Mapping the dsRNA World

Daniel P. Reich and Brenda L. Bass

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

Correspondence: bbass@biochem.utah.edu

SUMMARY

Long double-stranded RNAs (dsRNAs) are abundantly expressed in animals, in which they frequently occur in introns and 3' untranslated regions of mRNAs. Functions of long, cellular dsRNAs are poorly understood, although deficiencies in adenosine deaminases that act on RNA, or ADARs, promote their recognition as viral dsRNA and an aberrant immune response. Diverse dsRNA-binding proteins bind cellular dsRNAs, hinting at additional roles. Understanding these roles is facilitated by mapping the genomic locations that express dsRNA in various tissues and organisms. ADAR editing provides a signature of dsRNA structure in cellular transcripts. In this review, we detail approaches to map ADAR editing sites and dsRNAs genome-wide, with particular focus on high-throughput sequencing methods and considerations for their successful application to the detection of editing sites and dsRNAs.

Outline

1 Introduction2 The dsRNA world is a world governed by different properties than the RNA world3 Features of the dsRNAome4 Historical perspective on methods to find ADAR editing sites	5 Finding edited RNAs in high-throughput data sets6 Looking toward the futureReferences

1 INTRODUCTION

The ability of RNA to copy itself, or replicate, is a key feature of an RNA world, allowing Darwinian evolution and what we call life (Rich 1962; Gilbert 1986; Joyce 2002; Robertson and Joyce 2012). The RNA World hypothesis assumes that Watson–Crick base-pairing was key to replication in a primordial cell. Thus, by definition, replication of an RNA genome involved double-stranded RNA (dsRNA) intermediates, and evolution of modern-day cells likely occurred in the constant presence of dsRNA. Our goal in this review is to describe the unique properties dsRNA confers on a biological system and the techniques used for the genome-wide mapping of dsRNA (i.e., determining a dsRNAome).

In modern-day cells, nucleic acids play key and obvious roles, but they are also fundamental to self versus nonself discrimination (Crowl et al. 2017). The diverse RNA editing and modifications found in all kingdoms of life, in recent years referred to as the Epitranscriptome (Saletore et al. 2012), have myriad functions in modern-day cells, but it seems very plausible, and others have proposed (Eigenbrod et al. 2015; O’Connell et al. 2015), that they originated as a means to discriminate self from nonself. Selfish elements and viruses of the early RNA world also may have had an RNA genome that replicated through a dsRNA intermediate. Perhaps reflecting an ancient and ongoing conflict between cells and selfish elements, in modern cells, dsRNA-mediated pathways play critical roles in immune defense in response to viral infection.

Viruses have long been known to give rise to dsRNA (Ehrenfeld and Hunt 1971), which acts in the cytoplasm as a pathogen-associated molecular pattern (PAMP), to trig-

ger an immune response. The source of viral dsRNA is attributed to genomes of dsRNA viruses, replication intermediates and copy-back structures of single-stranded RNA viruses, and bidirectional transcription of certain DNA viruses (Schlee 2013). When assayed, this dsRNA is typically ≥ 100 base pairs (Pfaller et al. 2015) and thus distinct from smaller silencing RNAs, such as microRNAs (miRNAs) and small-interfering RNAs (siRNAs). It is now clear that our own cells encode and express long dsRNA, in both vertebrates and invertebrates (Whipple et al. 2015; Blango and Bass 2016; Reich et al. 2018), and that the adenosine deaminases that act on RNA (ADARs), by converting adenosine to inosine (A-to-I), mark a dsRNA as self (Hartner et al. 2009; Mannion et al. 2014; Liddicoat et al. 2015; Pestal et al. 2015; George et al. 2016; Reich et al. 2018). To understand how cells discriminate between self dsRNA and non-self viral dsRNA, it is important to understand the dsRNAome.

2 THE dsRNA WORLD IS A WORLD GOVERNED BY DIFFERENT PROPERTIES THAN THE RNA WORLD

There are several properties of the dsRNA world that are distinct from those of the RNA world. First, the RNAs that comprise the dsRNA world are rod-shaped molecules that can stretch for hundreds of base pairs with few branches (Fig. 1). Short dsRNA helices are the secondary structural elements that assemble to create the three-dimensional shapes of RNA molecules. However, with few exceptions (Rieder et al. 2013), dsRNA-binding proteins (dsRBPs) do not bind dsRNA helices that are buried in tertiary structure.

Figure 1. Representative double-stranded RNAs (dsRNAs) from three organisms. UNAFold-predicted RNA structures are shown for mouse *Sppl2a* 3' untranslated region (3' UTR), human *SPPL2A* 3' UTR, and *Caenorhabditis elegans eif-2α* pre-mRNA. Approximate lengths of highly base-paired regions are shown for scale, and minimum (most stable) predicted folding free energies (ΔG) are reported beneath structures.

ADAR and PKR do not bind transfer RNA (tRNA) (Bass and Weintraub 1987; Schmedt et al. 1995), and similarly, the A-to-I modifications created by ADARs are not observed in ribosomal RNA (rRNA) (Paul and Bass 1998). The preference for binding to dsRNA that is not buried in tertiary structure is likely important for self/nonself discrimination by dsRBPs. Although wild-type RIG-I, a dsRBP essential for the mammalian innate immune response, shows limited association with rRNA, a RIG-I mutation observed in Singleton–Merten syndrome allows the protein to associate with dsRNA expansion sequences in the large rRNA subunit (Lässig et al. 2015).

Because of the A-form structure of dsRNA, which includes a deep major groove, it is difficult for dsRBPs to recognize dsRNA sequence (Tian et al. 2004). In fact, although dsRBPs bind tightly to dsRNA, typically showing dissociation constants in the nanomolar range (Kim et al. 1994; Schmedt et al. 1995; Ohman et al. 2000; Ma et al. 2008; Parker et al. 2008; Sinha et al. 2015), they are not sequence-specific. This second property of the dsRNA world is key to functions of dsRBPs, which interact with dsRNAs of diverse sequences. For example, Dicer must bind pre-miRNAs of many different sequences (Daugaard and Hansen 2017), ADARs must edit numerous distinct dsRNAs (Bajad et al. 2017), and antiviral dsRBPs must bind diverse viral RNAs (Yoo et al. 2014). Consistent with the idea that dsRBPs have structural specificity for the A-form helix, disruptions to the helix, such as mismatches and loops, can establish a register for dsRBP binding (Lehmann and Bass 1999). Similarly, although dsRBPs cannot access functional groups of bases within the major groove of completely base-paired dsRNA, sequence-specific contacts in the minor groove can also establish register (Steffl et al. 2010).

