FUNCTIONS OF DOUBLE-STRANDED RNA AND

ADAR RNA EDITING ENZYMES IN

CAENORHABDITIS ELEGANS

by

Daniel Phillip Reich

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

The University of Utah

May 2018

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Daniel Phillip Reich**

has been approved by the following supervisory committee members:

Brenda L. Bass	, Chair	**3/16/2018** Date Approved
Demiàn Cazalla	, Member	**3/16/2018** Date Approved
Jared Rutter	, Member	**3/16/2018** Date Approved
Mark Metzstein	, Member	**3/16/2018** Date Approved
Andres Villu Maricq	, Member	**3/16/2018** Date Approved

and by **Wesley Sundquist** , Chair/Dean of

the Department/College/School of **Biochemistry**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Adenosine deaminases that act on RNA (ADARs) catalyze adenosine-to-inosine (A-to-I) conversion within double-stranded RNA (dsRNA). Since inosine prefers to base-pair with cytidine, it is read by cellular machinery like the ribosome as guanosine. Thus, A-to-I RNA editing can alter the translation of edited codons in cellular mRNAs. However, genome-wide A-to-I editing studies have demonstrated that editing in coding regions is exceedingly rare in most organisms. Instead, ADAR editing is abundant in noncoding sequences associated with protein-coding genes, particularly in introns and untranslated regions (UTRs). By extension, dsRNA structures must also be prevalent in such noncoding regions. These observations raise questions as to the purpose of RNA editing and dsRNA structure in gene-associated sequences. In this dissertation, I explore the physiological functions of ADAR enzymes and their noncoding dsRNA substrates in the nematode *Caenorhadbitis elegans*.

First, I describe in Chapter 2 how ADARs prevent processing and silencing of cellular dsRNAs by the antiviral RNA interference (RNAi) machinery. Using RNAseq, I defined ADAR-edited dsRNAs, or editing-enriched regions (EERs), expressed during four stages of *C. elegans* development. I found that, in *adr-1;adr-2* mutants, EERs gave rise to abundant ~23 nucleotide (nt) small interfering RNAs (siRNAs), and were involved in silencing their associated genes by an RNAi-dependent mechanism. Additionally, disruption of the 26G endogenous siRNA (endo-siRNA) pathway in *adr-1;adr-2* mutant

backgrounds caused a synthetic phenotype that was rescued by deleting factors involved in antiviral RNAi. These results suggest that ADARs limit RNAi activity against cellular dsRNAs, presenting a striking functional parallel to mammalian ADAR1, which prevents aberrant innate immune signaling by the antiviral dsRNA sensor MDA5.

Though the work in Chapter 2 suggests that gene-associated dsRNAs can effect transcriptional silencing, in Chapter 3, I detail analyses suggesting that dsRNA-associated genes in fact exhibit higher-than-expected expression. I used three computational methods to define genome-wide loci in *C. elegans* encoding dsRNA structures, observing their enrichment on autosome distal arms. Despite that genes in distal arm regions are overall less highly expressed and less likely to be essential than genes in autosome centers, dsRNAs in these regions were enriched within essential and highly expressed genes. These analyses could not explicitly determine if gene expression patterns correlated with dsRNA formation or another property common to these loci. However, they suggest that dsRNA structures may function as important gene regulatory elements. In Chapter 4, I propose additional experiments to test contributions of dsRNA structure to gene expression regulation.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

CHAPTER 1

INTRODUCTION

In the cell, RNAs perform diverse functions dictated by their distinct forms and sequences. RNA processing and modification activities often alter the sequence of a transcript, in turn modifying its structure and functions. For instance, splicing and polyadenylation dramatically modify RNA sequence by removing and adding large blocks of sequence. Other types of RNA modification may change only a single nucleotide (nt). Adenosine deaminases that act on RNA (ADARs) catalyze the covalent conversion of adenosine (A) to inosine (I) in double-stranded RNAs (dsRNAs), a type of RNA modification known as RNA editing (Nishikura, 2016). Inosine has similar base-pairing properties to guanosine (G), so A-to-I editing can alter RNA structure, binding partners, and translation, depending on where editing occurs. A-to-I conversion is the most common form of RNA editing in metazoans, impacting thousands of sites in diverse organisms (Bazak et al., 2014; Blango and Bass, 2016; Porath et al., 2017; Zhao et al., 2015). Given its wide range of effects and locations, ADAR editing is highly pleiotropic, and deciphering the mechanistic and functional impact of editing has remained a major challenge in the field.

In this dissertation, I describe studies of the physiological functions of ADAR editing in the model invertebrate *Caenorhabditis elegans*. I define endogenous edited

dsRNAs expressed during *C. elegans* development and show how editing prevents their recognition by antiviral RNA interference (RNAi) machinery. Further, I explore how edited dsRNAs correlate with properties of associated genes in different domains of the *C. elegans* genome. These studies shed light on the ancestral functions of ADAR and may suggest novel roles for cellular dsRNAs.

The ADAR protein family

To date, ADARs have been found in all multicellular animals with sequenced and annotated genomes, possibly excepting the basal metazoan *Trichoplax adhaerens* (Grice and Degnan, 2015). ADARs have a conserved domain organization that includes one or more N-terminal dsRNA-binding motifs (dsRBMs) and a C-terminal deaminase domain. The ADAR deaminase domain shows homology to adenosine deaminases that act on tRNA (ADATs), eukaryotic enzymes that convert adenosine 37 to inosine in tRNA^{Ala} (Gerber et al., 1998; Macbeth et al., 2005), suggesting that ADAR editing originated from a tRNA modifying activity. ADARs likely arose when an ancestral protein orthologous to yeast ADAT1 acquired one or more dsRBMs, conferring specificity for dsRNA substrates.

Mammals encode three ADAR genes, *ADAR1, ADAR2,* and *ADAR3,* and two ADAR-like genes, *TENR* and *TENRL* (also called *ADAD1* and *ADAD2*) (Nishikura, 2016). Of these, only *ADAR1* and *ADAR2* encode enzymatically active proteins, as no activity has been reported for ADAR3, while TENR and TENRL lack residues essential for catalysis (Chen et al., 2000; Nishikura, 2016; Oakes et al., 2017). *ADAR1* produces two major protein isoforms (Patterson and Samuel, 1995). The 110-kiloDalton (kD)

isoform, ADAR1p110, is expressed constitutively and contains three dsRBMs and the deaminase domain. Expression of the longer 150-kD isoform, ADAR1p150, is driven by an alternative upstream, interferon (IFN)-inducible promoter. In addition to the domains found in ADAR1p110, ADAR1p150 includes two N-terminal Z-DNA-binding domains. ADAR1 Z-DNA binding domains are poorly understood, but they have been implicated in localization to cytoplasmic stress granules (Ng et al., 2013). Other groups hypothesize that Z-DNA binding domains are important to recruit ADAR1p150 to regions of active transcription (Herbert and Rich, 1999) or promote transcription by maintaining open chromatin (Oh et al., 2002). ADAR2, the other enzymatically active ADAR protein in mammals, carries two dsRBMs and a deaminase domain (Nishikura, 2016), and is expressed broadly, with highest expression in the brain (Melcher et al., 1996). Consistent with its elevated expression in neural tissue, ADAR2 is required in mammals to recode a glutamine residue to an arginine (Q/R) in the *GRIA2* (also called *GluR-B* or *GluA2*) pre-mRNA, which encodes a subunit of neuronal glutamate receptors (Higuchi et al., 2000).

Other vertebrates express orthologs of mammalian ADAR proteins (Li et al., 2014), but nonvertebrate metazoa express different sets of ADAR enzymes. For instance, *Drosophila* encode a single ADAR, most similar in primary sequence to mammalian ADAR2. ADAR1-like enzymes containing N-terminal Z-DNA-binding domains have been observed in the genomes of sea urchin, sea anemone, and certain sponges (Grice and Degnan, 2015). *C. elegans* have two ADAR genes, *adr-1* and *adr-2* (Tonkin et al., 2002). ADR-2 contains a single dsRBM and is the only active editing enzyme of the two; its mutation abolishes all editing in the organism. The ADR-1 protein lacks residues critical for catalysis, but it contains two dsRBMs and binds dsRNA substrates to regulate editing

(Washburn et al., 2014), possibly by interacting with ADR-2 (H. Hundley, personal communication). Mutants lacking *adr-1* have lower overall editing, though editing at some sites increases in *adr-1* mutants (Tonkin et al., 2002; Zhao et al., 2015).

ADAR activity and specificity

Originally identified as a "dsRNA unwinding" activity (Bass and Weintraub, 1987; Rebagliati and Melton, 1987), ADAR editing acts specifically on dsRNA substrates. However, rather than unwinding its substrates, ADARs covalently modify them, weakening the association of complementary sequences by converting A-U (uridine) base-pairs to I-U mismatches (Bass and Weintraub, 1988; Wagner et al., 1989). ADAR dsRBMs provide high affinity for dsRNA with dissociation constants in the low nM range (Kim et al., 1994; Lai et al., 1995; Ohman et al., 2000). However, very high levels of dsRNA inhibit ADAR1 activity (Hough and Bass, 1994), though the mechanism of substrate inhibition is not well understood. Several studies have suggested that ADAR1p150 and ADAR2 homodimerize in order to edit dsRNA (Cho et al., 2003; Poulsen et al., 2006), but this is still subject to debate, since earlier work from our lab observed recombinant ADAR2 only in a monomeric state (Macbeth and Bass, 2007). Though editing destabilizes RNA duplexes by converting A-U base-pairs to I-U mismatches, ADARs can edit up to 50% of adenosines in a perfectly base-paired RNA (Nishikura et al., 1991), indicating ADARs still bind and edit dsRNAs containing unpaired regions. Indeed, many natural ADAR substrates contain mismatches, loops, and bulges (Higuchi et al., 1993; Morse et al., 2002; Rieder et al., 2013), which often direct site-specific editing. As is the case with other dsRNA-binding proteins (dsRBPs) (Ramos

et al., 2000; Ryter and Schultz, 1998; Tian et al., 2004), ADARs bind A-form dsRNA primarily through interactions with the phosphodiester backbone (Barraud et al., 2012; Matthews et al., 2016; Stefl et al., 2006). RNA-DNA hybrids also form A-form helices, and a recent report suggests ADARs can deaminate deoxyadenosine in RNA-DNA duplexes (Zheng et al., 2017). Since ADAR-substrate interactions are structure-dependent, ADARs bind their substrates sequence-indiscriminately. As an exception to this rule, human ADAR2 dsRBMs also make sequence-specific contacts within the minor groove of a model substrate, the *GRIA2* R/G hairpin (Stefl et al., 2010). Sequence-specific interactions, which contribute to binding and editing the R/G hairpin, are thought to promote a specific binding register to facilitate efficient editing. However, a structure of the ADAR2 deaminase domain in complex with the same substrate suggested the deaminase domain would sterically clash with dsRBMs making sequence-specific interactions, raising questions as to the relevance of these interactions *in vivo* (Matthews et al., 2016). Still, ADAR2 residues that mediate sequence-specific binding are not conserved in ADAR1, providing one potential explanation for the different substrate specificities of ADAR1 and ADAR2 (Lehmann and Bass, 2000; Tan et al., 2017). Chimeric proteins containing the ADAR1 deaminase domain and PKR dsRBMs also exhibit altered substrate specificity from wildtype ADAR1, demonstrating that ADAR dsRBMs confer distinct binding preferences (Liu et al., 2000).

The ADAR deaminase domain catalyzes conversion of adenosine to inosine by hydrolytic deamination (Polson et al., 1991). Within the deaminase domain, the catalytic pocket, containing a critical zinc ion, coordinates a water molecule for nucleophilic attack at adenine C6 (Macbeth et al., 2005; Matthews et al., 2016). To access this ordered water

within the catalytic pocket, target adenosines must completely flip out of the RNA duplex, a mechanism demonstrated in biochemical and structural studies (Kuttan and Bass, 2012; Matthews et al., 2016; Stephens et al., 2000; Yi-Brunozzi et al., 2001). ADAR inserts a small loop into the dsRNA minor groove to flip out the target adenosine and occupy its space. In doing so, the flipping loop makes contacts with neighboring bases, but minorly clashes with 5'G or C. These interactions and clashes with adjacent nucleotides result in editing site nearest neighbor preferences, causing ADARs to prefer to edit adenosines with a 5' U (or A) and 3' G, but disfavor those with a 5' G and 3' U or C (Eggington et al., 2011; Lehmann and Bass, 2000; Polson and Bass, 1994).

The combination of structural elements and neighboring nucleotides dictates ADAR editing patterns. Properties of dsRNA structures, specifically their length and presence of unpaired sequences, promote editing in either selective or nonselective modes (Deffit and Hundley, 2016). ADARs edit long, perfectly base-paired dsRNAs nonselectively, deaminating many adenosines largely at random, though influenced by nearest neighbor preferences. In contrast, short (15-40 base-pair) RNA duplexes often show editing at one or several specific adenosines, but not others. Loops, mismatches, and bulges often direct such selective editing by promoting ADAR binding and editing at specific sites (Higuchi et al., 1993; Rieder et al., 2013). Further, unpaired adenosines opposite cytosines can be preferentially edited to produce a more stable I-C base pair (Kallman et al., 2003; Wong et al., 2001). In many cases, structural elements direct selective editing at mRNA recoding sites (Higuchi et al., 1993; Wahlstedt and Ohman, 2011). Regulation of editing within coding sequences is important to promote specific, functional amino acid substitutions, but prevent promiscuous recoding that might

generate nonfunctional proteins.

Endogenous ADAR substrates

The first *in vivo* ADAR-edited RNAs were discovered serendipitously by Sanger sequencing cDNAs (Burns et al., 1997; Cattaneo et al., 1988; Kimelman and Kirschner, 1989; Sommer et al., 1991). These edited RNAs were all protein-coding transcripts, including those encoding measles virus proteins, *Xenopus* basic Fibroblast Growth Factor, and mammalian neuronal receptor proteins. Since the ribosome interprets inosine as guanosine, editing in these protein-coding RNAs was predicted, and in some cases shown, to cause amino acid substitutions. Later, unbiased methods to identify inosine in cellular RNAs without amplifying specific cDNAs recognized that ADAR editing frequently occurs in noncoding sequences, particularly pre-mRNA introns and 3' untranslated regions (UTRs) (Morse et al., 2002; Morse and Bass, 1999). Following publication of the human genome sequence, multiple groups in parallel used libraries of cDNAs and expressed sequence tags to identify thousands of A-to-I edited sites in human RNAs, observing abundant editing in repetitive *Alu* elements (Athanasiadis et al., 2004; Blow et al., 2004; Kim et al., 2004; Levanon et al., 2004). These early efforts to define endogenous ADAR-edited RNAs provided the framework and tools to understand where editing occurs in the cell.

Modern high-throughput sequencing technologies provide powerful methods to identify edited RNAs and describe patterns of editing. Like the ribosome, reverse transcriptases interpret inosine as guanosine, so ADAR-edited transcripts appear with A-to-G mismatches in RNA sequencing (RNAseq) data (Diroma et al., 2017; Eisenberg,

2012). Approaches to identify A-to-G changes by RNAseq have been used to define editing sites in more than two dozen metazoan species (Hung et al., 2017; Liscovitch-Brauer et al., 2017; Porath et al., 2014; Porath et al., 2017), and more than 50 human tissues (Tan et al., 2017). Because ADARs nonselectively edit highly base-paired dsRNAs, our laboratory and others have used clusters of A-to-G changes, called editing-enriched regions (EERs), to identify long cellular dsRNAs, mapping what we call the "dsRNAome" (Blango and Bass, 2016; Whipple et al., 2015).

RNAseq-based editing studies have validated many of the patterns observed by early methods. For instance, it is well established that the vast majority of editing occurs in noncoding sequences, especially introns and UTRs, while editing in coding sites is rare (Blango and Bass, 2016; Porath et al., 2017; Whipple et al., 2015; Zhao et al., 2015). Coleoid cephalopods pose an interesting exception to this observation: though editing still occurs more frequently outside coding sequences in these organisms, squid, octopuses, and cuttlefish show thousands of editing-dependent mRNA recoding events (Liscovitch-Brauer et al., 2017), which may underlie adaptation to different temperatures and environments (Garrett and Rosenthal, 2012). As observed in early studies of human cDNAs and ESTs, editing most often occurs in sequences derived from repetitive elements (Blango and Bass, 2016; Porath et al., 2017; Zhao et al., 2015). Because repetitive elements can be copied thousands or millions of times in metazoan genomes, complementary sequences from these elements often pair to generate dsRNA, such as when two nearby repeats occur in an inverted orientation (Bazak et al., 2014; Porath et al., 2017). Most human and *C. elegans* edited sequences are predicted to form intramolecular dsRNA structures in this manner (Blango and Bass, 2016; Whipple et al.,

2015), though separate sense and antisense RNAs can form intermolecular dsRNAs that are also ADAR substrates (Fischer et al., 2013; Kimelman and Kirschner, 1989).

ADAR regulation

Differences in editing patterns between cell types and conditions indicate that A-to-I editing is subject to regulation. For instance, editing frequency varies substantially between human tissues, and tissue-specific editing levels inversely correlate with expression of *AIMP2*, which encodes a negative regulator of ADAR editing (Tan et al., 2017). Temperature-dependent changes in editing patterns are thought to contribute to temperature adaptation in *Drosophila* and octopus (Buchumenski et al., 2017; Garrett and Rosenthal, 2012; Rosenthal, 2015), suggesting these organisms regulate ADAR in response to environmental perturbation. Additionally, misregulated editing may contribute to human disease, such as cancer, where alterations in both nonselective editing patterns and mRNA recoding occur (Baysal et al., 2017; Gallo et al., 2017). All of these observations emphasize that editing must be tightly regulated in the cell.

ADAR is subject to regulation by both transcriptional and post-translational mechanisms. Transcripts of both *C. elegans* ADAR genes, *adr-1* and *adr-2*, display developmental regulation marked by elevated levels in embryo stages (Hundley et al., 2008). As mentioned above, alternative promoters dictate mammalian ADAR1 isoform-specific expression, including induction of the long p150 isoform upon IFN stimulation (Patterson and Samuel, 1995). While both ADAR1p110 and ADAR1p150 shuttle between the nucleus and cytoplasm, the p110 isoform predominates in the nucleus while the long isoform primarily localizes to the cytoplasm due to a nuclear export signal

included only in the p150 isoform (Fritz et al., 2009; Poulsen et al., 2001). Regulation of ADAR1 isoform expression further underscores that ADAR subcellular localization is tightly controlled. ADAR2's nuclear localization is mediated by a phosphorylation-dependent interaction with the prolyl isomerase Pin1 and cytoplasmic ubiquitination by the E3 ligase WWP2 (Marcucci et al., 2011). ADAR nuclear localization likely facilitates editing of pre-mRNA substrates, including structures formed by intron-exon base-pairing (Aruscavage and Bass, 2000; Higuchi et al., 1993; Hoopengardner et al., 2003). Indeed, ADAR-dependent A-to-G changes are also observed in cDNAs from nascent *Drosophila* transcripts (Rodriguez et al., 2012), suggesting many editing events occur co-transcriptionally. Thus, regulation of ADAR subcellular localization impacts the transcripts it can access to edit.

Other regulatory mechanisms alter editing patterns by influencing ADAR substrate selection, but do not act on ADAR proteins directly. For instance, several mRNA recoding sites, including the *GRIA2* Q/R and *Gabra-3* I/M sites in mammals, use separate, proximal dsRNA structures to promote high-efficiency editing in RNA hairpins containing the recoding sites (Daniel et al., 2012; Daniel et al., 2017). These proximal structures, called editing-inducer elements (EIEs), are thought to increase ADAR local concentration around functionally important edited structures to facilitate binding and editing (Daniel et al., 2014; Ramaswami et al., 2015). Other factors regulate ADAR activity by competing for substrates or altering dsRNA structures to preclude ADAR binding. The *C. elegans* protein TDP-1, whose mammalian ortholog TDP-43 shows RNA strand dissociation activity, restricts dsRNA accumulation and limits editing at many sites *in vivo*, presumably by preventing dsRNA formation through duplex unwinding or

stabilizing single-stranded RNA (Saldi et al., 2014). N^6-methyladenosine (m6A) modification in cellular RNAs affects ADAR1 binding and editing patterns in m6A-modified transcripts (Xiang et al., 2018), though the mechanism is as-yet-undetermined. Cellular dsRNAs may also influence editing by competing for ADAR. The *C. elegans* noncoding RNA *rncs-1* forms a long, edited dsRNA hairpin and is robustly induced during starvation (Hellwig and Bass, 2008). Though existing studies suggest that *rncs-1* regulates Dicer/DCR-1 activity (Hellwig and Bass, 2008; Rybak-Wolf et al., 2014), no studies to date have investigated *C. elegans* editing patterns under starvation, leaving open the possibility that *rncs-1* induction impacts ADAR activity.

Genetic analyses of ADAR function

Loss-of-function genetic studies in metazoan model organisms have provided substantial insight into physiological functions of ADARs. ADAR loss-of-function mutants have been described in *C. elegans* (Tonkin et al., 2002), *Drosophila melanogaster* (Palladino et al., 2000), and mice (Higuchi et al., 2000; Wang et al., 2000). Additionally, zebrafish *adar2* has been studied using morpholino disruption (Li et al., 2014), and genetic studies have linked ADAR1 mutations to two human diseases, Aicardi-Goutières syndrome and dyschromatosis symmetrica hereditaria, caused by homozygous and heterozygous *ADAR1* loss-of-function, respectively (Miyamura et al., 2003; Rice et al., 2012). Phenotypes associated with ADAR loss-of-function are highly variable and their fitness effects range from very mild to severe and lethal (Nishikura, 2016). Understanding these pleiotropic effects has remained a challenge for the field, but ongoing studies into ADAR substrates and the cellular pathways impacted by A-to-I

editing provide a fuller picture of the roles that ADAR plays in the cell.

Editing-dependent mRNA recoding (Fig. 1.1A) is a critical ADAR function in several organisms. Many early-identified ADAR substrates were mRNAs encoding neuronal receptor subunits whose editing led to amino acid substitutions at functionally important residues (Burns et al., 1997; Hanrahan et al., 2000; Semenov and Pak, 1999; Sommer et al., 1991). For instance, editing of mouse *GRIA2* causes Q/R recoding in nearly 100% of transcripts (Sommer et al., 1991). In mice and *Drosophila*, genetic ablation of ADARs induces neuronal phenotypes, including frequent seizures in *ADAR2*[-/-] mice (Higuchi et al., 2000) and tremors, uncoordinated movement, temperature-sensitive paralysis, and brain degeneration in *Drosophila ADAR* mutants (Palladino et al., 2000). Strikingly, the phenotypes of murine *ADAR2*[-/-] mutants are rescued by genetically encoding an arginine residue at the *GRIA2* Q/R recoding site (Higuchi et al., 2000), demonstrating that mRNA recoding is an essential function of mouse ADAR2. Though *Drosophila ADAR* mutant phenotypes have not been traced to specific recoding deficits, ADAR recoding occurs in dozens of neuronal factor mRNAs (Hoopengardner et al., 2003), and pan-neuronal *ADAR* knockdown induces locomoter defects (Jepson and Reenan, 2009). Unlike in other organisms (discussed below), mutating antiviral response genes, like *dicer2* and *ago2*, fails to rescue *ADAR* mutant phenotypes in *Drosophila* (Jepson and Reenan, 2009). Thus, mRNA recoding is though to be one of *Drosophila* ADAR's central functions.

Though mRNAs are important ADAR substrates, most A-to-I editing occurs in noncoding regions, and genetic evidence indicates that ADARs edit these sequences to prevent activation of dsRNA-mediated immune mechanisms (Fig. 1.1B) (O'Connell et

al., 2015; Walkley and Li, 2017). In mice, loss-of-function mutations in *ADAR1* cause embryonic lethality between embryonic day E11.5 and E14.5, depending on the nature of the mutation (Hartner et al., 2004; Liddicoat et al., 2015). *ADAR1*$^{-/-}$ mutants robustly induce interferon (IFN)-stimulated genes, which are normally expressed during innate immune responses to pathogenic infection (Liddicoat et al., 2015; Mannion et al., 2014; Pestal et al., 2015). Increased IFN signaling is also observed in human patients with Aicardi-Goutières syndrome and dyschromatosis symmetrica hereditaria (Rice et al., 2012), so ADAR1 immune suppression is likely conserved among mammals. Importantly, deleting the innate immune genes *Mavs* or *Ifih1* (MDA5) rescues mouse *ADAR1*$^{-/-}$ mutants to birth, demonstrating that overactive immune signaling causes *ADAR1*$^{-/-}$ mutant lethality (Liddicoat et al., 2015; Mannion et al., 2014; Pestal et al., 2015). The requirement for MDA5, a RIG-I-like helicase that senses viral dsRNA, suggests that *ADAR1*$^{-/-}$ mutant mice activate immune signaling due to aberrant dsRNA sensing. Wildtype suppression of immune signaling is attributed to the IFN-induced ADAR1 isoform, ADAR1p150, since a deletion that specifically disrupts this isoform, but not ADAR1p110, induces lethality and IFN-stimulated gene expression (Pestal et al., 2015). However, *ADAR1*$^{-/-}$; *MAVS*$^{-/-}$ mice lacking both ADAR1 isoforms display kidney patterning defects not observed in *ADAR1p150*$^{-/-}$;*MAVS*$^{-/-}$ mutant mice, suggesting that ADAR1p110 contributes to kidney development. Likewise, patterning of intestine, spleen, and lymph node tissue is compromised in *ADAR1*$^{-/-}$;*MAVS*$^{-/-}$ and *ADAR1p150*$^{-/-}$;*MAVS*$^{-/-}$ neonates, suggesting that ADAR1p150 promotes normal development of these organs through an undetermined MAVS-independent mechanism.

Similar to mammalian *ADAR1* mutants, *C. elegans* mutants lacking ADARs

display phenotypes dependent on the antiviral RNAi pathway (Knight and Bass, 2002; Sebastiani et al., 2009; Tonkin and Bass, 2003), a central innate immune mechanism in invertebrates. Deleting either *adr-1* or *adr-2* in *C. elegans* causes transgene silencing, reduced lifespan, and chemotaxis defects. All of these phenotypes require the RNAi factors *rde-1* or *rde-4*, suggesting that loss of *C. elegans* ADARs causes aberrant RNAi activity that leads to mutant phenotypes. In the *C. elegans* antiviral RNAi response, a complex containing the RIG-I-like helicase DRH-1, the dsRBP RDE-4, and the endoribonuclease DCR-1 senses and processes viral dsRNAs into small interfering RNAs (siRNAs) that are loaded into the Argonaute protein RDE-1 and silence viral transcripts (Ashe et al., 2013; Guo and Lu, 2013). As detailed in Chapter 2, *C. elegans* ADARs edit dsRNAs in introns and UTRs of protein-coding genes. In *adr-1;adr-2* mutants, gene-associated dsRNAs are processed into siRNAs and cause silencing of their associated genes, analogous to the silencing of viral dsRNAs. This work indicates that *adr-1;adr-2* mutant phenotypes relate to silencing of specific genes containing dsRNA structures. Indeed, a recent study found that neuronal silencing of *clec-41*, which has an edited 3'UTR, causes the chemotaxis defects of *adr-1;adr-2* mutants, since neuronal *clec-41* re-expression rescues these defects (Deffit et al., 2017). Thus, analogous to mammalian ADAR1 antagonism of MDA5-dependent IFN signaling, *C. elegans* ADARs prevent cellular dsRNAs from being processed and silenced by antiviral RNAi.

Since dsRBPs bind their substrates without sequence specificity (Tian et al., 2004), antiviral dsRNA sensors like mammalian MDA5 and invertebrate Dicer cannot distinguish cellular (self) dsRNAs from viral (nonself) dsRNAs by their sequence. Mounting evidence indicates that ADARs provide a critical distinction between self and

nonself dsRNAs (Liddicoat et al., 2015; Mannion et al., 2014; Pestal et al., 2015; Reich et al., 2018). Since ADARs primarily act in the nucleus on nascent transcripts (Nishikura, 2016; Rodriguez et al., 2012), A-to-I editing should largely occur in cellular dsRNAs, but be absent from cytoplasmic viral dsRNAs, at least during initial stages of infection. Editing converts A-U base-pairs to I-U mismatches (Nishikura, 2016; O'Connell et al., 2015), so ADARs may reduce the double-stranded character of cellular dsRNAs to distinguish them from viral dsRNAs. Both MDA5 and Dicer have reduced activity towards I-U-containing dsRNAs, indicating these proteins do not mount robust responses to edited substrates (Scadden and Smith, 2001; Vitali and Scadden, 2010). Our work in *C. elegans,* together with studies of mammalian *ADAR1,* strongly indicates that a fundamental role of ADARs is to mark cellular dsRNAs as self to facilitate discrimination between self and nonself transcripts.

ADARs regulate other cellular RNA-mediated processes

By editing cellular transcripts, ADARs modulate RNA sequence- and structure-dependent interactions, thereby altering RNA properties and functions in diverse processes. This is clearly demonstrated in ADAR's essential roles in mRNA recoding and suppression of innate immune signaling. However, since essentially all RNA-mediated processes involve sequence- and/or structure-dependent interactions, it is no surprise that ADARs can also affect cellular activities such as pre-mRNA splicing, mRNA stability, and miRNA-mediated regulation (Nishikura, 2016; Walkley and Li, 2017).

Several lines of evidence suggest that A-to-I editing regulates alternative splicing of pre-mRNAs (Fig. 1.1C). Most directly, since splicing relies on sequence-dependent

interactions with spliceosomal snRNAs to identify splicing donor, acceptor, and branch point sites (Lee and Rio, 2015), editing-dependent sequence alterations can change existing splice sites or generate new ones. Indeed, editing of an intronic AA dinucleotide in the rat *ADAR2* pre-mRNA generates a new AI 3' splice acceptor site that facilitates inclusion of a 47-nt cassette (Rueter et al., 1999). Editing can likewise create or destroy splicing regulatory sequences to change where splicing factors bind, thereby altering splicing (Lev-Maor et al., 2007; Solomon et al., 2013). RNAseq of human *ADAR1* knockdown cells showed hundreds of genes with altered splicing patterns, which the authors attributed primarily to editing within splicing regulatory elements (Solomon et al., 2013). RNA structural elements likewise contribute to splicing regulation (Warf and Berglund, 2010), suggesting that ADARs could in theory impact splicing through effects on RNA secondary structure. In addition to splicing regulation, ADARs may promote alternative 3'UTR usage, as >2000 genes have altered 3'UTR lengths in human *ADAR1* knockdown cells (Bahn et al., 2015), and alternative polyadenylation sites are enriched among mouse and human EER-associated genes (Blango and Bass, 2016). While ADAR regulation of splicing and polyadenylation is not known to be essential *in vivo*, studies demonstrating that ADARs modulate pre-mRNA processing suggest that this could be a physiologically important function.

ADARs further impact mRNA stability and translation. For instance, editing in the 3'UTR of human *CTSS* recruits the HuR RNA-binding protein to promote mRNA stability (Stellos et al., 2016). How editing recruits HuR is less clear, though the authors propose that editing destabilizes dsRNA formed by *Alu* inverted repeats to uncover sequences containing HuR binding sites. Previous work from our own lab shows that *C.*

elegans mRNAs containing 3'UTR dsRNA structures are loaded with fewer ribosomes, suggesting that such structures decrease translation efficiency, though this effect was not impacted by ADARs (Hundley et al., 2008). ADAR-dependent gene expression differences are often attributed to effects on miRNA-mediated regulation (Fig. 1.1D) (Qi et al., 2017; Warf et al., 2012) through several proposed mechanisms (Nishikura, 2016). By editing 3'UTRs, ADARs can create or disrupt miRNA binding sites (Deffit and Hundley, 2016). Alternatively, editing within primary miRNA transcripts can inhibit processing by Drosha enzymes (Yang et al., 2006) or, if the miRNA precursor is still processed, alter the transcripts the mature miRNA targets (Kawahara et al., 2007). One study even suggested that mammalian ADAR1 promotes miRNA biogenesis by forming a complex with Dicer that more efficiently processes miRNA precursors (Ota et al., 2013). Thus, ADARs have the potential to impact gene expression through effects on RNA processing, stability, and translation, though it is not clear that such mechanisms are important to metazoan biology. At any rate, these activities further complicate the interpretation of ADAR mutant phenotypes, since they suggest that important ADAR functions could involve both direct and indirect mechanisms.

<div align="center">A central challenge to the field</div>

As described in this chapter, ADARs affect a wide range of substrates and activities through diverse mechanisms of action (Fig. 1.1). As a result, ADAR loss-of-function phenotypes can be highly pleotropic and difficult to interpret. Molecular and genetic studies have begun to reveal the processes, and in some cases the specific RNA substrates, that underlie ADAR mutant phenotypes, like mammalian *ADAR1* and *ADAR2*

mutant lethality and *C. elegans adr-1;adr-2* mutant chemotaxis deficiencies (Deffit et al., 2017; Higuchi et al., 2000; Liddicoat et al., 2015; Mannion et al., 2014; Pestal et al., 2015). However, researchers continue to describe new ways in which ADARs regulate additional RNA-dependent processes. It remains to be determined if ADAR activities observed in cultured cells have physiological relevance in whole organisms. Still, many ADAR mutant phenotypes, like murine *ADAR1$^{-/-}$* organ patterning defects, lack mechanistic explanation, presenting new avenues of future research.

Understanding how ADAR loss-of-function phenotypes and pathologies relate to discrete substrates and editing events represents a fundamental challenge, but it is a critical problem to address. By dissecting functionally important ADAR-substrate relationships, we can develop methods to manipulate particular aspects of ADAR biology without perturbing other core functions. This might facilitate, for instance, clinical treatments that address aberrant mRNA recoding in cancer without interrupting ADAR's role in distinguishing self and nonself dsRNAs. By studying the effects of A-to-I editing, we also learn more about the mechanisms that regulate cellular trancripts. In turn, understanding ADARs teaches us about life in the modern cell and in cells of the past.

In this work, we describe our efforts to determine ADAR substrates and functions in the nematode *C. elegans*. In Chapter 2, we define edited dsRNAs expressed during *C. elegans* development and demonstrate that ADARs are critical to prevent the recognition and processing of these dsRNAs by the antiviral RNAi machinery. In Chapter 3, we describe observations that dsRNA structures are associated with essential and highly expressed genes on distal autosome arms, suggesting potentially novel ADAR functions. Lastly, in Chapter 4, we discuss the implications of our findings and describe prospective

areas of future research.

References

Aruscavage, P.J., and Bass, B.L. (2000). A phylogenetic analysis reveals an unusual sequence conservation within introns involved in RNA editing. RNA *6*, 257-269.

Ashe, A., Belicard, T., Le Pen, J., Sarkies, P., Frezal, L., Lehrbach, N.J., Felix, M.A., and Miska, E.A. (2013). A deletion polymorphism in the Caenorhabditis elegans RIG-I homolog disables viral RNA dicing and antiviral immunity. Elife *2*, e00994.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. *2*, e391.

Bahn, J.H., Ahn, J., Lin, X., Zhang, Q., Lee, J.H., Civelek, M., and Xiao, X. (2015). Genomic analysis of ADAR1 binding and its involvement in multiple RNA processing pathways. Nat. Commun. *6*, 6355.

Barraud, P., Heale, B.S., O'Connell, M.A., and Allain, F.H. (2012). Solution structure of the N-terminal dsRBD of Drosophila ADAR and interaction studies with RNA. Biochimie *94*, 1499-1509.

Bass, B.L., and Weintraub, H. (1987). A developmentally regulated activity that unwinds RNA duplexes. Cell *48*, 607-613.

Bass, B.L., and Weintraub, H. (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell *55*, 1089-1098.

Baysal, B.E., Sharma, S., Hashemikhabir, S., and Janga, S.C. (2017). RNA editing in pathogenesis of cancer. Cancer Res. *77*, 3733-3739.

Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., Isaacs, F.J., Rechavi, G., Li, J.B., Eisenberg, E.*, et al.* (2014). A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res. *24*, 365-376.

Blango, M.G., and Bass, B.L. (2016). Identification of the long, edited dsRNAome of LPS-stimulated immune cells. Genome Res. *26*, 852-862.

Blow, M., Futreal, P.A., Wooster, R., and Stratton, M.R. (2004). A survey of RNA editing in human brain. Genome Res. *14*, 2379-2387.

Buchumenski, I., Bartok, O., Ashwal-Fluss, R., Pandey, V., Porath, H.T., Levanon, E.Y., and Kadener, S. (2017). Dynamic hyper-editing underlies temperature adaptation in Drosophila. PLoS Genet. *13*, e1006931.

Burns, C.M., Chu, H., Rueter, S.M., Hutchinson, L.K., Canton, H., Sanders-Bush, E., and Emeson, R.B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature *387*, 303-308.

Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., Meulen, V.t., and Billeter, M.A. (1988). Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. Cell *55*, 255-265.

Chen, C.-X., Cho, D.-S.C., Wang, Q., Lai, F., Carter, K.C., and Nishikura, K. (2000). A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. RNA *6*, 755-767.

Cho, D.S., Yang, W., Lee, J.T., Shiekhattar, R., Murray, J.M., and Nishikura, K. (2003). Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. J. Biol. Chem. *278*, 17093-17102.

Daniel, C., Silberberg, G., Behm, M., and Öhman, M. (2014). Alu elements shape the primate transcriptome by cis-regulation of RNA editing. Genome Biol. *15*, R28.

Daniel, C., Veno, M.T., Ekdahl, Y., Kjems, J., and Ohman, M. (2012). A distant cis acting intronic element induces site-selective RNA editing. Nucleic Acids Res. *40*, 9876-9886.

Daniel, C., Widmark, A., Rigardt, D., and Ohman, M. (2017). Editing inducer elements increases A-to-I editing efficiency in the mammalian transcriptome. Genome Biol. *18*, 195.

Deffit, S.N., and Hundley, H.A. (2016). To edit or not to edit: regulation of ADAR editing specificity and efficiency. Wiley Interdiscip. Rev. RNA *7*, 113-127.

Deffit, S.N., Yee, B.A., Manning, A.C., Rajendren, S., Vadlamani, P., Wheeler, E.C., Domissy, A., Washburn, M.C., Yeo, G.W., and Hundley, H.A. (2017). The C. elegans neural editome reveals an ADAR target mRNA required for proper chemotaxis. Elife *6,* e28625.

Diroma, M.A., Ciaccia, L., Pesole, G., and Picardi, E. (2017). Elucidating the editome: bioinformatics approaches for RNA editing detection. Brief. Bioinform. https://doi.org/10.1093/bib/bbx129.

Eggington, J.M., Greene, T., and Bass, B.L. (2011). Predicting sites of ADAR editing in double-stranded RNA. Nat. Commun. *2*, 319.

Eisenberg, E. (2012). Bioinformatic approaches for identification of A-to-I editing sites. Curr. Top. Microbiol. Immunol. *353*, 145-162.

Fischer, S.E., Pan, Q., Breen, P.C., Qi, Y., Shi, Z., Zhang, C., and Ruvkun, G. (2013). Multiple small RNA pathways regulate the silencing of repeated and foreign genes in C. elegans. Genes Dev. *27*, 2678-2695.

Fritz, J., Strehblow, A., Taschner, A., Schopoff, S., Pasierbek, P., and Jantsch, M.F. (2009). RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1. Mol. Cell. Biol. *29*, 1487-1497.

Gallo, A., Vukic, D., Michalik, D., O'Connell, M.A., and Keegan, L.P. (2017). ADAR RNA editing in human disease; more to it than meets the I. Hum. Genet. *136*, 1265-1278.

Garrett, S., and Rosenthal, J.J.C. (2012). RNA editing underlies temperature adaptation in K+ channels from polar octopuses. Science *335*, 848-851.

Gerber, A., Grosjean, H., Melcher, T., and Keller, W. (1998). Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. EMBO J. *17*, 4780-4789.

Grice, L.F., and Degnan, B.M. (2015). The origin of the ADAR gene family and animal RNA editing. BMC Evol. Biol. *15*, 4.

Guo, X., and Lu, R. (2013). Characterization of virus-encoded RNA interference suppressors in Caenorhabditis elegans. J. Virol. *87*, 5414-5423.

Hanrahan, C.J., Palladino, M.J., Ganetzky, B., and Reenan, R.A. (2000). RNA editing of the Drosophila para Na1 channel transcript: evolutionary conservation and developmental regulation. Genetics *155*, 1149-1160.

Hartner, J.C., Schmittwolf, C., Kispert, A., Muller, A.M., Higuchi, M., and Seeburg, P.H. (2004). Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. J. Biol. Chem. *279*, 4894-4902.

Hellwig, S., and Bass, B.L. (2008). A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes. Proc. Natl. Acad. Sci. USA. *105*, 12897-12902.

Herbert, A., and Rich, A. (1999). Left-handed Z-DNA: structure and function. Genetica *106*, 37-47.

Higuchi, M., Maas, S., Single, F.N., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R., and Seeburg, P.H. (2000). Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature *406*, 78-81.

Higuchi, M., Single, F.N., Kohler, M., Sommer, B., Sprengel, R., and Seeburg, P.H. (1993). RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-

exon structure determines position and efficiency. Cell *75*, 1361-1370.

Hoopengardner, B., Bhalla, T., Staber, C., and Reenan, R. (2003). Nervous system targets of RNA editing identified by comparative genomics. Science *301*, 832-836.

Hough, R.F., and Bass, B.L. (1994). Purification of the *Xenopus laevis* double-stranded RNA adenosine deaminase. J. Biol. Chem. *269*, 9933-9939.

Hundley, H.A., Krauchuk, A.A., and Bass, B.L. (2008). C. elegans and H. sapiens mRNAs with edited 3' UTRs are present on polysomes. RNA *14*, 2050-2060.

Hung, L.Y., Chen, Y.J., Mai, T.L., Chen, C.Y., Yang, M.Y., Chiang, T.W., Wang, Y.D., and Chuang, T.J. (2017). An evolutionary landscape of A-to-I RNA editome across metazoan species. Genome Biol. Evol. *10*, 521-537.

Jepson, J.E., and Reenan, R.A. (2009). Adenosine-to-inosine genetic recoding is required in the adult stage nervous system for coordinated behavior in Drosophila. J. Biol. Chem. *284*, 31391-31400.

Kallman, A.M., Sahlin M., and Ohman M. (2003). ADAR2 A->I editing: site selectivity and editing efficiency are separate events. Nucleic Acids Res. *31*, 4874-4881.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., and Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science *315*, 1137-1140.

Kim, D.D.Y., Kim, T.T.Y., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., and Gabriel, A. (2004). Widespread RNA editing of embedded Alu elements in the human transcriptome. Genome Res. *14*, 1719-1725.

Kim, U., Garner, T.L., Sanford, T., Speicher, D., Murray, J.M., and Nishikura, K. (1994). Purification and characterization of double-stranded RNA adenosine deaminase from bovine nuclear extract. J. Biol. Chem. *269*, 13480-13489.

Kimelman, D., and Kirschner, M.W. (1989). An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes. Cell *59*, 687-696.

