lists of genetic alleles and RNAi experiments causing lethality and sterility were downloaded from WormBase (http://www.wormbase.org/). We removed all mutant alleles and RNAi experiments that ambiguously targeted more than one gene. We further removed those causing male sterile phenotypes that were not also lethal or sterile in hermaphrodites. From the remaining alleles and RNAi experiments, we extracted Ensembl gene IDs of targeted genes and removed duplicates to generate a full list of essential genes.

For each developmental stage, we filtered out all genes that did not contain a region covered by five or more reads in our RNA-seq data from that stage. We then overlapped the lists of expressed essential genes and EAGs to determine the number of genes in common. The USeq application IntersectLists was used to determine overlapping genes and determine the significance of overlapping sets by χ^2 approximation with 10,000 randomized iterations.

Expression and chromatin analyses

A RefFlat table of *C. elegans* genes was downloaded from the UCSC Genome Browser (https://genome.ucsc.edu/). Expression data, calculated as FPKM values, were determined for exonic regions of *C. elegans* genes using DefinedRegionDifferentialSeq, a USeq (http://useq.sourceforge.net/) application shell for the R package DESeq2. Expression analyses in Figures 4 and 5 and Supplemental Figures S7 and S8 excluded genes with 0 FPKM.

For ChIP-chip analyses, we created .bed files of EERs, introns, IRs, and TRs present on autosome arm domains. A control file of length-matched random regions was made for EERs on autosome arms using the bedtools2 application shuffleBed, restricting possible locations to autosome arm regions covered by greater than or equal to five reads in developmental RNA-seq data from Reich et al. (2018) and using parameter "-f 0.5". Using the bedtools2 application intersectBed, we overlapped our regions of interest with .bed files reporting average ChIP-chip *Z*-scores of >50 nt genomic windows for each histone modification. We then calculated the average *Z*-score value over all regions in each group (i.e., EERs, random regions, structured introns, etc.) for each modification, and plotted a heatmap of the results in R.

RepeatMasker analysis

We downloaded a file of *C. elegans* repeats (ce10) from RepeatMasker (http://www.repeatmasker.org/), removed all low complexity and simple repeats, and converted the resulting file to .bed format. Using the bedtools2 application annotateBed, we calculated the number of repeat bases that overlapped noncoding regions of each gene, and then added these values to determine each gene's total noncoding repeat content.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

REFERENCES

Ahringer J, Gasser SM. 2018. Repressive chromatin in *Caenorhabditis elegans*: establishment, composition, and function. *Genetics* **208:** 491–511.

Bass BL, Hurst SR, Singer JD. 1994. Binding properties of newly identified *Xenopus* proteins containing dsRNA-binding motifs. *Curr Biol* **4:** 301–314.

Benson G. 1999. Tandem Repeats Finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27:** 573–580.

Billi AC, Fischer SE, Kim JK. 2014. Endogenous RNAi pathways in *C. elegans.* In *WormBook* (ed. The *C. elegans* Research Community), pp. 1–49. http://www.wormbook.org

Blango MG, Bass BL. 2016. Identification of the long, edited dsRNAome of LPS-stimulated immune cells. *Genome Res* **26:** 852–862.

Blazie SM, Babb C, Wilky H, Rawls A, Park JG, Mangone M. 2015. Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in *Caenorhabditis elegans* intestine and muscles. *BMC Biol* **13:** 4.

C. elegans Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282:** 2012–2018.

Carthew RW, Sontheimer EJ. 2009. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136:** 642–655.

Castella S, Bernard R, Corno M, Fradin A, Larcher JC. 2015. Ilf3 and NF90 functions in RNA biology. *Wiley Interdiscip Rev RNA* **6:** 243–256.

Castelnuovo M, Rahman S, Guffanti E, Infantino V, Stutz F, Zenklusen D. 2013. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* **20:** 851–858.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18:** 71–86.

Cutter AD, Dey A, Murray RL. 2009. Evolution of the *Caenorhabditis elegans* genome. *Mol Biol Evol* **26:** 1199–1234.

de Jong J, Akhtar W, Badhai J, Rust AG, Rad R, Hilkens J, Berns A, van Lohuizen M, Wessels LF, de Ridder J. 2014. Chromatin landscapes of retroviral and transposon integration profiles. *PLoS Genet* **10:** e1004250.