The final property of the dsRNA world is derivative of the first two. Because dsRBPs all bind dsRNA, and they are not sequence-specific, the dsRNA world is a world of constant competition. During its lifetime, a given dsRNA may interact with multiple dsRBPs.

3 FEATURES OF THE dsRNAome

This review focuses on methods used to determine dsRNAomes, and with few exceptions, these involve the determination of ADAR RNA editing sites. ADARs are a family of RNA editing enzymes, present in all animals (Defit and Hundley 2016; Nishikura 2016; Walkley and Li 2017). ADARs target only dsRNA, and the longer a dsRNA is, the more editing sites it will acquire. The observation of ADAR editing sites in an endogenous RNA is proof that the RNA is double-stranded *in vivo*, a gold standard used not only in determining dsRNAomes, but also as proof for specific dsRNA structures (Sijen and Plasterk 2003).

Figure 2 illustrates three dsRNAomes, with genomic locations that express dsRNA, sometimes called editing enriched regions (EERs), indicated on chromosomes of mouse, human, and *Caenorhabditis elegans*. Based on dsRNAomes determined to date (Table 1), dsRNA is predominantly encoded within protein-coding genes, rather than intergenic regions, with 5.9% of protein-coding genes encoding structures in *C. elegans*, 13% in human monocytes, and 1.3% in mouse bone marrow-derived macrophages. (Repetitive elements in mouse are more divergent, resulting in fewer pairing partners and lower free energies [Neeman et al. 2006; Blango and Bass 2016].) For protein-coding genes, structures mostly inhabit noncoding regions of mRNAs, introns and 3' untranslated regions (UTRs). Base-pairing usually occurs intramolecularly (Fig. 1), with complementary sequences within a single transcript folding back on themselves. The dsRNAs are predicted to fold into remarkably stable structures (Table 1) and, depending on the organism, encompass an average of 546–845 nucleotides (nt).

The chromosome maps of Figure 2 highlight interesting trends—some that are easily explained and others that are enigmatic, possibly hinting at EER functions. For example, human chromosome 19 has a dense concentration of EERs, likely because it is gene-rich, with a high number of repetitive elements (Grimwood et al. 2004). Consistent with their ability to base-pair, repeats, especially those that are inverted, inhabit many EERs (Fig. 3). Abundant repeat classes are represented in EERs, like *Alu* elements, which occur in more than a million copies in the human genome (Bazak et al. 2014) and DNA transposons, comprising 12.6% of the *C. elegans* genome (Ahringer and Gasser 2018). Comparisons of transcripts with the greatest number of editing sites, or “hyperedited” dsRNAs, in diverse metazoa, indicate that repeats, particularly transposon-derived repeats, give rise to most cellular dsRNAs (Porath et al. 2017).

In the *C. elegans* dsRNAome, EERs are enriched on distal arms, or chromosome ends, of autosomes (Fig. 2). This trend is not observed on the X chromosome, which overall has fewer EERs. Distal arms of *C. elegans* chromosomes show properties of heterochromatin and, compared with central regions, have fewer essential genes, more repetitive elements, lower levels of gene expression, and longer than average introns (Prachumwat et al. 2004; Liu et al. 2011; Ho et al. 2014; Ahringer and Gasser 2018). Intriguingly, although EERs also are enriched for repetitive elements and inhabit longer than average introns and 3' UTRs, they typically inhabit genes with higher than average expression (Fig. 4). Except for the correlation with repetitive elements, as observed on human chromosome 19, an enrichment of EERs on specific regions of mammalian chromosomes has not been observed.

Figure 2. Mouse, human, and *Caenorhabditis elegans* dsRNAomes. Vertical black lines denote positions of editing enriched regions (EERs) on chromosomes of mouse, human, and *C. elegans*. Chromosomes are not drawn to scale, so the horizontal black bars at the bottom display relative chromosome length. Maps of mouse and human dsRNAomes were generated with Idiographica and that for *C. elegans* with PhenoGram (Kin and Ono 2007; Wolfe et al. 2013). BMDMs, bone marrow–derived macrophages.

Although EERs are not conserved at a sequence level, of the 285 mouse genes that contain EERs, 74 (26%) have an orthologous human gene with an EER ($p < 0.0001$, χ^2 test) (Blango and Bass 2016). Similarly, whereas a *Caenorhabditis briggsae* dsRNAome has not been determined, of 1092 *C. elegans* genes with a structured intron ($\Delta G/\text{nt} < -0.5$ kcal/mole*nt at 20°C), 147 have an orthologous *C. briggsae* gene with a structured intron, a significant enrichment above the expected number ($p < 0.0001$, χ^2 test). Structured introns cluster on autosome distal arms in *C. briggsae*, as in *C. elegans* (Fig. 5). EERs and structured introns do not occur

in the same locations in orthologous genes, and an EER within an intron can be in a UTR in the ortholog. Regardless, the presence of EERs and structured introns in orthologous genes and chromosomal domains raises the possibility of a conserved function in gene regulation.

4 HISTORICAL PERSPECTIVE ON METHODS TO FIND ADAR EDITING SITES

ADARs were discovered when perfectly base-paired dsRNA was injected into *Xenopus laevis* embryos. Researchers no-

Table 1. Properties of long dsRNAs (EERs)

Organism (cell type)	EERs			EER-associated genes			
	#	Length (nt)	$\Delta G/\text{nt}^a$	#	% of all genes ^b	% with intronic EERs	% with 3'-UTR EERs
<i>Caenorhabditis elegans</i>	1962	563	-0.349	1196	5.9	65.0	41.6
Mouse (BMDMs)	342	546	-0.307	285	1.3	20.4	69.1
Human (monocytes)	3438	845	-0.346	2792	13.0	57.4	23.9

EERs, editing enriched regions; UTR, untranslated region; ΔG , predicted folding free energy; nt, nucleotide; BMDMs, bone marrow–derived macrophages.

^aPredicted at 37°C; kcal/mol*nt.

^bProtein-coding genes only.