Knight, S.W., and Bass, B.L. (2002). The role of RNA editing by ADARs in RNAi. Mol. Cell *10*, 809-817.

Kuttan, A., and Bass, B.L. (2012). Mechanistic insights into editing-site specificity of ADARs. Proc. Natl. Acad. Sci. USA *109*, E3295-E3304.

Lai, F., Drakas, R., and Nishikura, K. (1995). Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts. J. Biol. Chem. *270*, 17096-17105.

Lee, Y., and Rio, D.C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. Annu. Rev. Biochem. *84*, 291-323.

Lehmann, K.A., and Bass, B.L. (2000). Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry *39*, 12875-12884.

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

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D.*, et al.* (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. *22*, 1001-1005.

Li, I.C., Chen, Y.C., Wang, Y.Y., Tzeng, B.W., Ou, C.W., Lau, Y.Y., Wu, K.M., Chan, T.M., Lin, W.H., Hwang, S.P.*, et al.* (2014). Zebrafish Adar2 Edits the Q/R site of AMPA receptor Subunit gria2alpha transcript to ensure normal development of nervous system and cranial neural crest cells. PLoS One *9*, e97133.

Liddicoat, B.J., Piskol, R., Chalk, A.M., Ramaswami, G., Higuchi, M., Hartner, J.C., Li, J.B., Seeburg, P.H., and Walkley, C.R. (2015). RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. Science *349*, 1115-1120.

Liscovitch-Brauer, N., Alon, S., Porath, H.T., Elstein, B., Unger, R., Ziv, T., Admon, A., Levanon, E.Y., Rosenthal, J.J., and Eisenberg, E. (2017). Trade-off between transcriptome plasticity and genome evolution in cephalopods. Cell *169*, 191-202 e111.

Liu, Y., Lei, M., and Samuel, C.E. (2000). Chimeric double-stranded RNA-specific adenosine deaminase ADAR1 proteins reveal functional selectivity of double-stranded RNA-binding domains from ADAR1 and protein kinase PKR. Proc. Natl. Acad. Sci. USA *97*, 12541–12546.

Macbeth, M.R., and Bass, B.L. (2007). Large-scale overexpression and purification of ADARs from Saccharomyces cerevisiae for biophysical and biochemical studies. Methods Enzymol. *424*, 319-331.

Macbeth, M.R., Schubert, H.L., VanDemark, A.P., Lingam, A.T., Hill, C.P., and Bass, B.L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. Science *309*.

Mannion, N.M., Greenwood, S.M., Young, R., Cox, S., Brindle, J., Read, D., Nellaker, C., Vesely, C., Ponting, C.P., McLaughlin, P.J.*, et al.* (2014). The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. Cell Rep. *9*, 1482-1494.

Marcucci, R., Brindle, J., Paro, S., Casadio, A., Hempel, S., Morrice, N., Bisso, A.,

Keegan, L.P., Del Sal, G., and O'Connell, M.A. (2011). Pin1 and WWP2 regulate GluR2 Q/R site RNA editing by ADAR2 with opposing effects. EMBO J. *30*, 4211-4222.

Matthews, M.M., Thomas, J.M., Zheng, Y., Tran, K., Phelps, K.J., Scott, A.I., Havel, J., Fisher, A.J., and Beal, P.A. (2016). Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. Nat. Struct. Mol. Biol. *23*, 426-433.

Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P.H., and Higuchi, M. (1996). A mammalian RNA editing enzyme. Nature *379*, 460-464.

Miyamura, Y., Suzuki, T., Kono, M., Inagaki, K., Ito, S., Suzuki, N., and Tomita, Y. (2003). Mutations of the RNA-specific adenosine deaminase gene (DSRAD) are involved in dyschromatosis symmetrica hereditaria. Am. J. Hum. Genet. *73*, 693–699.

Morse, D.P., Aruscavage, P.J., and Bass, B.L. (2002). RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. Proc. Natl. Acad. Sci. USA *99*, 7906-7911.

Morse, D.P., and Bass, B.L. (1999). Long RNA hairpins that contain inosine are present in Caenorhabditis elegans poly(A)+ RNA. Proc. Natl. Acad. Sci. USA *96*, 6048-6053.

Ng, S.K., Weissbach, R., Ronson, G.E., and Scadden, A.D. (2013). Proteins that contain a functional Z-DNA-binding domain localize to cytoplasmic stress granules. Nucleic Acids Res. *41*, 9786-9799.

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

Nishikura, K., Yoo, C., Kim, U., Murray, J.M., Estes, P.A., Cash, F.E., and Liebhaber, S.A. (1991). Substrate specificity of the dsRNA unwinding/modifying activity. EMBO J. *10*, 3523-3532.

O'Connell, M.A., Mannion, N.M., and Keegan, L.P. (2015). The epitranscriptome and innate immunity. PLoS Genet. *11*, e1005687.

Oakes, E., Anderson, A., Cohen-Gadol, A., and Hundley, H.A. (2017). Adenosine deaminase that acts on RNA 3 (ADAR3) binding to glutamate receptor subunit B pre-mRNA inhibits RNA editing in glioblastoma. J. Biol. Chem. *292*, 4326-4335.

Oh, D.B., Kim, Y.G., and Rich, A. (2002). Z-DNA-binding proteins can act as potent effectors of gene expression in vivo. Proc. Natl. Acad. Sci. USA *99*, 16666-16671.

Ohman, M., Kallman, A.M., and Bass, B.L. (2000). In vitro analysis of the binding of ADAR2 to the pre-mRNA encoding the GluR-B R/G site. RNA *6*, 687-697.

Ota, H., Sakurai, M., Gupta, R., Valente, L., Wulff, B.E., Ariyoshi, K., Iizasa, H., Davuluri, R.V., and Nishikura, K. (2013). ADAR1 forms a complex with Dicer to promote microRNA processing and RNA-induced gene silencing. Cell *153*, 575-589.

Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A. (2000). A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity. Cell *102*, 437-449.

Patterson, J.B., and Samuel, C.E. (1995). Expression and regulation by interferon of a double-stranded RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. Mol. Cell. Biol. *15*, 5376-5388.

Pestal, K., Funk, C.C., Snyder, J.M., Price, N.D., Treuting, P.M., and Stetson, D.B. (2015). Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. Immunity *43*, 933-944.

Polson, A.G., and Bass, B.L. (1994). Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. EMBO J *13*, 5701-5711.

Polson, A.G., Crain, P.F., Pomerantz, S.C., McCloskey, J.A., and Bass, B.L. (1991). The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. Biochemistry *30*, 11507-11514.

Porath, H.T., Carmi, S., and Levanon, E.Y. (2014). A genome-wide map of hyper-edited RNA reveals numerous new sites. Nat. Commun. *5*, 4726.

Porath, H.T., Knisbacher, B.A., Eisenberg, E., and Levanon, E.Y. (2017). Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. Genome Biol. *18*, 185.

Poulsen, H., Jorgensen, R., Heding, A., Nielsen, F.C., Bonven, B., and Egebjerg, J. (2006). Dimerization of ADAR2 is mediated by the double-stranded RNA binding domain. RNA *12*, 1350-1360.

Poulsen, H., Nilsson, J., Damgaard, C.K., Egebjerg, J., and Kjems, J. (2001). CRM1 mediates the export of ADAR1 through a nuclear export signal within the Z-DNA binding domain. Mol. Cell. Biol. *21*, 7862-7871.

Qi, L., Song, Y., Chan, T.H.M., Yang, H., Lin, C.H., Tay, D.J.T., Hong, H., Tang, S.J., Tan, K.T., Huang, X.X.*, et al.* (2017). An RNA editing/dsRNA binding-independent gene regulatory mechanism of ADARs and its clinical implication in cancer. Nucleic Acids Res. *45*, 10436-10451.

Ramaswami, G., Deng, P., Zhang, R., Anna Carbone, M., Mackay, T.F., and Li, J.B. (2015). Genetic mapping uncovers cis-regulatory landscape of RNA editing. Nat. Commun. *6*, 8194.

Ramos, A., Grunert, S., Adams, J., R.Micklem, D., R.Proctor, M., Freund, S., Bycroft, M., Johnston, D.S., and Varani, G. (2000). RNA recognition by a Staufen double-stranded RNA-binding domain. EMBO J. *19*, 997-1009.

Rebagliati, M., and Melton, D. (1987). Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. Cell *48*, 599-605.

Reich, D.P, Tyc, K.M., and Bass, B.L. (2018). C. elegans ADARs antagonize silencing of cellular dsRNAs by the antiviral RNAi pathway. Genes Dev. *32,* 271-282.

Rice, G.I., Kasher, P.R., Forte, G.M., Mannion, N.M., Greenwood, S.M., Szynkiewicz, M., Dickerson, J.E., Bhaskar, S.S., Zampini, M., Briggs, T.A.*, et al.* (2012). Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. Nat. Genet. *44*, 1243-1248.

Rieder, L.E., Staber, C.J., Hoopengardner, B., and Reenan, R.A. (2013). Tertiary structural elements determine the extent and specificity of messenger RNA editing. Nat. Commun. *4*, 2232.

Rodriguez, J., Menet, J.S., and Rosbash, M. (2012). Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. Mol. Cell *47*, 27-37.

Rosenthal, J.J. (2015). The emerging role of RNA editing in plasticity. J. Exp. Biol. *218*, 1812-1821.

Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. Nature *399*, 75-80.

Rybak-Wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M., and Rajewsky, N. (2014). A variety of dicer substrates in human and C. elegans. Cell *159*, 1153-1167.

Ryter, J.M., and Schultz, S.C. (1998). Molecular basis of double-stranded RNA–protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. EMBO J. *17*, 7505-7513.

Saldi, T.K., Ash, P.E., Wilson, G., Gonzales, P., Garrido-Lecca, A., Roberts, C.M., Dostal, V., Gendron, T.F., Stein, L.D., 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.

Scadden, A.D.J., and Smith, C.W.J. (2001). RNAi is antagonized by A→I hyper-editing. EMBO Rep. *2*, 1107-1111.

Sebastiani, P., Montano, M., Puca, A., Solovieff, N., Kojima, T., Wang, M.C., Melista, E., Meltzer, M., Fischer, S.E., Andersen, S*., et al.* (2009). RNA editing genes associated with extreme old age in humans and with lifespan in C. elegans. PLoS One *4*, e8210.

Semenov, E.P., and Pak, W.L. (1999). Diversification of Drosophila chloride channel gene by multiple posttranscriptional mRNA modifications. J. Neurochem. *72*, 66-72.

Solomon, O., Oren, S., Safran, M., Deshet-Unger, N., Akiva, P., Jacob-Hirsch, J., Cesarkas, K., Kabesa, R., Amariglio, N., Unger, R*., et al.* (2013). Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR). RNA *19*, 591-604.

Sommer, B., Kohler, M., Sprengel, R., and Seeburg, P.H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. Cell *67*, 11-19.

Stefl, R., Oberstrass, F.C., Hood, J.L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R.B*., et al.* (2010). The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. Cell *143*, 225-237.

Stefl, R., Xu, M., Skrisovska, L., Emeson, R.B., and Allain, F.H. (2006). Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. Structure *14*, 345-355.

Stellos, K., Gatsiou, A., Stamatelopoulos, K., Perisic Matic, L., John, D., Lunella, F.F., Jae, N., Rossbach, O., Amrhein, C., Sigala, F*., et al.* (2016). Adenosine-to-inosine RNA editing controls cathepsin S expression in atherosclerosis by enabling HuR-mediated post-transcriptional regulation. Nat. Med. *22*, 1140-1150.

Stephens, O.M., Yi-Brunozzi, H.Y., and Beal, P.A. (2000). Analysis of the RNA-editing reaction of ADAR2 with structural and fluorescent analogues of the GluR-B R/G editing site. Biochemistry *39*, 12243-12251.

Tan, M.H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A.N., Liu, K.I., Zhang, R., Ramaswami, G., Ariyoshi, K*., et al.* (2017). Dynamic landscape and regulation of RNA editing in mammals. Nature *550*, 249-254.

Tian, B., Bevilacqua, P.C., Diegelman-Parente, A., and Mathews, M.B. (2004). The double-stranded-RNA-binding motif: interference and much more. Nat. Rev. Mol. Cell Biol. *5*, 1013-1023.

Tonkin, L., and Bass, B.L. (2003). Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. Science *302*, 1725.

Tonkin, L.A., Saccomanno, L., Morse, D.P., Brodigan, T., Krause, M., and L.Bass, B. (2002). RNA editing by ADARs is important for normal behavior in

Caenorhabditis elegans. EMBO J. *21*, 6025-6035.

Vitali, P., and Scadden, A.D. (2010). Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis. Nat. Struct. Mol. Biol. *17*, 1043-1050.

Wagner, R.W., Smith, J.E., Cooperman, B.S., and Nishikura, K. (1989). A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. Proc. Natl. Acad. Sci. USA *86*, 2647-2651.

Wahlstedt, H., and Ohman, M. (2011). Site-selective versus promiscuous A-to-I editing. Wiley Interdiscip. Rev. RNA *2*, 761-771.

Walkley, C.R., and Li, J.B. (2017). Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. Genome Biol. *18*, 205.

Wang, Q., Khillan, J., Gadue, P., and Nishikura, K. (2000). Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. Science *290*, 1765-1768.

Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem. Sci. *35*, 169-178.

Warf, M.B., Shepherd, B.A., Johnson, W.E., and Bass, B.L. (2012). Effects of ADARs on small RNA processing pathways in C. elegans. Genome Res. *22*, 1488-1498.

Washburn, M.C., Kakaradov, B., Sundararaman, B., Wheeler, E., Hoon, S., Yeo, G.W., and Hundley, H.A. (2014). The dsRBP and inactive editor ADR-1 utilizes dsRNA binding to regulate A-to-I RNA editing across the C. elegans transcriptome. Cell Rep. *6*, 599-607.

Whipple, J.M., Youssef, O.A., Aruscavage, P.J., Nix, D.A., Hong, C., Johnson, W.E., and Bass, B.L. (2015). Genome-wide profiling of the C. elegans dsRNAome. RNA *21*, 786-800.

Wong, S.K., Sato, S., and Lazinski, D.W. (2001). Substrate recognition by ADAR1 and ADAR2. RNA *7*, 846-858.

Xiang, J.F., Yang, Q., Liu, C.X., Wu, M., Chen, L.L., and Yang, L. (2018). N(6)-methyladenosines modulate A-to-I RNA editing. Mol. Cell *69*, 126-135 e126.

Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., and Nishikura, K. (2006). Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat. Struct. Mol. Biol. *13*, 13-21.

Yi-Brunozzi, H.Y., Stephens, O.M., and Beal, P.A. (2001). Conformational changes that occur during an RNA-editing adenosine deamination reaction. J. Biol. Chem. *276*,

37827-37833.

Zhao, H.Q., Zhang, P., Gao, H., He, X., Dou, Y., Huang, A.Y., Liu, X.M., Ye, A.Y., Dong, M.Q., and Wei, L. (2015). Profiling the RNA editomes of wild-type C. elegans and ADAR mutants. Genome Res. *25*, 66-75.

Zheng, Y., Lorenzo, C., and Beal, P.A. (2017). DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. Nucleic Acids Res. *45*, 3369-3377.
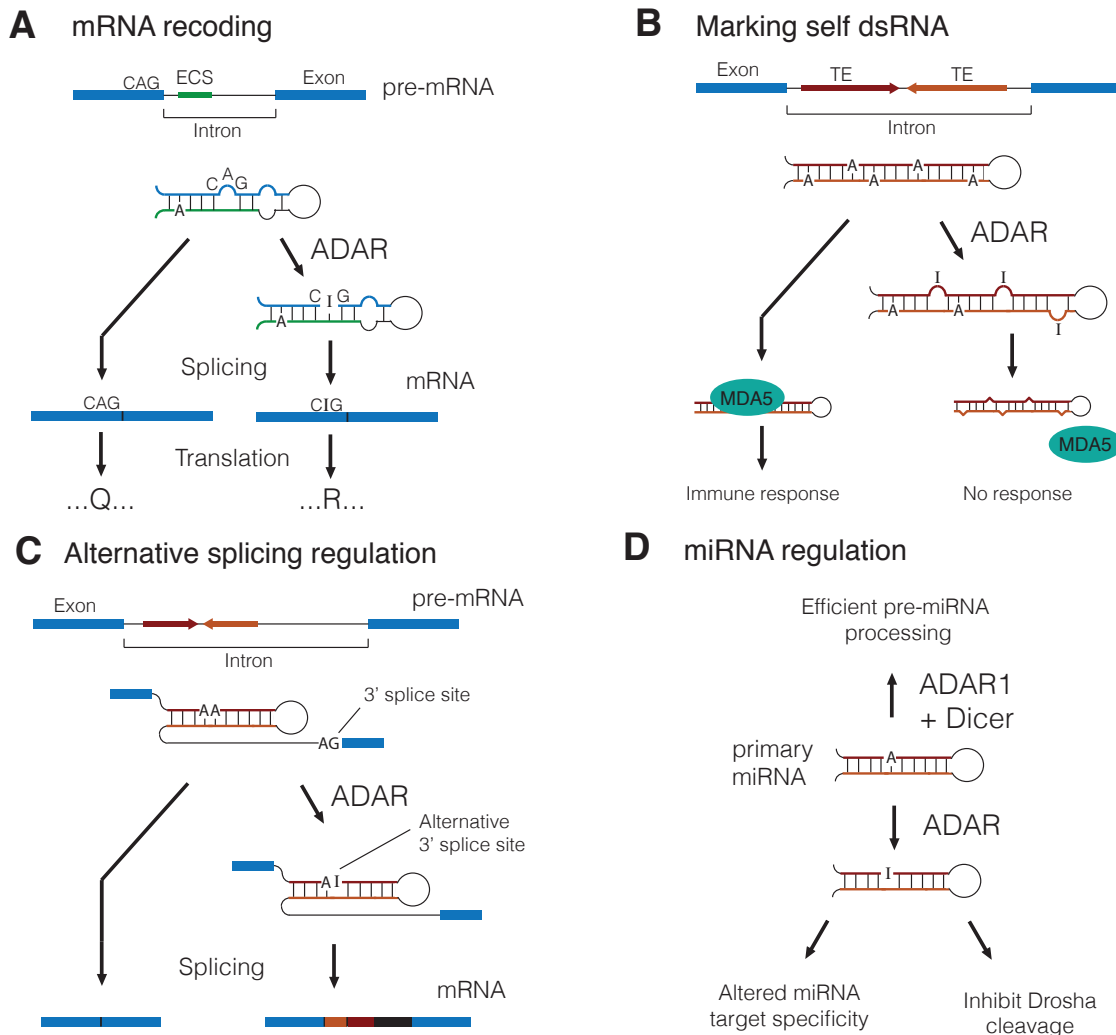
Figure 1.1 Mechanisms of ADAR function. (*A*) Coding sequence in a pre-mRNA base-pairs with an exon complementary sequence (ECS) in a proximal intron. ADAR edits an adenosine, converting a genomically encoded CAG codon to the edited form CIG. Following splicing, the unedited and edited codons are translated Q or R, respectively. (*B*) Transposable elements (TEs) in opposite orientations form highly base-paired dsRNA. ADAR nonselectively edits adenosines to inosines, introducing I-U mismatches. The unedited dsRNA, but not the edited duplex, triggers an immune response through the antiviral dsRNA sensor MDA5. (*C*) Duplex RNA in a pre-mRNA intron is edited by ADAR, converting an AA dinucleotide to AI and creating an alternative 3' splice site. Without editing, splicing occurs at the normal downstream AG 3' splice site, but the edited transcript is spliced at the alternative AI site, incorporating additional sequences (orange, red, black) in the mature mRNA. (*D*) ADARs impact miRNA regulation by several mechanisms. Editing in a primary miRNA typically inhibits Drosha processing, but it can also alter mature miRNA target specificity. Mammalian ADAR1 forms a complex with Dicer to promote efficient processing of pre-miRNAs.

CHAPTER 2

C. ELEGANS ADARS ANTAGONIZE SILENCING OF

CELLULAR DSRNAS BY THE ANTIVIRAL

RNAI PATHWAY*

C. elegans ADARs antagonize silencing of cellular dsRNAs by the antiviral RNAi pathway

Daniel P. Reich, Katarzyna M. Tyc,[1] and Brenda L. Bass

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

Cellular dsRNAs are edited by adenosine deaminases that act on RNA (ADARs). While editing can alter mRNA-coding potential, most editing occurs in noncoding sequences, the function of which is poorly understood. Using dsRNA immunoprecipitation (dsRIP) and RNA sequencing (RNA-seq), we identified 1523 regions of clustered A-to-I editing, termed editing-enriched regions (EERs), in four stages of *Caenorhabditis elegans* development, often with highest expression in embryos. Analyses of small RNA-seq data revealed 22- to 23-nucleotide (nt) siRNAs, reminiscent of viral siRNAs, that mapped to EERs and were abundant in *adr-1;adr-2* mutant animals. Consistent with roles for these siRNAs in silencing, EER-associated genes (EAGs) were down-regulated in *adr-1;adr-2* embryos, and this was dependent on associated EERs and the RNAi factor RDE-4. We observed that ADARs genetically interact with the 26G endogenous siRNA (endo-siRNA) pathway, which likely competes for RNAi components; deletion of factors required for this pathway (*rrf-3* or *ergo-1*) in *adr-1;adr-2* mutant strains caused a synthetic phenotype that was rescued by deleting antiviral RNAi factors. Poly(A)+ RNA-seq revealed EAG down-regulation and antiviral gene induction in *adr-1;adr-2;rrf-3* embryos, and these expression changes were dependent on *rde-1* and *rde-4*. Our data suggest that ADARs restrict antiviral silencing of cellular dsRNAs.

Adenosine deaminases that act on RNA (ADARs) catalyze conversion of adenosine (A) to inosine (I) within double-stranded regions of cellular RNAs (Hundley and Bass 2010; Nishikura 2016). Like guanosine (G), inosine prefers to pair with cytosine, and thus A-to-I RNA editing can alter mRNA-coding capacity. Recoding events are critical for normal function of the nervous system in vertebrates, squid, and *Drosophila* (Deffit and Hundley 2016; Nishikura 2016). However, in all organisms examined to date, the majority of A-to-I editing is outside of coding sequences, mostly in introns and 3′ untranslated regions (UTRs) (Whipple et al. 2015; Blango and Bass 2016; Walkley and Li 2017). In some cases, noncoding A-to-I changes in pre-mRNAs impact splicing and mRNA processing (Nishikura 2016). However, the function of most editing in noncoding regions is not well understood.

In mammals, ADARs are essential to prevent aberrant immune signaling by antiviral dsRNA sensor proteins (Mannion et al. 2014; Liddicoat et al. 2015; Pestal et al. 2015; George et al. 2016). Loss of mouse ADAR1 causes

embryonic lethality characterized by interferon (IFN) overproduction and up-regulation of IFN-inducible transcripts (Hartner et al. 2004; Mannion et al. 2014; Liddicoat et al. 2015). IFN-stimulated gene expression in $ADAR1^{-/-};p53^{-/-}$ mutant fibroblasts is partially rescued by an editing-deficient ADAR1 point mutant but more completely rescued by a catalytically active ADAR1, suggesting that both binding and editing contribute to ADAR1 antagonism of IFN signaling (Mannion et al. 2014; O'Connell et al. 2015). Strikingly, mutations in the immune signaling genes *Mavs* or *Ifih1* (MDA5) rescue $ADAR1^{-/-}$ mutant embryonic lethality and IFN hyperactivation (Mannion et al. 2014; Pestal et al. 2015). Thus, mammalian ADAR1 is thought to prevent cellular dsRNAs from activating RIG-I-like receptors, although it is unclear whether this is a conserved ADAR function.

The nematode *Caenorhabditis elegans* lacks an IFN response and instead uses RNAi to sense viral dsRNA and silence viral transcripts (Ashe et al. 2013; Guo et al. 2013). *C. elegans* ADARs inhibit RNAi-mediated

transgene silencing (Knight and Bass 2002) and small RNA biogenesis (Warf et al. 2012), suggesting that, by analogy to vertebrates, ADARs could mark cellular dsRNA as self. *C. elegans* have two genes encoding ADARs, *adr-1* and *adr-2*, and deletion of either gene causes chemotaxis defects, transgene silencing, and shortened life span (Knight and Bass 2002; Tonkin et al. 2002; Sebastiani et al. 2009). Consistent with the notion that these phenotypes relate to altered dsRNA-mediated silencing, these *adr* mutant phenotypes are rescued by additional loss of function of RNAi factors (Knight and Bass 2002; Tonkin and Bass 2003; Sebastiani et al. 2009).

The sole *C. elegans* Dicer enzyme, DCR-1, is required for biogenesis of microRNAs (miRNAs) as well as viral and endogenous siRNAs (endo-siRNAs) (Supplemental Fig. S1). Studies in *C. elegans* extracts indicate that DCR-1 cleaves dsRNA to produce siRNAs that are predominantly 23 nucleotides (nt) in length with a 5′ monophosphate (Welker et al. 2011). While such products arise during viral infection, they have not been observed among endo-siRNAs in wild-type *C. elegans* (Ruby et al. 2006; Ashe et al. 2013; Billi et al. 2014). Rather, characterized DCR-1-dependent endo-siRNAs are 26 nt, have a 5′ guanosine monophosphate (26G siRNAs), and are produced by DCR-1 acting in concert with the RNA-dependent RNA polymerase (RdRP) RRF-3 (Thivierge et al. 2012; Blumenfeld and Jose 2016). 26G siRNAs occur in embryos and germline tissues and exist in two classes bound to distinct Argonaute proteins. In sperm, 26G siRNAs bind ALG-3 and ALG-4, while ERGO-1 binds 26G siRNAs in oocytes and embryos. To effect silencing, 26G siRNAs and siRNAs from exogenous dsRNAs, such as those introduced by feeding, injection, or viral infection, trigger the RdRP-mediated production of 22-nt siRNAs with a 5′ triphosphorylated guanosine (22G siRNAs) antisense to target transcripts (Ruby et al. 2006; Pak and Fire 2007; Sijen et al. 2007; Vasale et al. 2010; Ashe et al. 2013; Billi et al. 2014).

Here we define the editing-enriched regions (EERs), ADAR-edited long dsRNAs, expressed in four stages of *C. elegans* development. EERs and their associated genes show highest expression in embryos and give rise to 23-nt 5′ monophosphorylated siRNAs that are abundant in *adr-1;adr-2* mutant animals. Using quantitative RT–PCR (qRT–PCR), we show that EER-associated genes (EAGs) are down-regulated in *adr-1;adr-2* mutant embryos but not L3 larvae. Down-regulation of EAGs in *adr-1;adr-2* embryos requires both RDE-4 and an associated double-stranded EER. Analysis of a siRNA-sensitive *GFP:: NRDE-3* transgene indicates that ADARs antagonize siRNA biogenesis independent of the 26G pathway. However, in *adr-1;adr-2* mutant backgrounds, 26G loss of function causes a synthetic phenotype dependent on the antiviral RNAi pathway. Transcriptomes of *adr-1;adr-2; rrf-3* mutant embryos reveal robust EAG down-regulation and increased expression of genes induced during Orsay virus infection, both of which are rescued by *rde-1* and *rde-4* deletion. Together, our results suggest a conserved role for *C. elegans* ADARs in antagonizing the antiviral response to self dsRNAs.

Results

Clustered ADAR-editing sites define dsRNAs expressed during C. elegans development

To identify dsRNAs expressed during *C. elegans* development, we performed high-throughput RNA sequencing (RNA-seq) on total and dsRNA-enriched rRNA-depleted RNA samples from embryos, early larval stages (L1 and L2), late larval stages (L3 and L4), and young adults. We used a previously developed pipeline (Whipple et al. 2015) to define dsRNAs by scanning for clusters of A-to-I RNA editing (Supplemental Fig. S2). From combined dsRNA immunoprecipitation (dsRIP) and input data sets of all stages, we defined 1523 EERs in total (Fig. 1A; Supplemental File S1). We found more than twice as many EERs in dsRIP samples than input, suggesting that the dsRIP enriched for dsRNA. Across developmental stages, we defined the greatest number of EERs in embryos, in contrast to a previous study that observed more clusters in L1 larvae (Zhao et al. 2015). This discrepancy may reflect differences in populations sequenced, since the previous study used L1s aged 4 h after diapause, while we sequenced mixed L1–L2 larvae aged 6–20 h after diapause (Materials and Methods). Many EERs were defined in multiple stages (Fig. 1B), with 81 EERs common to all four stages and 406 unique to one stage. Stage-specific EERs generally were most highly expressed at the stage in which they were defined (Supplemental Fig. S3). Since our pipeline required at least five reads to call an EER, expression primarily determined the stages where an EER was found.

The EERs defined in our study had properties similar to previously defined *C. elegans* editing clusters (Wu et al. 2011; Whipple et al. 2015; Zhao et al. 2015). EER loci were enriched on autosome distal arms (Supplemental Fig. S4) and largely derived from repetitive elements, particularly transposons (Supplemental Fig. S5). EERs mostly overlapped gene-associated noncoding sequences, while a control set of random expressed regions mapped to more exonic and intergenic sequences (Supplemental Fig. S6A). Specifically, 72.3% of EER nucleotides mapped to introns and 10.2% mapped to 5′ UTRs, 3′ UTRs, or regions within 1 kb of a gene on the same strand. Many EERs were predicted by UNAFold to form long stable intramolecular structures, in contrast to random regions (Supplemental Fig. S6B; Markham and Zuker 2008). Previously, we found significant overlap between human EERs and circular RNAs (circRNAs) (Blango and Bass 2016), and, similarly, in *C. elegans*, 78 EERs overlapped circRNAs (78 observed and 56 expected; $P = 0.0033$ by χ^2 test) (http://www. circbase.org; Supplemental File S2).

EER abundance, but not editing, varies during development

We next examined developmental patterns of EER abundance and editing. The heat map in Figure 1C shows relative abundance in each stage for the 250 most highly edited EERs. Of the 1523 EERs, 1336 overlapped or were within 1 kb of one of 955 unique EAGs (Supplemental File S3). For
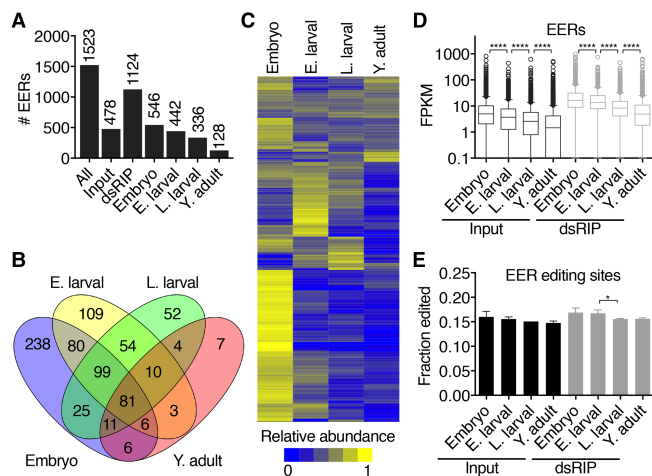
Figure 1. EER abundance, but not editing, changes during development. (E) Early; (L) late; (Y) young. (*A*) The number of EERs defined from each group of data sets. (*B*) Venn diagram of EERs defined in each developmental stage. (*C*) Heat map of relative abundance in input RNA-seq samples for the 250 EERs with the greatest number of edited windows. (*D*) Distribution of mean EER abundance in each stage and treatment. (****) $P < 0.0001$, Mann-Whitney *U*-test. (*E*) Fraction of all EER-editing sites that appeared as guanosines in each stage and treatment. While individual sites ranged from 1% to 99% edited, all sites together were ~15% edited in each sample. Error bars show standard deviation (SD) of three biological replicates. (*) $P < 0.05$, Student's *t*-test.

each gene-associated EER, we calculated a Pearson correlation coefficient (PCC) comparing EER abundance with EAG expression across all stages. We found that the mean EER–EAG PCC was 0.396 and the median PCC was 0.591, demonstrating that EER and EAG expression between developmental stages is correlated.

Strikingly, more than half of EERs (50.4%; 768 out of 1523) and EAGs (53.1%; 507 out of 955) displayed highest expression in embryos. We plotted the abundance of all 1523 EERs in each developmental stage and treatment (Fig. 1D), observing the highest collective expression in embryos, with decreased expression in subsequent stages. We compared EERs with length-matched random regions (Supplemental Fig. S7A) and observed greater differences in EER abundance between developmental stages than random regions (Supplemental Fig. S7B–D), suggesting that EER expression patterns are distinct from most transcripts.

Next, we assessed EER-editing levels during development. We made a list of adenosines within EERs edited >1% in all input and dsRIP samples pooled (referred to here as EER-editing sites). For each individual RNA-seq replicate, we determined the total number of A-to-G changes (#Ed) observed at EER-editing sites (Supplemental Fig. S8). In both input and dsRIP samples, we observed the greatest #Ed in embryos. For each replicate, we then counted the total reads covering each EER-editing site (#Tot) and calculated fraction editing as #Ed/#Tot (Materials and Methods). Finally, we calculated the average fraction editing across the three replicates in each stage and treatment (Fig. 1E). Editing frequency changed only minorly across development, in agreement with previous work (Zhao et al. 2015). We conclude that EER abundance changes during normal development, while editing within EERs remains stable.

We tested whether ADARs impacted development or viability by determining the fraction of wild-type and *adr-1;adr-2* mutant embryos that developed to adulthood over 3 d. We used three independently derived sets of *adr-*

1;adr-2 deletion alleles: two sets of previously described EMS-induced mutants (Tonkin et al. 2002; Hundley et al. 2008) and one that we created by injection of CRISPR/Cas9 ribonucleoparticles (Cho et al. 2013; Paix et al. 2015). Although the CRISPR mutant (*adr-1(uu49); adr-2(uu28)*) and *adr-1(tm668);adr-2(ok735)* strains were no different from wild type, the *adr-1(gv6);adr-2(gv42)* strain displayed a low-penetrance larval arrest phenotype (Supplemental Fig. S9), possibly due to background mutations. Since two mutant lines did not differ from wild type, we conclude that development occurs normally in *adr-1;adr-2* mutants, consistent with previous work (Tonkin et al. 2002; Sebastiani et al. 2009).

Abundant siRNAs mapping to EERs in adr-1;adr-2 mutants suggest EER processing by DCR-1

Since ADARs edit only dsRNAs, EERs must be double-stranded, and we hypothesized that EERs would also be substrates for other dsRNA-binding proteins (dsRBPs), particularly in the absence of ADARs. Like other Dicer enzymes, *C. elegans* DCR-1 cleaves dsRNA to produce primary siRNAs that have 5′ monophosphates and 3′ hydroxyls (Ruby et al. 2006). To distinguish primary siRNAs from secondary siRNAs that have 5′ triphosphates, small RNAs are typically sequenced using 5′ phosphate (5′P)-dependent protocols that capture only primary siRNAs and 5′P-independent protocols that capture both primary and secondary siRNAs (Ruby et al. 2006; Pak and Fire 2007). To determine whether EERs were DCR-1 substrates, we analyzed published 5′P-dependent small RNA-seq data sets from mixed-stage wild-type or *adr-1(tm668);adr-2 (ok735)* worms (Warf et al. 2012). We found that siRNAs mapped sense to 74.1% (1128 out of 1523) of EERs. Strikingly, 94.0% (1060 out of 1128) of these EERs showed increased siRNA abundance in *adr-1;adr-2* mutant animals compared with wild type (Fig. 2A,B; Supplemental File S1). Length-matched random regions did not show similar siRNA enrichment in *adr-1;adr-2* samples.
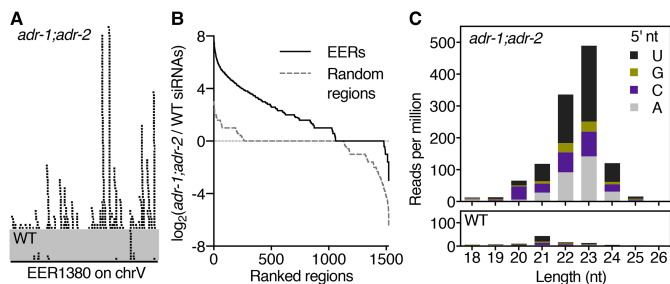
Reich et al.

Figure 2. EER-23H siRNAs are abundant in *adr-1; adr-2* double mutants. (*A*) Genome browser view of 5′P-dependent small RNA-seq reads from mixed-stage wild-type (WT) and *adr-1(tm668);adr-2(ok735)* mutant animals mapping sense to EER1380. (*B*) EER-23H siRNA enrichment in *adr-1(tm668);adr-2 (ok735)* mutants. Plots show the log_2 ratio of siRNA abundance in *adr-1;adr-2* mutants over wild type for EERs (black solid line) and control regions (gray dashed line). (*C*) Analysis of 5′ nucleotide and length distribution of all EER-23H siRNAs from *adr-1;adr-2* mutant and wild-type animals.

To determine the origin of EER-mapped 5′ monophosphate siRNAs, we analyzed their length distribution and 5′ nucleotide preferences (Fig. 2C). In wild-type worms, EER-mapped siRNAs showed a small peak at 21 nt with a preference for 5′ U, typical of *C. elegans* 21U-RNAs/piRNAs (Ruby et al. 2006). We suspect that we observed this 21U peak because 18 21U-RNA loci overlap EERs (Supplemental File S2), 21U-RNAs are abundantly expressed in embryonic and germline tissue, and 5′P-dependent protocols include these small RNAs (Batista et al. 2008). In contrast to wild type, EER-siRNAs from *adr-1; adr-2* animals showed a peak at 23 nt with a 5′ nucleotide bias against G (abbreviated H) (Cornish-Bowden 1985). This 5′ nucleotide preference is also seen in *C. elegans* miRNAs (Warf et al. 2011) and primary antiviral siRNAs (Ashe et al. 2013), suggesting that EER-mapped siRNAs in *adr-1;adr-2* mutants, which we refer to as EER-23H siRNAs, are produced by direct DCR-1 cleavage of EERs.

A previous study identified 454 regions (termed ADAR-modulated RNA loci) that give rise to abundant 23- to 24-nt primary siRNAs and 22-nt secondary siRNAs in *adr-1 (gv6);adr-2(gv42)* animals (Wu et al. 2011). Of these 454 loci, 93 overlapped EERs (85 EERs observed and four expected; $P < 0.0001$ by χ^2 test) (Supplemental File S2), suggesting that EERs and ADAR-modulated loci represent related but mostly distinct regions. ADAR-modulated RNA loci display markedly lower coverage than EERs in all stages (Supplemental Fig. S10), suggesting that these regions have different rates of transcription or degradation.

Like the 23- to 24-nt primary siRNAs from ADAR-modulated loci, we reasoned that EER-23H siRNAs may promote the production of secondary siRNAs. We thus analyzed siRNAs antisense to EERs from published 5′-independent small RNA-seq data sets from wild-type and *adr-1(gv6);adr-2(gv42)* embryos and L4 larvae (Supplemental Fig. S11; Wu et al. 2011). Although 5′P-independent siRNAs often mapped both sense and antisense to EERs when allowed to map multiple locations, we found that uniquely aligned reads mostly mapped antisense. In all samples, siRNAs antisense to EERs were primarily 22 nt with a 5′ G, suggesting an RdRP-dependent origin (Billi et al. 2014). We refer to these antisense secondary siRNAs as EER-22G siRNAs. Like EER-23H siRNAs, most EERs showed increased EER-22G siRNA abundance in *adr-1;adr-2* mutants relative to wild type in both embryo and L4 larval samples (Supplemental Fig. S12A,B;

Supplemental File S1). Control random regions did not show similar enrichment. The RNAi genes *rde-1* and *rde-4* promote secondary and primary siRNA biogenesis, respectively, in response to viral dsRNA (Ashe et al. 2013). In embryos, most EERs showed reduced EER-22G siRNA abundance in *adr-1(gv6);adr-2(gv42);rde-1(ne219)* and *adr-1(gv6);adr-2(gv42);rde-4(ne299)* triple mutants relative to *adr-1(gv6);adr-2(gv42)* double mutants (Supplemental Fig. S12C–F). However, EER-22G siRNA abundance in L4 larvae was reduced only marginally in *adr-1; adr-2;rde-4* and was no different from control regions in *adr-1;adr-2;rde-1*. Thus, *rde-1* and *rde-4* mediate EER-22G siRNA accumulation in embryos but not in L4 animals. These data suggest that abundant EER-23H siRNAs in *adr-1;adr-2* mutants promote the production of antisense secondary siRNAs.

In embryos, ADARs prevent down-regulation of EAGs via RNAi against EERs

We next sought to determine whether EER-siRNAs regulate expression of EAGs and first identified EAGs that were the best candidates for such regulation. Since *adr-1*, *adr-2*, and most EAGs are maximally expressed in embryos, we used input RNA-seq data to identify 452 EAGs with ≥50% higher gene expression in embryos than late stage larvae (Supplemental File S4). Using existing data sets from mixed-stage animals (Warf et al. 2012), we selected genes with more EER-23H siRNAs in *adr-1 (tm668);adr-2(ok735)* mutants than wild type and finally narrowed our list to genes down-regulated in *adr-1 (tm668);adr-2(ok735)* mutants relative to wild type by microarray analyses. This analysis revealed 53 EAGs as strong candidates for ADAR-mediated gene regulation.

From our 53 candidates, we assayed expression of eight EAGs in three *adr-1;adr-2* double-mutant strains relative to wild-type embryos (Supplemental Fig. S13). For all genes, we observed a modest reduction in at least two strains tested. The *adr-1(gv6);adr-2(gv42)* strain had more substantial differences compared with others tested, again suggesting that it harbors additional mutations. All further experiments were performed using *adr-1(uu49); adr-2(uu28)* deletions created for this study.

We next tested EAG expression in embryos and L3 larvae of *adr-1;adr-2* mutants with or without additional mutations in *rde-1* or *rde-4*. Six EAGs displayed significantly decreased expression in *adr-1;adr-2* embryos

relative to wild type (Fig. 3A). Although the differences were small (∼20% below wild type on average), we observed these differences reproducibly across many independent biological replicates ($n \geq 8$). A recent study observed small but reproducible down-regulation of pseudogenes and genes with edited 3′ UTRs in *adr-1;adr-2* embryos (Goldstein et al. 2017). When we measured EAG expression in L3 larvae, we observed no significant difference in expression between wild type and *adr-1;adr-2* mutants for all genes down-regulated in embryos (Fig. 3B). This suggests that ADARs promote EAG expression early in development but not at later stages. Importantly, deleting *rde-4* in *adr-1;adr-2* mutants with CRISPR protocols (Cho et al. 2013; Paix et al. 2015) completely or partially rescued the reduced gene expression in embryos for four of six EAGs, suggesting that these EAGs are down-regulated by RNAi in *adr-1;adr-2* embryos. Curiously, deleting *rde-1* did not strongly affect the expression of the EAGs tested. *C. elegans* encode 27 Argonaute proteins, and we speculate that some of these act redundantly with RDE-1 (Billi et al. 2014). Altogether, these data suggest that ADARs antagonize RNAi activity in embryos to promote normal EAG expression.