Dixon JR, Gorkin DU, Ren B. 2016. Chromatin domains: the unit of chromosome organization. *Mol Cell* **62:** 668–680.

Evans KJ, Huang N, Stempor P, Chesney MA, Down TA, Ahringer J. 2016. Stable *Caenorhabditis elegans* chromatin domains separate broadly expressed and developmentally regulated genes. *Proc Natl Acad Sci* **113:** E7020–E7029.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41:** 331–368.

Finerty PJ Jr, Bass BL. 1999. Subsets of the zinc finger motifs in dsRBP-ZFa can bind double-stranded RNA. *Biochemistry* **38:** 4001–4007.

Frøkjær-Jensen C, Jain N, Hansen L, Davis MW, Li Y, Zhao D, Rebora K, Millet JRM, Liu X, Kim SK, et al. 2016. An abundant class of non-coding DNA can prevent stochastic gene silencing in the *C. elegans* germline. *Cell* **166:** 343–357.

Garrigues JM, Sidoli S, Garcia BA, Strome S. 2015. Defining heterochromatin in *C. elegans* through genome-wide analysis of the heterochromatin protein 1 homolog HPL-2. *Genome Res* **25:** 76–88.

Gong C, Maquat LE. 2011. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **470:** 284–288.

Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, Fire A. 2012. Amplification of siRNA in *Caenorhabditis elegans* generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet* **44:** 157–164.

Heraud-Farlow JE, Kiebler MA. 2014. The multifunctional Staufen proteins: conserved roles from neurogenesis to synaptic plasticity. *Trends Neurosci* **37:** 470–479.

Herbert A, Rich A. 2001. The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. *Proc Natl Acad Sci* **98:** 12132–12137.

Hillier LW, Miller RD, Baird SE, Chinwalla A, Fulton LA, Koboldt DC, Waterston RH. 2007. Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. *PLoS Biol* **5:** e167.

Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A, et al. 2014. Comparative analysis of metazoan chromatin organization. *Nature* **512:** 449–452.

Hundley HA, Krauchuk AA, Bass BL. 2008. *C. elegans* and *H. sapiens* mRNAs with edited 3′ UTRs are present on polysomes. *RNA* **14:** 2050–2060.

Ikegami K, Egelhofer TA, Strome S, Lieb JD. 2010. *Caenorhabditis elegans* chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biol* **11:** R120.

Jih G, Iglesias N, Currie MA, Bhanu NV, Paulo JA, Gygi SP, Garcia BA, Moazed D. 2017. Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription. *Nature* **547:** 463–467.

Kamkina P, Snoek LB, Grossmann J, Volkers RJM, Sterken MG, Daube M, Roschitzki B, Fortes C, Schlapbach R, Roth A, et al. 2016. Natural genetic variation differentially affects the proteome and transcriptome in *Caenorhabditis elegans. Mol Cell Proteomics* **15:** 1670–1680.

Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* **9:** e1003470.

Laricchia KM, Zdraljevic S, Cook DE, Andersen EC. 2017. Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Mol Biol Evol* **34:** 2187–2202.

LeGendre JB, Campbell ZT, Kroll-Conner P, Anderson P, Kimble J, Wickens M. 2013. RNA targets and specificity of Staufen, a double-stranded RNA-binding protein in *Caenorhabditis elegans. J Biol Chem* **288:** 2532–2545.

Lev-Maor G, Sorek R, Levanon EY, Paz N, Eisenberg E, Ast G. 2007. RNA-editing-mediated exon evolution. *Genome Biol* **8:** R29.

Li F, Zheng Q, Ryvkin P, Dragomir I, Desai Y, Aiyer S, Valladares O, Yang J, Bambina S, Sabin LR, et al. 2012. Global analysis of RNA secondary structure in two metazoans. *Cell Rep* **1:** 69–82.

Liscovitch N, Bazak L, Levanon EY, Chechik G. 2014. Positive correlation between ADAR expression and its targets suggests a complex regulation mediated by RNA editing in the human brain. *RNA Biol* **11:** 1447–1456.

Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung MS, Ercan S, Ikegami K, Jensen M, Kolasinska-Zwierz P, et al. 2011. Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res* **21:** 227–236.

Mao H, Zhu C, Zong D, Weng C, Yang X, Huang H, Liu D, Feng X, Guang S. 2015. The Nrde pathway mediates small-RNA-directed histone H3 lysine 27 trimethylation in *Caenorhabditis elegans*. *Curr Biol* **25:** 2398–2403.

Markham NR, Zuker M. 2008. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* **453:** 3–31.

Nishikura K. 2016. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **17:** 83–96.

Ohta H, Fujiwara M, Ohshima Y, Ishihara T. 2008. ADBP-1 regulates an ADAR RNA-editing enzyme to antagonize RNA-interference-mediated gene silencing in *Caenorhabditis elegans*. *Genetics* **180:** 785–796.

Ortiz MA, Noble D, Sorokin EP, Kimble J. 2014. A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode *Caenorhabditis elegans*. *G3 (Bethesda)* **4:** 1765–1772.

Park E, Maquat LE. 2013. Staufen-mediated mRNA decay. *Wiley Interdiscip Rev RNA* **4:** 423–435.

Porath HT, Carmi S, Levanon EY. 2014. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat Commun* **5:** 4726.

Reich DP, Bass BL. 2019. Mapping the dsRNA world. In *RNA worlds: new tools for deep exploration* (ed. Cech TR, et al.), pp. 135–150. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Reich DP, Tyc KM, Bass BL. 2018. *C. elegans* ADARs antagonize silencing of cellular dsRNAs by the antiviral RNAi pathway. *Genes Dev* **32:** 271–282.

Rosa S, Duncan S, Dean C. 2016. Mutually exclusive sense-antisense transcription at FLC facilitates environmentally induced gene repression. *Nat Commun* **7:** 13031.

Saldi TK, Ash PE, Wilson G, Gonzales P, Garrido-Lecca A, Roberts CM, Dostal V, Gendron TF, Stein LD, Blumenthal T, et al. 2014. TDP-1, the *Caenorhabditis elegans* ortholog of TDP-43, limits the accumulation of double-stranded RNA. *EMBO J* **33:** 2947–2966.

Spencer WC, McWhirter R, Miller T, Strasbourger P, Thompson O, Hillier LW, Waterston RH, Miller DM III. 2014. Isolation of specific neurons from *C. elegans* larvae for gene expression profiling. *PLoS One* **9:** e112102.

Tian B, Bevilacqua PC, Diegelman-Parente A, Mathews MB. 2004. The double-stranded-RNA-binding motif: interference and much more. *Nat Rev Mol Cell Biol* **5:** 1013–1023.

Towbin BD, Gonzalez-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, Askjaer P, Gasser SM. 2012. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150:** 934–947.

Venkatesh S, Workman JL. 2013. Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. *Wiley Interdiscip Rev Dev Biol* **2:** 685–700.

Vergara IA, Tarailo-Graovac M, Frech C, Wang J, Qin Z, Zhang T, She R, Chu JS, Wang K, Chen N. 2014. Genome-wide variations in a natural isolate of the nematode *Caenorhabditis elegans*. *BMC Genomics* **15:** 255.

Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G. 2004. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res* **14:** 1861–1869.

Whipple JM, Youssef OA, Aruscavage PJ, Nix DA, Hong C, Johnson WE, Bass BL. 2015. Genome-wide profiling of the *C. elegans* dsRNAome. *RNA* **21:** 786–800.

Wolkowicz UM, Cook AG. 2012. NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic Acids Res* **40:** 9356–9368.

Yoshida J, Akagi K, Misawa R, Kokubu C, Takeda J, Horie K. 2017. Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Sci Rep* **7:** 43613.

Zeller P, Padeken J, van Schendel R, Kalck V, Tijsterman M, Gasser SM. 2016. Histone H3K9 methylation is dispensable for *Caenorhabditis elegans* development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet* **48:** 1385–1395.

Zhao HQ, Zhang P, Gao H, He X, Dou Y, Huang AY, Liu XM, Ye AY, Dong MQ, Wei L. 2015. Profiling the RNA editomes of wild-type *C. elegans* and ADAR mutants. *Genome Res* **25:** 66–75.