Figure 3. Repeat content of mouse, human, and *Caenorhabditis elegans* editing enriched regions (EERs). Pie charts depict percentage of total EER sequences that overlap RepeatMasker-annotated repeats (mouse: mm10; human: hg19; *C. elegans*: ce10). Major classes of repetitive elements (>2% of total EER sequence) are labeled, and classes comprising <2% are grouped as “Other.” Nonrepetitive sequences did not overlap any sequences annotated as repetitive by RepeatMasker. SINE, short interspersed nuclear element; LINE, long interspersed nuclear element; LTR, long terminal repeat; DNA TE, DNA transposable element.

ticed that after incubation in *Xenopus* cells, the dsRNA migrated aberrantly on native gels (Bass and Weintraub 1987) and showed altered sensitivity to single-strand-specific ribonucleases (Rebagliati and Melton 1987). Thus, ADARs were initially called an unwinding activity, and this activity was also observed in extracts of mammalian

cultured cells (Wagner and Nishikura 1988). Subsequent studies, using thin-layer chromatography or high-performance liquid chromatography (HPLC) to analyze nucleotides in treated dsRNA, revealed the activity involved covalent modification of adenosine to inosine (Bass and Weintraub 1988; Wagner et al. 1989). Only then was it realized that dsRNA strands were not being unwound. Rather, the RNA was becoming more single-stranded in character as AU base pairs were changed to IU mismatches. Although ADARs indeed change the stability of dsRNA, to date there is no example in which editing causes dsRNA to completely separate into two single strands. Subsequent studies revealed that A-to-I conversion occurred by hydrolytic deamination (Polson et al. 1991), and the activity was briefly called dsRAD or DRADA. In 1997, researchers agreed to rename the enzyme ADAR, based on recommendations by the HUGO committee (Bass et al. 1997).

Although ADAR activity in the above studies was detected using dsRNA prepared in vitro, it was not long before in vivo editing within naturally occurring transcripts was reported. The first example was identified serendipitously by Sanger sequencing of cloned cDNAs made from measles virus transcripts (Cattaneo et al. 1988; Bass et al. 1989). Inosine base-pairs like guanosine and prefers to pair with cytidine. Thus, reverse transcription of RNA containing

Figure 4. Expression of *Caenorhabditis elegans* editing enriched region (EER)-associated genes (EAGs). Tukey box plot shows gene expression, in fragments per kilobase*million reads (FPKM), for all expressed genes or EAGs in RNA-seq of four *C. elegans* developmental stages: embryo, early larval (E. larval; L1–L2), late larval (L. larval; L3–L4), and young adult (Y. adult) stages. ****, $P < 0.0001$, Mann-Whitney U -test.

Figure 5. Predicted intron structure along chromosome III for two nematode species. Length-normalized UNAFold-predicted folding free energies are plotted by relative position on chromosome III of *Caenorhabditis elegans* (blue) and *Caenorhabditis briggsae* (red). Trends observed on chromosome III are representative of all *C. elegans* autosomes. Average intronic $\Delta G/\text{nt}$ values were calculated by splitting chromosomes into 1000 equal-length segments and averaging $\Delta G/\text{nt}$ values of introns in each segment. Lower $\Delta G/\text{nt}$ values indicate presence of more stable intronic structures. ΔG , predicted folding free energy; nt, nucleotide.

A-to-I editing shows T-to-C changes in the first-strand cDNA, revealed as A-to-G changes in the second-strand cDNA. Initial examples of in vivo ADAR editing showed numerous editing events in individual, cloned cDNAs (Bass 1997), reiterating the nonselective hyperediting observed in the perfectly paired dsRNAs used in the first in vitro studies. In these early days of ADAR research, it was hard to imagine an important biological activity for ADARs.

This all changed with the observation of editing events in specific codons. The first reported example involved a specific codon in an α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) glutamate receptor, GRIA2 (previously called GluR-B) (Sommer et al. 1991). cDNA analyses indicated a specific amino acid was encoded as a glutamine or arginine, and the arginine was known to be important for ion transport properties of channels assembled with glutamate receptors. Researchers were puzzled when only a single gene could be identified for GRIA2, and its genomic sequence specified a glutamine codon. The investigators hypothesized that ADAR editing changed a glutamine (Q) codon to an arginine (R) codon (Q/R editing site). Indeed, subsequent studies revealed complementary sequence in the intron adjacent to the Q/R site that could fold back to encompass the edited codon (Higuchi et al. 1993; Egebjerg et al. 1994). The predicted structure was disrupted by mismatches and loops, and the idea that such disruptions could promote more selective editing began to emerge (Hurst et al. 1995; Bass 1997; Lehmann and

Bass 1999), as well as the realization that ADARs constituted a family of enzymes (Melcher et al. 1996), with some organisms having a single ADAR and others several (Bass 2002).

An astounding finding was that lethality of *ADAR2*^{-/-} mice could be rescued by replacing the unedited *GRIA2* allele with the edited one (Higuchi et al. 2000). These studies foreshadowed the many important ADAR editing events that occur in the nervous system of vertebrates and invertebrates (Morse and Bass 1999; Rosenthal 2015; Behm and Ohman 2016; Deffit and Hundley 2016; Nishikura 2016) and focused the field on finding and characterizing ADAR editing in codons. The nonselective editing observed in early studies was put aside by most.

Sanger sequencing of cloned cDNAs was instrumental in the identification of the first ADAR editing sites. However, this method required sequencing cDNAs on a case-by-case basis and was not amenable to the systematic discovery of ADAR editing sites. Toward this goal, a method for inosine-specific cleavage (Morse and Bass 1997) was developed and, when coupled with differential display, allowed the first unbiased identification of A-to-I edited RNAs in *C. elegans* (Morse and Bass 1999) and human brain (Morse et al. 2002). These efforts were facilitated by large-scale genome sequencing projects (Berks et al. 1995; *C. elegans* Sequencing Consortium 1998; Lander et al. 2001; Venter et al. 2001) that allowed identified transcripts to be cross-referenced with genomic sequences and specific genes. At the time, the field assumed that the primary role of ADARs was to edit codons, and it was perplexing that this “systematic” method identified inosine only in noncoding sequences, largely introns and 3′ UTRs, in which base-pairing often occurred between repetitive elements. This hinted that ADAR editing in codons, or mRNA recoding, is a rare event, something now known to be true for studied organisms, excepting some coleoid cephalopods (Liscovitch-Brauer et al. 2017). Although inosine-specific cleavage provided a systematic way to find editing sites, the method required isolation of differentially cleaved bands from a gel and was limited compared with the high-throughput protocols used today. Similar limitations applied to other methods developed to identify ADAR editing sites without Sanger sequencing. These included approaches to identify recoding sites by immunoprecipitating ADAR2 and detecting enriched transcripts by microarray analysis (Ohlson et al. 2005) and methods using high-resolution melt analysis, denaturing HPLC, or allele-specific polymerase chain reaction (PCR) to detect and measure editing in cDNA amplicons (Gallo et al. 2002; Chateigner-Boutin and Small 2007; Chen et al. 2008).