To confirm that EERs are required for EAG regulation by ADARs, we used CRISPR protocols (Cho et al. 2013; Paix et al. 2015) to remove intronic EER sequences (*ΔEER*) in three EAGs: *efa-6*, *ccb-1*, and *egl-8* (Fig. 3C,D; Supplemental Fig. S14). We chose genes with a single intronic EER that gave rise to abundant siRNAs in *adr-1;adr-2* strains (Supplemental File S4). All *ΔEER* mutants were viable without obvious morphological and developmental abnormalities, and we observed no expression differences in *ΔEER* mutants compared with wild type. If EERs are required to down-regulate associated EAGs in

adr-1;adr-2 mutant embryos, we predicted that EER deletion would abrogate this expression change. Indeed, for two of the three genes tested, EER deletion rescued EAG down-regulation in *adr-1;adr-2* embryos (Fig. 3C). We did not observe rescue of *ccb-1* expression upon EER deletion, although *ccb-1* expression was only slightly affected in *adr-1;adr-2* mutants, making it more difficult to establish significance. In L3 larvae, EAG mRNA levels in *adr-1; adr-2* double mutants and *adr-1;adr-2;ΔEER* triple mutants did not differ significantly from wild type (Fig. 3D). We conclude that ADARs antagonize RNAi-mediated down-regulation of EAGs via their EERs.

ADARs regulate a siRNA-sensitive reporter independent of the 26G endo-siRNA pathway

While ADARs impacted EAG expression, we were puzzled about why expression differences were so minor. Since EAG down-regulation in *adr-1;adr-2* mutants is RNAi-dependent, we considered that a parallel RNAi pathway might restrict EAG silencing by competing for common factors. The 26G endo-siRNA pathway was a prime candidate for such competition, since it uses factors required for robust RNAi, and its loss of function causes enhanced RNAi phenotypes (Vasale et al. 2010; Billi et al. 2014). In the 26G pathway, the ERI complex (containing DCR-1, RDE-4, and the RdRP RRF-3) couples dsRNA synthesis and cleavage to produce 26G siRNAs that promote secondary 22G siRNA biogenesis and silence target genes (Supplemental Fig. S1; Thivierge et al. 2012).

We first determined whether the 26G pathway was required for ADAR-antagonized siRNA biogenesis using a GFP::NRDE-3 reporter whose localization depends on
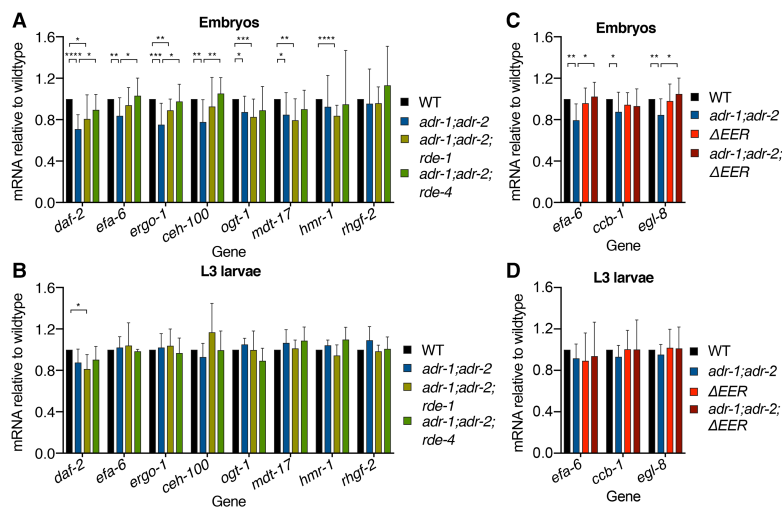

Figure 3. EAG expression decreases in *adr-1(uu49);adr-2(uu28)* embryos in an RNAi- and EER-dependent manner. Expression of EAGs in embryos (*A*; $n \geq 8$) and L3 larvae (*B*; $n = 5$) of four genotypes, measured by qRT–PCR. Expression levels for three EAGs in embryos (*C*; $n \geq 6$) and L3 larvae (*D*; $n = 5$) in strains where each EAG's sole EER was deleted by CRISPR (*ΔEER*). All panels show expression as mean. Error bars show SD. (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$; (****) $P < 0.0001$, Student's *t*-test.

22G secondary siRNA binding (Guang et al. 2008). When the 26G pathway is active, 26G siRNAs stimulate 22G siRNA production, causing NRDE-3 to localize to the nucleus. However, in *rrf-3* mutants, 26G siRNAs are absent, precluding downstream 22G synthesis, and NRDE-3 localizes to the cytoplasm. To test interactions between ADARs and the 26G pathway, we introduced the *GFP::NRDE-3* transgene into *adr-1;adr-2*, *rrf-3*, and *adr-1;adr-2;rrf-3* deletion strains. While GFP::NRDE-3 was primarily cytoplasmic in the *rrf-3* background, it localized to the nucleus in *adr-1;adr-2;rrf-3* animals (Fig. 4A), suggesting that ADARs antagonize siRNA production independent of the 26G pathway. Furthermore, these data support the conclusion that ADARs antagonize production of both primary and secondary siRNAs.

ADARs genetically interact with the 26G pathway in a manner dependent on antiviral RNAi

In testing GFP::NRDE-3 subcellular localization, we noticed that *adr-1;adr-2;rrf-3* mutants displayed defects not present in *adr-1;adr-2* and *rrf-3* mutants, suggesting a synthetic genetic interaction. While *adr-1;adr-2* double mutants are healthy, and *rrf-3* single mutants show temperature-sensitive sterility and reduced brood size (Simmer et al. 2002), *adr-1;adr-2;rrf-3* triple mutants displayed a phenotype marked by frequent adult bursting (Fig. 4B; Supplemental Fig. S15A). In addition, *adr-1;adr-2;rrf-3* mutants had markedly lower brood sizes than *rrf-3* single mutants (Fig. 4C). We confirmed the *adr-1;adr-2;rrf-3* genetic interaction using three independent *rrf-3* deletions and two sets of *adr-1;adr-2* deletions (Supplemental Fig. S15B,C), implying that it is specific to our genes of interest. Both *adr-1* and *adr-2* contributed to

the bursting phenotype, as 37.9% of *adr-1;adr-2;rrf-3* worms burst by day 5 after egg lay compared with 10.0% of *adr-2;rrf-3* and 0.4% of *adr-1;rrf-3* worms. Since ADR-1 binds dsRNA but lacks catalytic activity, while ADR-2 catalyzes A-to-I editing (Tonkin et al. 2002; Washburn et al. 2014), our observations suggest that both binding and editing contribute to ADAR functions in vivo. We observed the same phenotypes in *adr-1;adr-2;ergo-1* mutants lacking both ADARs and the ERGO-1 Argonaute required in the oocyte/embryo arm of the 26G pathway (Fig. 4B,C), suggesting that *C. elegans* ADARs genetically interact with the 26G pathway broadly, not with *rrf-3* alone.

Since ADARs limit RNAi against EAGs and since 26G loss of function causes enhanced RNAi, we hypothesized that the *adr-1;adr-2;rrf-3* synthetic phenotype resulted from increased RNAi activity against EAGs. To test this, we crossed *rde-1* and *rde-4* deletions into the *adr-1; adr-2;rrf-3* mutant background (Fig. 4D; Supplemental Fig. S15D). Loss of either *rde-1* or *rde-4* rescued the frequent bursting and reduced brood sizes of *adr-1;adr-2; rrf-3* mutants, suggesting that these genes are required for the *adr-1;adr-2;rrf-3* synthetic defects. Both RDE-1 and RDE-4 function in RNAi-mediated antiviral immunity (Ashe et al. 2013), so we tested additional antiviral components by deleting them with CRISPR in an *adr-1;adr-2; rrf-3* triple-mutant background. Intriguingly, *adr-1;adr-2; rrf-3* defects were fully rescued by mutating *drh-1*, which encodes a RIG-I-like helicase required for processing viral dsRNA into primary siRNAs, and largely rescued by deleting *rrf-1*, which encodes a somatic RdRP that makes secondary siRNAs (Ashe et al. 2013; Guo et al. 2013). Deleting *nrde-3*, the nucleocytoplasmic shuttling Argonaute, also rescued *adr-1;adr-2;rrf-3* bursting and brood
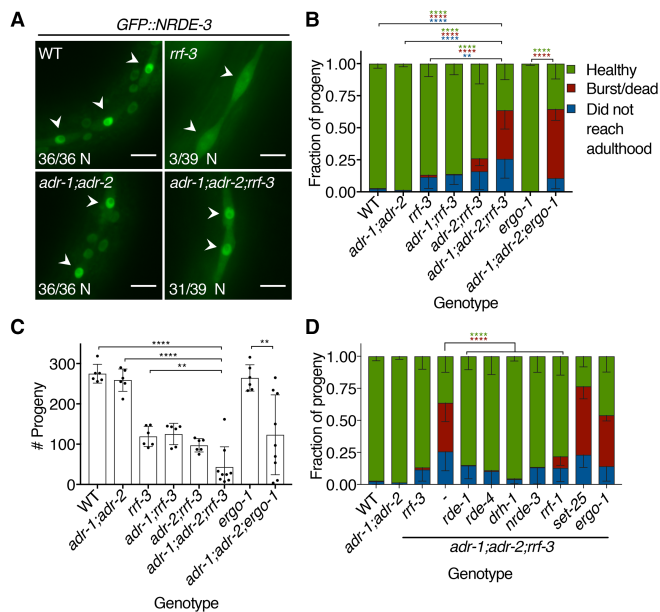

Figure 4. The 26G endo-siRNA pathway genetically interacts with ADARs. (*A*) GFP::NRDE-3 visualized in L3 larvae seam cells (arrowheads) of the indicated mutant genotypes. Numbers in the *bottom left* of each panel report the fraction of worms with nuclear-enriched (N) GFP::NRDE-3 in seam cells. Bar, 10 µm. (*B*) Bursting assay shows the fate of embryos laid by each genotype 5 d after egg lay. Error bars show SD. $n \geq 6$ assays. (**) $P < 0.01$; (****) $P < 0.0001$. Asterisk colors show categories compared by two-way ANOVA with Tukey's multiple comparisons correction. (*C*) Average brood size for each genotype in *B*, with individual broods shown as dots. Error bars show SD. $n \geq 6$ assays. (**) $P < 0.01$; (****) $P < 0.0001$, Student's *t*-test. (*D*) Developmental fates, as in *B*, of *adr-1(uu49);adr-2 (uu28);rrf-3(uu56)* mutant strains with additional mutations in genes encoding RNAi-related factors. Error bars show SD. $n \geq 6$ assays. (****) $P < 0.0001$. Asterisk colors show categories compared by two-way ANOVA with Tukey's multiple comparisons correction.

size defects. However, loss of *set-25*, which encodes a histone H3 Lys9 methylase, had no effect. Deleting *ergo-1* in the *adr-1;adr-2;rrf-3* mutant background also had no effect, consistent with *ergo-1* acting downstream from *rrf-3* in the 26G pathway. We conclude that antiviral RNAi activity causes bursting and small brood size when ADARs and the 26G pathway are inactive.

EAGs and virus-induced genes are misregulated in adr-1;adr-2;rrf-3 *mutants*

To gain insight into gene expression changes underlying *adr-1;adr-2;rrf-3* mutant phenotypes, we sequenced poly(A)$^+$ RNA from four biological replicates of embryos of six genotypes: wild type, *adr-1;adr-2*, *rrf-3*, *adr-1;adr-2;rrf-3*, *adr-1;adr-2;rrf-3;rde-1*, and *adr-1;adr-2;rrf-3;rde-4*. Using DESeq2 (Love et al. 2014), we analyzed differential gene expression between genotypes (Supplemental File S5).

Consistent with our predictions, collective EAG expression decreased in *adr-1;adr-2* mutant embryos and decreased further in *adr-1;adr-2;rrf-3* embryos in a manner dependent on both *rde-1* and *rde-4* (Fig. 5A). As observed by qRT–PCR (Fig. 3A), *rde-1* mutation did not rescue EAG expression as robustly as loss of *rde-4*. EAG expression also decreased in the *rrf-3* mutant, similar to the *adr-1;adr-2* mutant, indicating that the 26G pathway influences EAG expression even with ADARs present. Expression differences of most individual EAGs were small, <20% below wild type, and we hypothesize that the collective down-regulation of many EAGs—rescued by loss of *rde-1* or *rde-4*—causes *adr-1;adr-2;rrf-3* mutant

phenotypes. Indeed, gene ontology (GO) analysis on EAGs revealed enrichment for terms associated with morphogenesis and development, suggesting that EAG misregulation could result in developmental defects that cause bursting (Supplemental File S6). A recent RNA-seq analysis revealed expression changes in isolated neurons from *adr-1;adr-2* animals that are not observed in whole worms (Deffit et al. 2017), suggesting that EAG misregulation could be more substantial in specific tissues.

Our analyses involved 920 EAGs, and we considered the possibility that this large number of genes might mask certain trends. Thus, we divided EAGs into three groups: genes that were significantly down-regulated (231 genes), significantly up-regulated (50 genes), or not significantly changed (639 genes) in *adr-1;adr-2;rrf-3* triple mutants relative to wild type (Fig. 5B). Compared with all EAGs together, down-regulated EAGs displayed more robust silencing in all mutant genotypes relative to wild type (Fig. 5C). Up-regulated EAGs showed *rrf-3*-dependent increased expression but were largely unchanged in *adr-1;adr-2* double mutants (Fig. 5D). EAGs that were not significantly misexpressed in *adr-1;adr-2;rrf-3* mutants showed modest down-regulation in all mutant strains (Fig. 5E) in a pattern resembling that in Figure 5A. Potentially informing why down-regulated EAGs were robustly silenced, we found that these genes often had more than one EER-associated intron or UTR and had longer EERs than other genes (Supplemental Fig. S16A,B). In contrast, up-regulated EAGs tended to have a single shorter EER that was less likely to occur in an intron (Supplemental Fig. S16A–C).

We next determined differentially expressed genes (DEGs) in pairwise comparisons between genotypes. We
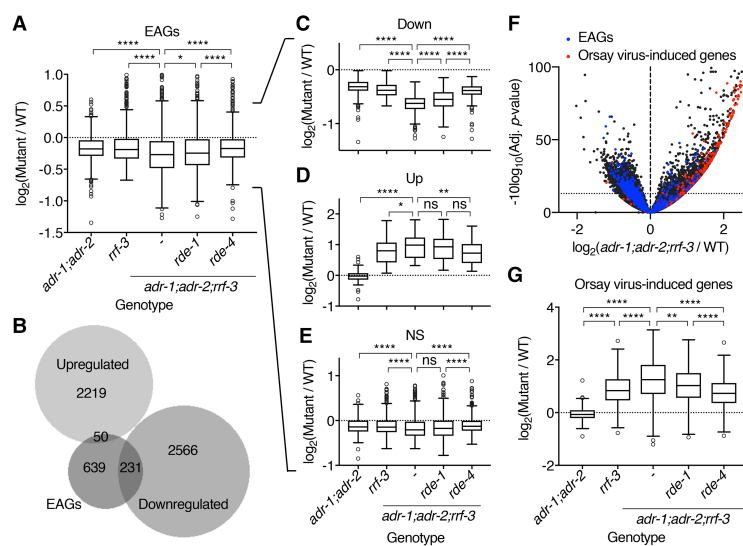

Figure 5. EAGs and Orsay virus-induced genes are misregulated when ADARs and the 26G pathway are disrupted. (*A*) Tukey box plots show distributions of log$_2$-(expression fold change compared with wild type) for EAGs in each mutant genotype analyzed by RNA-seq. (*) $P < 0.05$; (****) $P < 0.0001$, Mann-Whitney *U*-test. (*B*) Venn diagram showing the overlap between differentially expressed genes (DEGs) up-regulated and down-regulated in *adr-1 (uu49);adr-2(uu28);rrf-3(uu56)* mutants compared with wild type as well as EAGs expressed in RNA-seq samples (>10 reads total). Tukey box plots as in *A* show expression fold change in mutant genotypes for significantly down-regulated EAGs (down) (*C*), significantly up-regulated EAGs (up) (*D*), and EAGs not significantly changed (NS) (*E*) in *adr-1;adr-2;rrf-3* mutant embryos relative to wild type. Adjusted *P*-value cutoff was 0.05. (ns) $P > 0.05$; (*) $P < 0.05$; (**) $P < 0.01$; (****) $P < 0.0001$, Mann-Whitney *U*-test. (*F*) Genes analyzed by poly(A)$^+$ RNA-seq are plotted by log$_2$(expression fold

change compared with wild type) against −10log$_{10}$(adjusted *P*-value) in *adr-1;adr-2;rrf-3* mutants compared with wild type (i.e., higher *y*-values indicate more significant differences). The horizontal dotted line designates the adjusted *P*-value cutoff of 0.05 used to define DEGs. (*G*) Tukey box plots as in *A* showing Orsay-induced gene expression fold change in mutants. (**) $P < 0.01$; (****) $P < 0.0001$, Mann-Whitney *U*-test.

observed 2269 genes significantly up-regulated and 2797 down-regulated in *adr-1;adr-2;rrf-3* triple-mutant embryos compared with wild type (Fig. 5B,F). EAGs were significantly enriched among down-regulated DEGs and depleted from up-regulated DEGs ($P < 0.0001$, χ^2 test). In *adr-1;adr-2* double mutants compared with wild type, we observed only nine DEGs up and 15 down, excluding *adr-1* and *adr-2* (Supplemental File S5). The 15 down-regulated DEGs included four EAGs and other genes with properties suggesting that they form dsRNA: Two were transposons, three were antisense to other genes, and two were edited in introns covered by too few reads in our original analysis to be defined as EERs. We observed only three DEGs comparing *adr-1;adr-2;rrf-3* triple-mutant embryos with *adr-1;adr-2;rrf-3;rde-1* or *adr-1;adr-2;rrf-3;rde-4* quadruple mutants, and only one gene, the TURMOIL1 transposon *Y48G1BL.4*, was rescued by both *rde-1* and *rde-4* deletion (Supplemental Fig. S17A,B). As *Y48G1BL.4* is silenced in *adr-1;adr-2* mutants that do not burst, we conclude that it does not mediate *adr-1;adr-2;rrf-3* bursting. We speculated that the reason that we identified so few DEGs in *adr-1;adr-2* samples was that expression differences were too small or variable to achieve statistical significance. Thus, we performed gene set enrichment analysis (GSEA) (Subramanian et al. 2005) to find overrepresented gene classes altered in *rde-1-* and *rde-4*-rescued quadruple mutants relative to *adr-1;adr-2;rrf-3*. In both cases, the most enriched class of up-regulated genes was the 231 EAGs down in *adr-1;adr-2;rrf-3* relative to wild type (Supplemental Fig. S17C,D; Supplemental File S7), supporting the idea that *rde-1-* and *rde-4*-dependent EAG silencing mediates *adr-1;adr-2;rrf-3* defects.

We hypothesized that ADARs prevent antiviral pathways from recognizing self EERs as viral dsRNAs, so we next examined whether genes induced during viral infection were changed in *adr-1;adr-2;rrf-3* triple mutants. A previous study identified 320 genes (many with predicted functions in degradation and innate immunity) differentially expressed during Orsay virus infection, 298 of which increased in expression (Chen et al. 2017). We analyzed expression of these Orsay virus-induced genes and found that, of 268 genes expressed, 157 were significantly up-regulated in *adr-1;adr-2;rrf-3* mutant embryos (Fig. 5F). These differences were primarily *rrf-3*-dependent, as expression in *adr-1;adr-2* mutants was largely unchanged (Fig. 5G). Thus, unlike mammals, where unedited dsRNAs are thought to induce IFN-stimulated genes (Liddicoat et al. 2015), we did not see strong evidence that unedited dsRNAs drive Orsay virus-induced gene expression, as *rrf-3* deletion is not predicted to increase dsRNA levels. However, expression of virus-induced genes in *rrf-3* mutants was further increased in *adr-1;adr-2;rrf-3* mutants, suggesting that ADARs still impact their regulation. Loss of *rde-1* and *rde-4* in *adr-1;adr-2;rrf-3* mutants reduced expression of antiviral genes. Furthermore, by GSEA, Orsay-induced genes were the most enriched gene class down-regulated in quadruple mutants relative to *adr-1;adr-2;rrf-3* (Supplemental Fig. S17E,F; Supplemental File S7), suggesting that RNAi contributes to antiviral gene induction in *adr-1;adr-2;rrf-3* mutants.

Discussion

Here we present evidence that *C. elegans* ADARs prevent silencing of self dsRNAs by antiviral RNAi (Fig. 6). We identify 1523 EERs (edited structures in gene introns and 3′ UTRs) that give rise to abundant 23H siRNAs and promote RDE-4-dependent gene silencing when ADARs are absent. Genetic analyses suggest that the 26G endo-siRNA pathway restricts antiviral RNAi activity in *adr-1;adr-2* mutants. Robust EAG silencing and antiviral gene induction seen in *adr-1;adr-2;rrf-3* mutants emphasize that ADARs and the 26G pathway limit antiviral activity in the absence of viral infection.

ADARs and the 26G pathway limit antiviral RNAi responses to self dsRNAs

Antiviral RNAi is initiated when viral dsRNA is cleaved into primary siRNAs by DCR-1 in association with the dsRBP RDE-4 and the RIG-I-like helicase DRH-1 (Ashe et al. 2013). The Argonaute RDE-1 uses primary siRNAs
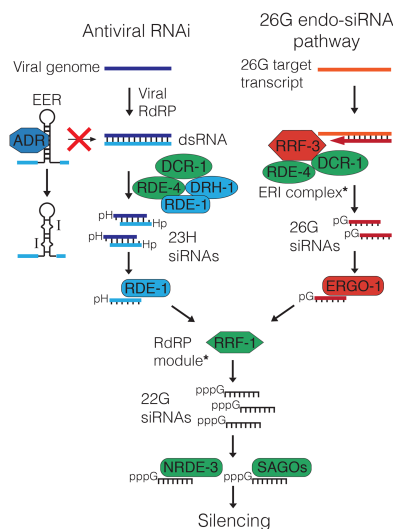

Figure 6. ADARs and the 26G pathway prevent antiviral RNAi-mediated silencing of self dsRNAs. During viral infection, viral replication generates dsRNAs that are processed into 23H siRNAs by a complex of DCR-1, RDE-1, RDE-4, and DRH-1. RDE-1 binds 23H siRNAs and stimulates 22G siRNA production by RRF-1. 22G siRNAs bind NRDE-3 and SAGO Argonautes to effect silencing. ADARs bind and edit EERs to prevent recognition as viral dsRNA and processing to 23H siRNAs by the antiviral DCR-1 complex. In the 26G pathway, the ERI complex (containing RRF-3, DCR-1, and RDE-4) generates 26G siRNAs, which bind ERGO-1, promote 22G siRNA synthesis, and silence targets through NRDE-3 and SAGO proteins. Thus, absent viral infection, antiviral RNAi is kept inactive by ADARs binding and editing self dsRNAs and by 26G pathway sequestration of common RNAi factors (green). For simplicity, we show only factors relevant to this study (see Supplemental Fig. S1); complexes containing additional components are noted with asterisks.

to promote production of antisense secondary 22G si-RNAs that silence viral transcripts (Pak and Fire 2007; Sijen et al. 2007; Guo et al. 2013). By analyzing small RNAs in wild-type and *adr-1;adr-2* mutant worms, we found that primary and secondary siRNAs mapping to EERs are more abundant in the absence of ADARs. Like primary siRNAs from viral dsRNA, sense primary EER-siRNAs are 22–23 nt with a bias against 5′ G (Ashe et al. 2013), suggesting that they result from DCR-1 cleavage of EERs. EER-23H siRNAs promote secondary siRNA production, as evidenced by increased EER-22G siRNAs in *adr-1;adr-2* mutants and nuclear GFP::NRDE-3 in *adr-1; adr-2;rrf-3* mutants. ADAR-antagonized EER-siRNAs are functional, since EAGs are down-regulated in *adr-1; adr-2* embryos, and this requires RDE-4 and EERs. Thus, ADARs prevent EER cleavage and EAG silencing by an RNAi mechanism analogous to the processing and silencing of viral dsRNAs.

The 26G pathway further limits antiviral RNAi activity. While *adr-1;adr-2, rrf-3,* and *ergo-1* mutant strains are largely healthy, deleting either *rrf-3* or *ergo-1* in the *adr-1;adr-2* deletion strain caused bursting and reduced brood size. These phenotypes required *rde-1, rde-4, drh-1,* and *rrf-1,* components also required for antiviral RNAi during Orsay virus infection, suggesting that antiviral RNAi activity causes *adr-1;adr-2;rrf-3* mutant phenotypes. The 26G and antiviral pathways compete for common sets of proteins (Vasale et al. 2010; Thivierge et al. 2012; Ashe et al. 2013), including the Argonautes NRDE-3, SAGO-1, and SAGO-2 that are limiting factors in RNAi (Yigit et al. 2006; Zhuang et al. 2013). The bursting of *adr-1;adr-2;ergo-1* triple mutants suggests that the 26G pathway primarily restricts antiviral RNAi activity by competing for these downstream factors rather than DCR-1 and RDE-4, which would still act upstream in 26G biogenesis in *adr-1;adr-2;ergo-1* mutants.

ADAR antagonism of *C. elegans* antiviral RNAi clearly parallels mammalian ADAR1 antagonism of MDA5-dependent IFN signaling (Mannion et al. 2014; Liddicoat et al. 2015; Pestal et al. 2015). We conclude that ADARs perform a conserved role in preventing antiviral responses to self dsRNAs. The role of the 26G pathway in further limiting antiviral RNAi underscores the importance of restricting immune signaling to appropriate contexts.

How do ADARs prevent recognition of cellular dsRNAs as nonself?

Editing converts A–U base pairs in dsRNA to I–U mismatches (Hundley and Bass 2010; Nishikura 2016), making dsRNAs less "double-stranded" and less ideal substrates for Dicer and other dsRBPs. Indeed, edited dsRNAs are poorly processed into siRNAs in *Drosophila* extracts (Scadden and Smith 2001), while, in HeLa cells, dsRNAs containing I–U mismatches fail to activate MDA5 like control dsRNA (Vitali and Scadden 2010). We found that ADR-1, which binds but does not edit dsRNA (Washburn et al. 2014), restricts antiviral RNAi, since *adr-2;rrf-3* double mutants that lack all editing burst less frequently than *adr-1;adr-2;rrf-3* triple mutants. Still,

adr-1;rrf-3 double mutants retain ADR-2 editing and rarely burst, suggesting that the loss of editing is more deleterious than losing ADR-1 binding. Abundant intron editing suggests that ADARs act in the nucleus, consistent with nuclear localization of most mammalian ADARs (Nishikura 2016) and A-to-I editing of nascent transcripts (Rodriguez et al. 2012). Nuclear localization likely allows ADARs to bind and edit EERs soon after transcription to preempt processing by DCR-1.

How does EER-mediated silencing occur?

EERs are predominantly intronic and thus likely nuclear, while DCR-1 and RRF-1 act primarily in the cytoplasm (Aoki et al. 2007; Drake et al. 2014). However, we observed primary and secondary siRNAs mapping to intronic EERs, and our expression analyses indicate that EAGs can be silenced via intronic EERs (Fig. 3C). Thus, either DCR-1 and RRF-1 act in the nucleus or EERs go to the cytoplasm. In somatic and germline tissues, secondary siRNA production occurs in perinuclear foci (Phillips et al. 2012; Yang et al. 2014) that conceivably could facilitate nucleocytoplasmic exchange of RNAi factors and/or EERs. Alternatively, mitotic nuclear breakdown may provide a window for EER processing and secondary amplification. A mitosis-dependent mechanism may explain why EAGs are down-regulated in *adr-1;adr-2* mutant embryos but not L3 larvae, since nonproliferative larval cells would not generate EER-siRNAs. Still, EER-22G siRNAs are more abundant in *adr-1;adr-2* mutants than wild type at both the embryo and L4 stages. Perhaps *adr-1;adr-2* L3 larvae do not silence EAGs because Argonautes at this stage bind fewer EER-siRNAs either because they comprise a smaller proportion of total siRNAs or because Argonautes become less abundant or tissue-restricted. Our future work aims to establish precise mechanisms of EER silencing.

What causes adr-1;adr-2;rrf-3 mutant phenotypes?

Like bursting and brood size defects, EAG misregulation in *adr-1;adr-2;rrf-3* mutants is *rde-1-* and *rde-4*-dependent. GO terms associated with development and morphogenesis are enriched among EAGs, and we suspect that EAG misexpression compromises vulval morphogenesis, although we did not test this. Bursting is a frequent phenotype of miRNA mutants, attributed to *lin-41* misregulation by *let-7* (Parry et al. 2007; Ecsedi et al. 2015), and *adr-1;adr-2* mutants show altered miRNA networks (Warf et al. 2012). Still, we saw no evidence of *lin-41* misexpression in *adr-1;adr-2;rrf-3* mutant embryos, although we did not test later stages. Importantly, the fact that disrupting RNAi factors rescued *adr-1;adr-2;rrf-3* bursting suggests that siRNA, not miRNA, regulation is perturbed.

Orsay virus-induced genes are robustly induced in *adr-1;adr-2;rrf-3* triple mutants, partly through *rde-1* and *rde-4,* suggesting that activation of an antiviral transcriptional program could contribute to bursting. While we predicted that unedited EERs in *adr-1;adr-2* mutants might induce antiviral gene transcription, we instead saw Orsay-induced genes induced in *rrf-3* but not *adr-1;adr-2* mutants,

suggesting that 26G inactivity is associated with antiviral gene expression. Still, expression of virus-induced genes further increases in *adr-1;adr-2;rrf-3* mutants, suggesting that ADARs limit their induction. Although the 26G pathway has not been implicated in the antiviral response, we suspect that its inhibition could help combat viral infection by relieving competition with antiviral RNAi.

What triggers the virus-induced transcriptional program remains unclear. Generalized stresses such as heat shock, oxidative stress, and translation inhibition trigger innate immune signaling (Kim and Ewbank 2015), so *adr-1;adr-2;rrf-3* mutant antiviral gene induction could indicate nonspecific cellular dysfunction. However, a subset of genes expressed during Orsay infection is not induced in *drh-1* mutant JU1580 worms (Sarkies et al. 2013), suggesting the intriguing possibility that DRH-1 may activate antiviral gene transcription in response to viral dsRNA.

Materials and methods

C. elegans *maintenance and strains used in this study*

All *C. elegans* strains were cultured at 20°C under standard conditions (Brenner 1974). Strains used in the study are listed in Supplemental Table S1.

Sample collection and RNA isolation

Embryos were obtained by sodium hypochlorite treatment (Emmons et al. 1979) of well-fed worms grown 4–5 d in S-Complete liquid medium, washed three times in M9 buffer, and then collected or hatched overnight without food. Synchronized L1 larvae were filtered over two layers of Miracloth (Calbiochem) and cultured in S-Complete liquid medium to mid-L1 (6–8 h), L1/L2 molt (12–14 h), mid-L2 (18–20 h), mid-L3 (28–30 h), L3/L4 molt (34–36 h), mid-L4 (40–42 h), or young adulthood (50–54 h). Early larval samples were prepared by mixing mid-L1, mid-L2, and L1/L2 molt populations at a 5:5:2 volumetric ratio, respectively. Late larval samples were similarly prepared with mid-L3, mid-L4, and L3/L4 molt populations.

Samples were lysed by three freeze–thaw cycles in Trizol reagent. RNA was extracted with chloroform, ethanol-precipitated, treated for 1 h at 37°C with TURBO DNase, and isolated using Zymo Research RNA clean and concentrator columns.

RNA-seq and data preparation

For developmental RNA-seq, we prepared three biological replicates of each stage as described above. For each sample, we treated 40 µg of total RNA with Ribo-Zero human–mouse–rat rRNA removal kit and took 10% of the output for input cDNA libraries. The remaining RNA was incubated overnight at 4°C with 10 µg of J2 anti-dsRNA antibody (English Scientific and Consulting Kft.) in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, and 1% NP-40. J2-bound RNA was collected with Protein-A/G agarose beads for 4 h at 4°C and isolated with Trizol. cDNA libraries were prepared with the Illumina TruSeq stranded total RNA sample preparation LS protocol modified by addition of 2.5% DMSO to reverse transcription reactions. cDNA libraries were sequenced by paired-end 101-cycle sequencing on an Illumina HiSeq 2000 platform by the Microarray and Genome Analysis

Core Facility at the University of Utah Huntsman Cancer Institute.

For poly(A)[+] RNA-seq of the *adr-1;adr-2;rrf-3* triple-mutant and related strains, we collected four biological replicates of embryos of each genotype. Each library was prepared from 1 µg of total RNA by the University of Utah Microarray and Genome Analysis Core Facility using the Illumina TruSeq stranded mRNA library preparation kit. cDNA libraries were sequenced by paired-end 125-cycle sequencing on an Illumina HiSeq 2500 platform.

The Novoalign alignment package (http://www.novocraft.com) was used to trim adaptor sequences and align reads to the *C. elegans* genome (ce10/WS220). Reads were filtered to allow up to four mismatches with the USeq (http://useq.sourceforge.net) application SamTranscriptomeParser using parameters "-a 120 -n 1 –p –r –b." The USeq AligmentEndTrimmer application was used to remove reads with more than one non-A-to-G mismatch and trim read ends of low-quality bases.

Detection of EERs

We detected EERs as described in Whipple et al. (2015). Sequence variants were called with SAMtools mpileup (http://samtools.sourceforge.net). USeq applications RNAEditingPileupParser, RNAEditingScanSeqs, and EnrichedRegionMaker were used to define 50-nt strand-specific windows covered by five or more sequencing reads that contained three or more adenosines edited in >1% of reads. Overlapping edited windows and windows within 1 kb were merged to define EERs. EERs comprised of a single 50-nt window were excluded.

RNA-seq expression and editing analyses

A RefFlat table of *C. elegans* genes (ce10) was downloaded from the University of California at Santa Cruz (UCSC) genome browser (https://genome.ucsc.edu), and EER coordinates were added. Expression was quantified for RefFlat table entries using the USeq application DefinedRegionDifferentialSeq, which also identified DEGs via the R package DESeq2 (Love et al. 2014).

For editing analyses, we determined A-to-G mismatches from .bar files outputted by the USeq application RNAEditingPileupParser. Restricting our analyses to adenosines edited >1% and <99% within EERs, for each RNA-seq replicate, we calculated total A-to-G mismatches (Supplemental Fig. S8) and divided by the total reads covering each base (i.e., if a read covered X edited bases, it was counted X times, once for each base).

Small RNA-seq analyses

Published small RNA-seq data sets, alignment parameters, read filtering, and analyses are described in the Supplemental Material.

qRT–PCR

For each sample, 2 µg of total RNA was reverse-transcribed with the Applied Biosystems high-capacity cDNA reverse transcription kit. qPCR was performed with Roche LightCycler 480 SYBR Green master mix on a Roche LightCycler 480 platform using primers listed in Supplemental Table S2. Transcript abundance was determined by the $\Delta\Delta C_t$ method and normalized to the geometric mean of *Y45F10D.4*, *cdc-42*, *ama-1*, and *pmp-3* mRNAs. Relative values were determined by normalizing to wild type in each individual trial. Outliers were determined by the ROUT method with $Q = 1\%$ and excluded (Motulsky and Brown 2006).

CRISPR/Cas9 genome editing

Protocols for recombinant Cas9 purification and design and synthesis of guide RNAs and homology-directed repair (HDR) templates are described in the Supplemental Material. For each targeting event, 8 mg/mL recombinant Cas9 was complexed for 15 min at 37°C with an equal volume of a 2:2(:2):1:1 mixture of 4 µg/µL target sgRNA 1, 4 µg/µL target sgRNA 2 (0.5 µg/µL HDR template DNA; *efa-6* only), 4 µg/µL *dpy-10* sgRNA, and 40 µM *dpy-10* HDR DNA oligonucleotide to induce the Roller phenotype (Paix et al. 2015). Cas9 complexes were then injected into the distal gonads of 20 wild-type young adults. After 3–4 d, Rol F1s were isolated and screened by PCR (primers in Supplemental Table S3) for the desired genome modifications. All mutations were confirmed by Sanger sequencing (Supplemental Table S4).

Bursting assay

For each sample, eight gravid adults were placed on a seeded plate of nematode growth medium (NGM) to lay eggs for 90 min, after which adults were removed and the eggs laid were counted. Plates were incubated for 5 d at 20°C followed by counting of healthy, burst, or dead adults on the plate. The difference between total number of adults and the original egg count was recorded as number of progeny that did not reach adulthood.

Brood size assay

For each trial, one healthy L4 larva was placed on a seeded NGM plate and allowed to develop for 5 d at 20°C. Hatched larvae were picked off plates and counted before they reached reproductive age.

Accession numbers

RNA-seq data and reanalyses of published small RNA data were deposited in the NCBI Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) under superseries accession number GSE89890.

Acknowledgments

References

Aoki K, Moriguchi H, Yoshioka T, Okawa K, Tabara H. 2007. In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J* **26:** 5007–5019.

Ashe A, Belicard T, Le Pen J, Sarkies P, Frezal L, Lehrbach NJ, Felix MA, Miska EA. 2013. A deletion polymorphism in the *Caenorhabditis elegans* RIG-I homolog disables viral RNA dicing and antiviral immunity. *Elife* **2:** e00994.

Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S, et al. 2008.

PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell* **31:** 67–78.

Billi AC, Fischer SE, Kim JK. 2014. Endogenous RNAi pathways in *C. elegans*. *WormBook* (ed. The *C. elegans* Research Community). *WormBook* doi: 10.1895/wormbook.1.170.1, 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.

Blumenfeld AL, Jose AM. 2016. Reproducible features of small RNAs in *C. elegans* reveal NU RNAs and provide insights into 22G RNAs and 26G RNAs. *RNA* **22:** 184–192.

Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77:** 71–94.

Chen K, Franz CJ, Jiang H, Jiang Y, Wang D. 2017. An evolutionarily conserved transcriptional response to viral infection in *Caenorhabditis* nematodes. *BMC Genomics* **18:** 303.

Cho SW, Lee J, Carroll D, Kim JS, Lee J. 2013. Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics* **195:** 1177–1180.

Cornish-Bowden A. 1985. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res* **13:** 3021–3030.

Deffit SN, Hundley HA. 2016. To edit or not to edit: regulation of ADAR editing specificity and efficiency. *Wiley Interdiscip Rev RNA* **7:** 113–127.

Deffit SN, Yee BA, Manning AC, Rajendren S, Vadlamani P, Wheeler EC, Domissy A, Washburn MC, Yeo GW, Hundley HA. 2017. The *C. elegans* neural editome reveals an ADAR target mRNA required for proper chemotaxis. *Elife* **6:** e28625.

Drake M, Furuta T, Suen KM, Gonzalez G, Liu B, Kalia A, Ladbury JE, Fire AZ, Skeath JB, Arur S. 2014. A requirement for ERK-dependent Dicer phosphorylation in coordinating oocyte-to-embryo transition in *C. elegans*. *Dev Cell* **31:** 614–628.

Ecsedi M, Rausch M, Grosshans H. 2015. The let-7 microRNA directs vulval development through a single target. *Dev Cell* **32:** 335–344.

Emmons SW, Klass MR, Hirsh D. 1979. Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*. *Proc Natl Acad Sci* **76:** 1333–1337.

George CX, Ramaswami G, Li JB, Samuel CE. 2016. Editing of cellular self-RNAs by adenosine deaminase ADAR1 suppresses innate immune stress responses. *J Biol Chem* **291:** 6158–6168.

Goldstein B, Agranat-Tamir L, Light D, Ben-Naim Zgayer O, Fishman A, Lamm AT. 2017. A-to-I RNA editing promotes developmental stage-specific gene and lncRNA expression. *Genome Res* **27:** 462–470.

Guang S, Bochner AF, Pavelec DM, Burkhart KB, Harding S, Lachowiec J, Kennedy S. 2008. An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* **321:** 537–541.

Guo X, Zhang R, Wang J, Ding SW, Lu R. 2013. Homologous RIG-I-like helicase proteins direct RNAi-mediated antiviral immunity in *C. elegans* by distinct mechanisms. *Proc Natl Acad Sci* **110:** 16085–16090.

Hartner JC, Schmittwolf C, Kispert A, Muller AM, Higuchi M, Seeburg PH. 2004. Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *J Biol Chem* **279:** 4894–4902.

Hundley HA, Bass BL. 2010. ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends Biochem Sci* **35:** 377–383.

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.

Reich et al.

Kim DH, Ewbank JJ. 2015. Signaling in the innate immune response. *WormBook* (ed. The *C. elegans* Research Community). *WormBook*, doi: 10.1895/wormbook.1.83.2, http://www.wormbook.org.

Knight SW, Bass BL. 2002. The role of RNA editing by ADARs in RNAi. *Mol Cell* **10:** 809–817.

Liddicoat BJ, Piskol R, Chalk AM, Ramaswami G, Higuchi M, Hartner JC, Li JB, Seeburg PH, Walkley CR. 2015. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349:** 1115–1120.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

Mannion NM, Greenwood SM, Young R, Cox S, Brindle J, Read D, Nellaker C, Vesely C, Ponting CP, McLaughlin PJ, et al. 2014. The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep* **9:** 1482–1494.

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

Motulsky HJ, Brown RE. 2006. Detecting outliers when fitting data with nonlinear regression—a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7:** 123.

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

O'Connell MA, Mannion NM, Keegan LP. 2015. The epitranscriptome and innate immunity. *PLoS Genet* **11:** e1005687.

Paix A, Folkmann A, Rasoloson D, Seydoux G. 2015. High efficiency, homology-directed genome editing in *C. elegans* using CRISPR/Cas9 ribonucleoprotein complexes. *Genetics* **201:** 47–54.

Pak J, Fire A. 2007. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315:** 241–244.

Parry DH, Xu J, Ruvkun G. 2007. A whole-genome RNAi Screen for *C. elegans* miRNA pathway genes. *Curr Biol* **17:** 2013–2022.

Pestal K, Funk CC, Snyder JM, Price ND, Treuting PM, Stetson DB. 2015. Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity* **43:** 933–944.

Phillips CM, Montgomery TA, Breen PC, Ruvkun G. 2012. MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the *C. elegans* germline. *Genes Dev* **26:** 1433–1444.

Rodriguez J, Menet JS, Rosbash M. 2012. Nascent-seq indicates widespread cotranscriptional RNA editing in *Drosophila*. *Mol Cell* **47:** 27–37.

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127:** 1193–1207.

Sarkies P, Ashe A, Le Pen J, McKie MA, Miska EA. 2013. Competition between virus-derived and endogenous small RNAs regulates gene expression in *Caenorhabditis elegans*. *Genome Res* **23:** 1258–1270.

Scadden ADJ, Smith CWJ. 2001. RNAi is antagonized by A→I hyper-editing. *EMBO Rep* **2:** 1107–1111.

Sebastiani P, Montano M, Puca A, Solovieff N, Kojima T, Wang MC, Melista E, Meltzer M, Fischer SE, Andersen S, et al. 2009. RNA editing genes associated with extreme old age in humans and with lifespan in *C. elegans*. *PLoS One* **4:** e8210.

Sijen T, Steiner FA, Thijssen KL, Plasterk RHA. 2007. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315:** 244–247.

Simmer F, Tijsterman M, Parrish S, Koushika SP, Nonet ML, Fire A, Ahringer J, Plasterk RHA. 2002. Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr Biol* **12:** 1317–1319.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102:** 15545–15550.

Thivierge C, Makil N, Flamand M, Vasale JJ, Mello CC, Wohlschlegel J, Conte D Jr, Duchaine TF. 2012. Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. *Nat Struct Mol Biol* **19:** 90–97.

Tonkin L, Bass BL. 2003. Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science* **302:** 1725.

Tonkin LA, Saccomanno L, Morse DP, Brodigan T, Krause M, Bass BL. 2002. RNA editing by ADARs is important for normal behavior in *Caenorhabditis elegans*. *EMBO J* **21:** 6025–6035.

Vasale JJ, Gu W, Thivierge C, Batista PJ, Claycomb JM, Youngman EM, Duchaine TF, Mello CC, Conte D Jr. 2010. Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci* **107:** 3582–3587.

Vitali P, Scadden AD. 2010. Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis. *Nat Struct Mol Biol* **17:** 1043–1050.

Walkley CR, Li JB. 2017. Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. *Genome Biol* **18:** 205.

Warf MB, Johnson WE, Bass BL. 2011. Improved annotation of *C. elegans* microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. *RNA* **17:** 563–577.

Warf MB, Shepherd BA, Johnson WE, Bass BL. 2012. Effects of ADARs on small RNA processing pathways in *C. elegans*. *Genome Res* **22:** 1488–1498.

Washburn MC, Kakaradov B, Sundararaman B, Wheeler E, Hoon S, Yeo GW, Hundley HA. 2014. The dsRBP and inactive editor ADR-1 utilizes dsRNA binding to regulate A-to-I RNA editing across the *C. elegans* transcriptome. *Cell Rep* **6:** 599–607.

Welker NC, Maity TS, Ye X, Aruscavage PJ, Krauchuk AA, Liu Q, Bass BL. 2011. Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol Cell* **41:** 589–599.

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.

Wu D, Lamm AT, Fire AZ. 2011. Competition between ADAR and RNAi pathways for an extensive class of RNA targets. *Nat Struct Mol Biol* **18:** 1094–1101.

Yang H, Vallandingham J, Shiu P, Li H, Hunter CP, Mak HY. 2014. The DEAD box helicase RDE-12 promotes amplification of RNAi in cytoplasmic foci in *C. elegans*. *Curr Biol* **24:** 832–838.

Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC. 2006. Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127:** 747–757.

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.

Zhuang JJ, Banse SA, Hunter CP. 2013. The nuclear argonaute NRDE-3 contributes to transitive RNAi in *Caenorhabditis elegans*. *Genetics* **194:** 117–131.

Reich *et al.*

Supplemental Material: *C. elegans* ADARs antagonize silencing of cellular dsRNAs by the

antiviral RNA interference pathway

Daniel P. Reich, Katarzyna M. Tyc, Brenda L. Bass

Supplemental Inventory:

Supplemental Materials and Methods – Materials and methods related to data presented in

Supplemental Figures, Tables, or Files, or not included in the primary Materials and Methods

section due to lack of space.

Supplementary References – References in Supplemental Material that are not included in

primary References section.

Supplemental Fig. S1 – Model of Dicer-dependent siRNA pathways in *C. elegans*.

Supplemental Fig. S2 – Schematic of the EER detection pipeline.

Supplemental Fig. S3 – Developmental expression of stage-specific EERs.

Supplemental Fig. S4 – EER genomic locations.

Supplemental Fig. S5 – EER repeat content and associated transposon classes.

Supplemental Fig. S6 – EER genomic annotation and predicted structural stabilities.

Supplemental Fig. S7 – EER expression in each developmental stage compared to random

regions.

Supplemental Fig. S8 – Average EER editing events per developmental stage.

Supplemental Fig. S9 – Viability and development of three *adr-1;adr-2* double mutant strains.

Supplemental Fig. S10 – EER abundance compared to ADAR-modulated RNA loci in each

developmental stage.

Supplemental Fig. S11 – Length and 5' nucleotide preferences of 5'P-independent siRNAs

from wildtype and *adr-1(gv6);adr-2(gv42)* mutant animals that map to EERs.

Reich *et al.*

Supplemental Fig. S12 – Relative EER-22G siRNA abundance for EERs and random regions in wildtype, *adr-1(gv6);adr-2(gv42), adr-1(gv6);adr-2(gv42);rde-1(ne219),* and *adr-1(gv6);adr-2(gv42);rde-4(ne299)* strains.

Supplemental Fig. S13 – Expression of eight EAGs in embryos of three independent *adr-1;adr-2* double mutant strains.

Supplemental Fig. S14 – Cas9 targeting schemes used to generate *ΔEER* mutations.

Supplemental Fig. S15 – Additional data on bursting and brood size phenotypes of *adr-1;adr-2;rrf-3* triple mutant lines, not shown in Figure 4.

Supplemental Fig. S16 – Analyses of EER properties for three group of EAGs.

Supplemental Fig. S17 – Differential expression analysis and GSEA of *adr-1(uu49);adr-2(uu28);rrf-3(uu56);rde-1(uu51)* and *adr-1(uu49);adr-2(uu28);rrf-3(uu56);rde-4(uu53)* quadruple mutant embryos compared to *adr-1(uu49);adr-2(uu28);rrf-3(uu56)* triple mutants.

Supplemental Table S1 – Strains used in this study

Supplemental Table S2 – Primers used in qRT-PCR analyses.

Supplemental Table S3 – Primers used for sgRNA synthesis and genotyping of new mutations generated for this study.

Supplemental Table S4 – Mutations generated by CRISPR protocols for this study.

Supplemental File S1 – List of EERs, with genomic properties and siRNA abundance measurements.

Supplemental File S2 – Lists of circRNAs, ADAR-modulated RNA loci, and 21U/piRNA loci that overlap EERs.

Supplemental File S3 – Table describing EER-EAG expression correlation across developmental stages.

Reich *et al.*

Supplemental File S4 – List of EAGs with associated EER-23H siRNA abundance and

expression data.

Supplemental File S5 – Differential gene expression in *adr-1(uu49);adr-2(uu28);rrf-3(uu56)*

triple mutant embryos and related strains.

Supplemental File S6 – EAG enriched GO categories.

Supplemental File S7 – Gene set enrichment analysis comparing *adr-1(uu49);adr-2(uu28);rrf-3(uu56);rde-1(uu51)* and *adr-1(uu49);adr-2(uu28);rrf-3(uu56);rde-4(uu53)* quadruple mutants

to *adr-1(uu49);adr-2(uu28);rrf-3(uu56)* triple mutants.

Supplemental Materials and Methods

EER annotation

Tables of annotated *C. elegans* (ce10/WS220) protein coding genes, introns, 3' UTRs, 5'

UTRs, ncRNAs, and pseudogenes and their genomic coordinates were downloaded from

UCSC Genome Browser in .bed format (https://genome.ucsc.edu/). We merged ncRNA and

pseudogene annotations into a single "ncRNA" annotation. The bedtools2

(https://github.com/arq5x/bedtools2) application annotateBed defined the number of

overlapping bases shared between EERs and annotated features. To annotate EERs within or

further than 1kb from genes, we combined annotations of protein-coding genes and ncRNAs,

and then extended the start and stop coordinates 1000 nt away from gene boundaries.

EER overlap with other annotated regions

The USeq application IntersectRegions was used to overlap EERs with .bed files for *C.

elegans* circRNAs (circBase.org) or ADAR-modulated RNA loci (from Wu et al. 2011) and

calculate enrichment significance by X^2 approximation. A .bed file of genomic regions covered

Reich *et al.*

by at least 5 reads in combined developmental RNAseq datasets (i.e. genomic space with sufficient coverage to define EERs) was used with parameters "-r" and "-n 10000" to make 10,000 randomized expressed regions used to approximate the likelihood of random intersection with circRNAs or ADAR-modulated RNA loci.

Annotation of EER-associated genes (EAGs)

EAGs were defined as genes containing one or more EERs within the gene on the same strand or else the closest gene within 1kb of an EER on the same strand. Using gene tables for protein coding genes and pseudogenes downloaded from UCSC (see Methods), we used the bedtools application intersectBed to identify genes overlapping EERs, and the Useq application FindNeighboringGenes, to find the closest gene to EERs within 1kb.

Small RNAseq data preparation and analysis

Mixed-stage 5'P-dependent small RNAseq datasets were downloaded from Gene Expression Omnibus record GSE28888. Small RNAseq 5'P-independent datasets of wildtype, *adr-1(gv6);adr-2(gv42), adr-1(gv6);adr-2(gv42);rde-1(ne219),* and *adr-1(gv6);adr-2(gv42);rde-4(ne299)* embryos and L4 larvae from Wu et al. (2011) were provided by Diane Wu and Andrew Fire. Small RNAseq reads were aligned to the *C. elegans* ce10/WS220 genome with Novoalign, parameters: -o SAM -a ATCTCGTATGCCGTCTTCTGCTTG -r All). We used the USeq application SamTranscriptomeParser (parameters -n 1000000 –a 30) to parse alignments into bam files. For uniquely mapping reads, we parsed reads with the samtools application view (parameters –F 4 –q 10). Small RNA libraries were normalized to the total number of mapped reads in each alignment file. For Supplemental Fig. S11, reads mapping sense or antisense to EERs were extracted with the bedtools2 application intersectBed

Reich *et al.*

(https://github.com/arq5x/bedtools2) and analyzed by custom bash scripts. Repetitively

mapped reads were used for relative siRNA abundance plots shown in Supplemental Fig. S12.

Viability assays

To quantify development of *adr-1;adr-2* mutants, we modified the bursting assay described in

Materials and Methods as follows. After adults were removed and eggs laid on Day 0 were

counted, embryos were allowed to hatch and mature for 24 hrs, after which unhatched eggs

were counted. 64-70 hrs after egg lay, worms on the plate were scored and counted as L3-

adults or L1-L2 larvae. If the sum of unhatched eggs, L1-L2 larvae, and L3-adults was less

than the initial egg count, the difference was added to the subset of unhatched eggs. All steps

and incubations were performed at 20°C.

Recombinant Cas9 purification

To express Cas9 in *E. coli*, we used a human codon-optimized *Streptococcus pyogenes* Cas9

gene with N-terminal HA-SV40, NLS, and TEV protease site cloned into pET28b with N- and

C-terminal 6xHis tags (a generous gift of Dr. Jin-Soo Kim, Seoul National University). pET28b-

Cas9-N3T (Kan[R]) was transformed into BL21(DE3) cells. A 500 mL culture of LB-Kan (50 mg/L

Kan) was inoculated with a 5 mL overnight culture and grown at 37°C to an OD of 0.4-0.5. We

added IPTG to 0.5 mM and induced Cas9 expression for 4 hrs at 25°C.

Cells were pelleted, resuspended in lysis buffer (50 mM NaH_2PO_4 pH 8.0, 10 mM

imidazole, 300 mM NaCl, 10% glycerol, 1 mM β-mercaptoethanol), and lysed by

homogenization followed by sonication. Clarified lysate was batch bound to Qiagen Ni-NTA

agarose resin for 45 min at 4°C. Resin was washed 3x in 1 M NaCl, 1x in 300 mM NaCl, and

protein was eluted with 125 mM imidazole. Protein was dialyzed into 20 mM HEPES pH 7.5,

Reich *et al.*

150 mM KCl, 1 mM DTT, 10% glycerol, and concentrated to 8 mg/mL. Protein was stored long

term in 25% glycerol at -80°C.

sgRNA design and synthesis

We used the Broad Institute sgRNA Design Tool (http://portals.broadinstitute.org/gpp/public

/analysis-tools/sgrna-design-v1) to identify *S. pyogenes* Cas9 target sites in regions of interest.

We chose guide sequences with a 5'G (to facilitate T7 RNA polymerase *in vitro* transcription)

scoring >0.1 and BLASTed them against the *C. elegans* genome, eliminating candidates with

<4 mismatches that contained the NGG protospacer adjacent motif needed for cleavage.

For each CRISPR/Cas9 mutation, we designed two guide RNAs. For whole-gene

disruptions, we selected sgRNAs situated ~1 kb or more apart to induce large deletions. To

delete the EERs in *ccb-1* and *egl-8*, we chose guide RNAs to target Cas9 to sequences

flanking each EER within the intron far enough from 5' and 3' splice sites and consensus

splicing signals to prevent disruption. The EER-containing intron of *efa-6* was too repetitive to

delete without likely off-target mutations, so we directed Cas9 to cleave within the *efa-6* exonic

sequences flanking the intron, and provided a single-stranded DNA template to replace the

entire intron with a 60 nt synthetic intron lacking the predicted double-stranded structure

(Materials and Methods; Supplemental Fig. S14C; Supplemental Table S3).

Primers of the form TAATACGACTCACTATA-N_{19-22}-GTTTTAGAGCTAGAAATAG,

where N_{19-22} represented the guide target sequence, were used with the reverse primer

CRISPR_sgRNA_R1 (Supplemental Table S3) to amplify the sgRNA transcription template

from the plasmid DR274 (Hwang et al. 2013). *In vitro* transcription reactions were incubated

overnight at 37°C in 100 μl containing 10 μg PCR template, 20 U T7 RNA polymerase, 3 mM

Reich *et al.*

ATP, 3 mM UTP, 3 mM CTP, 3 mM GTP, 40 mM Tris pH 7.8, 2 mM spermidine, 15 mM DTT,

and 15 mM MgCl$_2$. Reactions were DNase-treated 1 hr at 37°C, and RNA was extracted in

phenol-chloroform, ethanol precipitated, and resuspended in 10 mM Tris, 1 mM EDTA to a

concentration of ~4 µg/µl.

Imaging GFP::NRDE-3 localization

L3 worms of each genotype were anesthetized in M9 buffer containing 5 mM levamisole and

mounted on 2% agarose pads. Images in Figure 4A and Supplemental Fig. S15A were taken

using a QImaging Retiga 2000R camera on a fluorescent Zeiss Axioskop 2 MOT microscope.

Gene Ontology (GO) analysis

A background list of genes was determined by calculating all genes overlapping a .bed file of

genomic regions covered by at least 5 reads in our developmental RNAseq experiment (i.e. all

genes with sufficient coverage to define an EER). The lists of background genes and EAGs

(see Supplemental File S3) were provided to the GOMiner web interface

(https://discover.nci.nih.gov/gominer/htgm.jsp) to calculate enriched GO categories with a

False Discovery Rate < 0.05 (Zeeberg et al. 2003; Zeeberg et al. 2005).

Gene set enrichment analysis (GSEA)

Gene sets for 877 biological process high quality GO annotations were downloaded from

http://www.go2msig.org/cgi-bin/prebuilt.cgi?taxid=6239, to which were added gene sets of 298

Orsay virus-induced genes and 231 EAGs significantly downregulated in *adr-1;adr-2;rrf-3*

Reich *et al.*

mutants relative to wildtype (Fig. 5B). GSEA software was downloaded from http://software.broadinstitute.org/gsea/index.jsp. Poly(A)+ RNAseq rlog-normalized expression values outputted by DESeq2 were used for analysis. Default parameters were used for GSEA, with the exception that we used permutation type "gene_sets," since "phenotype" permutation recommends seven samples per condition. Note that we also included the full set of 965 EAGs, but GSEA was unable to calculate normalized enrichment scores for a gene set this large.

Supplementary References

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**: 227-229.

Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S et al. 2003. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* **4**: R28.

Zeeberg BR, Qin H, Narasimhan S, Sunshine M, Cao H, Kane DW, Reimers M, Stephens RM, Bryant D, Burt SK et al. 2005. High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics* **6**: 168.

Dicer-dependent siRNA pathways
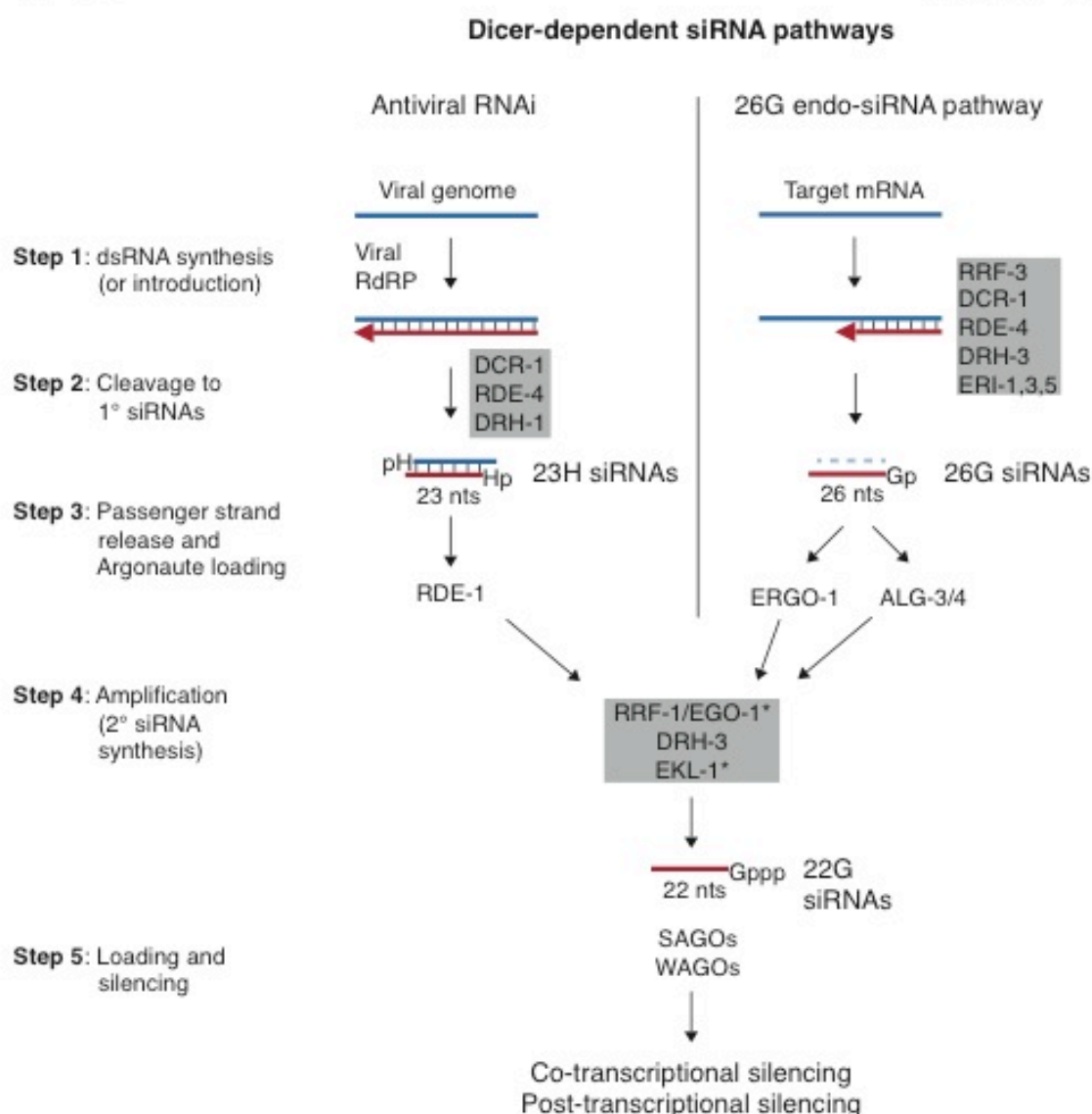
Supplemental Figure S1. Model of *C. elegans* Dicer-dependent siRNA pathways. *LEFT*: The antiviral RNAi pathway (Ashe et al. 2013; Guo et al. 2013) acts on viral dsRNA produced during viral genome replication by a viral RNA-dependent RNA polymerase (RdRP; sense strand, blue; antisense strand, red). In Step 2 of this pathway, DCR-1, the dsRNA binding protein RDE-4, and the Dicer-related helicase DRH-1, act together to produce double-stranded siRNAs with 5' phosphates and an overhang at the 3' end, as expected for a Dicer cleavage product. Each strand of the viral siRNA is 23 nts with a 5' nucleotide that is more often A, C, or U than G (23H siRNAs, see main text). Primary 23H siRNAs are loaded into the Argonaute RDE-1 to promote the production of secondary siRNAs (Step 4) by the *C. elegans* RdRP RRF-1. Secondary siRNAs are predominantly 22 nt long, contain a 5' triphosphorylated guanosine (22G siRNAs), and are antisense to target RNAs. Factors required in Steps 4 and 5 of the antiviral pathway are inferred to be the same as for the 26G pathway, but roles for some
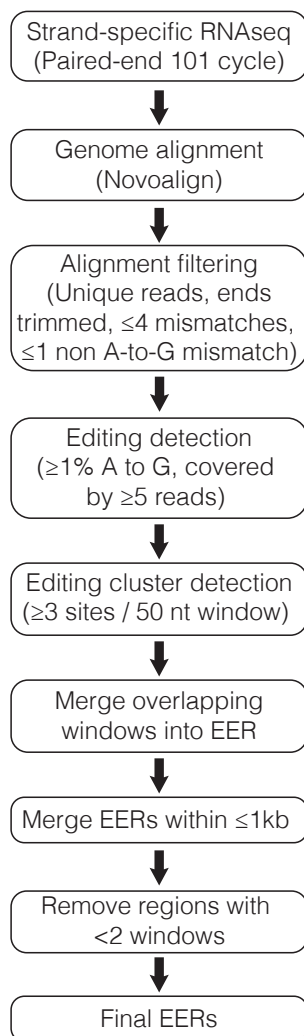
factors have not been tested (noted by asterisks). *RIGHT*: In a poorly understood manner, certain mRNAs are targeted for silencing by the 26G endogenous siRNA pathway (Vasale et al. 2010; Thivierge et al. 2012; Billi et al. 2014). In contrast to the antiviral pathway, Steps 1 and 2 likely occur concomitantly (Blumenfeld and Jose, 2016), with synthesis of dsRNA by the RdRP RRF-3 followed closely by its cleavage by DCR-1. RDE-4, DRH-3 and ERI-1, 3 and 5 are also required for Steps 1 and 2 to produce 26 nt siRNAs, which have a 5' monophosphorylated guanosine (26G siRNAs) and are primarily antisense to the mRNA. 26G siRNAs are loaded into tissue-specific Argonautes (ERGO-1 or ALG-3 or ALG-4, see main text) that promote production of secondary 22G siRNAs, by a complex containing the RdRP RRF-1 or EGO-1 (in the soma and germline, respectively). For both pathways, loading and silencing (Step 5) involves additional Argonautes (WAGO and SAGO).
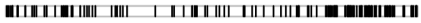
Fig

Strand-specific RNAseq
(Paired-end 101 cycle)

↓

Genome alignment
(Novoalign)

↓

Alignment filtering
(Unique reads, ends
trimmed, ≤4 mismatches,
≤1 non A-to-G mismatch)

↓

Editing detection
(≥1% A to G, covered
by ≥5 reads)

↓

Editing cluster detection
(≥3 sites / 50 nt window)

↓

Merge overlapping
windows into EER

↓

Merge EERs within ≤1kb
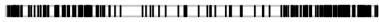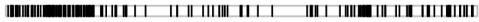
↓

Remove regions with
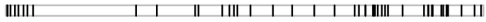<2 windows

↓

Final EERs

Supplemental Figure S2. Flowchart outlining the bioinformatics pipeline used to identify Editing-Enriched Regions, adapted from Whipple et al. (2015). Additional details are provided in Materials and Methods.
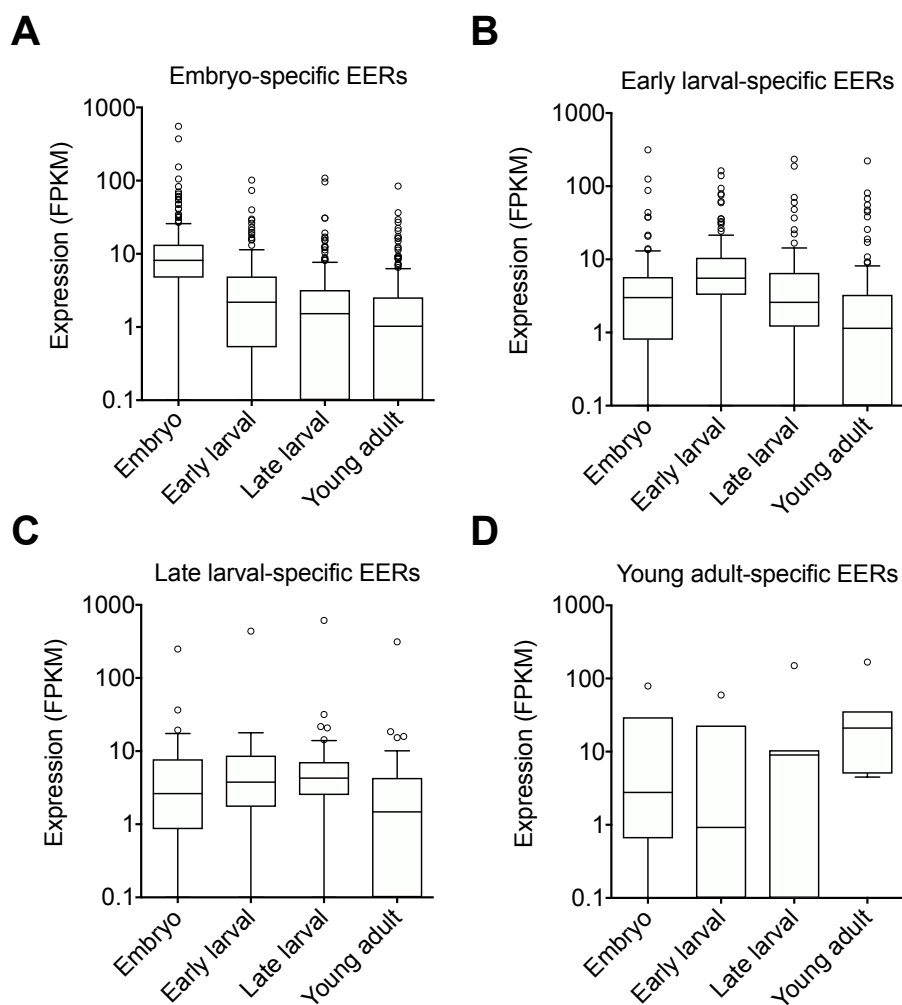
**Supplemental Figure S3.** Tukey boxplots show expression in each developmental stage of (*A*) embryo-specific EERs, (*B*) early larval-specific EERs, (*C*) late larval-specific EERs, and (*D*) young adult-specific EERs. For all panels, expression values are calculated from input RNAseq samples only. Note for all plots that quartile boundaries with value of 0 are set at 0.1 FPKM due to the logarithmic scale.

(#EERs)

Chr I	▌▌▌▌▌▌▌▌▌▌▌▌▌▌ ▐ ▌▐▐▐▐▐ ▐▐▌ ▐ ▌▐▐▌ ▐ ▐▐▐▐ ▐▌ ▐▌▌▌▐▐▐▌	443
Chr II	▌▐▐▌▐ ▌▐▌ ▐▐▌▌ ▐▌▌ ▐ ▌ ▐▌▐ ▐▌▌ ▐▌▌▌▐▌▌ ▐▌ ▌▌ ▌▌▐▐▌▌▌ ▐ ▐▌	191
Chr III	▌▌▌ ▐▐▐▌ ▌▌▐▌▌▌ ▐▐▐▌▌ ▐▐▌▌ ▐ ▌ ▌▐ ▐ ▐ ▌▐▐ ▌▌▌▌▌▐▌▌▌▌▌▌▌	322
Chr IV	▌▌▌▌▌▌▌▌▌▌▌▌▌ ▐▌▐▐▐ ▌ ▐▌ ▐ ▐▌▐▐▐ ▐▐ ▐ ▐▐▐ ▐ ▐▐▌▐ ▐▌ ▐▌ ▐▐▐▐▐▐▌	273
Chr V	▌▐▐▐▌▐▐▌ ▐ ▌▐ ▐ ▐▌▐▌▌▐ ▌▐ ▐▐▌ ▐ ▐▌▌ ▐▐▌▐ ▐ ▐ ▐▌▐▐▌ ▐▐▐▌▐ ▐ ▐ ▐▐▌ ▌▐▐▌▌▌▐▐▐▐ ▌	244
Chr X	▌▌▌▌▌▌ ▐ ▐ ▐ ▐▌ ▐▐▐▌ ▐ ▐ ▐ ▐ ▐ ▐▐▐▐▌ ▐ ▐▐ ▐▐ ▐▐▐	50

Supplemental Figure S4. Genomic locations of all 1523 EERs. Vertical black lines indicate EER locations, while the total number of EERs on each chromosome are shown to the right.

A

EER repeat content

B

EER transposon repeats
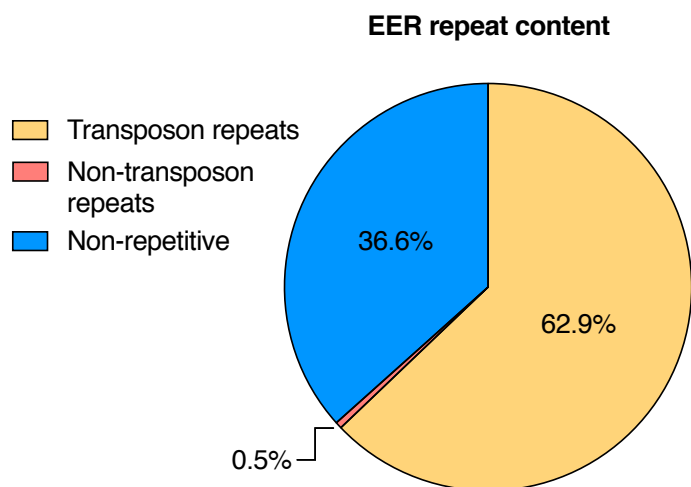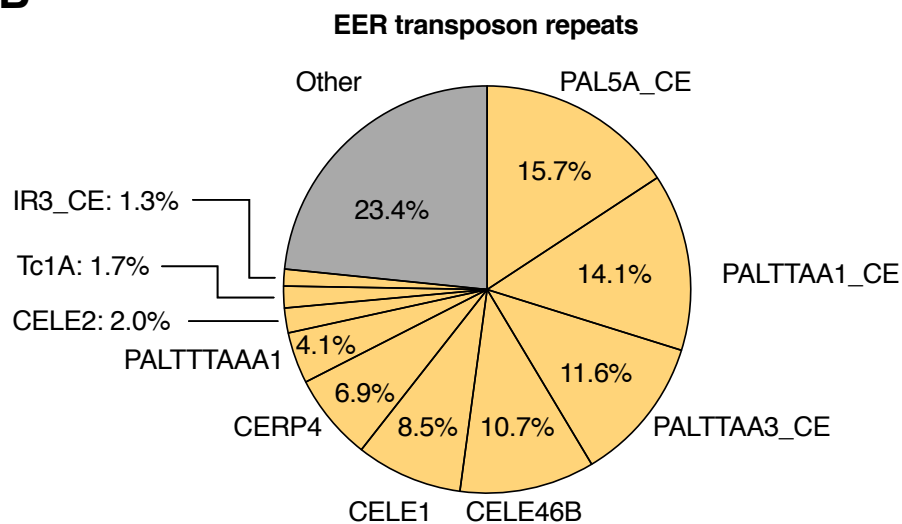
Supplemental Figure S5. EER repeat content. (*A*) Classification of EER sequences by the % of EER nucleotides that overlapped annotated RepeatMasker repetitive sequences. (*B*) Classification of the ten most abundant transposon classes represented in EER sequences.
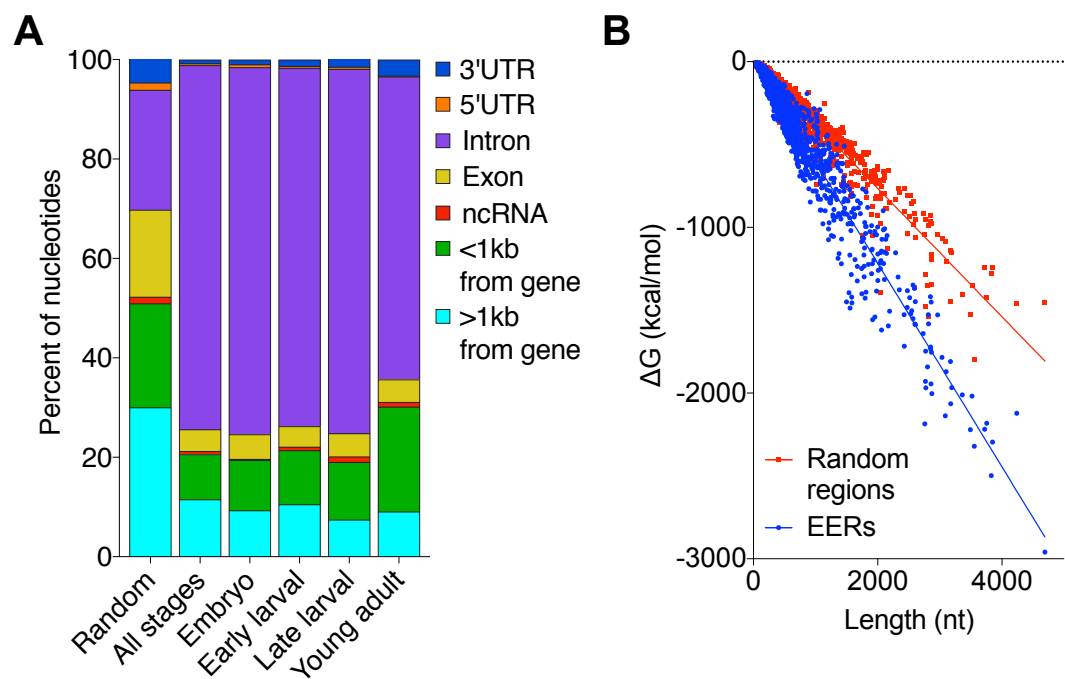
**Supplemental Figure S6.** EER genomic annotation and predicted structure stabilities. (*A*) The percent of nucleotides from random regions or EERs from each set of developmental stages that overlap specified genomic annotations. (*B*) UNAFold-predicted folding free energies (ΔG) of sequences from EERs (blue) or length-matched expressed random regions (red) plotted against length. A slope test on the linear regressions of each set of regions shows they are significantly different ($p<0.0001$).
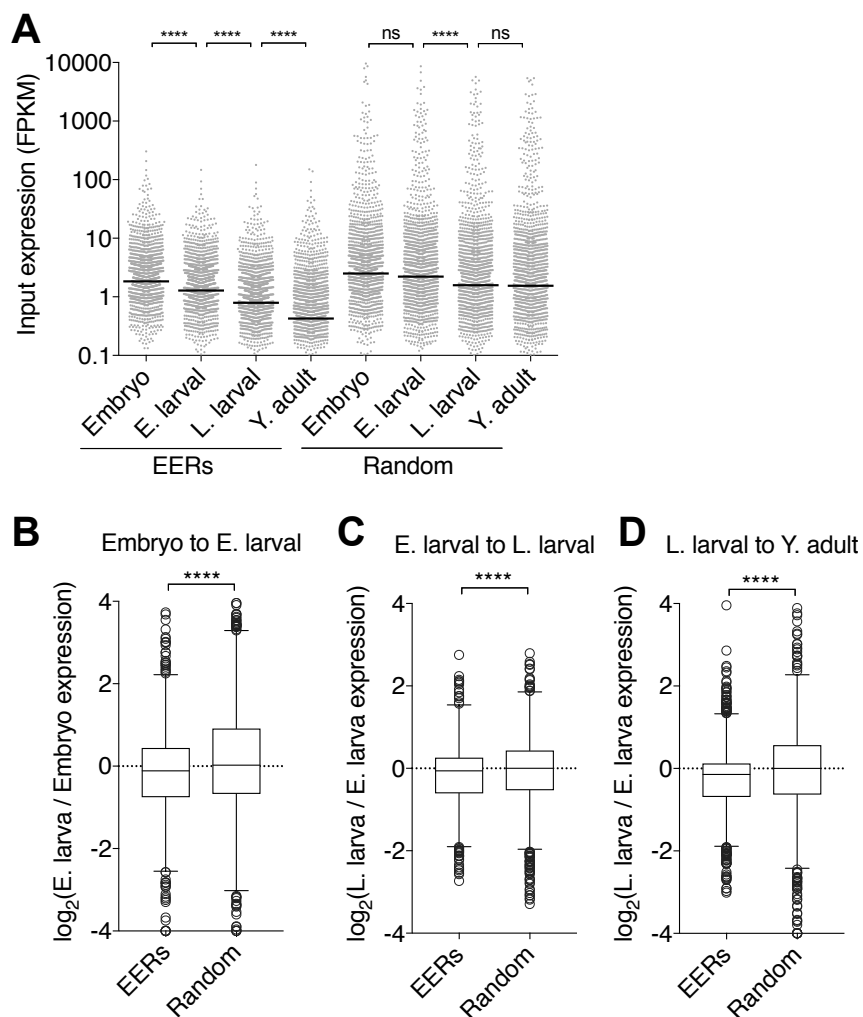
Supplemental Figure S7. EERs display expression patterns distinct from random regions. (*A*) Distributions of EER abundance and length-matched expressed random region (Random) abundance in each stage of development measured. Solid black lines report median abundance. (*B-D*) Tukey boxplots show log$_2$(abundance fold-change) of all EERs and random regions in the transitions between each developmental stage and the subsequent stage. ns: not significant; ****: $p < 0.0001$; Mann-Whitney U test.
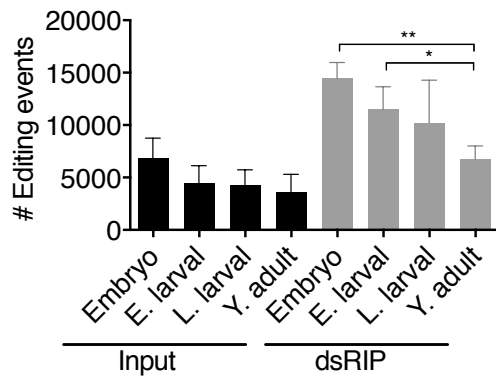
Supplemental Figure S8. EER editing events per stage. Plot shows the average number of A-to-G changes observed in reads covering all EER editing sites (edited >1% and <99%) in each developmental stage for each treatment (three biological replicates each). Error bars reflect standard deviation of three RNAseq replicates; *: $p<0.05$, **: $p<0.01$; Student's T-test.
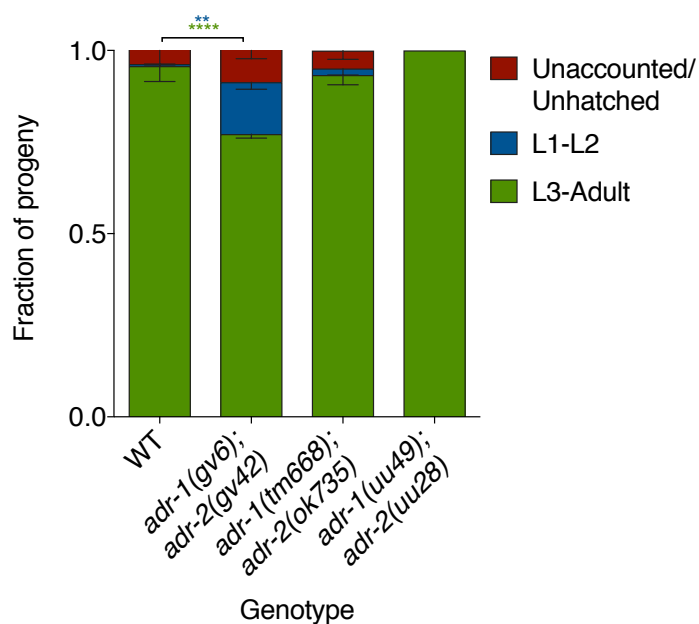
Supplemental Figure S9. Viability and development of progeny in three independent *adr-1;adr-2* mutant strains. Shown are the fraction of embryos laid on Day 0 that were found on Day 3 as L3-adults (green), L1-L2 larvae (blue), or which were lost or failed to hatch (red). Error bars represent standard deviation over n=3 assays; **: $p<0.01$; ****: $p<0.0001$, significance determined by two-way ANOVA with Tukey's multiple comparisons correction, asterisk colors indicate the categories compared between genotypes.
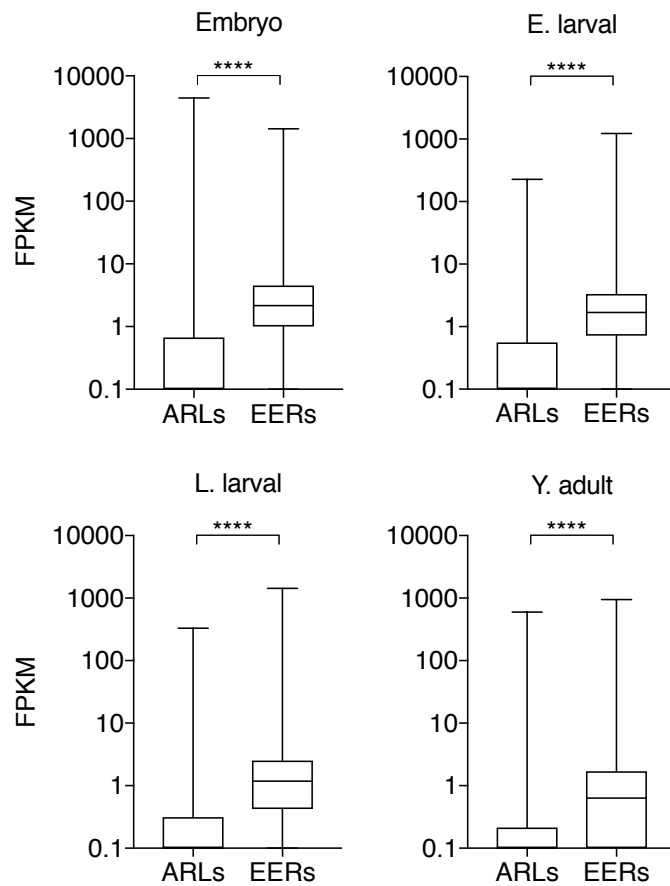
Supplemental Figure S10. EERs are more abundant than ADAR-modulated RNA loci in wildtype animals. Tukey boxplots show distributions of ADAR-modulated RNA loci (ARL) and EER abundance (FPKM) in input RNAseq samples from each developmental stage (E: Early, L: Late, Y: Young). ****: $p < 0.0001$; Mann-Whitney U test.
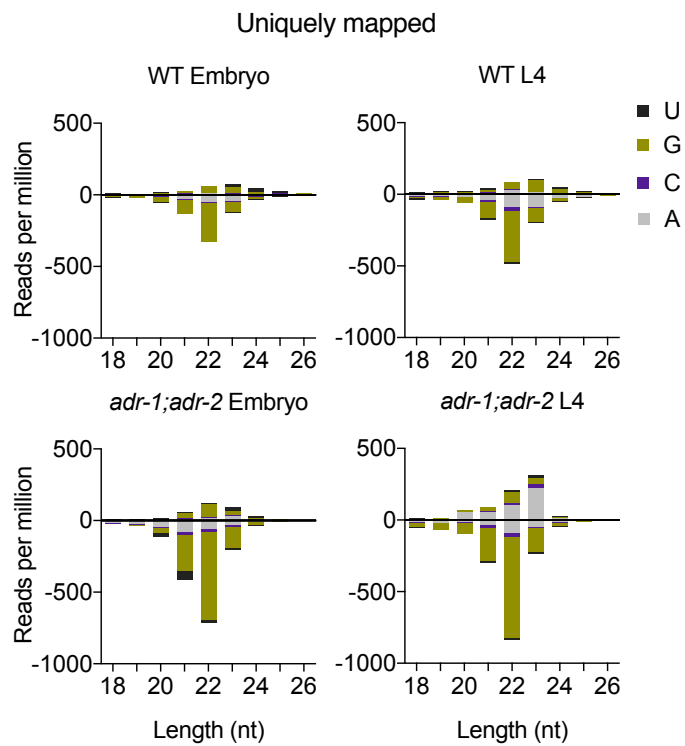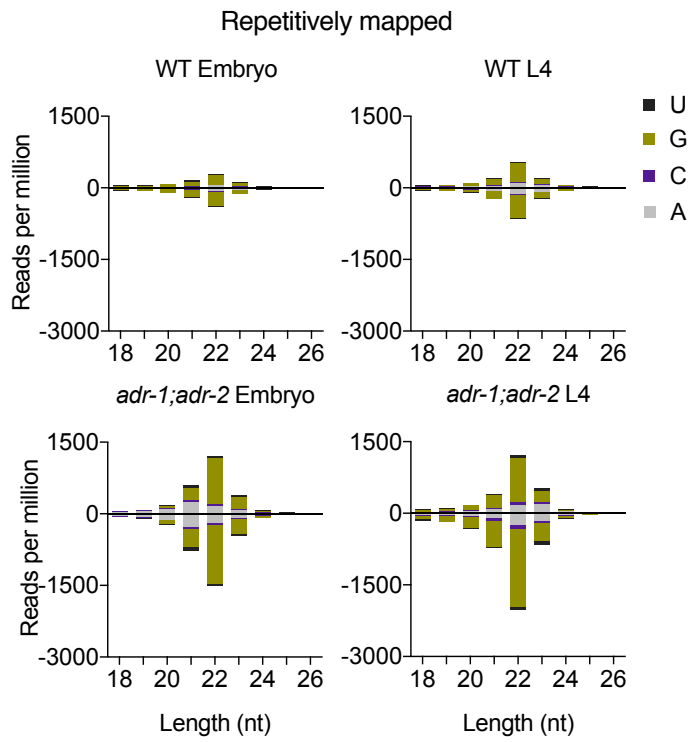
Supplemental Figure S11. EER-22G siRNAs in wildtype and *adr-1(gv6);adr-2(gv42)* mutant animals. Sense (positive) and antisense (negative) EER-mapped siRNAs in wildtype and *adr-1(gv6);adr-2(gv42)* embryo and L4 stage 5'P-independent small RNAseq samples from Wu et al. (2011) are plotted by length (nt) and 5' nt (see color key in top right panels). Small RNA reads were aligned either to allow reads to map in multiple locations (repetitively mapped) or to only include reads that mapped to a single location (uniquely mapped).