The first attempt to identify inosine-containing transcripts focused on *C. elegans*, because at the time it had the most complete genome sequence (*C. elegans* Sequenc-

ing Consortium 1998). As sequencing of other genomes progressed (Adams et al. 2000; Lander et al. 2001; Venter et al. 2001; Mouse Genome Sequencing Consortium et al. 2002), comparative analyses became feasible. Phylogenetic comparisons of glutamate receptor pre-mRNAs from six vertebrates revealed surprising conservation of exonic and intronic sequences around the R/G recoding site, which results in the conversion of an arginine (R) to a glycine (G) (Aruscavage and Bass 2000). Similar regions of strong conservation between different *Drosophila* species allowed identification of 16 genes edited in coding sequences (Hoopengardner et al. 2003), and comparative analyses in mammals identified a handful of previously unrecognized editing sites (Clutterbuck et al. 2005; Levanon et al. 2005).

Around this time, large numbers of cDNAs and cDNA fragments, many as expressed sequence tags (ESTs), were sequenced and mapped to genomic DNA sequences (Boguski et al. 1993; Kikuno et al. 2002; Ota et al. 2004). Several research groups recognized that cDNA and EST libraries, when compared with genomic sequences, could be used to identify editing-dependent RNA–DNA differences (Athanasiadis et al. 2004; Blow et al. 2004; Kim et al. 2004; Levanon et al. 2004) and found tens of thousands of unique editing sites in human transcripts. These studies indicated that A-to-G conversions, the most abundant class of RNA–DNA mismatches, were abundant in repetitive sequences, especially *Alu* retroelements, and displayed nearest neighbor preferences similar to those observed for ADAR in vitro (Polson and Bass 1994; Lehmann and Bass 2000). Approaches to align and compare libraries of cDNA sequences proved powerful, but ultimately were limited by the number of cDNAs and ESTs available. As described below, using modern sequencing technology, one can sequence millions of cDNAs per sample and detect editing using rapid and robust pipelines.

5 FINDING EDITED RNAs IN HIGH-THROUGHPUT DATA SETS

Although early cDNA and EST sequencing projects compiled tens to hundreds of thousands of cellular sequences, high-throughput technologies like RNA-seq rapidly produce tens to hundreds of millions of sequences per sample (Wang et al. 2009). RNA-seq facilitates genome-wide profiling of RNA expression and editing patterns and can easily be applied to different organisms, cell types, and conditions, for comparative analyses. Several previous reviews provide useful information for detecting edited RNAs with high-throughput protocols (Eisenberg et al. 2010; Ramaswami and Li 2016).

Like Sanger sequencing and cDNA/EST approaches, RNA-seq is used to identify editing sites by searching for

A-to-G differences between RNA-derived cDNA libraries and matched genomic sequences. First and foremost, finding RNA editing sites requires that edited RNAs are well represented among the transcripts sequenced in an experiment. Samples dominated by abundant transcripts like ribosomal RNA (rRNA) will not have sufficient sequencing coverage of edited transcripts to reliably distinguish editing sites from sequencing errors and genomic variants. Equally important is the approach used to align sequencing reads to the genome. By definition, edited reads will not perfectly align to the genome, and mismatch parameters for alignment and read filtering must be carefully chosen to avoid discarding edited transcripts (Lee et al. 2013).

Editing sites that occur at low frequency are particularly problematic. A site naturally edited in only 10% of cellular transcripts on average will appear as G in only one of ten reads. Increasing the number of sequencing reads will increase the average number of times a particular nucleotide is represented, or the depth of coverage (Sims et al. 2014). Given a site edited at frequency f , one can calculate the number of reads, n , that provide probability P of observing editing at that site using the formula

$$n = \log_{1-f}(1 - P).$$

For instance, one needs approximately 18 reads to have a 90% chance of observing editing at a site edited in 12% of endogenous transcripts, whereas coverage of approximately five reads only provides an ~50% chance. In practice, costs of high-throughput sequencing are balanced with coverage demands. Regardless, coverage of edited transcripts will increase by removing abundant unedited transcripts or by biochemically enriching for edited RNAs, without significant increase in costs. Thus, this is standard protocol in analyses of edited transcripts.

5.1 Enriching for Edited Transcripts

High-abundance transcripts, especially rRNAs, present the major obstacle to obtaining deep coverage of edited RNAs. Because rRNA makes up 80%–90% of most total RNA samples (O’Neil et al. 2013), its inclusion in sequencing libraries substantially reduces the fraction of RNA-seq reads that contain editing information. Typically, rRNA is removed from an RNA sample before library preparation, either by selecting polyadenylated RNAs with oligo(dT) or depleting rRNAs with bead-conjugated antisense oligos (ribo-minus) (O’Neil et al. 2013; Zhao et al. 2014). Although oligo(dT) capture removes rRNA, it also depletes transcripts that lack a poly(A) tail, including nascent pre-mRNAs that have not yet acquired a poly(A) tail, spliced intron lariats, and abundant polymerase III transcripts like

tRNAs. Ribo-minus protocols, although they remove rRNA less efficiently than oligo(dT) capture, do not deplete intronic and intergenic sequences and other abundant non-coding RNAs (Cui et al. 2010). If the goal is to look for editing sites in mature mRNAs, poly(A) selection is an excellent method. However, if the goal is a comprehensive determination of editing sites, including those that frequently occur in introns (Table 1) (Whipple et al. 2015; Zhao et al. 2015; Blango and Bass 2016), ribo-minus treatment to remove rRNA is preferable.

In addition to increasing the complexity of RNA samples by removing abundant, unedited transcripts, biochemical approaches can be used to enrich for ADAR substrates in RNA-seq libraries. An immunoprecipitation approach to enrich for dsRNA with the J2 anti-dsRNA antibody, raised against L-dsRNA, from a “killer” virus of *Saccharomyces cerevisiae* (Schönborn et al. 1991), has been used in studies of human Dicer (Kaneko et al. 2011), TDP-1, the *C. elegans* ortholog of TDP-43 (Saldi et al. 2014), and in RNA-seq approaches to map dsRNAs using ADAR editing sites (Whipple et al. 2015; Blango and Bass 2016; Reich et al. 2018). Although this immunoprecipitation approach can improve coverage of dsRNAs, its reported efficiency is modest. In *C. elegans*, edited sequences were enriched about twofold (Reich et al. 2018), and transposon sequences, which often form inverted repeat structures, were enriched four- or fivefold (Saldi et al. 2014). When applied to mouse RNAs, J2 immunoprecipitation improved the number of EERs detected by at most 1.8-fold, suggesting it may have limited benefit in certain systems (Blango and Bass 2016). Other research groups have immunoprecipitated ADARs to isolate bound RNAs and compare substrate specificity of different ADAR isoforms and paralogs (Wang et al. 2013). Although not all ADAR-bound RNAs were edited, these groups observed edited RNAs enriched among ADAR-bound transcripts and in one case identified thousands of novel editing sites (Wang et al. 2013). Still, immunoprecipitation methods run risks. For example, it is unknown whether hyperedited transcripts with numerous IU mismatches are efficiently immunoprecipitated with the J2 antibody. Nonetheless, although not essential to define edited sequences, enrichment protocols can improve coverage of edited transcripts.

Sequencing RNA from genetic mutants that either accumulate dsRNA or lack proteins that compete with ADARs for binding dsRNA can also increase coverage of edited sequences and/or elevate editing levels. Several *C. elegans* studies used mutant strains, including *dcr-1(mg375)*, which contains a mutation in the helicase domain of *C. elegans* Dicer that results in deficient endo-siRNA processing (Whipple et al. 2015), and *tdp-1(ok803)*, which accumulates excess cellular dsRNA (Saldi et al. 2014). Indeed, the *tdp-1*

mutant showed increased editing in ~40% of well-represented edited RNAs. In mammalian cells, reduction in *N*⁶-methyladenosine RNA modification by *METTL3* or *METTL14* knockdown likewise results in elevated A-to-I editing at many sites (Xiang et al. 2018). Performing RNA-seq on mutants with elevated editing will enable determination of a more comprehensive dsRNAome. However, if the goal is to determine biologically relevant editing sites, it is important to remember that such genetic perturbations will identify editing sites that may not exist in the wild-type context.

A-to-I editing only occurs in dsRNA structures, which may impede efficient reverse transcription. The use of thermostable reverse transcriptases at high temperatures, or addition of organic solvents like dimethylsulfoxide (DMSO), can relax secondary structure to promote cDNA synthesis from long RNA duplexes (Yasukawa et al. 2010; Mohr et al. 2013; Whipple et al. 2015; Nottingham et al. 2016). However, these adaptations are usually unnecessary, because RNA fragmentation during library preparation reduces the length, and thus the stability, of base-paired regions, and further, dsRNAs containing many IU mismatches are easier to reverse transcribe with conventional enzymes (Whipple et al. 2015). Importantly, library preparation protocols should incorporate high-fidelity enzymes to maintain low error rates during cDNA synthesis (Lee et al. 2013).

5.2 Sequencing and Alignment Protocols

Most RNA editing studies use Illumina sequencing platforms because of their ability to sequence relatively long reads at deep coverage, providing around 200 million reads per lane (Diroma et al. 2017). The high sequence information content of long reads (>100 nt) enables accurate mapping to a reference genome, especially across splice junctions and in repetitive regions. Paired-end protocols effectively double the information per read by sequencing from both 5' and 3' ends (Chhangawala et al. 2015). Strand-ed library construction protocols further refine mapping information and distinguish between A-to-G and T-to-C conversions (Mills et al. 2013). We typically perform stranded, paired-end RNA-seq with >100-nt reads for editing detection studies, because paired-end reads with ~200-nt total sequence rarely align to more than one genomic location, even if they include repetitive sequences.

Early studies used cloned cDNAs to determine the number of editing sites per transcript and defined transcripts with A-to-G changes at >10% of adenosines as non-selectively edited or hyperedited (Bass 1997). Although a number of viral transcripts were found to be hyperedited, with a few exceptions (Morse and Bass 1999; Morse et al.

2002), endogenous transcripts showed editing at <10% of the adenosines in a single base-paired region. Information about the number of editing sites in a single transcript is often limited when using Illumina RNA-seq protocols, because individual reads are comparatively short. Compilation of all reads for a given transcript indicates the number of potential editing sites, but not the number of sites that can occur in a single transcript. High-throughput protocols suited for sequencing longer reads, such as the 300-nt paired-end reads provided by Illumina MiSeq platforms, have been used to determine how many editing events occur in a single transcript (Wheeler et al. 2015). Such studies provide information about how editing at one position influences editing at other sites.

After library construction and sequencing, quality control software such as FastQC (see bioinformatics.babraham.ac.uk/projects/fastqc/) assess base quality and sequence complexity in the resulting RNA-seq reads. Preprocessing programs like Cutadapt (Martin 2011) trim reads to remove adapter sequences, poly(A) tails, and low-quality bases, after which reads are aligned to a reference genome. Alignment algorithms differ in run time and accuracy (Ruffalo et al. 2011; Borozan et al. 2013). Many aligners detect exon-intron junctions to map spliced RNA-seq reads, although aligners without this feature map spliced reads effectively if provided a table of splice junctions (Borozan et al. 2013; Diroma et al. 2017).

Two ways are commonly used to align reads so as to avoid discarding those with ADAR editing sites. The first uses standard alignment algorithms, but increases the number of mismatches allowed in a single read, whereas the second uses “editing aware” alignment algorithms to align cDNA sequences with either an A or G to a genomic A. Using a standard aligner, one must allow for mismatched bases to map edited reads, because the edited sequence varies from the genomic sequence. The optimal number of mismatches allowed depends on read length, because a 100-nt read with four possible mismatches will generally align to fewer locations than a 50-nt read with the same number of mismatches. Because most *C. elegans* edited RNAs have four or fewer editing events per molecule (Wheeler et al. 2015), we permit four mismatches when using a standard aligner to map 100-nt paired-end reads from *C. elegans*. For other organisms, different mismatch parameters should be tested to find those that accurately map edited reads while minimizing mapping to multiple locations.

The second approach to mapping ADAR-edited sequences relies on alignment programs like GSNAP and GNUMAP that include “editing aware” modes. These aligners are particularly adept at mapping highly edited sequences (Clement et al. 2010; Wu and Nacu 2010;

Hong et al. 2013). In one comparison, GNUMAP identified more editing sites and editing clusters than the “editing unaware” aligner, Novoalign, which aligned 100-nt reads with up to four mismatches (Whipple et al. 2015). Applying a similar “editing aware” principle, several research groups used a “three-base” approach, so-called because it takes highly edited reads unmapped by conventional alignment and realigns them after converting all As to Gs in read and reference genome sequences, thus restricting the genome to three bases: G, C, and T (Wu et al. 2011; St Laurent et al. 2013; Porath et al. 2014; Zhao et al. 2015). Three-base methods involve additional alignment and computational steps and risk mapping reads that contain G-to-A mismatches (Porath et al. 2014). Thus, editing-aware algorithms (e.g., GNUMAP) that accurately map highly edited sequences in a single alignment step are usually preferable.