5'P-independent siRNAs antisense to EERs
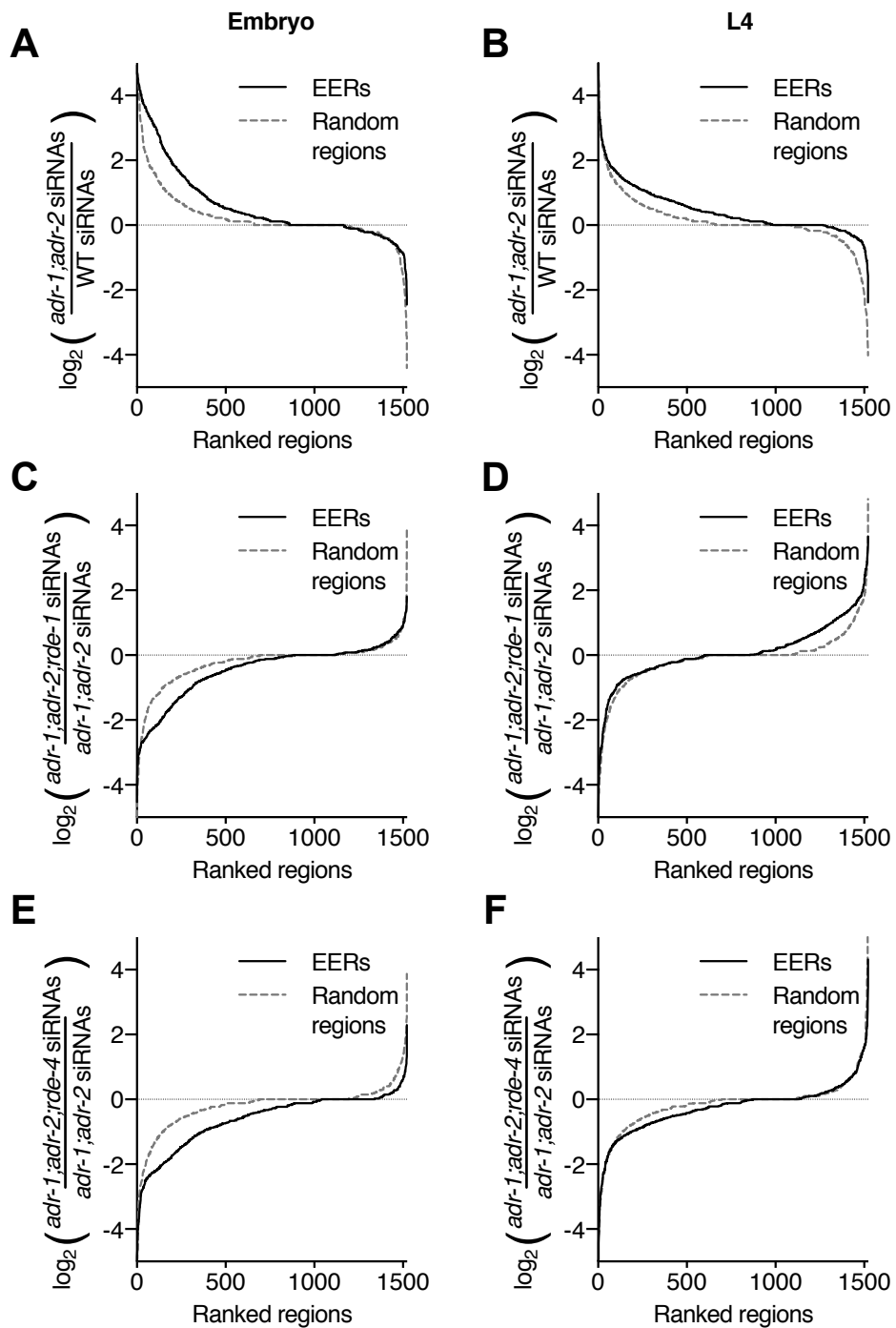
Supplemental Figure S12. EER-22G siRNAs require *rde-1* and *rde-4* for embryonic accumulation in *adr-1;adr-2* double mutant animals. In all panels, abundance of siRNAs antisense to 1523 EERs (black line) or control random regions (dotted grey line) are plotted as $\log_2$ ratio of siRNA reads in one genotype over siRNA reads in the control genotype. For top panels, EER-22G siRNA enrichment in *adr-1(gv6);adr-2(gv42)* double mutants are plotted relative to wildtype in (*A*) embryo or (*B*) L4 larval stages. In middle panels, EER-22G siRNAs in *adr-1(gv6);adr-2(gv42);rde-1(ne219)* triple mutants are plotted relative to *adr-1(gv6);adr-2(gv42)* double mutants in (*C*) embryo or (*D*) L4 larval stages. In the two bottom panels, EER-22G siRNA abundance in *adr-1(gv6);adr-2(gv42);rde-4(ne299)* triple mutants is plotted relative to *adr-1(gv6);adr-2(gv42)* double mutants in (*E*) embryo or (*F*) L4 larval stages.
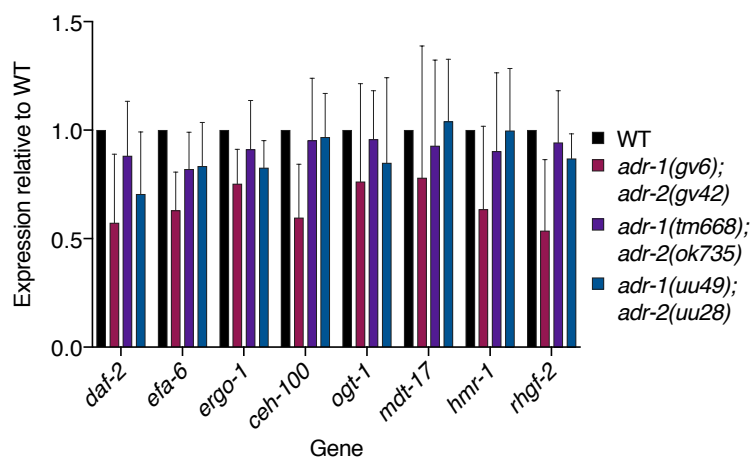
Supplemental Figure S13. EAG expression is decreased in *adr-1;adr-2* double mutant strains. Plots show EAG expression, determined by qRT-PCR, in three independent sets of *adr-1;adr-2* deletions. Each EAG was normalized to the geometric mean of four control genes (*Y45F10D.4*, *cdc-42*, *pmp-3*, and *ama-1*). Plot shows mean expression relative to the expression of the wildtype strain in each experiment (n=8); error bars represent SD.
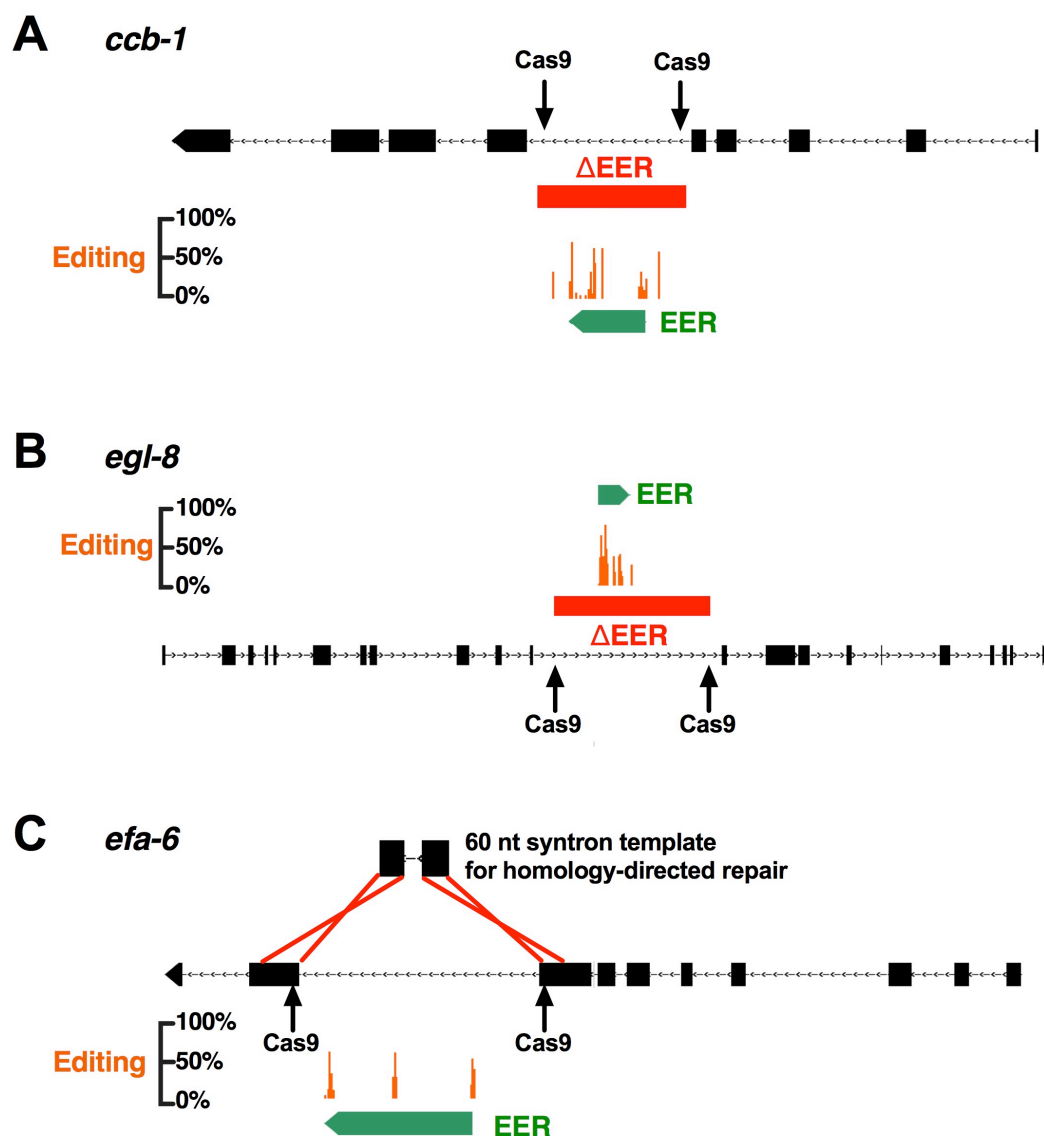
Supplemental Figure S14. Cas9 targeting schemes used to generate three *ΔEER* mutations. Genome browser views of (*A*) *ccb-1,* (*B*) *egl-8*, and (*C*) *efa-6* are shown with the location of EERs (green) in each gene and the editing frequency (orange) of edited adenosines in all developmental stages combined. Regions targeted for cleavage by Cas9 are marked with arrows. For (*A*) and (*B*), deleted sequences are shown in red. A schematic of the HDR template and recombination pattern (red crossed lines) are shown for *efa-6(ΔEER)* in (*C*). See Supplemental Table S4 for specific information on each mutation.
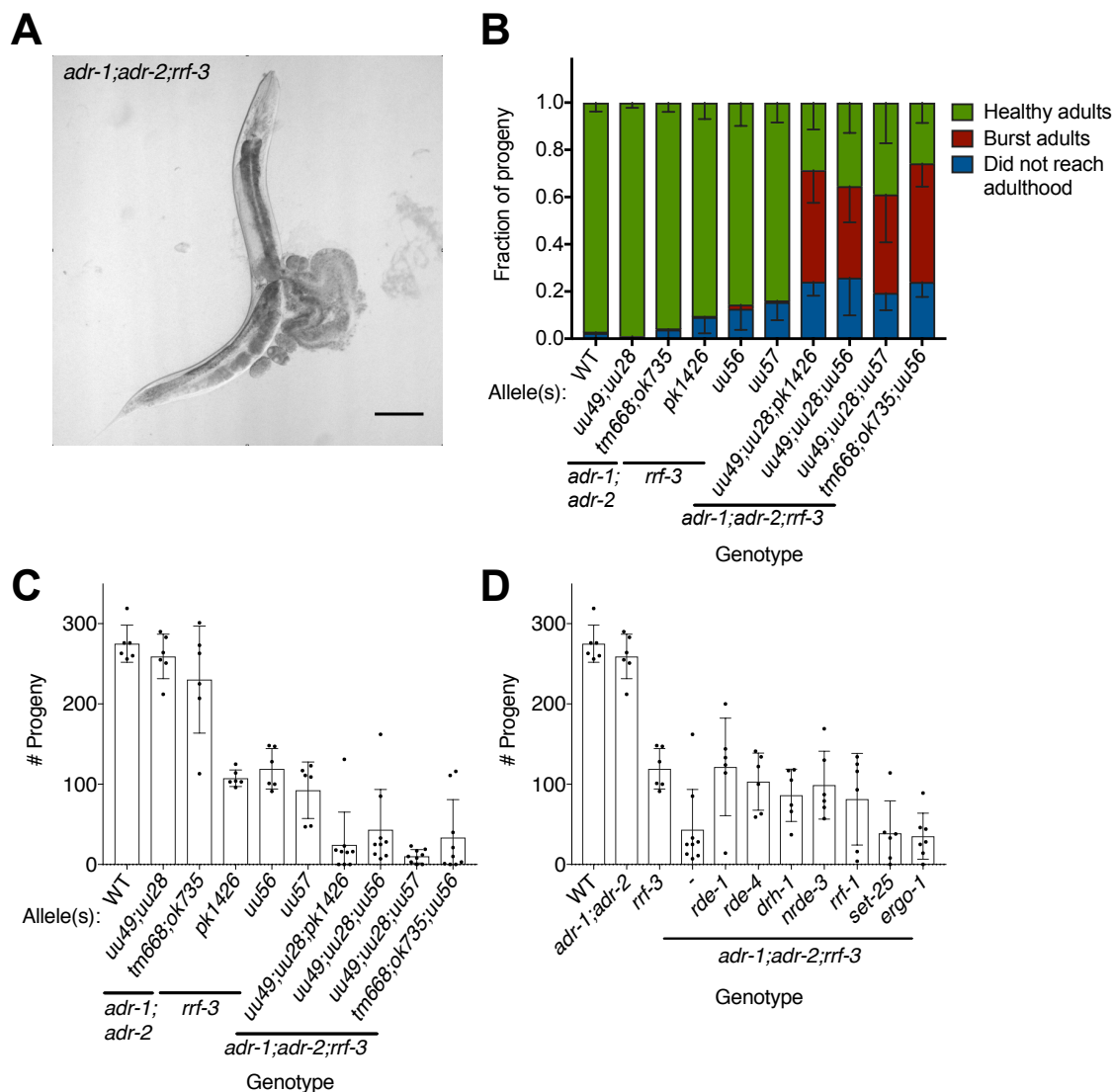
Supplemental Figure S15. Synthetic phenotypes of *adr-1;adr-2;rrf-3* mutants are consistent across multiple alleles and two assays. (*A*) A representative burst *adr-1(uu49);adr-2(uu28);rrf-3(uu56)* adult. Scale bar = 100 µm. (*B*) Bursting assay of *adr-1;adr-2;rrf-3* triple mutant strains and parent strains using multiple independently-derived deletions of each gene. Allele designations for each gene are listed in the order of the genes listed below them (i.e. the first listed allele corresponds to the first listed gene below it). See Supplemental Table S4 for specific information on *uu* alleles; n ≥ 6 assays. (*C*) Brood sizes of the lines used in (*B*); n ≥ 6 assays. (*D*) Brood sizes of *adr-1;adr-2;rrf-3* rescue strains characterized in Fig. 5D; n ≥ 6 assays. Individual broods are shown as dots. In all panels, error bars represent SD.
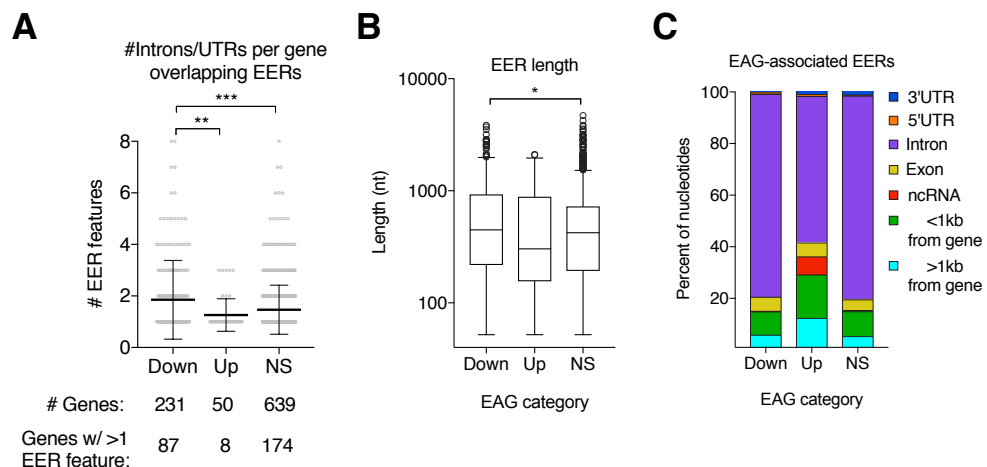
Supplemental Figure S16. Characteristics of EERs associated with misregulated EAGs. (*A*) The number of introns or UTRs that overlap EERs in each gene (EER features; gray circles) are plotted for EAG categories shown in Fig. 5C-E. Note that we counted strand-specific unannotated regions <1kb from a gene as a UTR, but only if an annotated UTR did not also overlap an EER. Horizontal lines plot the average #EER features per gene and error bars show SD. Mann-Whitney U test: **, $p < 0.01$; ***, $p < 0.001$. (*B*) Tukey boxplots show the lengths of EERs associated with EAGs in each category. Mann-Whitney U test: *, $p < 0.05$. (*C*) Annotation of EER nucleotides associated with EAGs in each category.
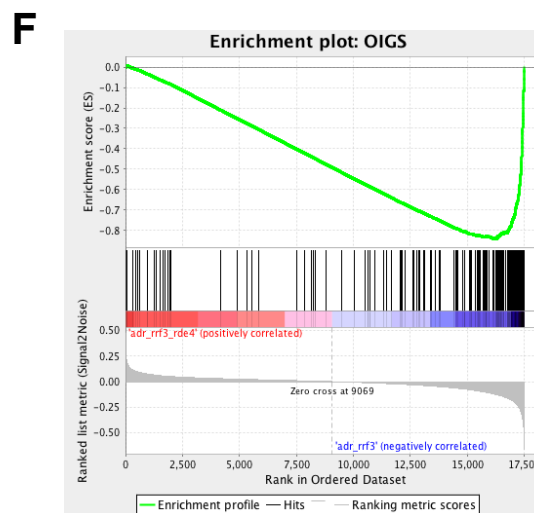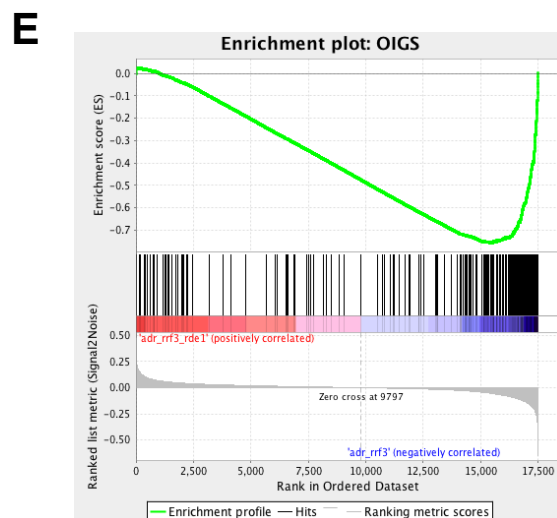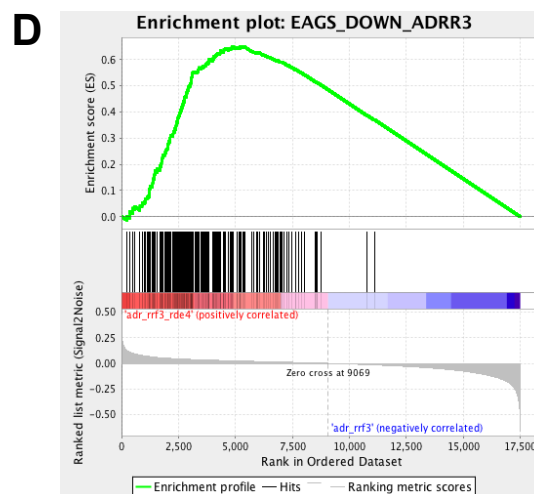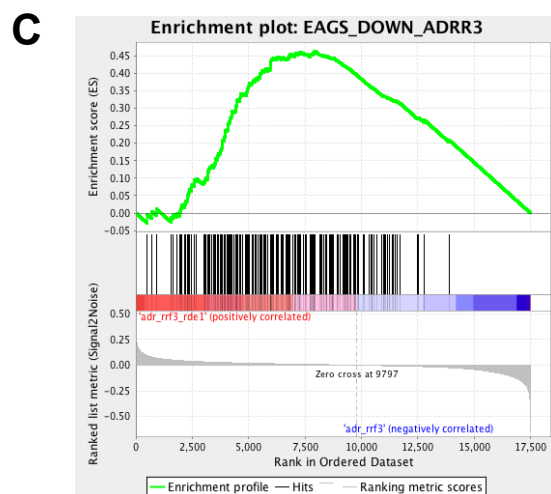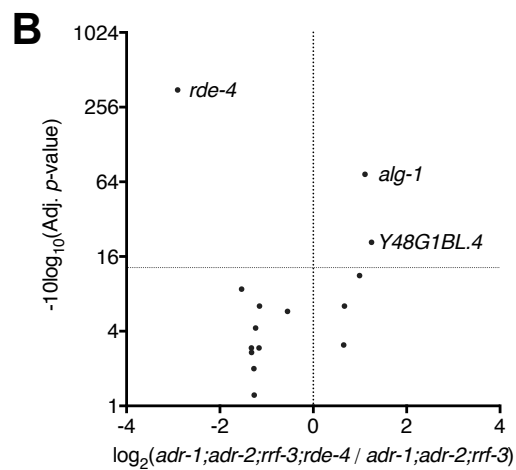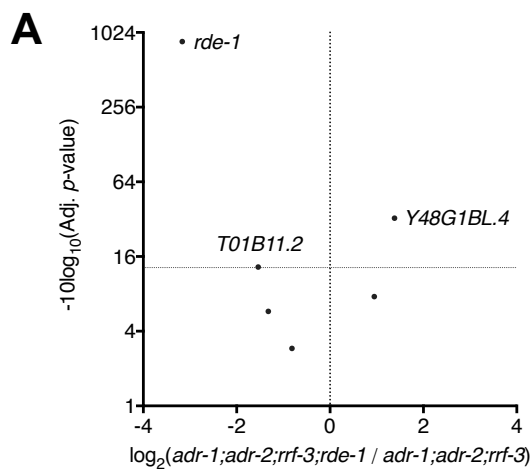
Supplemental Figure S17. EAG and Orsay virus-induced gene misregulation in *adr-1(uu49);adr-2(uu28);rrf-3(uu56)* triple mutants partly depends on *rde-1* and *rde-4.* Volcano plots show the fold change and significance of gene expression changes in (*A*) *adr-1(uu49);adr-2(uu28);rrf-3(uu56);rde-1(uu51)* quadruple mutant embryos and (*B*) *adr-1(uu49);adr-2(uu28);rrf-3(uu56);rde-4(uu53)* quadruple mutant embryos relative to *adr-1(uu49);adr-2(uu28);rrf-3(uu56)* triple mutant embryos. Since the vast majority of gene expression changes had -10log$_{10}$(adjusted p-value) close to 0, most genes are not shown. Horizontal dotted line represents adjusted p-value cutoff of 0.05, where genes above this line were considered significantly differentially expressed. GSEA enrichment plots for 231 EAGs downregulated in *adr-1;adr-2;rrf-3* to wildtype show expression enrichment in (*C*) *adr-1;adr-2;rrf-3;rde-1* and (*D*) *adr-1;adr-2;rrf-3;rde-4* embryos compared to *adr-1;adr-2;rrf-3.* Vertical black lines (center of each plot) indicate the position of EAGs in a gene list sorted highest to lowest by expression in *adr-1;adr-2;rrf-3;rde-1* (or *adr-1;adr-2;rrf-3;rde-4*) relative to expression in *adr-1;adr-2;rrf-3*, with the most upregulated genes to the left and most downregulated genes to the right. Green traces show the enrichment score, described in Subramanian *et al.* (2005), calculated for the EAG gene set. (*E-F*) Orsay virus-induced gene enrichment plots are shown for the same genotype comparisons as in (*C-D*).

Supplemental Table S1. Strains used in this study. Citations are listed in the References section of the main text.

Strain	Genotype	Citation
Bristol N2	Wildtype	(Brenner 1974)
BB4	*adr-1(gv6);adr-2(gv42)*	(Tonkin et al. 2002)
BB21	*adr-1(tm668);adr-2(ok735)*	(Hundley et al. 2008)
BB204	*ccb-1(uu35)*	This study
BB234	*efa-6(uu46; Δefa-6 intron 8 + 60 nt syntron)*	This study
BB235	*egl-8(uu47)*	This study
BB239	*adr-1(uu49);adr-2(uu28)*	This study
BB242	*adr-1(uu49);adr-2(uu28);rde-1(uu51)*	This study
BB244	*adr-1(uu49);adr-2(uu28);rde-4(uu53)*	This study
BB245	*adr-1(uu49);adr-2(uu28);ccb-1(uu35)*	This study
BB246	*adr-1(uu49);adr-2(uu28);efa-6(uu46; Δefa-6 intron 8 + 60 nt syntron)*	This study
BB247	*adr-1(uu49);adr-2(uu28);egl-8(uu47)*	This study
BB250	*rrf-3(uu56)*	This study
BB251	*rrf-3(uu57)*	This study
BB259	*adr-1(uu49);adr-2(uu28);ggIs1[nrde-3p::3xFlag::gfp::nrde-3 CDS + unc-119(+)]*	This study
BB260	*adr-1(uu49);adr-2(uu28);rrf-3(pk1426)*	This study
BB261	*adr-1(uu49);adr-2(uu28);rrf-3(uu56)*	This study
BB265	*adr-1(uu49);adr-2(uu28);rrf-3(uu57)*	This study
BB266	*adr-1(uu49);rrf-3(uu56)*	This study
BB267	*adr-2(uu28);rrf-3(uu56)*	This study
BB270	*adr-1(uu49);adr-2(uu28);;rrf-3(uu56);rde-1(uu51)*	This study
BB272	*adr-1(uu49);adr-2(uu28);rrf-3(uu56);rde-4(uu53)*	This study
BB273	*adr-1(uu49);adr-2(uu28);rrf-3(uu56);drh-1(uu60)*	This study
BB277	*rrf-3(uu56); ggIs1[nrde-3p::3xFlag::gfp::nrde-3 CDS + unc-119(+)]*	This study
BB278	*adr-1(uu49);adr-2(uu28);rrf-3(uu56);ggIs1[nrde-3p::3xFlag::gfp::nrde-3 CDS + unc-119(+)]*	This study
BB279	*adr-1(uu49);adr-2(uu28);rrf-3(uu56);nrde-3(uu64)*	This study
BB280	*adr-1(uu49);adr-2(uu28);rrf-3(uu56);rrf-1(uu65)*	This study
BB283	*adr-1(uu49);adr-2(uu28);ergo-1(uu68)*	This study
BB286	*adr-1(uu49);adr-2(uu28);rrf-3(uu56);set-25(uu66)*	This study
BB288	*ergo-1(uu68)*	This study
BB289	*adr-1(uu49);adr-2(uu28);rrf-3(uu56);ergo-1(uu68)*	This study
NL2099	*rrf-3(pk1426)*	(Simmer et al. 2002)
YY178	*ggIs1[nrde-3p::3xFlag::gfp::nrde-3 CDS + unc-119(+)]*	(Guang et al. 2008)

Supplemental Table S2. Primers used for qRT-PCR analysis.

Primer name	Sequence
ama1_RTPCR_F1	GTCAATGATGGGACATCGTGTC
ama1_RTPCR_R1	GTGATGAGTTGTCTCGGCACC
Y45F10D4_RTPCR_F1	CGAGAACCCGCGAAATGTCGGA
Y45F10D4_RTPCR_R1	CGGTTGCCAGGGAAGATGAGGC
cdc42_RTPCR_F1	AGCCATTCTGGCCGCTCTCG
cdc42_RTPCR_R1	GCAACCGCTTCTCGTTTGGC
pmp3_RTPCR_F2	TGGAATTGTTTCACGGAATGC
pmp3_RTPCR_R2	TTCAGCTCTTCGTGAAGTTCC
daf2_RTPCR_F1	GCTACTATACGCCTGATCCTC
daf2_RTPCR_R1	TTGTGTAATGCGTGAGGTCTC
hmr1_RTPCR_F1	ATGTGTTCTCCGTTCAGGTCG
hmr1_RTPCR_R1	TGAGTCCATTCAGTTGAGCTG
efa6_RTPCR_F1	GTTGATCCAGATTCAGTTGTC
efa6_RTPCR_R1	TCTGTGTCAAAGTAGAGAACG
mdt17_RTPCR_F1	GAACTCGATGATGAGAACGTC
mdt17_RTPCR_R1	CTGATTTAGGGATTCATGCAG
ceh100_RTPCR_F1	ACCAGAAATCGAGGAACTTCC
ceh100_RTPCR_R1	GGAACCATTCGCGTATCTGTG
ergo1_RTPCR_F1	TCCACACTCAAGGAATTCTCG
ergo1_RTPCR_R1	GTTCCGACTTTTCCGAGCAC
ogt1_RTPCR_F1	CGTTGTGCTTCGATAAGGTTC
ogt1_RTPCR_R1	TCTACTCGACTCCTATCATGC
rhgf2_RTPCR_F1	CCACTCGATGAGTATGGAAGG
rhgf2_RTPCR_R1	GAGCTTCGGTGCATGTAGTTC
ccb1_RTPCR_F1	ACTCGAATTTCAGTCATCACG
ccb1_RTPCR_R1	CCACCGGTTTATACTTGGCTC
egl8_RTPCR_F1	CAGACGTGTTCTTCAAGGACG
egl8_RTPCR_R1	TTGGACAGCAGCATATCTCC

Supplemental Table S3. Primers used for CRISPR/Cas9 gene targeting and genotyping. For primers encoding sgRNA transcription templates used for gene disruptions, sgRNA target sequences are shown in lowercase letters.

Primer name	Sequence
CRISPR_sgRNA_R1	AAA AGC ACC GAC TCG GTG CCA C
T7_sgRNA_ccb1_Int5_A	TAATACGACTCACTATAgttattttgagatagggaagGTTTTAGAGCTAGAAATAG
T7_sgRNA_ccb1_Int5_B	TAATACGACTCACTATAgtaagacccccaagtgagaaGTTTTAGAGCTAGAAATAG
ccb1_Ex5_F1	GAT GCC AAG AAG TGG ATC ACG
ccb1_Ex6_R1	CTT TTG TTG GAG GCG TCG TAA C
ccb1_Int5_F2	TTG CAG TGC AAG ACG ATT ACC
T7_sgRNA_efa6_Ex8	TAATACGACTCACTATAgtatgttagagagactgatggGTTTTAGAGCTAGAAATAG
T7_sgRNA_efa6_Ex9	TAATACGACTCACTATAgctccattcggacgtcgcagtGTTTTAGAGCTAGAAATAG
efa6_Int8_F1	AAC ACC ATT CCC TAG TGA GTG
Syntron_sense	GTA AGT TTA AAC AGT TCG GTA CTA ACT AAT CCA TGG ACA TAG ATA TCT TTA AAT TTT CAG
efa6_Ex8_Synt_F1	GGA TTT CTT ATG CGA AAA TAT GTT AGA GAa ACc GAc GGT GGA AAG AGT AAG TTT AAA CAG
efa6_Ex9_Synt_R1	CGA AGA CGA GCG TAT ACC ATT CTC CAg gaG CGA CGT CCG AAT GGA GCT GAA AAT TTA AAG
T7_egl8_Int11_sgRNA_A_F1	TAATACGACTCACTATAggcagcagccaacacccataGTTTTAGAGCTAGAAATAG
T7_egl8_Int11_sgRNA_B_F1	TAATACGACTCACTATAggagttacaggaaatacaaaGTTTTAGAGCTAGAAATAG
egl8_Ex11_F1	CCA GGG AAA ACG ACG AAG CAC
egl8_Ex12_R1	GCG AAA ATC CGC TCC TCT TC
egl8_Int11_F1	TCG AAA ATG TGG GAA ATG CTC
T7_sgRNA_adr2_5p_A	TAATACGACTCACTATAggaacaaaaagtccacatgGTTTTAGAGCTAGAAATAG
T7_sgRNA_adr2_3p_B	TAATACGACTCACTATAggtgattcgagttacttatctGTTTTAGAGCTAGAAATAG
adr2_5p_F2	GTTCACTAGTCGATGTTGCTC
adr2_3p_R2	AATCACATGGGTCACTGATGC
adr2_WT_R1	ACAGTTTCCTTCACAAAGTCG
T7_adr1_Start_sgRNA_A	TAATACGACTCACTATAggtaatttatttgactacgaaaGTTTTAGAGCTAGAAATAG
T7_adr1_Ex11_sgRNA	TAATACGACTCACTATAggccgttgtgaattatgcgaGTTTTAGAGCTAGAAATAG
adr1_5p_F1	GTGTCTACTTAAGAACGTGGAG
adr1_Ex12_R1	CGAAAGCAGCAAGAGTGAAGG
T7_rde1_5p_sgRNA_A	TAATACGACTCACTATAggacatgtttcatcactttgGTTTTAGAGCTAGAAATAG

T7_rde1_Ex10_sgRNA_B	TAATACGACTCACTATAggaattgtgaacccatcatcGTTTTAGAGCTAGAAATAG
rde1_5p_F1	AGA GTG GTT CTG CAA ACA CG
rde1_WT_F2	CTA CGT GTT AGT CAT GAT GAG C
rde1_Ex11_R1	CTA GCA GAG AGA AAA GCA AGT C
T7_rde4_Ex1_sgRNA_A	TAATACGACTCACTATAggtactagaagaggctgctaGTTTTAGAGCTAGAAATAG
T7_rde4_Ex4_sgRNA_B	TAATACGACTCACTATAggtctggagaaactagacgcGTTTTAGAGCTAGAAATAG
rde4_Ex1_F1	AAG CGT TTT CGG TGG ATC AG
rde4_WT_F2	AAG ACG GTA TCG AAT CTC TGG
rde4_Ex4_R1	ACA AGC ACA CTG TTT AGC CTC
T7_rrf3_sgRNA_A	TAATACGACTCACTATAgctcaaatctcgcatacgagGTTTTAGAGCTAGAAATAG
T7_rrf3_Ex8_sgRNA_C	TAATACGACTCACTATAgttgaacctgacattgaaggGTTTTAGAGCTAGAAATAG
rrf3_Ex4_F1	CGA TTG CGA TTG GAA ACT GC
rrf3_WT_R1	GCA CGT TTC CAT ATT GAG AAC C
rrf3_Ex9_R3	TTG TGA TCC TTC TGT GAG AGC
T7_drh1_sgRNA_A	TAATACGACTCACTATAgcctctgctcggagagcaagGTTTTAGAGCTAGAAATAG
T7_drh1_sgRNA_B	TAATACGACTCACTATAggtatcttctctcaggattcgGTTTTAGAGCTAGAAATAG
drh1_Ex4_F1	GTT TCC TAC GCT TCA TTG AAC
drh1_WT_R1	GCA GTT CCT AAA TAG ACC ATC
drh1_Ex18_R2	ACT TCA ATC AAC TGA CCA AGC
T7_sgRNA_nrde3_Ex1_A	TAATACGACTCACTATAggatctcctagacaaagtaatGTTTTAGAGCTAGAAATAG
T7_sgRNA_nrde3_Ex11_B	TAATACGACTCACTATAggtcattgcatttagatcgtGTTTTAGAGCTAGAAATAG
nrde3_5p_F1	CAT TCC TTT GCT GTG CGA CTG
nrde3_3p_R1	CAA GTG AAA TCC CTG GTA AAC C
nrde3_WT_R2	CGA CCT CCA AGA GAT CCT TGC
T7_rrf1_sgRNA_A	TAATACGACTCACTATAggcagtcactatccatacaacGTTTTAGAGCTAGAAATAG
T7_rrf1_sgRNA_B	TAATACGACTCACTATAgtcagaaagaaagctgttgccGTTTTAGAGCTAGAAATAG
rrf1_Ex_F1	ACG GTT CGA TTG TGA TTG GAG
rrf1_Ex17_R2	GAT TTG CTC CAC CAA TCT GC
rrf1_WT_F2	CAA CGG ACA ACT CAG GGT TAG
T7_set25_Ex2_sgRNA_A	TAATACGACTCACTATAggcactcgggcttcagtgtgGTTTTAGAGCTAGAAATAG
T7_set25_Ex7_sgRNA_B	TAATACGACTCACTATAgacttcgacgaacaccgagctGTTTTAGAGCTAGAAATAG
set25_Ex1_F1	TAC AGA AGC GAC AGC ATC TC
set25_Ex7_R1	CAT TTG ACA CTC GAC CGT TTC

set25_WT_R2	ATC TCC TGC GAT TGC TTT GAG
T7_ergo1_Ex1_sgRNA_A	TAATACGACTCACTATAggttatcgtggatacaaccaGTTTTAGAGCTA GAAATAG
T7_ergo1_Ex6_sgRNA_B	TAATACGACTCACTATAggagagttcatagatcacacGTTTTAGAGCT AGAAATAG
ergo1_Ex1_F1	GGA CAA TCG CTA CGA TGA TCG
ergo1_Ex6_R1	CAC GTA TCG TGA AGC ACA TAG
ergo1_WT_F2	TCG ACG TTT CTC ATC CAT CG

Supplemental Table S4: Novel mutations generated by CRISPR/Cas9 for this study.

Mutation	Coordinates (ce10/WS220)	Nature of mutation	Inserted sequence
adr-2(uu28)	chrIII:7230936-7232861	-1926 nt deletion	-
ccb-1(uu35; ΔEER)	chrI:3644035-3644883	-849 nt deletion	-
efa-6(uu46; Δefa-6 intron 8 + 60 nt syntron)	chrIV:12607422-12609861	+97/-2440 nt insertion/deletion	AACCGACGGTGGAAAGAGTAAG TTTAAACAGTTCGGTACTAACT AATCCATGGACATAGATATCTT TAAATTTTCAGCTCCATTCGGA CGTCGCTCC
egl-8(uu47; ΔEER)	chrV:30802-34717	-3916 nt deletion	-
adr-1(uu49)	chrI:7773430-7777062	+1/-3633 nt insertion/deletion	T
rde-1(uu51)	chrV:9988461-9991606	-3146 nt deletion	-
rde-4(uu53)	chrIII:10217478-10218402	-925 nt deletion	-
rrf-3(uu56)	chrII:8163812-8165364	-1553 nt deletion	-
rrf-3(uu57)	chrII:8163805-8165366	+22/-1562 nt insertion/deletion	GTATTCTGGTGGCCACCAGACA
drh-1(uu60)	chrIV:6608226-6612649	+54/-4424 nt insertion/deletion	CATGTATACAATTTGGGAAAAAG CACTTGCCTTGGAGAGAGAAGA TACCTTGGA
nrde-3(uu64)	chrX:372267-376618	+64/-4352 nt insertion/deletion	CACATGTGCCATGTGCACTCCAT GCCATGTGCACATGCCATGTGC ACTCCAAGTGGTGCACAGG
rrf-1(uu65)	chrI:7645169-7648452	+20/-3284 nt insertion/deletion	TCATTCATACTCATATTGAA
set-25(uu66)	chrIII:13287014-13292728	+15/-5715 nt insertion/deletion	GCTGTGACCCCAGCT
ergo-1(uu68)	chrV:1009051-1016719	+1/-7669 nt insertion/deletion	T

CHAPTER 3

DOUBLE-STRANDED RNA STRUCTURES ARE ASSOCIATED

WITH ESSENTIAL AND HIGHLY EXPRESSED GENES

ON *C. ELEGANS* AUTOSOME DISTAL ARMS

Introduction

Interactions between long double-stranded RNAs (dsRNAs) and dsRNA-binding proteins (dsRBPs) mediate critical cellular processes. The best 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) to silence target transcripts (Carthew and Sontheimer, 2009). However, dsRNAs are also important substrates of other dsRBPs, like the regulatory protein Staufen (Heraud-Farlow and Kiebler, 2014) and Adenosine Deaminases that act on RNA (ADARs), RNA editing enzymes that convert adenosine to inosine in dsRNA (Nishikura, 2016). dsRBPs interact with the phosphodiester backbone of A-form dsRNA, and so dsRBPs bind their substrates regardless of sequence (Tian et al., 2004). Complementary RNA sequences that form dsRNA often arise from transposon-derived repetitive sequences, which are abundant in metazoan genomes (Feschotte and Pritham, 2007; Nishikura, 2016). Thus, duplex structures are common genomic features, and indeed, human and *C. elegans* each express thousands of ADAR-edited dsRNAs (Blango and Bass, 2016; Reich et al., 2018).

However, despite the ubiquity of dsRNA structures and the essential roles of many dsRBPs, the functions of most cellular dsRNAs remain largely uncharacterized.

Recent work to define cellular dsRNAs found that most occur within introns and 3'UTRs of protein-coding genes (Blango and Bass, 2016; Reich et al., 2018; Whipple et al., 2015), suggesting that these structures may function to regulate gene expression. Indeed, a growing body of evidence indicates that dsRNAs impact expression of their associated genes through various mechanisms. In *C. elegans*, ADARs edit dsRNAs to prevent their processing and silencing by antiviral RNAi mechanisms (Reich et al., 2018). Duplex structures also impact translation efficiency, since *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 to trigger mRNA decay (Gong and Maquat, 2011), suggesting that dsRNA formation regulates transcript stability. Together, these studies demonstrate that dsRNAs regulate gene expression by multiple mechanisms.

The organization of eukaryotic genomes into distinct domains also influences gene expression patterns. Genomic domains can differ widely by gene transcription levels, nucleosome organization, and post-translational histone modification patterns (Ahringer and Gasser, 2018). The *C. elegans* genome is comprised of the sex chromosome, Chromosome X, and five autosomal chromosomes that are 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 well-conserved (Consortium, 1998; Cutter et al., 2009), in part due to low meiotic recombination

frequency in autosome centers. In contrast, autosome distal arms are characterized by frequent recombination, 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) (Ho et al., 2014; Liu et al., 2011; Towbin et al., 2012). Autosome arms are also physically associated with the nuclear lamina, seen in chromatin immunoprecipitation (ChIP) experiments with the nuclear lamina component LEM-2 (Ikegami et al., 2010). However, distal arms are not completely silent, as they contain many essential, expressed genes and display subdomains with marks of active chromosome that are not lamina-associated (Ikegami et al., 2010; Liu et al., 2011). Though these genomic domains clearly exhibit distinct properties, the mechanisms that establish and regulate chromosomal domains and subdomains are far from understood.

In this report, we characterize the genomic distribution of *C. elegans* dsRNAs and report their association with essential and highly expressed genes on the distal arms of autosomes. Using three independent methods to identify dsRNAs, we confirm that dsRNA loci cluster on autosome arms and most often occur in gene introns. Of the genes that occur on autosome arms, those associated with dsRNAs are often essential; over 30% of essential genes on Chromosome I and V distal arms contain edited dsRNA structures. Further, dsRNA-associated genes exhibit elevated expression compared to other distal arm genes at four developmental stages. Histone modifications associated with transcriptional elongation are enriched over dsRNA loci. Because we find that repetitive sequence content is also positively associated with gene expression on autosome arms, we cannot conclude that the effects observed relate explicitly to dsRNA structure.

Limited analyses of transgenic reporters containing duplex structures further indicate that these structures are not sufficient to increase gene expression. Nonetheless, our observations identify a property common to many highly expressed, essential genes on autosome arms, offering potential insight into their regulation and possibly suggesting a novel function of dsRNA.

Results

Predicted dsRNA structures correlate with EERs

We previously reported that *C. elegans* editing-enriched regions (EERs), expressed regions containing clustered A-to-I editing, are enriched on the distal arms of autosomes (Reich et al., 2018; Whipple et al., 2015). This genomic pattern is striking, so we sought to understand if all *C. elegans* dsRNAs exhibit this organization or if it was restricted to the edited transcripts we found in earlier analyses. To this end, we compiled experimental datasets of *C. elegans* dsRNAs by three independent methods.

Our first dataset comprised the 1523 EERs previously determined from RNAseq of *C. elegans* populations at four development stages: embryos, L1-L2 larval stages, L3-L4 larval stages, and young adults (Fig. 3.1A) (Reich et al., 2018). Because these regions are edited by ADAR, we know they form dsRNA *in vivo*. Like guanosine, inosine prefers to pair with cytidine, so ADAR editing sites appear as A-to-G transitions in cDNA. We defined EERs by identifying regions containing clustered A-to-G changes in aligned RNAseq reads. In this analysis, we restricted our EER pipeline to genomic regions covered by ≥5 reads, and thus missed any dsRNAs expressed below this threshold. As a control dataset, we randomly permuted EERs around genomic regions covered by ≥5

reads, generating a set of length-matched, expressed random regions. However, because we could not define dsRNAs in lowly expressed regions, we sought to identify dsRNAs in a manner that did not rely on gene expression.

Since most EERs occur in the introns of protein coding genes (Reich et al., 2018; Whipple et al., 2015), we used UNAFold to predict the folding free energy of *C. elegans* introns to determine their likelihood of forming duplex structures at 20°C (Fig. 3.1B; (Markham and Zuker, 2008). Introns ≥9000 nt, the maximum length that UNAFold would analyze, were excluded from analysis. Further, we did not include introns ≤40 nt, the minimum length we predicted would form dsRNA long enough to bind ADAR (~15 base-pairs) (Herbert and Rich, 2001) that did not encompass consensus splicing sequences. This dataset comprised 99.5% of all *C. elegans* introns annotated by 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 us a total of 113,740 introns with ΔG/nt values that varied from 0.03 to -1.01 kcal/mol*nt (where more negative values equate to more stable introns). Using the ΔG/nt values, we also determined the most stable intron for each of 20,735 intron-containing genes. For the vast majority of genes (75.6%), the minimal intron ΔG/nt value fell between -0.2 and -0.4 kcal/mol*nt. However, a set of 1521 genes carried at least one exceptionally structured intron (ΔG/nt < -0.5 kcal/mol*nt), including 579 EER-associated genes (EAGs) containing edited dsRNA structures. Further, 710 EERs (46.6%), but only 105 length-matched random regions (6.9%), overlapped an intron with ΔG/nt < -0.5 kcal/mol*nt, demonstrating that intron folding free energy is a good predictor of dsRNA formation.

Most EERs form dsRNA through intramolecular interactions of inverted repeat (IR) sequences, so we next used the publicly available Inverted Repeats Finder tool (Warburton et al., 2004) to predict IRs in the *C. elegans* genome (Fig. 3.1C). Only regions containing repeated sequences of ≥20 base-pairs within <2kb of each other were defined as IRs. Unlike EERs and stable introns, we restricted our list of IRs to include only sequences predicted to form dsRNA; we excluded loop sequences not predicted to base-pair. As a control dataset for IRs, we also used the Tandem Repeat Finder (Benson, 1999) to predict tandem repeats (TRs) in the *C. elegans* genome, which we predict would not fold into intramolecular structures. To avoid low complexity sequences (i.e. dinucleotide tracts, telomeric repeats, etc.), we only considered TRs comprised of repeated units of unique sequences ≥20 nt long. We reasoned that TR sequences might also form dsRNA if they occurred nearby an inverted copy of the TR sequence. Therefore, to separate control TR regions from the list of IRs, we removed TR sequences predicted to overlap IRs, since these would have potential to form dsRNA. Demonstrating that these methods effectively predicted dsRNAs, we found that 1476 EERs (96.9%) overlapped IRs, while 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* chromosome to a control set of length-matched expressed random regions (Reich et al., 2018). We divided each chromosome into ten equal segments and determined the number of EERs and control random regions in each chromosomal segment (Fig. 3.2A). While 89.5% of EERs occurred outside of a

chromosome's central four segments, only 59.8% of control random regions were found in the same distal arm segments, a significantly different distribution than EERs ($P <$ 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 the distal arms of *C. elegans* autosomes (Whipple et al., 2015).

To determine if EER locations reflected the distribution of dsRNAs globally, we next analyzed the chromosomal positions of genomic IRs, as well as TR controls. When we mapped predicted IR and TR regions to chromosomal segments shown in Fig. 3.2A, we found that 77.6% of IRs and 75.3% of TRs occurred in autosome distal arm regions (Fig. 3.2B), 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 proximal (<1000 nts) or distal (>1000 nts) to genes (Fig. 3.2C). Similar to EERs, 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%), though a higher fraction of TR sequences occurred outside annotated genes than IRs. Our observations suggest that both IR and TR sequences are enriched on distal autosome arms and largely occur in noncoding regions of protein-coding genes, similar to EERs.

Next, we plotted intron $\Delta G/nt$ values along each chromosome (Fig. 3.2D). In concordance with EER and IR patterns, intron $\Delta G/nt$ decreased in the distal arms of all five *C. elegans* autosomes, but not Chromosome X, suggesting stable RNA structures

cluster in these domains. The differences in intronic ΔG/nt between arm and center domains were remarkably clear in some cases (for example, on Chromosomes I and III). These large chromosomal differences in intron ΔG/nt allowed us to calculate more refined boundaries of "arm" and "center" domains for each chromosome (dotted vertical lines in Fig. 3.2D; see Materials and Methods), which we used in all subsequent analyses.

We wanted to understand if dsRNA clustering on autosome arms was conserved in a related nematode species, *Caenorhabditis briggsae*, which exhibit chromosome arm and center domains with respect to recombination rates, gene density, and repeat density (Hillier et al., 2007). Though we have not determined ADAR editing sites or EERs in *C. briggsae,* we were able to predict intron ΔG/nt values for *C. briggsae* as we had done 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. 3.2D). From our combined analyses, we conclude that gene-associated dsRNAs cluster on distal autosome arms in two *Caenorhabditis* species.

RNA structures are enriched in essential genes

We previously showed that in *adr-1;adr-2* mutant worms, EERs associated with protein-coding genes were cleaved by Dicer to generate siRNAs, which induced gene silencing by RNAi (Reich et al., 2018). However, this work did not identify a function of EERs in wildtype animals, so we sought to understand if dsRNAs show signatures of functional importance. Mammalian EERs show almost no sequence conservation between mouse and human, but mouse EAGs have orthologous human genes containing

EERs at a greater fraction than expected by random chance (Blango and Bass, 2016). To assess if similar sets of orthologous *Caenorhabditis* genes were associated with dsRNA, we compared *C. elegans* genes containing a structured intron (ΔG/nt < -0.5 kcal/mol*nt) to genes whose *C. briggsae* ortholog had a structured intron. We found that 147 of 1521 *C. elegans* genes containing structured introns (9.6%) also 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 suggests that dsRNA-associated genes maintain an association with duplex structures over many generations, indicating potential functional relevance. However, structured introns did not always occur in the same intron in orthologous genes (data not shown), and this was also found in comparison of mouse and human EAGs (Blango and Bass, 2016). Further, when we compared orthologous *C. elegans* and *C. briggsae* genes containing structured introns, we found no correlation between minimum intron ΔG/nt in each species (Pearson correlation coefficient = -0.03). 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 perform some 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. When we overlapped this list of essential genes with the 955 EAGs defined in our previous publication (Reich et al., 2018), we observed 213 essential EAGs, more than double the number expected by chance (Fig. 3.3A). Further, essential genes were significantly enriched among EAGs defined from individual developmental stages,

except young adults, where we defined only 128 EERs. Our results were particularly surprising because EERs and EAGs cluster on autosome distal arms, while 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. 3.3B). Over 30% of the essential genes on the arms of Chromosomes I and V were associated with EERs. 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 genes on autosome arms contained EERs than essential genes. These results indicate that EERs occur within functionally important genes on distal autosome arms.

A possible explanation for EER enrichment in essential genes could be differences in expression of essential and nonessential genes. If essential genes had higher coverage than nonessential genes in RNAseq datasets used to define EERs, we may have simply identified EERs in essential genes but missed editing in nonessential genes due to low coverage. In overlapping EAGs and essential genes, we had limited comparisons to genes with enough read coverage depth to define an EER ($\geq$5 reads); however, we still worried that gene expression differences might have biased our results. To avoid complications from differences in gene expression, we analyzed ΔG/nt of the most stable intron in essential and nonessential genes on chromosome arms and centers (Fig. 3.3C). Both essential and nonessential genes on distal arms had more stable introns than genes in chromosome centers. However, on autosome arms, but not Chromosome X, essential genes exhibited more stable introns overall than nonessential genes. Thus, independent of gene expression, we observed an association between essential genes and

dsRNA structures.

We next determined if IRs or TRs were enriched within essential genes. We first determined the number of IR and TR nucleotides in noncoding regions (i.e. introns and UTRs) of each gene on autosome arms. Then, we separated genes into four categories based on their IR and TR content and determined the fraction of genes in each category that were essential and nonessential (Fig. 3.3D). Consistent with earlier observations, we found that IRs were associated with essential genes, since 18.2% of essential genes on autosome arms contained >1kb of IR sequence, compared to only 6.2% of nonessential genes. TRs were also associated with essential genes, though to a lesser extent than IRs, as 15.5% of essential genes carried >1 kb of TR sequence, compared to 5.6% of nonessential 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). Our previous work showed that EERs can also cause gene silencing either when ADARs are absent or the 26G endo-siRNA pathway is inactive (Reich et al., 2018). Thus, we hypothesized that EAGs, particularly those on autosome arms, would exhibit decreased expression compared to other genes. To test this hypothesis, we separated genes by their position on chromosome arms or centers, and then plotted embryo-stage expression of EAGs and all expressed genes (Fig. 3.4A). As observed previously, genes on autosome arms exhibited lower overall expression than genes in autosome centers. However, to our surprise, we found that EAGs showed the opposite

trend; EAGs on autosome arms were more highly expressed than those in autosome centers. Furthermore, EAGs on autosome arms displayed much 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 showed increased expression relative to all autosome arm genes at every stage (Fig. 3.4B).

Like the association with essential genes, we worried that our observations might simply reflect a bias towards identifying EERs within highly expressed genes. To avoid confounding effects from defining dsRNAs based on expression, we classified dsRNA-associated genes as those containing a structured intron (ΔG/nt < -0.5 kcal/mol*nt) and plotted expression of all genes and structured genes in each chromosomal domain (Fig. 3.4C). As observed with EAGs, genes with structured introns exhibited 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.

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, with three or more distinct introns or UTRs (5' or 3') that overlapped EERs. Included among these highly structured genes were examples such as *epc-1* and *ssl-1,* which contained eight and twelve edited introns/UTRs, respectively. Of the 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 $\geq$3 EER-associated introns/UTRs displayed significantly higher gene

expression than the set of all autosome arm genes (Fig. 3.4D). 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 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 often correlates with histone modification patterns, which have been mapped in detail along the *C. elegans* genome (Evans et al., 2016; Ho et al., 2014; Liu et al., 2011; Towbin et al., 2012). Modifications like H3K4me3 and acetylated H3K27 (H3K27ac) typically mark active promoters, while H3K36me marks (mono-, di-, and trimethylation) often correlate with transcriptional elongation (Ahringer and Gasser, 2018; Evans et al., 2016; Venkatesh and Workman, 2013). In contrast, H3K9me3 and H3K27me3 are enriched over transcriptionally silent regions (Ho et al., 2014; Liu et al., 2011; Towbin et al., 2012), and are deposited on targets of nuclear RNAi (Gu et al., 2012; Mao et al., 2015). Since 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 chromatin immunoprecipitation and microarray (ChIP-chip) data to determine the frequency of 19 histone modifications at dsRNA or

control loci on autosome arms (Fig. 3.4E). 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 modifications patterns over stable introns ($\Delta G/nt < -0.5$ kcal/mol*nt) and all other autosome arm introns, we observed that chromatin marks over stable introns closely resembled those over EERs. IRs displayed more subtle enrichment of H3K36me marks and H3K9me1 and me2, but, like EERs and stable introns, did not exhibit high levels of the silencing marks H3K9me3 and H3K27me3. Overall, modification patterns over IRs were very similar to those at TR loci, suggesting these regions are regulated by similar mechanisms. We note that a large fraction of IRs and TRs are not gene-associated (Fig. 3.2B), which may explain why chromatin patterns at IRs and TRs differ in magnitude from those at EERs and stable introns. Our analyses indicate that gene-associated dsRNAs are associated with marks of transcriptional elongation, rather than gene silencing.

Intronic dsRNA structure is associated with elevated expression of

genes with long introns and abundant periodic A_n/T_n clusters

In human, mouse, and *C. elegans*, EERs occur in particularly long introns (Blango and Bass, 2016; Whipple et al., 2015). EER length does not correlate with intron length, suggesting EER-containing introns are not long solely because they contain an EER. To further examine the association between intron structure and length, we took each gene's most stable intron and plotted $\Delta G/nt$ against intron length (Fig. 3.5A). We found that

intron stability and length were moderately correlated ($r^2 = 0.334$) with exceptionally stable introns mostly >500 nt. Like dsRNAs, long introns cluster on the distal arms of *C. elegans* autosomes (Prachumwat et al., 2004). Therefore, we wondered if long introns are associated with gene expression on autosome arms and if this depended on intronic dsRNA structures. We divided autosome arm genes first by whether they had an intron >500 nt long, and then by whether they had an intron with $\Delta G/nt$ < -0.5 kcal/mol*nt. Then we plotted the embryonic expression of genes in each group (Fig. 3.5B). Regardless of intron $\Delta G/nt$, genes with long introns exhibited higher expression than genes without long introns. However, for genes containing introns >500 nt, the presence of a stable intron correlated with significantly higher gene expression, suggesting that stable introns are not associated with highly expressed genes solely because they occur in genes with long introns.

In *C. elegans,* 10-nt periodic A_n/T_n clusters (PATCs) promote germline expression of genes in repressive chromatin (Frokjaer-Jensen et al., 2016). PATC elements frequently occur in long introns on autosome distal arms, much like EERs and other predicted dsRNAs. To discern if genes containing dsRNAs also contained abundant PATC sequences, we plotted each autosome arm gene's minimum intron $\Delta G/nt$ against its maximum intron PATC score (Fig 3.5C), which measures how well a sequence conforms to the PATC sequence pattern (higher score = more PATC-like). Intron stability and PATC score were weakly correlated ($r^2 = 0.176$), suggesting genes with high PATC content were slightly more likely to contain an intronic dsRNA structure. As we had done with long introns, we plotted embryonic expression of autosome arms genes divided into those with low (<50) or high (>50) PATC scores, and those with or without

structured introns (Fig. 3.5D). High PATC score correlated with increased gene expression overall. However, among genes with intronic PATC > 50, the presence of a stable intron correlated with significantly higher expression. We conclude that the association between dsRNA structure and elevated gene expression on autosome arms is not caused by long introns or high PATC density alone.

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 (Reich et al., 2018; Zhao et al., 2015). 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. 3.6A,B). 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, while we observed higher 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. 3.6C), again we observed that higher repeat content correlated with higher expression. We cannot rule out that repeat content correlates with expression due to increased prevalence of dsRNA structures. However, due to the close relationship between duplex structure and repetitive sequences,

we cannot conclude that associations we describe relate explicitly to RNA structure.

dsRNA structures are not sufficient to increase gene expression

To gain further insight into association between EERs and highly expressed genes on distal autosome arms, we designed a set of two *GFP* reporter genes driven by the same ubiquitous *Pbaf-1* promoter (Fig. 3.7A). The control reporter contained three ~60 nt unstructured introns and the *unc-54* 3'UTR, which lacks predicted dsRNA structure. However, in place of the second *GFP* intron and *unc-54* 3'UTR, the EER reporter gene substituted structured, edited elements from two EAGs: intron 1 from *ssl-1* and the 3'UTR of *eif-2alpha*. We constructed four transgenic strains using CRISPR protocols to integrate a single copy of either the control or EER reporter at defined sites on the distal arm or center of Chromosome III (Fig. 3.7B). We chose integration sites without annotated genes >1kb in either direction in hopes that reporters would not be influenced by neighboring genes. Since autosome arms exhibit high H3K9me3 and H3K27me3 deposition relative to autosome centers (Liu et al., 2011), we chose an arm site within an annotated heterochromatic region (Ho et al., 2014). We hypothesized that the EER reporter would exhibit higher expression than the control reporter at the arm site, but not the center site. Contrary to our expectations, both the control and EER reporters integrated on Chromosome III's arm exhibited fluorescence barely above background, likely due to integration within heterochromatin (Fig. 3.7C,D). In contrast, at the Chromosome III center site, we observed higher fluorescence of the EER reporter, which correlated with mRNA levels by quantitative RT-PCR (Fig. 3.7E). EAGs are downregulated via their EERs in *adr-1;adr-2* mutants that lack A-to-I editing (Reich et

al., 2018). To test if editing impacted our EER reporters, we deleted *adr-2*, encoding the only active editing enzyme in *C. elegans* in strains containing *GFP* reporters (Fig. 3.7F). Loss of *adr-2* had no effect on any of the reporter lines tested, suggesting that A-to-I editing does not strongly influence their expression.

Differences in intronic and 3'UTR sequences of our reporters made results difficult to interpret. For one, we did not know if increased expression of the EER reporter derived from the *ssl-1* intron, *eIF-2alpha* 3'UTR, or both. To test this, we constructed a third reporter containing the *eIF-2alpha* 3'UTR without the edited *ssl-1* intron, and integrated it at the Chromosome III center site. This 3'UTR EER reporter exhibited fluorescence between the original control and EER reporters (Fig. 3.7G), suggesting that both the *ssl-1* intron and *eIF-2alpha* 3'UTR contribute to higher expression of the EER reporter. However, we could still not rule out that expression differences of the control and 3'UTR EER reporters were not simply intrinsic to their 3'UTR sequences, rather than their structures. The *eIF-2alpha* 3'UTR contains an inverted repeat sequence derived from a PALTTAA3_CE DNA transposon. To test if the dsRNA structure in the *eIF-2alpha* 3'UTR promoted expression of the 3'UTR EER reporter, we reversed the orientation of the second repeated element in the PALTTAA3 inverted repeat to make a tandem repeat. The tandem repeat reporter should have the same length and similar sequence as the 3'UTR EER reporter, but should not form dsRNA (Fig. 3.7H). When we integrated this reporter into the Chromosome III center site, we found that the tandem repeat reporter was expressed nearly twice as highly as the reporter containing the inverted repeat in its 3'UTR (Fig. 3.7I). Thus, we concluded that *eIF-2alpha* 3'UTR does not promote gene expression due to the dsRNA structure.

Discussion

In this work, we used three methods to identify *C. elegans* dsRNA-producing loci and correlated these loci with essential genes and gene expression patterns. We find that dsRNAs are enriched on autosome distal arms within essential genes and genes with higher than expected expression, a correlation that is not dependent on intron length and PATC content. While dsRNA structures are not sufficient to increase expression of a reporter gene, our results suggest that RNA structures may positively regulate gene expression in *C. elegans*.

How well do the three methods identify dsRNAs?

We used three different approaches to identify *C. elegans* dsRNAs, specifically the type of long, highly base-paired duplexes that would be substrates for ADARs and other dsRBPs. Previous work established EERs as long dsRNAs that we know to be ADAR substrates (Reich et al., 2018; Whipple et al., 2015), so we used this dataset as our baseline. The computationally predicted datasets, stable introns and predicted IRs, exhibited substantial overlap with EERs, suggesting our approaches effectively predicted known dsRNAs. However, many stable introns and IRs did not overlap EERs, suggesting the three methods identified different sets of regions. Read coverage requirements used to define EERs provide the simplest explanation of these differences, since many dsRNAs may be expressed at low levels and lack sufficient RNAseq coverage to identify clustered editing sites. Of the five most stable *C. elegans* introns, only two overlap EERs, but the other three did not meet the 5-read coverage threshold needed to define EERs (data not shown). Still, these three stable introns, despite that they did not overlap EERs, all

exhibited several aligned reads with A-to-G changes, indicating that they are edited. Additionally, we found that 53.4% of EERs did not overlap a stable intron, perhaps because our stability cutoff of -0.5 kcal/mol*nt was too stringent. Since 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, our dataset of IRs overlapped nearly every EER (96.9%). Thus, almost all EERs can form intramolecular duplexes, suggesting that edited intermolecular dsRNAs may be rare in *C. elegans*. EERs overlapped only 14.4% of total IR sequences, demonstrating that IR prediction identified many putative dsRNAs not found by editing detection. Again, expression likely explains differences between IR and EER datasets, since almost half of IRs are intergenic and may be lowly expressed. It is also possible that many IRs form structures simple too short to be nonselectively edited.

What is the important property of dsRNA loci?

While our work describes an observed correlation between dsRNA-producing loci and highly expressed and essential genes, we cannot be certain that this correlation requires transcription of loci into dsRNAs. In fact, dsRNA loci have several other characteristics that could underlie the associations we describe.

For one, most dsRNA loci are repetitive, with ~63% of EER sequences derived from DNA transposons (Reich et al., 2018). Similarly, 27.9% of stable introns and 25.9% of IRs map to RepeatMasker-annotated transposons (comprising ~11.8% of the total genome). If specific transposon sequences promote gene expression, this could explain

the 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 (Chuong et al., 2017; Lev-Maor et al., 2007), so it is conceivable that transposon sequences might promote gene expression (Fig. 3.8A). Indeed, we observed a positive correlation between the amount of noncoding RepeatMasker transposon sequence in a gene and its relative expression (Fig. 3.5C). However, we speculate that transposons with different sequences would have different regulatory effects, some increasing expression and some silencing. Thus, we predict that only a subset of transposons would associate with highly expressed genes, but no single transposon class occurs in most EERs (the most abundant type, PAL5A_CE, comprise just 15.7% of transposon-derived EER sequences) (Reich et al., 2018). Nonetheless, repetitive sequences may contribute to elevated expression of dsRNA-associated genes.

The association between dsRNAs and highly expressed genes might involve duplex structures formed by DNA sequences, rather than RNA. During transcription of a dsRNA-producing locus, inverted repeat sequence in the nontemplate DNA strand can fold back to pair with itself, forming an intramolecular DNA duplex structure. It is not clear that DNA hairpin structures would necessarily promote gene expression. We speculate, however, that such structures might interfere with nucleosome reassembly following transcription (Fig 3.8B). Defects in nucleosome deposition can lead to spurious transcription (Venkatesh and Workman, 2015), providing a potential mechanism that links IR sequences to elevated gene expression. In future work, we hope to establish whether this, or another mechanism, underlie the correlation between dsRNA loci and highly expressed genes.

How might dsRNA promote gene expression?

Though we cannot rule out repetitive sequences or DNA structures in the association between RNA structures 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. We do not believe the association between dsRNA structure and highly expressed gene relates to different rates of RNA decay, because our gene expression analyses only measured coverage of mature mRNA exons, excluding intronic sequences where most duplex structures reside. Perhaps intronic dsRNA structures promote efficient splicing by folding to bring 5' and 3' splice sites in close proximity (Fig. 3.8C). Alternatively, nascent dsRNA structures might bind nuclear dsRBPs that in turn recruit transcription machinery to promote expression (Fig. 3.8D).

If dsRNAs promote gene expression through interactions with dsRBPs, which proteins might mediate this phenomenon? While Dicer primarily acts on dsRNA to silence genes (Billi et al., 2014; Carthew and Sontheimer, 2009), the dsRBPs ADAR, Staufen, and NF90 are candidate factors due to well-established roles in gene regulation (Bass et al., 1994; Castella et al., 2015; Heraud-Farlow and Kiebler, 2014; Nishikura, 2016). We previously showed that ADARs promote EAG expression by antagonizing RNAi silencing (Reich et al., 2018). Because Dicer-dependent RNAi mechanisms presumably do not silence genes lacking dsRNA, expression differences between structured and unstructured genes are not likely due to RNAi antagonism by ADARs. However, we note that while EAGs were downregulated in *adr-1;adr-2* embryos lacking the 26G factor *rrf-3*, deleting the RNAi factor *rde-4* in this background partially, but not completely, rescued EAG expression. Perhaps EAGs remain slightly downregulated in

adr-1;adr-2;rrf-3;rde-4 mutants because ADARs also promote gene expression by an RNAi-independent mechanism. Expression of ADAR-edited dsRNAs positively correlates with ADAR expression in human brain tissue (Liscovitch et al., 2014), suggesting that ADAR may promote dsRNA expression in humans as well as *C. elegans*.

Besides ADAR, 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, suggesting it could perform additional functions like supporting dsRNA expression. There is significant overlap between EAGs and transcripts bound by the sole *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), suggesting that STAU-1 may regulate many of the highly expressed, dsRNA-associated genes we identified. In mammals, the dsRBP NF90 acts with its obligate binding partner NF45 to both positively and negatively regulate gene expression through effects on transcription, mRNA stability, and translation (Castella et al., 2015). *C. elegans* lack an ortholog of NF90, but nematodes do encode an NF45 ortholog and an NF90-related gene orthologous to mammalian ZFR. ZFR proteins bind NF45 and contain three zinc-fingers domains with homology to dsRNA-binding zinc fingers in the *Xenopus* protein dsRBP-ZFa (Finerty and Bass, 1999; Wolkowicz and Cook, 2012). Thus, the *C. elegans* ZFR ortholog, *Y95B8A.8*, may have uncharacterized functions in dsRNA regulation, and could underlie the correlation between dsRNAs and highly expressed genes on autosome arms.

Materials and methods

Published datasets

The coordinates of 1523 *C. elegans* EERs are reported in Supplemental File S1 of Reich et al. (2018). RNAseq datasets used to define these EERs and used for expression analyses in this report are available from Gene Expression Omnibus (GSE79375). Parameters for alignment and read filtering are described in Reich et al. (2018).

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

Intron stability analysis

A .bed file of annotated *C. elegans* introns (ce10/WS220) was downloaded from UCSC Genome Browser. To compile a list of unique introns, we collapsed introns with the same start and stop coordinates into a single entry. If two or more introns overlapped, we removed the largest intron and determined if the remaining overlapped, repeating this 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 we had done with *C. elegans*. For both sets of introns, we removed those <40 nt and >9000 nt.

To predict the folding free energy of intronic sequences, we first used the bedtools2 (https://github.com/arq5x/bedtools2) application fastaFromBed to determine the genomic sequences of each intron. The ΔG of each intron 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 < 20 or sequence entropy ≤ 1 were discarded. TRs were merged using the bedtools2 application mergeBed, and sequences that overlapped predicted IRs were removed using the bedtools2 application subtractBed.

To calculate the number of IR and TR nucleotides in each gene's intron and UTRs, we first determined noncoding regions associated with each gene. We downloaded a table of annotated gene transcription start and stop sites and a table of coding exons. 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 the 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 $\Delta G/nt$, we first determined the $\Delta G/nt$ of each gene's most stable intron. Then, we separated each chromosome into 100 kb segments, counted the number of genes in each segment, and determined the fraction of genes in each 100 kb region with an intron $\Delta G/nt < -0.5$ kcal/mol*nt (% structured genes). Starting at the center-most segment of each chromosome, we moved outward toward the left arm until we encountered three consecutive 100 kb segments with $\geq 20\%$ structured genes. Of the three consecutive segments, we took the margin closest to the chromosome center as the boundary of the left distal arm. We then repeated this same process to define the boundary of the right arm of each chromosome.

Essential gene analysis

Lists of genetic alleles and RNAi phenotypes 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 genetic manipulations that only induced male sterile phenotypes, but were not also lethal or sterile in hermaphrodites. From the remaining alleles and RNAi experiments, we extracted Ensembl gene IDs of affected genes and removed duplicates to generate a full list of essential genes.

For each developmental stage, we first filtered out all genes in the lists of

essential genes and EER-associated genes that did not contain a region covered by five or more reads in RNAseq data from that stage. We then overlapped the lists of expressed essential genes and EER-associated genes 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 Chi-square approximation with 10,000 randomized iterations.

Expression and chromatin analyses

Expression data, calculated as fragments per kilobase million reads (FPKM) values, were determined for *C. elegans* genes with the USeq (http://useq.sourceforge.net/) application DefinedRegionDifferentialSeq, which uses the R package DESeq2. A RefFlat table of *C. elegans* genes was downloaded from UCSC Genome Browser and DefinedRegionDifferentialSeq was used to determine FPKM over exonic gene regions from each input RNAseq sample.

For ChIP-chip analyses, we created .bed files of EERs, introns, IRs, and TRs present on autosome arm domains. We made a list of random regions using the bedtools2 application shuffleBed, restricting their possible locations to autosome arm regions covered by ≥5 reads in combined developmental RNAseq datasets from Reich et al. (2018). Using the bedtools2 application intersectBed, we overlapped our regions of interest with .bed files reporting average ChIP-chip Z-scores over 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, stable introns, etc.) for each modification, and plotted a heatmap of the results in R.

Intron PATC analysis

We downloaded a BigWig file of balanced PATC scores previously calculated for the *C. elegans* genome at 25 nt resolution (Frokjaer-Jensen et al., 2016). Using the UCSC Genome Browser utility bigWigAverageOverBed, we calculated the average PATC score for each unique *C. elegans* intron. We then determined the highest intron PATC score for each intron-containing gene.

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 genes' total noncoding repeat content.

GFP reporter construction

All cloning was performed using the Gibson assembly method. Primers are listed in Table 3.1. To build the control GFP reporter cassette, the *GFP* coding sequence, containing three 60 nt introns, and *unc-54* 3'UTR was amplified from pPD95.67. The *baf-1* promoter, consisting of a 287 nt fragment upstream of the *baf-1* coding sequence, was amplified from genomic DNA, and assembled with the *GFP::unc-54 3'UTR* fragment in a pBluescript SK(+) cloning vector. A synthetic intron containing the *C. briggsae unc-119* gene (in the reverse orientation the *GFP* gene) flanked by loxP sites

was amplified from pMLS252 (a generous gift of Matthew Schwartz and Eric Jorgensen) and cloned into *GFP* in place of the third 60 nt intron. To create the EER 3'UTR reporter, we amplified the *eif-2alpha* 3'UTR and cloned this into the control reporter in place of the *unc-54* 3'UTR. Then, to make the full EER reporter, we amplified intron 1 of *ssl-1* from genomic DNA, and inserted it into the EER 3'UTR reporter cassette in place of the second 60 nt intron. All inserted sequences were verified by Sanger sequencing.

To build *Pbaf-1::GFP* cassettes suitable for homologous recombination into the genome, we first amplified a 2.3 kb genomic fragment (coordinates ChrIII:6106174 – 6108438) containing the ChrIII center site and cloned this into pBluescript SK(+). Control or EER reporter cassettes were inserted into the center of homology sequences, within the 20 nt Cas9-sgRNA targeted sequence, 3-6 nts proximal to the PAM. We then cloned a 2.3 kb genomic fragment (coordinates ChrIII:385139 – 387463) containing the ChrIII arm site into Bluescript SK(+), and similarly inserted the control or EER *Pbaf-1::GFP* reporter cassettes.

CRISPR/Cas9-mediated single copy transgene integration

We cloned Cas9 guide RNA target sequences in the *PU6:sgRNA* expression plasmid pMLS134 as described in Schwartz and Jorgensen (2016).

For each integration, we used standard microinjection protocols to inject *unc-119(ed3)* (EG6249) young adult gonads with mixed plasmids as follows. Each injection mix included the following: 30 ng/µl *Peft-3::Cas9-SV40_NLS::tbb-2 3'UTR* (Addgene plasmid 46168), 30 ng/µl *PU6::sgRNA*, 30 ng/µl *Pbaf-1::GFP* reporter gene repair template, 4 ng/µl pCFJ104 (*Pmyo-3::mCherry::unc-54 3'UTR*), 4 ng/µl pGH8 (*Prab-*

3::mCherry::unc-54 3'UTR), and 2 ng/µl pCFJ90 (*Pmyo-2::mCherry::unc-54 3'UTR*). Injected P_0s were grown at 22°-25°C for 8-10 days before screening. Progeny were screened for non-Unc worms lacking mCherry fluorescence. Insertions were verified by PCR using primer sets that included one primer within the transgene and one primer outside the cloned homology region.

LoxP-flanked *unc-119(+)* cassettes were removed by injecting lines homozygous for each integration with 50 ng/µl, 2 ng/µl pCFJ90, and 48 ng/µl pBluescipt SK(+). F_1 progeny expressing mCherry were picked onto new plates and Unc F_2 progeny were collected and genotyped by PCR. Those containing the integrated reporter were crossed N2 males to remove the *unc-119(ed3)* mutation.

Acknowledgements

References

Ahringer, J., and Gasser, S.M. (2018). Repressive chromatin in Caenorhabditis elegans: establishment, composition, and function. Genetics *208*, 491-511.

Bass, B.L., Hurst, S.R., and Singer, J.D. (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, A.C., Fischer, S.E., and Kim, J.K. (2014). Endogenous RNAi pathways in C. elegans. WormBook (ed. The C. elegans Research Community).WormBook doi:

10.1895/wormbook.1.170.1, http://www.wormbook.org.

Blango, M.G., and Bass, B.L. (2016). Identification of the long, edited dsRNAome of LPS-stimulated immune cells. Genome Res. *26*, 852-862.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and mechanisms of miRNAs and siRNAs. Cell *136*, 642-655.

Castella, S., Bernard, R., Corno, M., Fradin, A., and Larcher, J.C. (2015). Ilf3 and NF90 functions in RNA biology. Wiley Interdiscip. Rev. RNA *6*, 243-256.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71-86.

Consortium, T.C.e.S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012-2018.

Cutter, A.D., Dey, A., and Murray, R.L. (2009). Evolution of the Caenorhabditis elegans genome. Mol. Biol. Evol. *26*, 1199-1234.

Evans, K.J., Huang, N., Stempor, P., Chesney, M.A., Down, T.A., and Ahringer, J. (2016). Stable Caenorhabditis elegans chromatin domains separate broadly expressed and developmentally regulated genes. Proc. Natl. Acad. Sci. USA *113*, E7020-E7029.

Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. *41*, 331–368.

Finerty, P.J.J., and Bass, B.L. (1999). Subsets of the zinc finger motifs in dsRBP-ZFa can bind double-stranded RNA. Biochemistry *38*, 4001-4007.

Frokjaer-Jensen, C., Jain, N., Hansen, L., Davis, M.W., Li, Y., Zhao, D., Rebora, K., Millet, J.R.M., Liu, X., Kim, S.K.*, et al.* (2016). An abundant class of non-coding DNA can prevent stochastic gene silencing in the C. elegans germline. Cell *166*, 343-357.

Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature *470*, 284-288.

Gu, S.G., Pak, J., Guang, S., Maniar, J.M., Kennedy, S., and 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, J.E., and Kiebler, M.A. (2014). The multifunctional Staufen proteins: conserved roles from neurogenesis to synaptic plasticity. Trends Neurosci. *37*, 470-479.

Herbert, A., and 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. USA *98*, 12132-12137.

Hillier, L.W., Miller, R.D., Baird, S.E., Chinwalla, A., Fulton, L.A., Koboldt, D.C., and Waterston, R.H. (2007). Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. *5*, e167.

Ho, J.W., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.A., Minoda, A., Tolstorukov, M.Y., Appert, A.*, et al.* (2014). Comparative analysis of metazoan chromatin organization. Nature *512*, 449-452.

Hundley, H.A., Krauchuk, A.A., and Bass, B.L. (2008). C. elegans and H. sapiens mRNAs with edited 3' UTRs are present on polysomes. RNA *14*, 2050-2060.

Ikegami, K., Egelhofer, T.A., Strome, S., and Lieb, J.D. (2010). Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. Genome Biol. *11*, R120.

LeGendre, J.B., Campbell, Z.T., Kroll-Conner, P., Anderson, P., Kimble, J., and 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, E.Y., Paz, N., Eisenberg, E., and Ast, G. (2007). RNA-editing-mediated exon evolution. Genome Biol. *8*, R29.

Liscovitch, N., Bazak, L., Levanon, E.Y., and 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, T.A., Vielle, A., Latorre, I., Cheung, M.S., 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., and Guang, S. (2015). The Nrde pathway mediates small-RNA-directed histone H3 lysine 27 trimethylation in Caenorhabditis elegans. Curr. Biol. *25*, 2398-2403.

Markham, N.R., and 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.

Park, E., and Maquat, L.E. (2013). Staufen-mediated mRNA decay. Wiley Interdiscip. Rev. RNA *4*, 423-435.

Prachumwat, A., DeVincentis, L., and Palopoli, M.F. (2004). Intron size correlates positively with recombination rate in Caenorhabditis elegans. Genetics *166*, 1585-1590.

Reich, D.P., Tyc, K.M., and Bass, B.L. (2018). C. elegans ADARs antagonize silencing of cellular dsRNAs by the antiviral RNAi pathway. Genes Dev. *32*, 271-282.

Schwartz, M.L., and Jorgensen, E.M. (2016). SapTrap, a toolkit for high-throughput CRISPR/Cas9 gene modification in Caenorhabditis elegans. Genetics *202*, 1277-1288.

Tian, B., Bevilacqua, P.C., Diegelman-Parente, A., and Mathews, M.B. (2004). The double-stranded-RNA-binding motif: interference and much more. Nat. Rev. Mol. Cell Biol. *5*, 1013-1023.

Towbin, B.D., Gonzalez-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., Askjaer, P., and Gasser, S.M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. Cell *150*, 934-947.

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

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nat. Rev. Mol. Cell Biol. *16*, 178-189.

Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y., and 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, J.M., Youssef, O.A., Aruscavage, P.J., Nix, D.A., Hong, C., Johnson, W.E., and Bass, B.L. (2015). Genome-wide profiling of the C. elegans dsRNAome. RNA *21*, 786-800.

Wolkowicz, U.M., and Cook, A.G. (2012). NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. Nucleic Acids Res. *40*, 9356-9368.

Zhao, H.Q., Zhang, P., Gao, H., He, X., Dou, Y., Huang, A.Y., Liu, X.M., Ye, A.Y., Dong, M.Q., and Wei, L. (2015). Profiling the RNA editomes of wild-type C. elegans and ADAR mutants. Genome Res. *25*, 66-75.