5.3 Tools and Approaches to Identify Editing Sites

A growing number of bioinformatics tools and pipelines are publicly available for detecting ADAR editing sites (Porath et al. 2014; Whipple et al. 2015; Blango and Bass 2016; Deffit and Hundley 2016; Diroma et al. 2017). Editing detection pipelines use variant calling to identify A-to-G conversions in RNA-seq reads, and then assess if A-to-G changes represent true editing events. Assessment approaches aim to distinguish editing sites from genetic polymorphisms and sequencing errors, typically through statistical analyses or filtering based on location and frequency of editing (Porath et al. 2014; Deffit and Hundley 2016; Wang et al. 2016; Diroma et al. 2017; John et al. 2017). Our laboratory uses applications in the USeq (see github.com/HuntsmanCancerInstitute/USeq) and SAMtools (see github.com/samtools/) sequencing analysis suites to identify edited sites and define editing clusters (Whipple et al. 2015; Blango and Bass 2016; Reich et al. 2018). In addition to variant-calling (SAMtools mpileup) and editing-detection functions (USeq RNAEditingPileupParser), these analysis suites provide useful, intuitive programs to work with sequencing data sets and compare editing sites to other genomic features. Other pipelines are specifically for detecting RNA editing, and a recent review evaluated five of these for accuracy and sensitivity (Diroma et al. 2017). Results varied based on alignment algorithm used, but overall the RNAEditor, JACUSA, and REDIttools pipelines predicted the most editing sites, whereas GIREMI predicted fewer sites but with a lower false discovery rate. All-inclusive programs like RNAEditor and RES-Scanner are useful for those with little experience to quickly and easily identify editing sites, as they incorporate alignment, detection, and filtering steps to define editing sites from raw sequencing files (Wang et al. 2016; John et al. 2017). However, for

determination of dsRNAomes, we prefer the EER pipeline (described below), which defines editing clusters to identify long dsRNAs (Whipple et al. 2015; Blango and Bass 2016; Reich et al. 2018).

In addition to RNA-seq-based approaches that identify editing through A-to-G changes, high-throughput sequencing can be applied to inosine-specific detection methods. Inosine-specific cleavage with RNase T1, an early tool for editing detection (Morse and Bass 1999), was paired with RNA-seq to identify 665 editing sites in mouse brain, many of which were novel (Cattenoz et al. 2013). Inosine reactivity with acrylonitrile underlies another chemical detection method, termed inosine chemical erasing (ICE) (Sakurai et al. 2010). High-throughput sequencing improves the scale and sensitivity of these methods (Cattenoz et al. 2013; Sakurai et al. 2014; Suzuki et al. 2015), which distinguish inosine from A-to-G changes caused by genetic polymorphisms or sequencing errors. However, these methods rely on chemical conversion methods that are not 100% efficient; the recommended conditions for ICE convert only 80%–90% of inosines at the highly edited *GRIA2* Q/R site (Suzuki et al. 2015). Further, because inosine-specific detection methods require complex analysis pipelines and suffer from the same read coverage issues as conventional editing detection pipelines (Cattenoz et al. 2013; Sakurai et al. 2014; Suzuki et al. 2015), they do not provide a strong advantage over other approaches.

5.4 Excluding False Positives

Effective editing pipelines must distinguish true RNA editing events from editing-independent mismatches that arise because of single-nucleotide polymorphisms (SNPs) and errors in sequencing and alignment. Controlling for SNPs requires accurately identifying genetic variants and removing them from analysis. Sequencing genomic DNA from the same samples used for RNA-seq provides a definitive solution (Picardi and Pesole 2013; Wang et al. 2016) by providing a more accurate genomic sequence than a published reference. Alternatively, RNA-seq analyses of ADAR mutant strains can validate that identified A-to-G conversions require ADAR (Bahn et al. 2012; Zhao et al. 2015). Although these approaches effectively control for genomic variation, they are expensive and resource-intensive, and SNPs typically make up a tiny fraction of A-to-G changes identified (<0.1% in Zhao et al. 2015). A cheap alternative is to simply remove SNPs recorded in public variant databases from potential editing sites (Ramaswami et al. 2013; Bazak et al. 2014; Whipple et al. 2015; Blango and Bass 2016; Diroma et al. 2017; Tan et al. 2017). Several research groups have shown that true editing sites can be accurately identified using only RNA-seq data, without the need for sample-

matched genomic sequence (Ramaswami et al. 2013; Zhu et al. 2013; Zhang and Xiao 2015). However, without an available reference sequence to align reads, DNA-seq is important to determine RNA–DNA differences (Alon et al. 2015).

Filtering out sequencing errors is critical for accurate editing detection. Higher Illumina error rates at read ends, and mismatches introduced by random hexamer priming, lead to clustering of sequencing errors at the ends of reads (Minoche et al. 2011; van Gurp et al. 2013). Thus, 5' and 3' read ends are typically trimmed before variant calling either by a fixed amount or according to mismatch density (Bazak et al. 2014; Whipple et al. 2015; Zhao et al. 2015; Wang et al. 2016). Reads with multiple non-A-to-G mismatches are also typically filtered out, because additional mismatches are an indication of poor sequence quality or inaccurate mapping (Blango and Bass 2016; Deffit et al. 2017). Repetitive genomic regions, in particular, risk issues of mismapping (Eisenberg 2012). Although reads that map to multiple repetitive regions can be removed, editing is most abundant in repetitive sequences (Bazak et al. 2014; Whipple et al. 2015; Blango and Bass 2016; Porath et al. 2017). Filtering out repetitively mapped reads removes sequences carrying true editing sites. Sequencing protocols that provide longer reads facilitate more accurate mapping, including within repetitive sequences, so these protocols reduce the number of reads needed to be filtered due to ambiguous alignment. Even with long reads, we recommend including repetitively mapped reads, as they are rich sources of editing information.

Several approaches can be used to validate that alignment and detection parameters are appropriate for editing discovery. The most straightforward is to choose candidate edited regions for editing validation by another method, typically cDNA amplification and Sanger sequencing (Li et al. 2009). This approach is accurate and sensitive, but it can be time-consuming to test many candidates. We recommend initially interrogating data sets for the presence of known ADAR targets. By comparing known editing sites to experimentally determined patterns, one can identify and address potential issues in bioinformatic pipelines. For instance, the absence of editing within well-expressed ADAR substrates might indicate that edited reads are being discarded during alignment or filtering. Publicly available databases, including DARNED, RADAR, and REDiportal, provide curated information on RNA editing sites, incorporating annotation, editing level, and tissue-specific editing information (Kiran et al. 2013; Ramaswami and Li 2014; Picardi et al. 2017). However, these databases do not always include up-to-date information; whereas DARNED and RADAR only provide data for human, mouse, and *Drosophila*, REDiportal currently includes only human infor-

mation (Kiran et al. 2013; Ramaswami and Li 2014; Picardi et al. 2017). Without curated data, we have used published editing targets (Morse et al. 2002; Hellwig and Bass 2008) to validate parameters for detecting editing in *C. elegans* data sets (Whipple et al. 2015). Once the pipeline is considered optimal, it is standard practice to verify editing sites in a subset of newly identified targets using cDNA amplification and Sanger sequencing, especially those chosen for more in-depth study.

5.5 Defining Editing Clusters and Determining a dsRNAome

In highly base-paired dsRNAs, ADARs nonselectively deaminate many adenosines to inosine (Nishikura 2010; Samuel 2011; Deffit and Hundley 2016), resulting in clusters of A-to-G changes in sequencing reads. Defining editing clusters provides increased sensitivity to identify true editing events, because sequencing errors or genomic variants are unlikely to result in clusters of a single variant (A-to-G) (Bazak et al. 2014; Zhao et al. 2015). Importantly, editing clusters also indicate the presence of long, highly base-paired dsRNA (Fig. 1) and, when mapped on a genome-wide scale, allow the determination of a dsRNAome (Fig. 2). We have determined dsRNAomes for specific cell types of human and mouse, as well as *C. elegans* (Whipple et al. 2015; Blango and Bass 2016).

Our EER-detection pipeline uses a window-scanning approach to find genomic regions covered by RNA-seq reads carrying clustered A-to-G changes (Whipple et al. 2015; Blango and Bass 2016; Reich et al. 2018). Genomic regions with sufficient read coverage (≥ 5 reads) are scanned in small windows, typically 50 nt, to identify windows containing ≥ 3 sites with A-to-G changes in $>1\%$ of reads. We chose read coverage and editing thresholds to optimize the sensitivity of EER detection without markedly increasing the false discovery rate. Alternative parameters could be used to predict greater numbers of clusters, or to identify clusters with extremely low false-positive rates.

Once edited windows are determined, overlapping windows are merged, and then combined with other merged windows separated by a predetermined “gap” distance to define the EER. The gap parameter connects complementary regions of an intramolecular dsRNA that may be separated by unedited, intervening sequences. The optimal gap distance varies between organisms. For *C. elegans*, a gap of 1 kb merges complementary sequences of a single dsRNA without also combining dsRNAs from separate, independent transcripts (Whipple et al. 2015), whereas mouse and human dsRNAomes require a longer 2.5-kb gap (Blango and Bass 2016). We typically determine optimal gap distance by testing several lengths and manually surveying

several dozen EERs in a genome browser to determine if single structures are merged without also merging separate transcripts. Including a gap parameter occasionally causes closely juxtaposed independent structures (as in *eIF-2 α* pre-mRNA in Fig. 1) to be merged into one EER; however, it is rare that structures from different genes are merged (Whipple et al. 2015; Blango and Bass 2016).

Once EERs are defined, additional methods are used to validate that dsRNAs have been accurately determined (Whipple et al. 2015; Blango and Bass 2016). RNA secondary structure prediction algorithms like UNAFold (Markham and Zuker 2008) provide a measure of the thermodynamic stability of predicted EER structures. Length-matched random regions provide a control to confirm that predicted dsRNAs are more structured than expected by chance. Length-matched control sets should sample properties of the transcriptome that meet the same criteria used to define dsRNAs. For instance, if the pipeline dictates that EERs are defined only in regions covered by more than five reads, control regions should be restricted to the same read coverage threshold. We use the BEDtools2 (see github.com/arq5x/bedtools2) application `shuffleBed` to make length-matched control regions, because it can randomly permute regions across the genome or within specified areas. For edited sequences that do not have predicted duplex structure, we use the BLAT sequence alignment tool to find complementary genomic sequences that may facilitate intermolecular dsRNA formation (Whipple et al. 2015). Similar approaches to detect proximal (within 10 kb) complementary sequences with the *bl2seq* algorithm have been used in other editing cluster analyses (Porath et al. 2014; 2017). Examining nearest neighbor preferences of EER editing sites validates that sites show characteristics of ADAR editing (Whipple et al. 2015; Blango and Bass 2016).

6 LOOKING TOWARD THE FUTURE

6.1 Determining Other dsRNAomes

To date, ADAR editing sites have been determined for approximately 23 metazoan species (Liscovitch-Brauer et al. 2017; Porath et al. 2017) and 53 human tissues (Tan et al. 2017). In all cases, editing shows similar properties (Table 1). It predominates in noncoding sequences, particularly introns and 3' UTRs (Whipple et al. 2015; Blango and Bass 2016; Liscovitch-Brauer et al. 2017; Porath et al. 2017), and frequently occurs in mobile element-derived repetitive sequences predicted to form stable intramolecular structures (Fig. 3) (Bazak et al. 2014; Whipple et al. 2015; Blango and Bass 2016; Porath et al. 2017).

In most cases, editing information from these studies has not been used to map structures and determine a

dsRNAome. Such analyses would be straightforward because existing data can simply be mined and analyzed with available pipelines. EERs and editing sites can be viewed in a genome browser, allowing researchers to evaluate their favorite genes. BED files of existing dsRNAomes (Whipple et al. 2015; Blango and Bass 2016; Reich et al. 2018) are freely available in the Gene Expression Omnibus repository.