A Editing-enriched region analysis

B Intron stability analysis

C Inverted and tandem repeat prediction
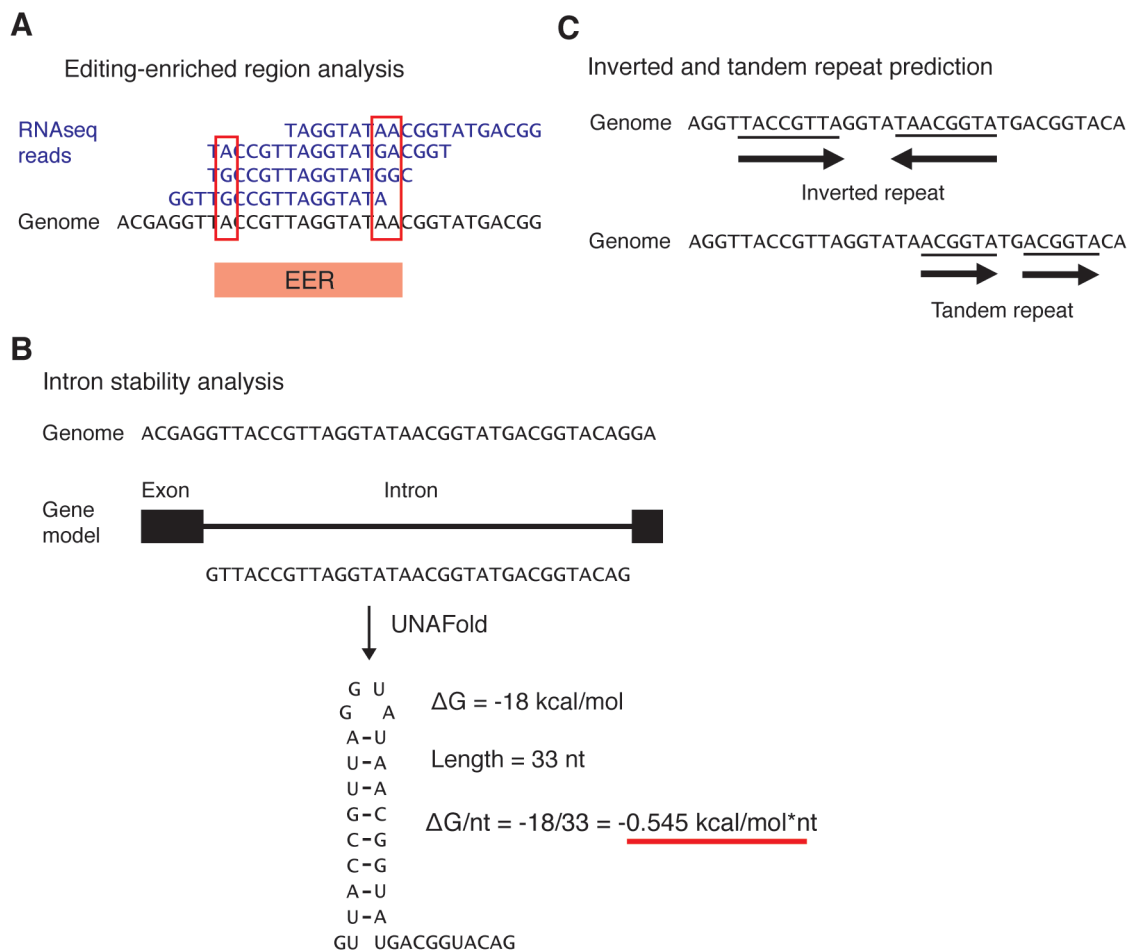
Figure 3.1 Methods to identify dsRNA loci. (*A*) Clustered A-to-G changes in RNAseq reads defined editing-enriched regions, transcribed regions that form long dsRNA *in vivo* and are edited by ADARs. (*B*) Intronic duplex structures were predicted using UNAFold to determine the folding free energy of intronic sequences (Markham and Zuker, 2008). 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 and tandem repeat sequences were determined computationally from the *C. elegans* genome sequence using previously published tools (Materials and Methods) (Benson, 1999; Warburton et al., 2004).

Figure 3.2 dsRNAs cluster on the distal arms of two *Caenorhabditis* species. (*A*) The distribution of EERs (black) and random expressed regions (grey) over the six major *C. elegans* chromosomes divided into ten 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*) The distribution of inverted repeats (IRs; black) and tandem repeats (TRs; grey) across relative chromosomal segments. (*C*) The percentage of the entire *C. elegans* genome, IR regions, and TR regions that overlap annotated gene features. (*D*) Smoothed distributions of average intron ΔG/nt across *C. elegans* (blue) and *C. briggsae* (red) chromosomes. More stable introns have lower (more negative) ΔG/nt. Vertical dotted lines depict chromosome arm and center domains, determined by intron ΔG/nt (Materials and Methods).
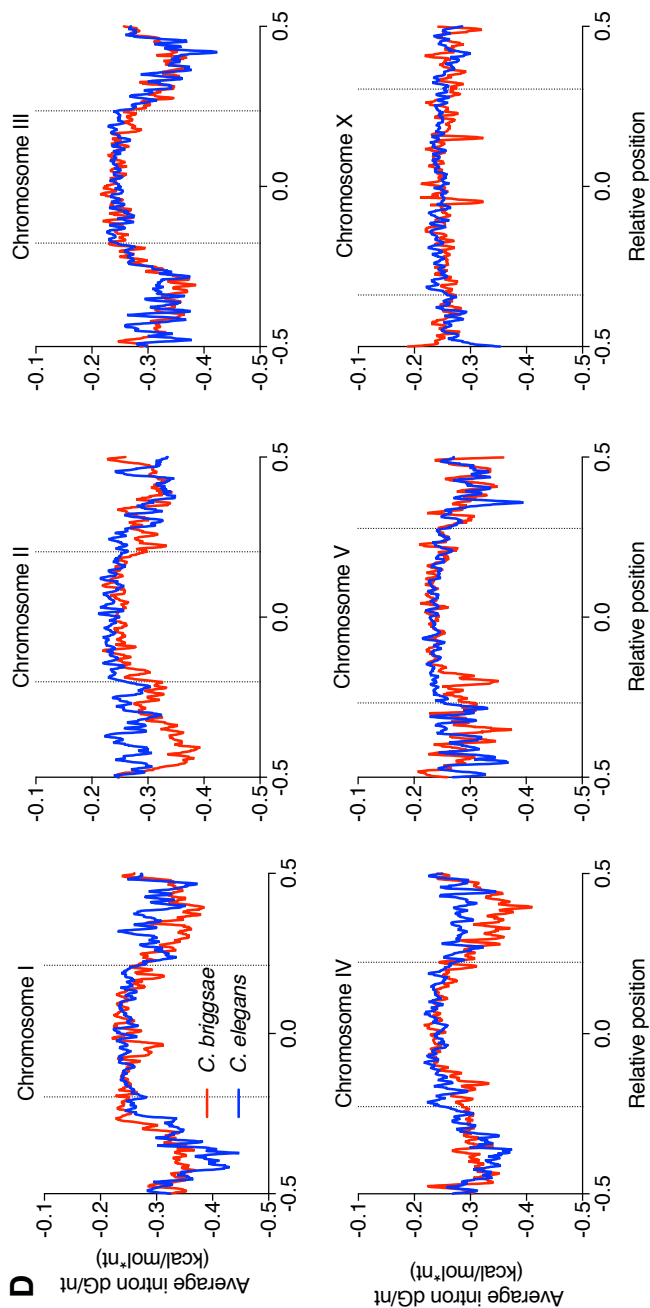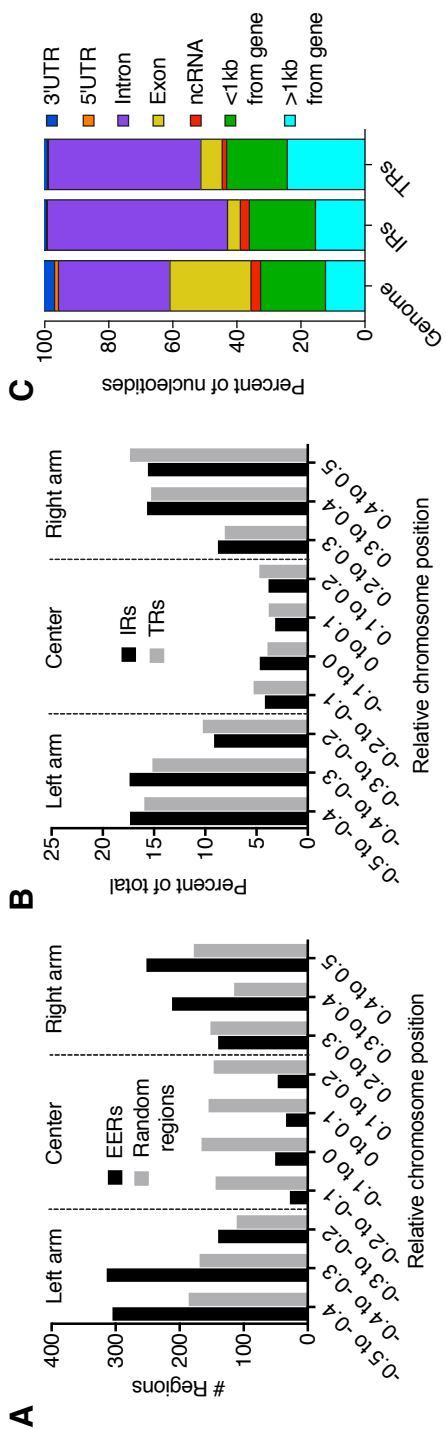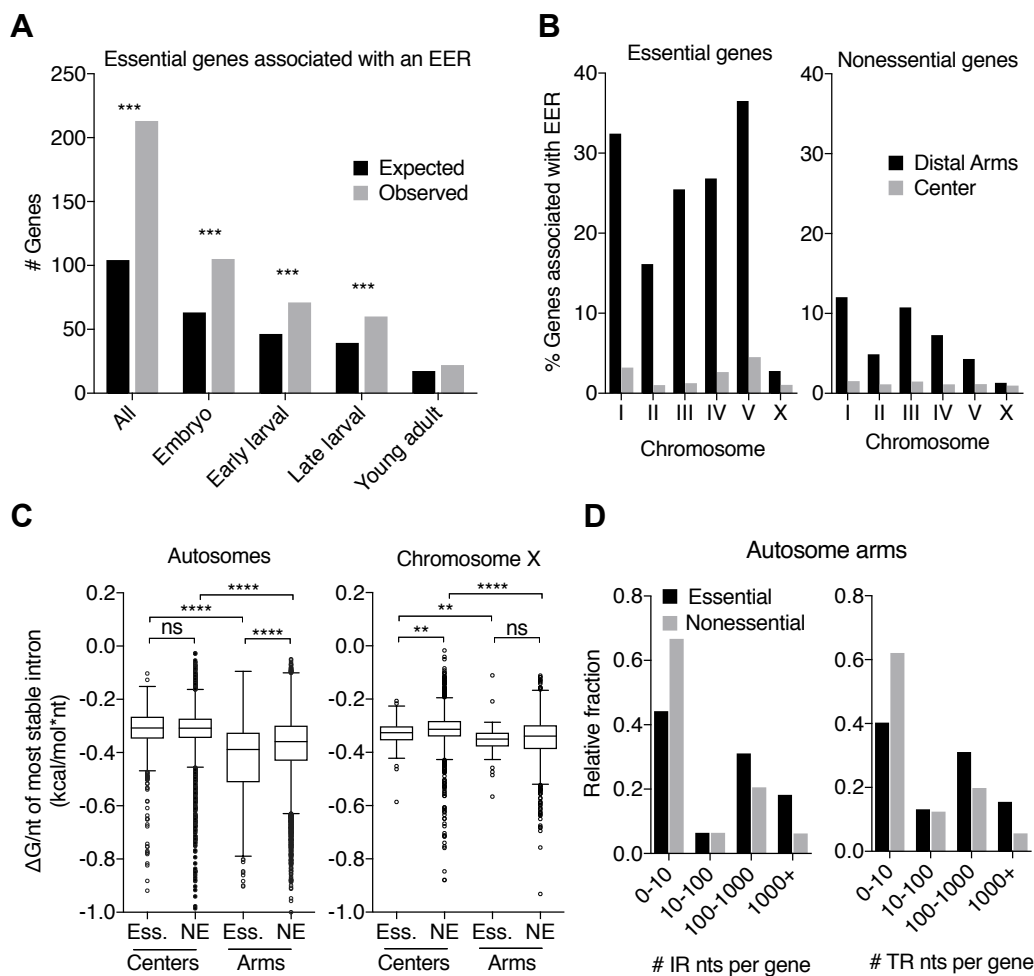
Figure 3.3 Duplex RNA structures are enriched in essential genes. (*A*) The number of essential genes expected by chance to overlap genes associated with EERs from all stages or each development stage, compared to the actual observed number of essential genes that overlapped EAGs. ***: $P < 0.001$, χ^2 test. (*B*) The percentage of essential (left) and nonessential (right) genes associated with an EER on the distal arm and center domains of each chromosome. (*C*) Distributions of ΔG/nt values of the most stable intron in essential and nonessential genes in center and arm domains of *C. elegans* autosomes (left) and Chromosome X (right). (*D*) Relative fractions of essential and nonessential genes on autosome arms, binned by the amount of IR (left) or TR (right) sequence in each gene's noncoding elements (i.e. introns, UTRs).

Figure 3.4 dsRNAs are associated with highly expressed genes and marks of transcriptional elongation. (*A*) Tukey box-plots show the distributions of gene expression in embryos for all genes or EAGs on distal arm or center domains of autosomes (left) or Chromosome X. ns: $P > 0.05$, *: $P < 0.05$; **: $P < 0.01$, ****: $P < 0.0001$, Mann-Whitney *U*-test. (*B*) Tukey box-plot, as in (*A*), showing expression in each developmental stage of all genes or EAGs on autosome distal arms. ****: $P < 0.0001$, Mann-Whitney *U*-test. (*C*) Tukey box-plots as in (*A*) showing embryo-stage expression of all genes or genes containing a structured intron (Struct.; intron $\Delta G/nt < -0.5$ kcal/mol*nt), broken down by chromosomal domain. *: $P < 0.05$; ****: $P < 0.0001$, Mann-Whitney *U*-test. (*D*) Tukey box-plot, as in (*A*), showing embryo-stage expression of all genes, EAGs, or genes containing 3 or more EER-associated introns and/or UTRs (3+ EER) present on autosome arms. (*E*) Heatmap displays the relative ChIP-chip signal for 19 histone modifications over EERs, random regions, structured ($\Delta G/nt < -0.5$ kcal/mol*nt) or unstructured introns, IR loci, and TR loci present on autosome distal arms.
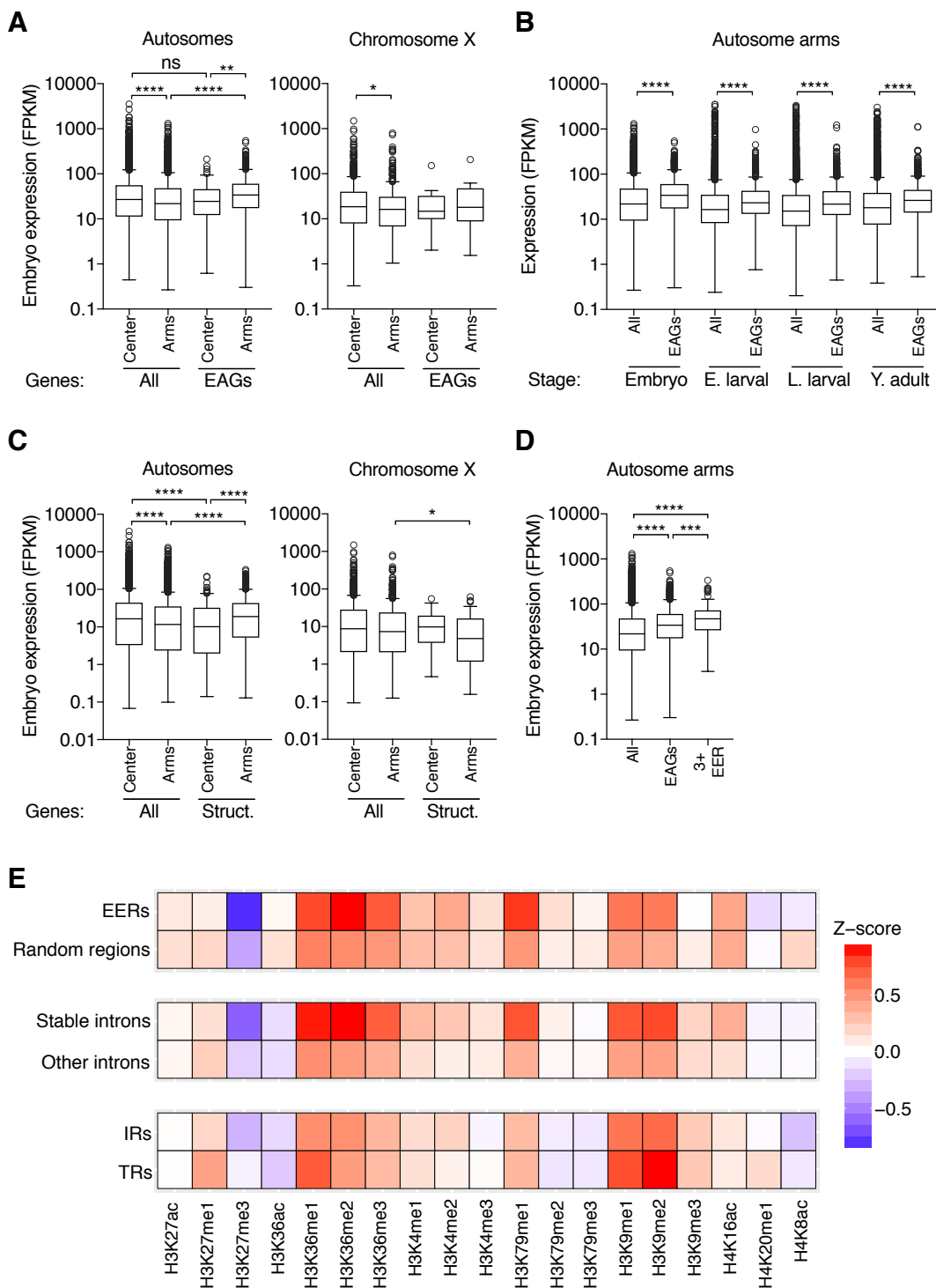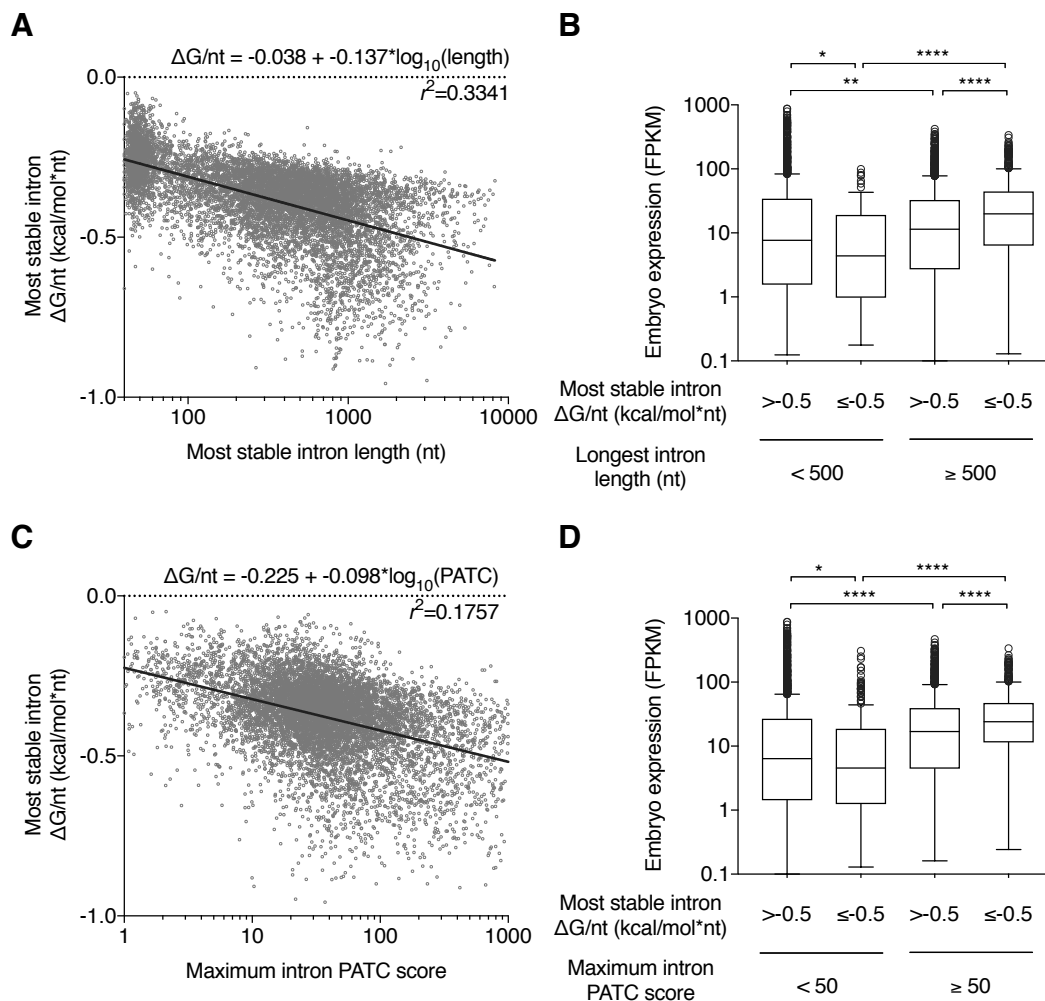
Figure 3.5 Correlations of intron structure, length, and PATC content with gene expression. (*A*) Scatter plot shows autosome arm genes by the length and ΔG/nt of their most stable intron. Semi-log regression analysis (black line; formula at top) reveals a moderate correlation. (*B*) Tukey box-plots showing embryo-stage expression of autosome arm genes divided by the length of their longest intron and ΔG/nt of their most stable intron. *: $p < 0.05$; **: $p < 0.01$; ****: $p < 0.0001$, Mann-Whitney *U*-test. (*C*) Scatter plot, as in (*A*), shows genes by most stable intron ΔG/nt and maximum intron PATC score (see Materials and Methods), with a semi-log regression in black (formula at top). (*D*) Tukey box-plot, as in (*B*), shows embryo-stage expression of genes divided by most stable intron ΔG/nt and maximum intron PATC score. *: $p < 0.05$; ****: $p < 0.0001$, Mann-Whitney *U*-test.
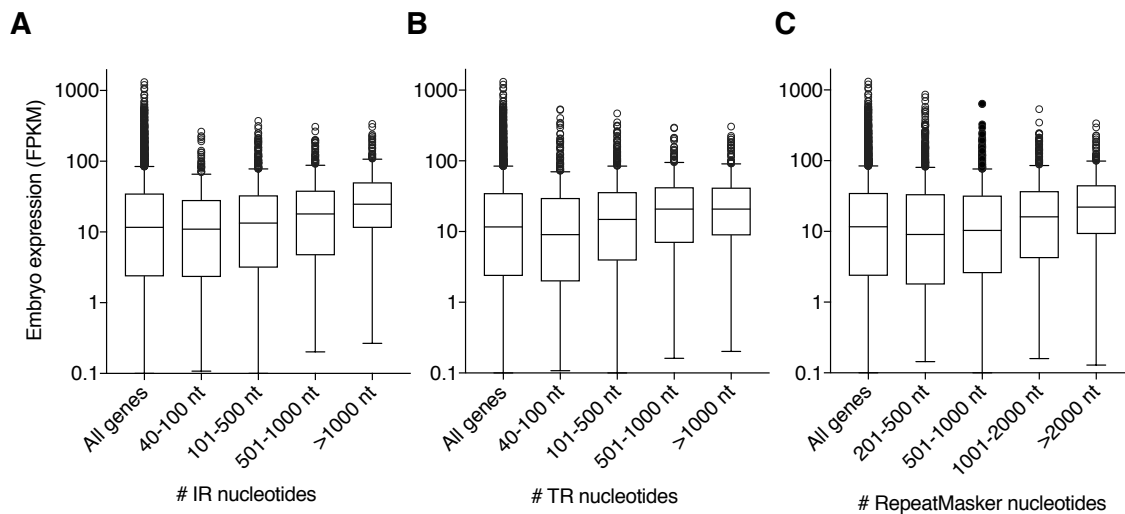
Figure 3.6 Expression of autosome arm genes by repeat content. Tukey box-plots show embryo-stage expression of autosome arm genes separated by the amount of noncoding sequence in each gene that overlaps (*A*) IRs, (*B*) TRs, or (*C*) annotated RepeatMasker transposon repeats.

Figure 3.7 Effects of EERs on *GFP* reporter expression. (*A*) Gene schematic of control (top) and EER-containing *GFP* reporter cassettes. (*B*) Chromosome III sites where reporter genes were integrated. (*C*) Fluorescent images of adults carrying *GFP* reporter transgenes in each integration site. Scale bar = 100 μm. (*D*) Quantification of adult-stage fluorescence in control (-) and EER-containing (+) reporters. Error bars reflect standard error of the mean, four experiments of n>10 worms each. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$, Student's t-test. (*E*) Quantitative RT-PCR analysis of *GFP* mRNA normalized to the geometric mean of *ndk-1, Y45F10D.4*, and *cdc-42* mRNAs, show expression relative to the control reporter at the ChrIII center site. Error bars show standard deviation, n = 4 biological replicates. (*F*) Adult-stage fluorescence of GFP reporters in wildtype or *adr-2* mutant backgrounds. ND: not determined. Error bars show standard deviation, n ≥ 12 worms. (*G*) Adult-stage fluorescence of GFP reporters integrated at the ChrIII center site containing EERs in a 3'UTR, intron, or neither. Error bars show standard deviation, n ≥ 10 worms. ****: $p < 0.0001$, Student's t-test. (*H*) Gene schematic of *GFP* reporter cassettes containing an inverted repeat sequence or a tandem repeat sequence in the 3'UTR. (*I*) Adult-stage fluorescence of GFP reporters containing inverted or tandem repeats in the 3'UTR sequence. Error bars show standard deviation, n = 14 worms. ****: $p < 0.0001$, Student's t-test.

A Control reporter

Pbaf-1 — Synthetic intron — *unc-54* 3'UTR

EER reporter

Pbaf-1 — EER — *ssl-1* intron 1 — EER — *eif-2alpha* 3'UTR

B

Arm site / Center site

Left arm — ChrIII center — Right arm

C Control reporter | EER reporter

Center site

Arm site

D Fluorescence

Fluorescence above background (AU/pixel)

EERs: - + - +
Site: Center Arm

E mRNA

Relative *gfp* mRNA

EERs: - - + - +
Site: - Center Arm

F

Fluorescence above background (AU/pixel)

Wildtype / *adr-2⁻/⁻*

EERs: - + - +
Site: Center Arm

ND

G

Fluorescence above background (AU/pixel)

Intron EER: - - +
3'UTR EER: - + +

H

Inverted

Pbaf-1 — EER

Tandem

Pbaf-1

I

Fluorescence above background (AU/pixel)

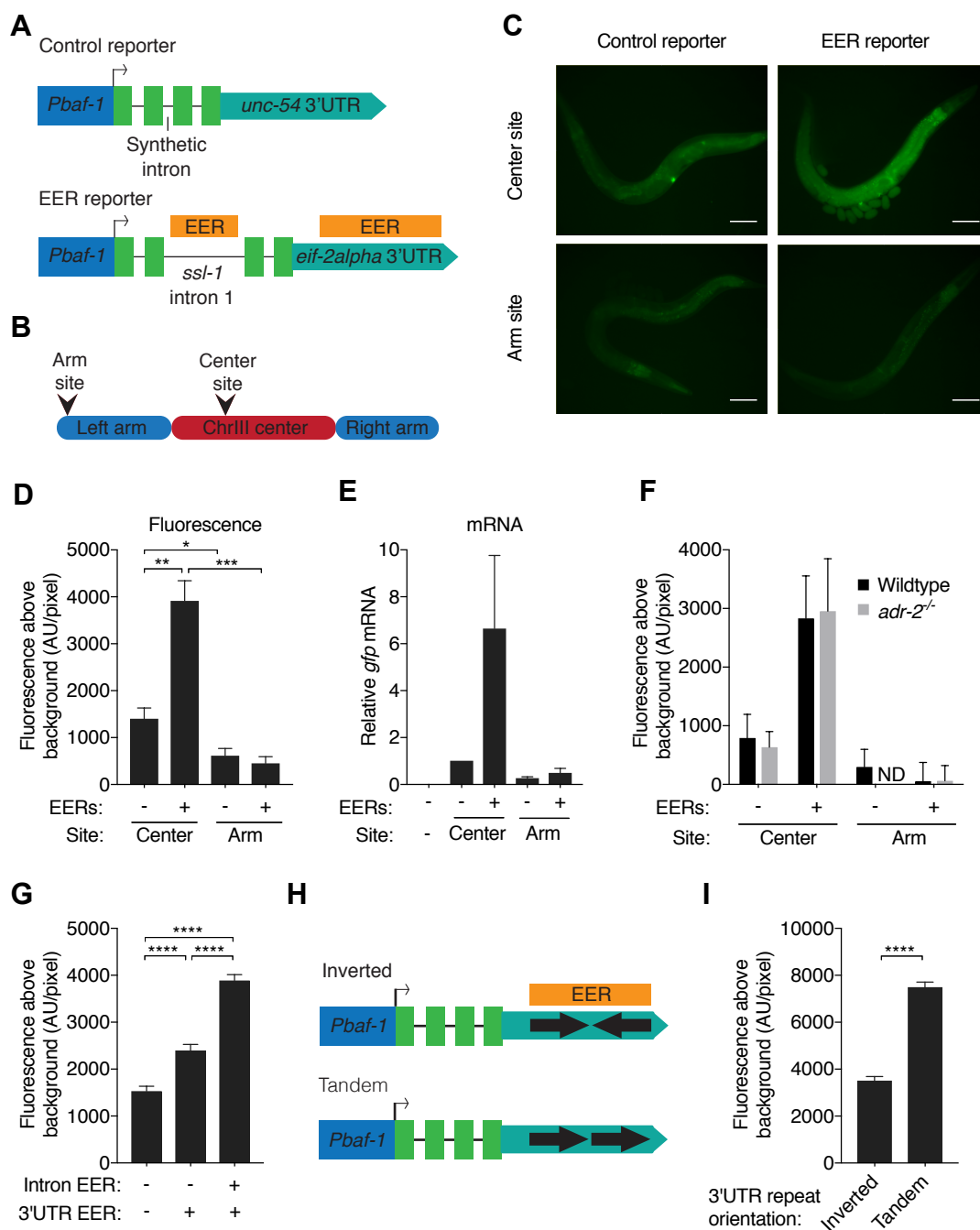3'UTR repeat orientation: Inverted / Tandem

Figure 3.8 Models to explain dsRNA association with elevated gene expression. (*A*) A dsRNA-containing gene houses two inverted transposable elements (TEs) in an intron. Each TE copy contains a transcription factor (TF) binding site. TF binding to TE sequences leads to transcriptional activation, promoting gene expression. (*B*) Nucleosomes (gray shapes) at a gene locus prevent spurious transcription. Upon transcription, nucleosomes are disassembled in front of the RNA polymerase (Pol II) complex and reassembled behind it, but formation of a dsDNA duplex in the non-template strand prevents nucleosome reassembly. As a result, nucleosomes are not reestablished after transcription, leading to spurious expression of the locus. (*C*) A locus containing inverted TEs in its intron is transcribed to RNA. The U1 and U2 snRNPs bind sequences around the 5' and 3' splice sites. Fold back of complementary sequences in the intron brings U1 and U2 snRNPs in close proximity, promoting their association and recruiting the U4/U5/U6 tri-snRNP to catalyze splicing. Efficient splicing due to dsRNA formation promotes elevated gene expression. (*D*) Transcription of a dsRNA locus leads to dsRNA formation in the nascent transcript. An undetermined dsRBP binds the dsRNA structure and recruits Pol II and other components of the transcription machinery to further promote gene expression.
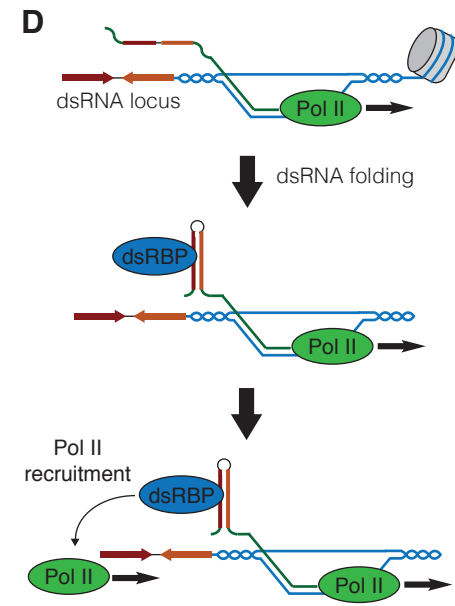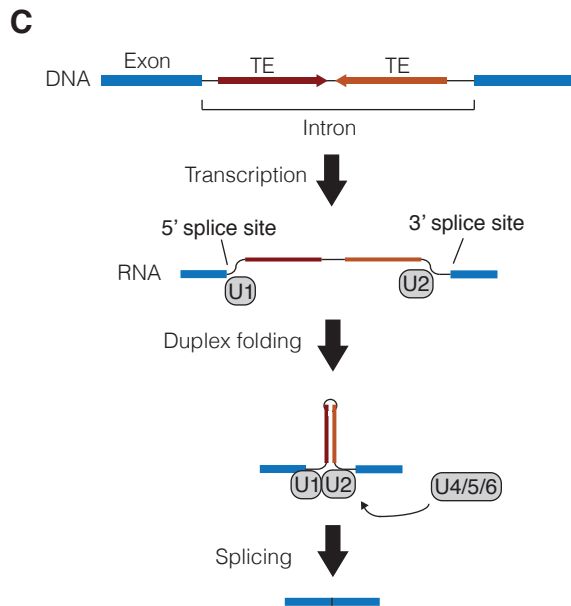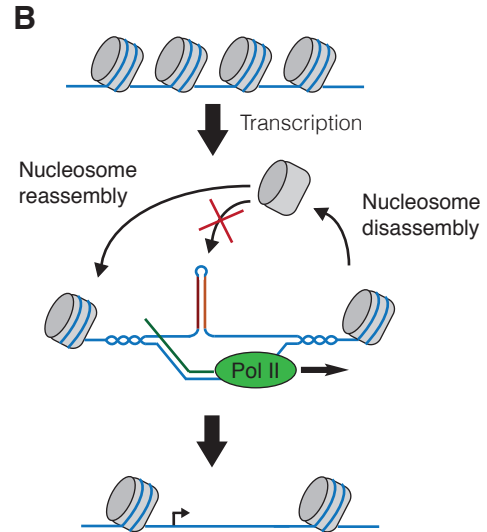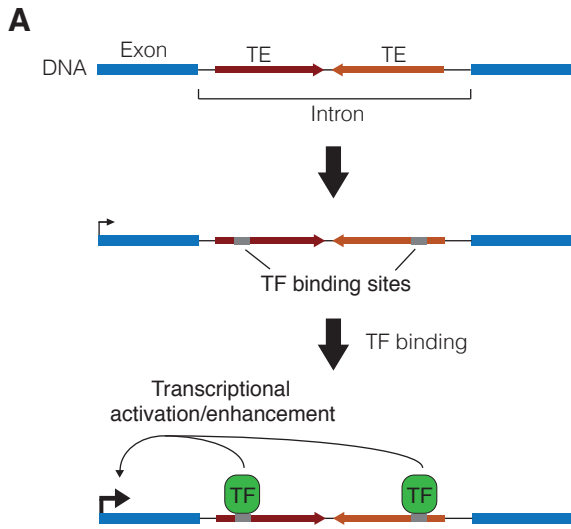
Table 3.1: Primers used in this study

Primer name	Sequence
pBlue_Pbaf1_F1	CGAGGTCGACGGTATCGATAAGCTTGATATCGTGTCAGCGACATACGAATGAATCG
GFP_Pbaf1_R1	CAGTGAAAAGTTCTTCTCCTTTACTCATGGTTTCTGAAACACAAAATAATTACATTCTTG
Pbaf1_GFP_F2	CAAGAATGTAATTATTTTGTGTTTCAGAAACCATGAGTAAAGGAGAAGAACTTTTCACTG
Pbaf1_pBlue_R2	CGATTCATTCGTATGTCGCTGACACGATATCAAGCTTATCGATACCGTCGACCTCG
GFP_Y37E_3UTR_F1	GCATGGATGAACTATACAAATAGTTATTTGAAAATTGAAAATTGAAAATTCCCATTG
pBlue_Y37E_3UTR_R1	CTAGAACTAGTGGATCCCCCGGGCTGCAGGCGAATTGGGAGCCCACAAAAATGG
Y37E_3UTR_pBlue_F2	CCATTTTTGTGGGCTCCCAATTCGCCTGCAGCCCGGGGGATCCACTAGTTCTAG
Y37E_3UTR_GFP_R2	CAATGGGAATTTTCAATTTTCAATTTTCAAATAACTATTTGTATAGTTCATCCATGC
ssl1_Int1_F2	GTACGATTTTTTAAATTTAATTACTTTCCTCAAATCC
ssl1_Int1_R2	CTGAAAAACATTAATTCATAATTTTGAAATGTAAC
ssl1_Int1_GFPEx3_F3	GTTACATTTCAAAATTATGAATTAATGTTTTTCAGGTGCTGAAGTCAAGTTTGAAGGTG
ssl1_Int1_GFPEx2_R3	GGATTTGAGGAAAGTAATTAAATTTAAAAAATCGTACGTGTCTTGTAGTTCCCGTCATC
pBlue_chr3arm_frag_F1	CGAGGTCGACGGTATCGATAAGCTTGATATCGCCTAAAGAACTAAGTCCACAGCGAGC
pBlue_chr3arm_frag_R1	CACTAAAGGGAACAAAAGCTGGAGCTCGGTACACCTTCTGATTTCACCAAAAATCTCGTC
chr3arm_Pbaf1_F1	CGAGGAGTCCAAAAAACTTTGTAAGCATGGATGCCGCTGTCAGCGACATACGAATGAATC
chr3arm_Y37E_3UTR_R1	GTCTAGACCCTTATCCGACCAGGTTTGGCTCCGCCGAATTGGGAGCCCACAAAAATGG
chr3arm_unc54_3UTR_R2	GTCTAGACCCTTATCCGACCAGGTTTGGCTCCGCGAAACAGTTATGTTTGGTATATTGG
pBlue_chr3body_frag_F1	CATTCGCCATTCAGGCTGCGCAACTGCTATATTGACCAGGACTGTCTACTTTTAAAGTTG
pBlue_chr3body_frag_R1	GAAGCGGAAGAGCGCCCAATACGCAAACGTCTCCTGTACTAACCTCGTCTTTTTTGTAG
chr3_body_Pbaf1	CAGGAAATGGATAGTAGACAGGGGAGGCAGTGTCAGCGACATACGAATGAATCG
chr3body_Y37E_3UTR_R1	CAAATTTCCGTTTTCGGCACCCAAAGCAGCAGCGAATTGGGAGCCCACAAAAATGG
chr3body_unc54_3UTR_R2	CAAATTTCCGTTTTCGGCACCCAAAGCAGCAGGAAACAGTTATGTTTGGTATATTGG
Y37E_3UTR_rep2_F1	GTAACTTTTATTTGTCCCCCAC

Table 3.1 continued

Primer name	Sequence
Y37E_3UTR_rep2 _R1	AATAGAAGGTCAAAGGTGGAG
Y37E_3UTR_loop _rep2rev_F1	CATTGGGGCACAAAAATGTAACTTTTATCAAAGGT GGAGTATCGAAACC
Y37E_3UTR_3pEnd _rep2rev_R1	GGATGGGAGACAATAAATAGAAGGTTTGTCCCCC ACTGGTCG
Y37E_3UTR_rep2rev _3pEnd_F2	CGACCAGTGGGGGACAAACCTTCTATTTATTGTCT CCCATCC
Y37E_3UTR_rep2rev _loop_R2	GGTTTCGATACTCCACCTTTGATAAAAGTTACATTT TTGTGCCCCAATG
3Arm_pMLS134_F	TTGGATGCCGCGGAGCCAAACC
3Arm_pMLS134_R	GGTTTGGCTCCGCGGCATCCAA
3Body_pMLS134_F	TTGGAATTGGAATGAAACAGAC
3Body_pMLS134_R	GTCTGTTTCATTCCAATTCCAA
gfp_RTPCR_F1	TACCTGTTCCATGGCCAACAC
gfp_RTPCR_R1	ACCTTCAAACTTGACTTCAGC
ndk1_RTPCR_F1	GAGTCCACCGGAGTCCACCG
ndk1_RTPCR_R1	CTCAAGATGCTCAAGATG
Y45F10D4_RTPCR_F1	CGAGAACCCGCGAAATGTCGGA
Y45F10D4_RTPCR_R1	CGGTTGCCAGGGAAGATGAGGC
cdc42_RTPCR_F1	AGCCATTCTGGCCGCTCTCG
cdc42_RTPCR_R1	GCAACCGCTTCTCGTTTGGC

CHAPTER 4

PERSPECTIVES

Work described in the preceding chapters explored functions of ADAR A-to-I editing enzymes and their dsRNA substrates in the nematode *C. elegans*. In Chapter 2, I characterized edited dsRNAs expressed in different *C. elegans* developmental stages and showed that ADARs prevent processing and silencing of their substrates by antiviral RNAi. This work complements growing evidence that mammalian ADAR1 prevents cellular dsRNAs from triggering MDA5-dependent antiviral signaling, demonstrating that ADARs have a conserved role in distinguishing endogenous self dsRNAs from viral nonself dsRNAs. In Chapter 3, I described how *C. elegans* dsRNA loci cluster on distal autosome arms and are associated with essential and highly expressed genes, suggesting that duplex structures might serve important regulatory roles though an undescribed mechanism. In this section, the broader implications of these studies will be discussed, with particular focus paid to future avenues of research based on these findings, including preliminary data from relevant experiments.

What dsRNAs are relevant to ADAR mutant phenotypes?

Several lines of evidence argue that ADARs are necessary to prevent immune activity against endogenous dsRNAs. First, mouse *ADAR1* mutations cause IFN

activation and embryonic lethality, and these phenotypes are rescued by mutating either the innate immune effector gene *MAVS*, or *IFIH1*, which encodes the antiviral dsRNA sensor MDA5 (Liddicoat et al., 2015; Mannion et al., 2014; Pestal et al., 2015). Second, deleting *adr-1* and *adr-2* in *C. elegans* leads to processing of endogenous dsRNAs into siRNAs with characteristics similar to DCR-1-dependent antiviral siRNAs (Ashe et al., 2013; Reich et al., 2018; Wu et al., 2011). Finally, components of the antiviral RNAi machinery are required for bursting and low brood size phenotypes that result from *adr-1;adr-2* deletion in *rrf-3* or *ergo-1* mutant backgrounds (Reich et al., 2018).

While these results show that ADARs are required to prevent aberrant immune activity in response to cellular dsRNA, they offer little insight into which *specific* dsRNAs ADARs must bind and edit to prevent harmful immune responses. Because dsRBPs bind their substrates sequence-nonspecifically, all cellular dsRNAs likely have the potential to activate dsRNA-responsive immune mechanisms. Indeed, the fact that 74% of EERs gave rise to monophosphorylated siRNAs suggests that ADARs prevent processing of most *C. elegans* dsRNAs by the antiviral DCR-1 complex (Reich et al., 2018). EAGs are enriched for Gene Ontology terms associated with vulval morphology and development, indicating that global EAG downregulation in *adr-1;adr-2;rrf-3* mutants might result in vulval defects and bursting. However, misregulation of a single gene, *lin-41*, can cause bursting in *C. elegans* (Ecsedi et al., 2015), suggesting that *adr-1;adr-2;rrf-3* mutant bursting could relate to silencing of *lin-41* or a limited number of EAGs. In mammals, some dsRNAs may be expressed too lowly to trigger an IFN response. Other dsRNAs, like those in introns, may remain primarily in the nucleus, sequestered away from cytosolic dsRNA sensors that initiate immune signaling. Thus, the

question remains, which ADAR substrates mediate the immune-relevant phenotypes of ADAR mutant animals?

<u>Defining spatiotemporal requirements of *adr-1;adr-2*-dependent</u>

<u>phenotypes</u>

To approach this question in *C. elegans*, I designed a series of experiments to identify transcripts relevant to the bursting phenotype of *adr-1;adr-2;rrf-3* mutants. First, the tissues and developmental stages where ADAR normally functions to prevent bursting must be determined. Defining spatiotemporal requirements of *adr-1;adr-2;rrf-3* mutant bursting will narrow down where and when relevant transcripts are silenced, allowing a more directed search for candidate transcripts. To elucidate when and where ADARs act to prevent bursting, I constructed a set of genetic tools to control ADR-1 and ADR-2 stability in a spatiotemporal manner. These tools employ the auxin-inducible degron (AID) system derived from plants (Nishimura et al., 2009) and recently applied in *C. elegans* (Zhang et al., 2015). The AID system uses the *Arabidopsis*-derived E3 ubiquitin ligase TIR1 to direct ubiquitination and subsequent proteosomal degradation of proteins carrying a 44-amino acid degron sequence, but only when the small molecule 3-indole acetic acid, also known as the plant hormone auxin, is present (Nishimura et al., 2009). To apply the AID system to *C. elegans* ADAR proteins, I first used CRISPR protocols to insert a 174-nt sequence encoding a 1x-Flag tag, 44-amino acid degron sequence, and 6-amino acid Gly-Ser linker just after the start codons of the endogenous *adr-1* and *adr-2* genes (Fig. 4.1A). I refer to these insertions as *adr-1(AID)* and *adr-2(AID)*. Next, *adr-1(AID);adr-2(AID)* mutants were crossed to one of four *C. elegans*

strains carrying a single-copy transgene expressing *TIR1* from a tissue-specific promoter, either *Peft-3* (pan-somatic), *Psun-1* (germline), *Pmyo-2* (pharyngeal muscle), or *Pges-1* (intestine) (Zhang et al., 2015). *TIR1*-expressing worms carrying the *adr-1(AID);adr-2(AID)* insertions should rapidly degrade both ADR-1 and ADR-2 in the presence of auxin, which *C. elegans* readily take up (Zhang et al., 2015).

To validate that the ADAR-AID system degrades ADARs in a tissue-specific manner, I crossed the repetitive, integrated transgene *uuIs1[Psur-5::sur-5::GFP; Phsp-16.2::GFP(IR)]* into *adr-1(AID);adr-2(AID)* backgrounds carrying *Peft-3::TIR1, Pmyo-2::TIR1*, or *Pges-1::TIR1*. The *uuIs1* transgene constitutively expresses nuclear GFP in somatic tissues, but because it is repetitive, it gives rise to dsRNA and is silenced by RNAi in *adr-1;adr-2* mutant backgrounds (Knight and Bass, 2002). When I grew *adr-1(AID);adr-2(AID);Peft-3::TIR1;uuIs1* worms on media containing 1 mM auxin, I observed robust silencing of GFP in all somatic tissues, while worms grown without auxin strongly expressed GFP (Fig. 4.1B). This suggests that ADR-1 and ADR-2 are degraded in the presence of auxin and TIR1, leading to transgene silencing. Since GFP is expressed in *adr-1(AID);adr-2(AID);Peft-3::TIR1* worms without auxin, *adr-1(AID);adr-2(AID)* insertions do not interfere with wildtype ADAR functions. I next grew *adr-1(AID);adr-2(AID);Pmyo-2::TIR1;uuIs1* and *adr-1(AID);adr-2(AID);Pges-1::TIR1; uuIs1* worms on 1 mM auxin, observing GFP silencing only in tissues where *TIR1* was expressed, pharyngeal muscle and intestine, respectively. This suggests that ADAR degradation is tissue-specific and that *adr-1;adr-2*-dependent transgene silencing does not spread between tissues, as some other RNAi mechanisms do in *C. elegans* (Jose et al., 2011). Most importantly, our results demonstrate that the AID system can be used for

tissue-specific, auxin-dependent, ADAR loss-of-function.

To begin to define spatiotemporal requirements of *adr-1;adr-2;rrf-3* mutant bursting, I crossed the *rrf-3(uu56)* deletion with *adr-1(AID);adr-2(AID);Psun-1::TIR1*. Because other *TIR1*-expressing transgenes are integrated at a Chromosome II locus (oxTi179) only ~1 centiMorgan from *rrf-3*, I could not easily cross these strains with existing *rrf-3* deletions. Thus, I used CRISPR protocols to delete *rrf-3* exons 4-8 in *adr-1(AID);adr-2(AID);Peft-3::TIR1* using the same guide RNAs used to generate *rrf-3(uu56)* (Reich et al., 2018). Neither *adr-1(AID);adr-2(AID);rrf-3;Psun-1::TIR1* nor *adr-1(AID);adr-2(AID);rrf-3;Peft-3::TIR1* strains exhibited bursting in the absence of auxin. However, when these strains were grown in the presence of auxin from embryo stages, adults expressing somatic *TIR1*, but not germline *TIR1*, burst at frequencies similar to *adr-1;adr-2;rrf-3* mutants (Fig. 4.1C) (Reich et al., 2018). Notably, *adr-1(AID);adr-2(AID);rrf-3;Psun-1::TIR1,* which expresses germline *TIR1*, never burst, even when grown on auxin for several generations (data not shown). This suggests that *adr-1;adr-2;rrf-3* bursting is a somatic phenotype and does not involve germline ADAR loss-of-function. Because transgene silencing, chemotaxis, and longevity phenotypes of *adr-1;adr-2* mutants do not suggest germline ADAR activity (Knight and Bass, 2002; Sebastiani et al., 2009; Tonkin et al., 2002), this result is not wholly unexpected. However, ADARs may still have germline functions, since both *adr-1* and *adr-2* mRNAs are expressed in germline tissues (Ortiz et al., 2014), and *adr-1;adr-2* mutants have altered levels of 21U-RNAs/piRNAs (Warf et al., 2012), which are highly expressed in the germline (Billi et al., 2014).

To define when ADARs normally function to prevent bursting, I grew *adr-*

1(AID);adr-2(AID);rrf-3;Psun-1::TIR1 and *adr-1(AID);adr-2(AID);rrf-3;Peft-3::TIR1* strains on normal media, and then transferred worms to media containing 1 mM auxin at different developmental stages before measuring bursting on Day 5 after egg lay (Fig. 4.1C). Somatic *TIR1*-expressing *adr-1(AID);adr-2(AID);rrf-3;Peft-3::TIR1* worms burst roughly as frequently as *adr-1;adr-2;rrf-3* mutants when added to auxin in embryo, L1, or L2 stages. However, worms added to auxin at L3 stages showed markedly lower levels of bursting, while auxin-treated L4s almost never burst. Again, *adr-1(AID);adr-2(AID);rrf-3;Psun-1::TIR1* never exhibited bursting. These findings suggest that ADARs are required in the soma between L2 and L3 larval stages to prevent bursting in *rrf-3* mutants. Thus, the dsRNAs that mediate bursting are likely expressed in a somatic tissue at the same time. Since bursting requires RNAi activity, relevant transcripts are likely silenced after this time period. Depending on the time required for dsRNA processing, siRNA amplification, and target silencing, relevant transcripts may not be silenced until L3-L4 stages, when vulval patterning and morphogenesis occur (Sternberg, 2005).

The timeframe involved in bursting can be further narrowed by the inclusion of more experimental time-points and larger sample sizes. However, the above experiments suggest only that ADAR functions somatically to suppress bursting, offering few clues as to the specific tissue(s). Though requirements in the intestine and pharyngeal muscle can be tested using existing transgenes, testing other tissues will require the generation of transgenes that express *TIR1* under other promoters. The primary tissue of interest is the developing vulva, since I hypothesize that aberrant vulval development causes loss of structural integrity and subsequent bursting. Expressing *TIR1* in the P6.p vulval precursor cell under the *Plag-2* or *Pegl-17* promoters may facilitate ADAR loss-of-function in

vulval cells, to test if ADAR acts in these cells to prevent bursting.

Approaches to identify endogenous dsRNAs relevant to

adr-1;adr-2 phenotypes

The end goal of these experiments is to identify the gene(s) whose silencing causes *adr-1;adr-2;rrf-3* bursting. If possible, individual dsRNA structures required for this phenotype would also be defined. Towards this purpose, I propose performing cell-specific RNAseq to determine edited transcripts expressed in cells of interest within the timeframe involved in bursting. Heather Hundley's group has used cell-specific RNAseq on isolated *C. elegans* neurons to identify neural-specific editing and expression patterns (Deffit et al., 2017), and a similar approach could be applied to find bursting-relevant ADAR targets. Expression of a fluorescent protein would mark cells of interest, which could be isolated from wildtype and *adr-1;adr-2;rrf-3* mutant worms by fluorescence-activated cell sorting after dissociating worms into single cells. With relevant cells isolated, RNA would be extracted and analyzed by RNAseq. Then, candidate genes would be identified as those edited in wildtype tissues and silenced in *adr-1;adr-2;rrf-3* samples. Focusing on genes with known functions in vulval development could further narrow down the list of candidates. Edited dsRNA structures in candidate genes would be identified using the EER analysis pipeline (Blango and Bass, 2016; Reich et al., 2018; Whipple et al., 2015). To show that relevant genes and dsRNA structures are involved in bursting, one would overexpress candidate genes or delete candidate dsRNA structures in *adr-1;adr-2;rrf-3* mutants and test if these manipulations rescue bursting. Identification of the genes or dsRNAs involved in bursting would facilitate further characterization of

adr-1;adr-2;rrf-3 mutants by relieving their fitness defects without also disrupting RNAi function.

Using ADARs to study immunity

<u>Finding additional suppressors of *adr-1;adr-2;rrf-3* bursting</u>

Aberrant immune activity causes the deleterious phenotypes of *ADAR1* mutant mice and *adr-1;adr-2;rrf-3* mutant *C. elegans* (Liddicoat et al., 2015; Mannion et al., 2014; Pestal et al., 2015; Reich et al., 2018). Because disrupting immune-relevant genes rescues these phenotypes, one could in theory define factors required for the immune response by screening for mutations that rescue ADAR mutant phenotypes. Though some genetic screens are feasible, mice are not ideal for large-scale forward genetic screens due to their complex internal development, long generation times, low brood sizes, and costs required to house and breed large numbers of animals (Kile and Hilton, 2005). In contrast, given their prolific reproductive capacity, low costs of maintenance, and hermaphroditic genetics, *C. elegans* provide a powerful system for genome-wide screens, particularly suppressor screens (i.e. screens for mutations that suppress a phenotype) (Jorgensen and Mango, 2002). Thus, additional components of the *C. elegans* antiviral response could be identified by screening for suppressors of the *adr-1;adr-2;rrf-3* mutant bursting phenotype.

In practice, a screen for suppressors of *adr-1;adr-2;rrf-3* bursting could be performed in two ways. The first way would be to treat P_0 *adr-1;adr-2;rrf-3* mutants with a mutagen such as ethyl methane sulfonate, isolate F_2 progeny, and look for lines where no adults burst. This F_2 screen would be simple, unbiased, and would presumably

generate many suppressor mutations in a short period of time. However, mapping and identifying the mutations isolated from an F₂ screen would be time-consuming, even using high-throughput sequencing to find candidate genetic lesions. Further, since at least five genes in the antiviral RNAi pathway suppress *adr-1;adr-2;rrf-3* bursting, I suspect many hits in a simple F₂ suppressor screen would occur in the same genes or pathway and thus a large proportion of hits would likely be uninformative. The second approach, a genome-wide RNAi screen, simplifies the problem of candidate identification. Using this method, one would simply grow *adr-1;adr-2;rrf-3* mutants on *E. coli* expressing dsRNA to silence each gene in the genome and screen for genes whose silencing rescues bursting. Because *adr-1;adr-2;rrf-3* mutants presumably exhibit the same enhanced RNAi (Eri) phenotype as *rrf-3* single mutants (Simmer et al., 2002), gene silencing should be highly effective in this strain. However, before screening, one would need to ensure the feasibility of this approach by testing if RNAi against known suppressor genes like *rde-1*, *rde-4*, or *drh-1* rescues bursting. If so, a genome-wide RNAi screen could be a useful approach to identify additional factors involved in *adr-1;adr-2;rrf-3* bursting.

A candidate screen for RNAi factors that mediate *adr-1;adr-2*

transgene silencing

As proof-of-principle for the feasibility of screening *adr-1;adr-2*-dependent phenotypes, I performed a candidate RNAi screen to identify factors involved in *adr-1;adr-2* mutant transgene silencing (Table 4.1). I began by identifying candidate genes involved in endogenous (Billi et al., 2014) or exogenous RNAi (Grishok, 2005), including factors that mediate co-transcriptional (Holoch and Moazed, 2015) or post-

transcriptional silencing (Fischer, 2010; Fischer et al., 2013). Roughly 15 candidates were not represented in the available *C. elegans* RNAi library and thus were not tested. I grew *adr-1(gv6);adr-2(gv42);uuIs1* worms on *E. coli* expressing dsRNA to silence each of 212 candidate genes. The test strain carries a silenced *sur-5::GFP* transgene, which becomes re-expressed when RNAi factors required for silencing are disrupted (Habig et al., 2008; Knight and Bass, 2002). Worms were scored by the strength of GFP re-expression on a scale of zero to four, with zero indicating no GFP expression above background and four indicating GFP re-expression equal to that seen by *dcr-1(RNAi)* (see Table 4.1 notes). Each RNAi clone was screened twice, and clones with an average score above 0.5 after the first two rounds were re-screened two additional times. Including factors known to mediate *adr-1;adr-2* transgene silencing, like *dcr-1*, *rde-1*, and *rde-4,* a total of 36 genes were recovered from this screen. Among the strongest hits were the RIG-I-like helicase encoded by *drh-1* (Ashe et al., 2013), as well as *nrde-3* and *C04F12.1*, genes encoding Argonaute proteins (Yigit et al., 2006). I also observed roles for chromatin-related factors, like the H3K9 trimethylase gene *set-25* (Towbin et al., 2012), as well as *sin-3* and *hda-2,* factors involved in histone deacetylation (Cui et al., 2006). Using CRISPR protocols, I validated that *drh-1* deletion in *adr-1(gv6);adr-2(gv42);uuIs1* also rescues GFP expression (data not shown). Further, GFP expression was rescued by deleting *rrf-1*, a somatic RdRP required for secondary siRNA biogenesis (Billi et al., 2014), which was not represented in the RNAi library. Oddly, genetic deletion of *lin-61* did not rescue GFP expression as observed on *lin-61(RNAi)*, suggesting the effect from *lin-61(RNAi)* may be nonspecific.

The results of this candidate screen informed the genes tested in Chapter 2 for

rescue of *adr-1;adr-2;rrf-3* bursting. Of the factors tested, most genes that rescued *adr-1;adr-2* transgene silencing also rescued *adr-1;adr-2;rrf-3* bursting (Reich et al., 2018). However, disruption of two genes, *C04F12.1* (data not shown) and *set-25*, did not rescue bursting, though they partially rescued transgene silencing. This suggests that transgene silencing and bursting phenotypes proceed through related but not identical RNAi mechanisms. Differences between these mechanisms may reflect differences in pathways used to silence dsRNA from distinct sources. For instance, silencing of cytoplasmic viral dsRNA may rely solely on post-transcriptional silencing mechanisms, while silencing transposable elements that give rise to dsRNA could involve chromatin modification. In uncovering such differences, this candidate screen demonstrates how ADAR mutant phenotypes can be leveraged to study immune-related biology.

Defining the Intracellular Pathogen Response

While *adr-1;adr-2;rrf-3* mutants burst due to aberrant RNAi activity, poly(A)+ mRNA profiling of *adr-1;adr-2;rrf-3* embryos also revealed activation of a poorly characterized antimicrobial transcriptional response. As described in Chapter 2, hundreds of genes induced during Orsay virus infection (Chen et al., 2017) are upregulated in *adr-1;adr-2;rrf-3* embryos and other *rrf-3* mutant strains. Though Orsay virus-induced genes are not upregulated in *adr-1;adr-2* mutant embryos, their induction is higher in *adr-1;adr-2;rrf-3* triple mutants compared to *rrf-3* single mutants, suggesting ADARs influence this transcriptional response. Genes induced during Orsay infection include factors involved in protein ubiquitination and degradation, and they correlate well with genes induced during infection with the intracellular fungus *Nematocida parisii*

(Bakowski et al., 2014). At least one induced gene, *cul-6*, facilitates resistance to both *N. parisii* and Orsay virus infection. Because this transcriptional program is induced by intracellular pathogens, but not extracellular pathogens, it has been recently named the Intracellular Pathogen Response (IPR) (Reddy et al., 2017). However, mechanisms underlying IPR activation remain a mystery. For instance, factors required to sense intracellular pathogens and transduce this signal are entirely undetermined. Recent reports implicate *pals-22,* a gene of unknown function, as a negative regulator of the IPR, since *pals-22* mutants induce IPR genes (Leyva-Díaz et al., 2017; Reddy et al., 2017). Intriguingly, *pals-22* mutants exhibit an Eri phenotype similar to *rrf-3* mutants, which exhibited IPR activation in our studies (Reich et al., 2018). Thus, there appears to be some connection between IPR activation and Eri phenotypes, possibly due to reduced function of the 26G endo-siRNA pathway. Since the 26G pathway uses factors involved in the antiviral RNAi response (Ashe et al., 2013; Billi et al., 2014), inhibiting the 26G pathway could be an important aspect of the antiviral response. Fully understanding how 26G activity relates to the IPR will require additional study, and *adr-1;adr-2;rrf-3* mutants may provide a useful means to investigate IPR factors.

IPR-related factors could be determined using a genome-wide RNAi screen to identify genes required for IPR reporter expression in *adr-1;adr-2;rrf-3* mutants. To perform such a screen, one would first need to identify or generate an IPR-dependent fluorescent reporter whose expression increased in *adr-1;adr-2;rrf-3* mutants. Because both *adr-1;adr-2* mutants and *rrf-3* mutants silence repetitive transgenes (Knight and Bass, 2002; Simmer et al., 2002), reporter expression should not be driven from a simple transgenic array that produces high amount of dsRNA. The reporter would also need to

express a second fluorescent protein under an IPR-independent promoter to control for nonspecific transgene silencing. The integrated transgenic array *jyIs8[Ppals-5::GFP, Pmyo-2::mCherry]* might be a suitable reporter, since it induces IPR-dependent GFP expression, even in *pals-22* mutants that silence repetitive transgenes, and also constitutively expresses mCherry in pharyngeal muscle (Leyva-Díaz et al., 2017; Reddy et al., 2017). After crossing the reporter into *adr-1;adr-2;rrf-3* background and validating elevated GFP fluorescence, one would use a genome-wide RNAi library to knock down genes in *adr-1;adr-2;rrf-3* worms carrying the reporter and identify RNAi clones that decrease GFP expression, but not mCherry expression. Ideally, this approach would identify factors involved in IPR activation, potentially including sensor proteins that respond to pathogen-associated molecular patterns and the transcription factor(s) required for IPR gene induction.

Duplex RNA structures in gene regulation and evolution

Do dsRNA structures promote gene expression?

In Chapter 3, I describe a correlation between dsRNA structures and essential and highly expressed genes on autosome distal arms. Though this correlation is intriguing, I was not able to demonstrate that dsRNA causes elevated gene expression. Limited transgenic analyses (Fig. 3.7) suggested that dsRNA structures are not sufficient to increase gene expression, though these experiments did not rule out the possibility that duplex structures promote expression under certain circumstances. One problem with the transgenic analyses was a small number of conditions tested. I only examined contributions of one structured intron, *ssl-1* intron 1, and one structured 3'UTR, the *eif-*

2alpha 3'UTR, making it difficult to discern if the observed effects involved dsRNA structures or regulatory properties specific to these elements. Furthermore, reporter constructs containing control or structured elements were integrated at one of only two genomic loci, limiting the ability to interpret effects from distinct chromosomal domains.

In these experiments, control and dsRNA reporters were integrated in precise locations using CRISPR/Cas9 to stimulate homologous recombination of reporter genes at specific loci. While effective, this approach is biased by the selection of certain loci and its ease of application is limited by the need to include flanking sequences homologous to the locus of interest. I propose an alternative approach to study gene regulation by dsRNA structure using *Mos1* transposon-based insertion to integrate dsRNA-containing reporters into hundreds of random genomic sites. The miniMos system can mobilize and integrate a minimal 550 nt *Mos1* transposon sequence carrying up to 7.5 kb of cargo DNA randomly throughout the genome, generating potentially hundreds of independent lines carrying single insertions of interest (Frokjaer-Jensen et al., 2014). Using the miniMos system, one could integrate control or dsRNA-containing reporters around the genome, determine insertion sites by inverse PCR, and measure reporter expression in diverse genomic contexts. Such an approach was previously employed to study the effects of intronic PATC sequences on gene expression (Frokjaer-Jensen et al., 2016). In this context, this method could provide experimental support to the hypothesis that dsRNA structures in introns and 3'UTRs promote elevated gene expression from autosome arm domains. Though one could not necessarily integrate control and dsRNA reporters at the same sites, analysis of hundreds of independent reporter lines would hopefully provide clear expression differences between control and

experimental reporters. Alternatively, one could design a single reporter gene carrying a dsRNA structure flanked by loxP sites and integrate it throughout the genome. After isolating integrants, injecting a *Cre* recombinase transgene could be used to excise the dsRNA structure, allowing comparison of the same reporter in the same locations, with and without dsRNA. Because the miniMos system can be used to generate and isolate dozens or hundreds of insertions over the course of several weeks (Frokjaer-Jensen et al., 2014), one could construct many reporters containing different dsRNA or control sequences and integrate these genome-wide in a relatively short time.

What mechanism underlies increased expression of
dsRNA-containing genes?

Though the association between duplex structures and highly expressed genes is intriguing, the lack of a clear mechanism of action makes directed studies of this phenomenon difficult. I propose two future directions to clarify potential mechanisms explaining the observed correlations described in Chapter 3.

First, detailed bioinformatics analyses should further investigate the possibility that specific transposon-derived sequences enriched in dsRNA structures promote expression of associated genes. Since transposable elements (TEs) contain sequences that can act as regulatory elements, like transcription factor binding sites, TE insertion near or within protein coding genes can influence gene expression patterns (Chuong et al., 2017). EER sequences are ~63% derived from DNA transposons (Reich et al., 2018), suggesting that EAG expression patterns might relate to TE content. Though we have determined the predominant transposon classes within EERs, EAG expression patterns have not been

correlated with the presence of particular transposons. If EAGs, or perhaps all genes, were separated based on the presence of specific TE classes in gene introns, UTRs, or proximal sequences, gene expression analyses could potential identify specific TEs associated with elevated gene expression. While straightforward in theory, such analyses could be complicated by that fact that TE sequences are often highly fragmented and divergent from consensus sequences. Further, if functional sequences occur in several different TE classes, analysis of highly expressed genes might not reveal association with a single TE. Sequence motif analysis, using software suites like MEME (http://meme-suite.org/), might identify common TE sequences enriched within introns/UTRs of highly expressed EAGs or other autosome arm genes. Genome editing approaches could then be used to test if enriched sequences are necessary and/or sufficient for elevated gene expression.

Identifying proteins bound to highly expressed dsRNA-associated transcripts could offer additional insight into dsRNA-mediated mechanisms of gene regulation. In this approach, one would pull down highly expressed EAG transcripts (i.e. *eif-2alpha* or *ssl-1*) from *C. elegans* extracts using either biotinylated antisense oligos or RNA aptamers knocked into the gene of interest. Proteins bound to target RNAs would then be isolated and identified by tandem mass spectrometry. This approach would aim to identify candidate factors involved in regulation of highly expressed dsRNA-associated genes. Candidate factors could then be disrupted using RNAi or genetic mutation. Using qRT-PCR or RNAseq to assay EAG mRNA levels, one would determine which candidate factors influence EAG expression, but not control genes.

Duplex RNA structures as platforms of gene evolution

A central focus of this work is the idea that dsRNA-dsRBP interactions mediate post-transcriptional gene expression regulation. However, dsRNA loci have two additional properties that suggest that these elements also impact gene regulation through effects on sequence diversification and evolution. First, dsRNA sequences are largely transposon-derived (Porath et al., 2017; Reich et al., 2018; Zhao et al., 2015), suggesting they can be readily introduced by TE insertion. Active transposons are typically silenced in the *C. elegans* germline by RNAi, since their transcription gives rise to dsRNA (Sijen and Plasterk, 2003). However, since Tc1 transposon copy numbers vary between strains from ~20-30 copies in Bristol N2 to ~300-500 copies in Bergerac (Bessereau, 2006), germline TE transposition likely occurs at some rate over time. Second, long inverted repeat sequences are highly recombinogenic and unstable in eukaryotic genomes (Gordenin et al., 1993; Waldman et al., 1999; Wang and Leung, 2006), suggesting they can promote frequent sequence alterations within and around dsRNA-containing genes. Thus, dsRNAs introduced into genes by TE insertion may stimulate recombination and lead to new genetic variants that in turn affect gene expression and functions.

If true, I predict genes containing dsRNA structures exhibit greater variation in sequence and expression than unstructured genes. Indeed, most *C. elegans* inverted repeats occur in distal arm domains, where genes have lower levels of conservation (Consortium, 1998) and more divergent expression between wild isolate strains (Denver et al., 2005). However, it is not clear that these properties of distal arms relate explicitly to dsRNA structures. To gain a better understanding of dsRNA functions in *C. elegans* genome evolution, a population genetics approach could be used to define how dsRNA-

containing genes vary in sequence and expression. Given the highly recombinogenic nature of dsRNA loci, these sequences likely change rapidly over evolutionary time, complicating phylogenetic comparisons of dsRNAs in different species. However, by analyzing many different populations of the same species, one can more easily compare sequences of even rapidly evolving loci, since the populations are more closely related. The genomes of hundreds of natural *C. elegans* isolates from around the world have already been sequenced and are freely available from the *Caenorhabditis elegans* Natural Diversity Resource (https://www.elegansvariation.org/) (Cook et al., 2017). Genome sequences contain genetic variant information showing how gene sequences differ between populations. By analyzing genetic variation among dsRNA-associated genes and control genes, one could test the hypothesis that dsRNA structures are associated with higher rates of genetic variance. Further, functionally important dsRNA structures could be identified by searching for dsRNAs containing compensatory mutations that maintain base-pairing. RNAseq of wild isolate strains could further define how dsRNA structures correlate with variation in gene expression patterns. Such analyses may broaden our understanding of the role of dsRNA not only in direct gene regulation but also in gene and genome evolution.

References

Ashe, A., Belicard, T., Le Pen, J., Sarkies, P., Frezal, L., Lehrbach, N.J., Felix, M.A., and Miska, E.A. (2013). A deletion polymorphism in the Caenorhabditis elegans RIG-I homolog disables viral RNA dicing and antiviral immunity. Elife *2*, e00994.

Bakowski, M.A., Desjardins, C.A., Smelkinson, M.G., Dunbar, T.L., Lopez-Moyado, I.F., Rifkin, S.A., Cuomo, C.A., and Troemel, E.R. (2014). Ubiquitin-mediated response to microsporidia and virus infection in C. elegans. PLoS Pathog. *10*, e1004200.

Bessereau, J.-L. (2006). Transposons in *C. elegans*. WormBook (ed. The C. elegans Research Community), WormBook doi/10.1895/wormbook.1891.1870.1891, http://www.wormbook.org.

Billi, A.C., Fischer, S.E., and Kim, J.K. (2014). Endogenous RNAi pathways in C. elegans. WormBook (ed. The C. elegans Research Community). WormBook doi: 10.1895/wormbook.1.170.1, http://www.wormbook.org.

Blango, M.G., and Bass, B.L. (2016). Identification of the long, edited dsRNAome of LPS-stimulated immune cells. Genome Res. *26*, 852-862.

Chen, K., Franz, C.J., Jiang, H., Jiang, Y., and Wang, D. (2017). An evolutionarily conserved transcriptional response to viral infection in Caenorhabditis nematodes. BMC Genomics *18*, 303.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71-86.

Consortium, T.C.e.S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012-2018.

Cook, D.E., Zdraljevic, S., Roberts, J.P., and Andersen, E.C. (2017). CeNDR, the Caenorhabditis elegans natural diversity resource. Nucleic Acids Res. *45*, D650-D657.

Cui, M., Kim, E.B., and Han, M. (2006). Diverse chromatin remodeling genes antagonize the Rb-involved SynMuv pathways in C. elegans. PLoS Genet. *2*, e74.

Deffit, S.N., Yee, B.A., Manning, A.C., Rajendren, S., Vadlamani, P., Wheeler, E.C., Domissy, A., Washburn, M.C., Yeo, G.W., and Hundley, H.A. (2017). The C. elegans neural editome reveals an ADAR target mRNA required for proper chemotaxis. Elife *6*, e28625.

Denver, D.R., Morris, K., Streelman, J.T., Kim, S.K., Lynch, M., and Thomas, W.K. (2005). The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. Nat. Genet. *37*, 544-548.

Ecsedi, M., Rausch, M., and Grosshans, H. (2015). The let-7 microRNA directs vulval development through a single target. Dev. Cell *32*, 335-344.

Fischer, S.E. (2010). Small RNA-mediated gene silencing pathways in C. elegans. Int. J. Biochem. Cell Biol. *42*, 1306-1315.

Fischer, S.E., Pan, Q., Breen, P.C., Qi, Y., Shi, Z., Zhang, C., and Ruvkun, G. (2013). Multiple small RNA pathways regulate the silencing of repeated and foreign genes in C. elegans. Genes Dev. *27*, 2678-2695.

Frokjaer-Jensen, C., Davis, M.W., Sarov, M., Taylor, J., Flibotte, S., LaBella, M., Pozniakovsky, A., Moerman, D.G., and Jorgensen, E.M. (2014). Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. Nat. Methods *11*, 529-534.

Frokjaer-Jensen, C., Jain, N., Hansen, L., Davis, M.W., Li, Y., Zhao, D., Rebora, K., Millet, J.R.M., Liu, X., Kim, S.K.*, et al.* (2016). An abundant class of non-coding DNA can prevent stochastic gene silencing in the C. elegans germline. Cell *166*, 343-357.

Gordenin, D., Lobachev, K., Degtyareva, N., Malkova, A., Perkins, E., and Resnick, M. (1993). Inverted DNA repeats: a source of eukaryotic genomic instability. Mol. Cell. Biol. *13*, 5315-5322.

Grishok, A. (2005). RNAi mechanisms in Caenorhabditis elegans. FEBS Lett. *579*, 5932-5939.

Habig, J.W., Aruscavage, P.J., and Bass, B.L. (2008). In C. elegans, high levels of dsRNA allow RNAi in the absence of RDE-4. PLoS One *3*, e4052.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nat. Rev. Genet. *16*, 71-84.

Jorgensen, E.M., and Mango, S.E. (2002). The art and design of genetic screens: caenorhabditis elegans. Nat. Rev. Genet. *3*, 356-369.

Jose, A.M., Garcia, G.A., and Hunter, C.P. (2011). Two classes of silencing RNAs move between Caenorhabditis elegans tissues. Nat. Struct. Mol. Biol. *18*, 1184-1188.

Kile, B.T., and Hilton, D.J. (2005). The art and design of genetic screens: mouse. Nat. Rev. Genet. *6*, 557-567.

Knight, S.W., and Bass, B.L. (2002). The Role of RNA Editing by ADARs in RNAi. Mol. Cell *10*, 809-817.

Leyva-Díaz, E., Stefanakis, N., Carrera, I., Glenwinkel, L., Wang, G., Driscoll, M., and Hobert, O. (2017). Silencing of repetitive DNA is controlled by a member of an unusual C. elegans gene family. Genetics *207*, 529-545.

Liddicoat, B.J., Piskol, R., Chalk, A.M., Ramaswami, G., Higuchi, M., Hartner, J.C., Li, J.B., Seeburg, P.H., and Walkley, C.R. (2015). RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. Science *349*, 1115-1120.

Mannion, N.M., Greenwood, S.M., Young, R., Cox, S., Brindle, J., Read, D., Nellaker, C., Vesely, C., Ponting, C.P., McLaughlin, P.J.*, et al.* (2014). The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. Cell Rep. *9*, 1482-1494.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods *6*, 917-922.

Ortiz, M.A., Noble, D., Sorokin, E.P., and Kimble, J. (2014). A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode Caenorhabditis elegans. G3 (Bethesda) *4*, 1765-1772.

Pestal, K., Funk, C.C., Snyder, J.M., Price, N.D., Treuting, P.M., and Stetson, D.B. (2015). Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. Immunity *43*, 933-944.

Porath, H.T., Knisbacher, B.A., Eisenberg, E., and Levanon, E.Y. (2017). Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. Genome Biol. *18*, 185.

Reddy, K.C., Dror, T., Sowa, J.N., Panek, J., Chen, K., Lim, E.S., Wang, D., and Troemel, E.R. (2017). An intracellular pathogen response pathway promotes proteostasis in C. elegans. Curr. Biol. *27*, 3544-3553 e3545.

Reich, D.P., Tyc, K.M., and Bass, B.L. (2018). C. elegans ADARs antagonize silencing of cellular dsRNAs by the antiviral RNAi pathway. Genes Dev. *32*, 271-282.

Sebastiani, P., Montano, M., Puca, A., Solovieff, N., Kojima, T., Wang, M.C., Melista, E., Meltzer, M., Fischer, S.E., Andersen, S.*, et al.* (2009). RNA editing genes associated with extreme old age in humans and with lifespan in C. elegans. PLoS One *4*, e8210.

Sijen, T., and Plasterk, R.H.A. (2003). Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. Nature *426*, 310-314.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S.P., Nonet, M.L., Fire, A., Ahringer, J., and Plasterk1, R.H.A. (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi. Curr. Biol. *12*, 1317-1319.

Sternberg, P.W. (2005). Vulval development. WormBook (ed. The C. elegans Research Community). WormBook, doi/10.1895/wormbook.1.6.1, www.wormbook.org.

Tonkin, L.A., Saccomanno, L., Morse, D.P., Brodigan, T., Krause, M., and L.Bass, B. (2002). RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans. EMBO J. *21*, 6025-6035.

Towbin, B.D., Gonzalez-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., Askjaer, P., and Gasser, S.M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. Cell *150*, 934-947.

Waldman, A.S., Tran, H., Goldsmith, E.C., and Resnick, M.A. (1999). Long inverted repeats are an at-risk motif for recombination in mammalian cells. Genetics *153*, 1873-1883.

Wang, Y., and Leung, F.C. (2006). Long inverted repeats in eukaryotic genomes: recombinogenic motifs determine genomic plasticity. FEBS Lett. *580*, 1277-1284.

Warf, M.B., Shepherd, B.A., Johnson, W.E., and Bass, B.L. (2012). Effects of ADARs on small RNA processing pathways in C. elegans. Genome Res. *22*, 1488-1498.

Whipple, J.M., Youssef, O.A., Aruscavage, P.J., Nix, D.A., Hong, C., Johnson, W.E., and Bass, B.L. (2015). Genome-wide profiling of the C. elegans dsRNAome. RNA *21*, 786-800.

Wu, D., Lamm, A.T., and Fire, A.Z. (2011). Competition between ADAR and RNAi pathways for an extensive class of RNA targets. Nat. Struct. Mol. Biol. *18*, 1094-1101.

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell *127*, 747-757.

Zhang, L., Ward, J.D., Cheng, Z., and Dernburg, A.F. (2015). The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. Development *142*, 4374-4384.

Zhao, H.Q., Zhang, P., Gao, H., He, X., Dou, Y., Huang, A.Y., Liu, X.M., Ye, A.Y., Dong, M.Q., and Wei, L. (2015). Profiling the RNA editomes of wild-type C. elegans and ADAR mutants. Genome Res. *25*, 66-75.
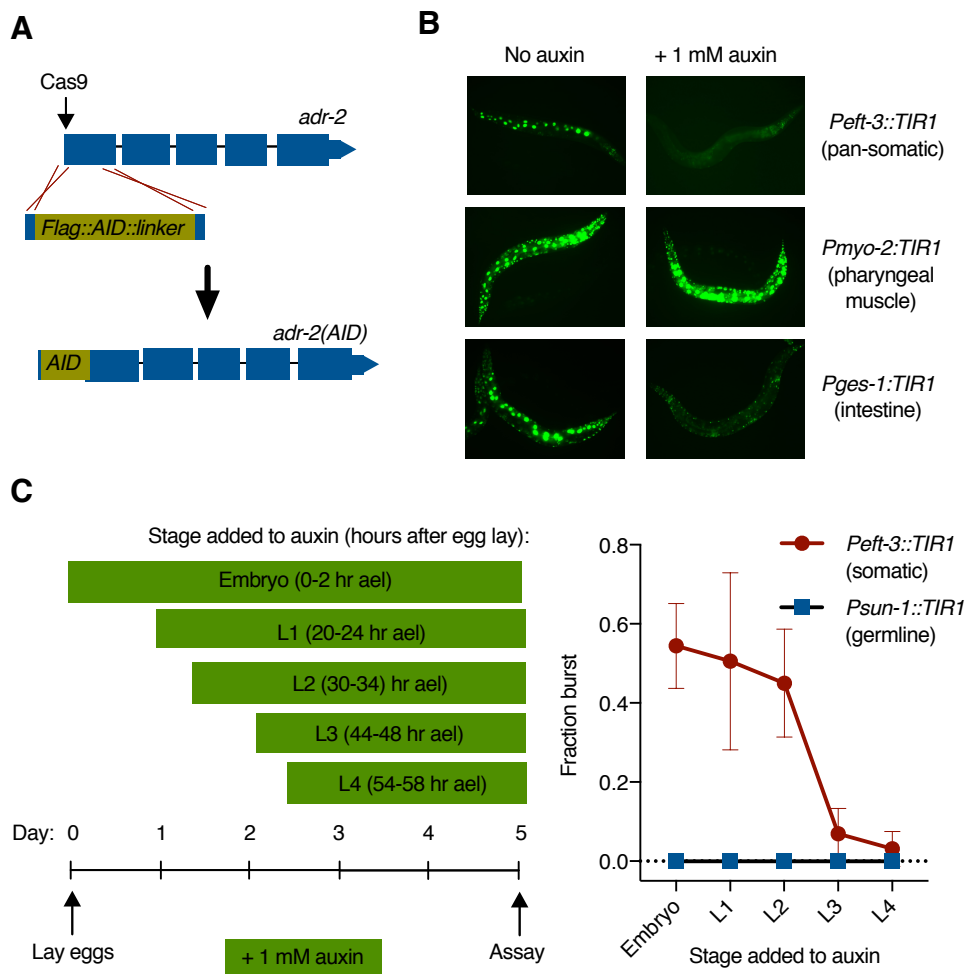
Figure 4.1 Modulating ADAR levels in space and time. (*A*) Cas9-mediated DNA cleavage proximal to the *adr-2* start codon was used to stimulate recombination (red lines) with a homology-directed repair donor sequence encoding a 1x Flag tag, the 44 amino acid auxin-inducible degron (AID) and Gly-Ser linker. This approach was used to insert the *Flag::AID::linker* sequence in-frame into the N-termini of the endogenous *adr-1* (not shown) and *adr-2* loci. (*B*) Auxin treatment of *adr-1(AID);adr-2(AID)* strains expressing *TIR1* under different tissue-specific promoters induces transgene silencing in tissues where *TIR1* is expressed. (*C*) Auxin treatment (green boxes) of *adr-1(AID);adr-2(AID);rrf-3* strains beginning at different stages (scheme at left) induces bursting at frequencies dependent on the stage of addition and tissue of *TIR1* expression (plotted on right).

Table 4.1 RNAi genes required for *adr-1;adr-2* transgene silencing.

Gene	Score	Description
dcr-1	4.0	Endoribonuclease required in RNAi
drh-1	4.0	RIG-I like helicase; promotes viral dsRNA processing
rde-4	3.3	dsRBP and DCR-1 binding partner
rde-1	2.5	Argonaute; required for antiviral RNAi
nrde-3	2.0	Argonaute; transports 22G siRNAs from cytosol to nucleus
inst-1	2.0	Integrator complex subunit
C04F12.1	2.0	Argonaute; involved in exogenous RNAi
ima-3	1.8	Importin alpha protein
sin-3	1.8	Histone deacetylase subunit
lin-61	1.8	H3K9me-binding protein
rsp-7	1.5	SR protein; splicing factor
set-25	1.5	Histone methylase; deposits H3K9me3
cid-1	1.4	Poly(U)-polymerase; limits siRNA accumulation
T25G3.3	1.3	Predicted 60S ribosome export factor
hda-2	1.2	Histone deacetylase
ipgm-1	1.1	Phosphoglycerate mutase
gfl-1	1.0	Chromatin remodeling factor
prp-17	1.0	Splicing factor
kin-10	0.9	Protein kinase
arp-6	0.9	Actin-related protein; possible chromatin remodeling factor
T23D8.3	0.8	Predicted 40S ribosome export factor
mut-15	0.8	RNAi factor present in Mutator foci
pgl-1	0.8	P-granule component
set-16	0.8	Histone methylase; deposits H3K4me1
cpsf-2	0.8	mRNA 3' end processing factor
rsd-6	0.7	Tudor-domain protein involved in RNAi spreading
F55F8.3	0.7	Ribosome maturation factor
cit-1.2	0.7	Cyclin T; involved in RNA Pol II regulation
zfp-1	0.7	Involved in chromatin remodeling and RNA Pol II pausing
ZK1127.5	0.7	RNA 3' terminal phosphate cyclase
imb-2	0.7	Importin beta protein
dao-5	0.7	Involved in rRNA transcription and nucleolar structure
zfp-3	0.6	Zinc-finger protein
ain-1	0.6	Involved in miRNA-mediated gene regulation
C55B7.11	0.6	Worm-specific protein of unknown function
T08G11.4	0.6	Trimethylguanosine synthase

Genes were scored 0 to 4 based on GFP expression in *adr-1(gv6);adr-2(gv42);uuIs1*:

0 – No GFP fluorescence above background.

1 – At least 50% of progeny had GFP signal higher than background.

2 – At least 50% of progeny had GFP signal 25-50% of *dcr-1(RNAi)*.

3 – 100% of progeny had GFP signal 50%-75% of *dcr-1(RNAi)*.

4 – 100% of progeny had GFP 75-100% of *dcr-1(RNAi)*.