An unanswered question is whether all long dsRNAs are edited. Most dsRNAomes were determined using ADAR editing sites as an *in vivo* signature of dsRNA, so dsRNAs without editing sites were not identified. Possibly, certain dsRNAs are protected from ADAR editing by other dsRBPs, and indeed, there are clear examples of dsRBPs competing for the same substrates (Warf et al. 2012; Elbarbary et al. 2013; Sakurai et al. 2017). In a few cases, cellular dsRNAs were identified by other criteria, including accessibility to nuclease (Li et al. 2012), chemical probes (Lucks et al. 2011), or *in vivo* cross-linking to dsRBPs (Rybak-Wolf et al. 2014). About 18% of *C. elegans* EERs (Whipple et al. 2015) overlap 9972 ssRNA nuclease-resistant sites (Li et al. 2012), validating the double-stranded character of EERs. Nonoverlapping sites could include dsRNA that is protected from ADAR editing, but the latter study was not focused on long dsRNAs, and many nonoverlapping sites are short dsRNA regions buried in tertiary structure that are inaccessible to ADARs. About 26% of human EER-associated genes overlap DICER-bound human genes (Rybak-Wolf et al. 2014; Blango and Bass 2016), consistent with the idea that dsRBPs compete for dsRNA structures. These two studies were performed in different cell types, so again, further analyses are required to evaluate whether nonoverlapping sites are dsRNAs that are protected from editing. In future studies, it will be interesting to obtain definitive information about competition between dsRBPs and determine if competition is affected by dsRBP tissue-specificity, abundance, or intracellular localization.

6.2 What Is the Function of the dsRNAome?

The dsRNAomes determined so far indicate that EERs do not show a predilection for the type of gene they inhabit; they are typically enriched in genes expected to be expressed in the particular cell type or condition being studied (Blango and Bass 2016). Is the dsRNAome just a vestige of an ongoing battle with mobile elements? Silencing mobile element expression is essential for viability (Friedli and Trono 2015), and studies in *C. elegans* (Reich et al. 2018) and *D. melanogaster* (Savva et al. 2013) are consistent with the idea that ADARs exist to allow expression of mRNAs that contain repetitive elements and otherwise would be silenced.

Selfish elements that integrate into a genome over time can become recognized as “self,” so the above idea is closely tied with studies indicating that ADARs function to mark cellular dsRNA as self (Hartner et al. 2009; Mannion et al. 2014; Liddicoat et al. 2015; Pestal et al. 2015; George et al. 2016; Reich et al. 2018). The simple idea is that ADARs, which are typically in the nucleus, target cellular dsRNA, whereas viruses, which often replicate in the cytoplasm, are protected from deamination. However, the actual situation is clearly more complicated. Some viruses replicate in the nucleus, such as HDV, which uses editing for regulating its life cycle (Polson et al. 1996). The ADAR1 isoform most closely associated with an immune response, ADAR1p150, is expressed in response to interferon and is found in both the nucleus and cytoplasm (Patterson and Samuel 1995). Mounting evidence indicates additional regulation during viral infection. ADAR1p110 is degraded in response to interferon (Li et al. 2016), whereas the timing of interferon induction of ADAR1p150 is balanced with induction of MDA5 and other proteins that mediate the mammalian immune response (Ahmad et al. 2018). As discussed earlier, some viral transcripts are edited by ADARs (Pfaller et al. 2015), and ADARs can have both proviral and antiviral effects (Samuel 2011). Repurposing ADARs as antiviral factors emphasizes their roles in innate immunity. Indeed, *ADAR1* shows evidence of positive selection, suggesting it has adapted through genetic conflict with viruses (Forni et al. 2015).

ADARs likely arose, and continue to function, for these global roles, but have also been co-opted for additional purposes, such as mRNA recoding (Higuchi et al. 2000; Jepson and Reenan 2009; Garrett and Rosenthal 2012; Lis-covitch-Brauer et al. 2017), creating or destroying splice sites or altering splicing regulatory sequences (Rueter et al. 1999; Solomon et al. 2013). A key issue in future studies will be to definitively connect observed phenotypes with specific editing events. In *C. elegans* lacking ADARs, most unedited dsRNAs, rather than a small subset, are processed into siRNAs by antiviral RNAi machinery (Reich et al. 2018). Similarly, in ADAR1-deficient human cells, MDA5 oligomerizes on hundreds of inverted-repeat *Alu* elements and triggers downstream immune signaling (Ahmad et al. 2018). However, in *C. elegans* and mammals, it is unclear if all, or only some, dsRNAs are relevant to mutant phenotypes. The early lethality of *ADAR2*^{-/-} mutant mice and chemotaxis defects of *adr-1;adr-2* mutant *C. elegans* are largely ascribed to editing of single targets (Higuchi et al. 2000; Deffit et al. 2017). Understanding if the same applies to immune-relevant ADAR-dependent pathologies could allow potential treatments to specifically target relevant substrates without disrupting global ADAR functions.

REFERENCES

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Ahmad S, Mu X, Yang F, Greenwald E, Park JW, Jacob E, Zhang C-Z, Hur S. 2018. Breaching self-tolerance to Alu duplex RNA underlies MDA5-mediated inflammation. *Cell* **172**: 797–810.e13.
- Ahringer J, Gasser SM. 2018. Repressive chromatin in *Caenorhabditis elegans*: Establishment, composition, and function. *Genetics* **208**: 491–511.
- Alon S, Garrett SC, Levanon EY, Olson S, Graveley BR, Rosenthal JJC, Eisenberg E. 2015. The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. *eLife* **4**: e05198.
- Aruscavage PJ, Bass BL. 2000. A phylogenetic analysis reveals an unusual sequence conservation within introns involved in RNA editing. *RNA* **6**: 257–269.
- Athanasiadis A, Rich A, Maas S. 2004. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* **2**: e391.
- Bahn JH, Lee J-H, Li G, Greer C, Peng G, Xiao X. 2012. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res* **22**: 142–150.
- Bajad P, Jantsch MF, Keegan L, O'Connell M. 2017. A to I editing in disease is not fake news. *RNA Biol* **14**: 1223–1231.
- Bass BL. 1997. RNA editing and hypermutation by adenosine deamination. *Trends Biochem Sci* **22**: 157–162.
- Bass BL. 2002. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* **71**: 817–846.
- Bass BL, Weintraub H. 1987. A developmentally regulated activity that unwinds RNA duplexes. *Cell* **48**: 607–613.
- Bass BL, Weintraub H. 1988. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* **55**: 1089–1098.
- Bass BL, Weintraub H, Cattaneo R, Billeter MA. 1989. Biased hypermutation of viral RNA genomes could be due to unwinding/modification of double-stranded RNA. *Cell* **56**: 331.
- Bass BL, Nishikura K, Keller W, Seeburg PH, Emeson RB, O'Connell MA, Samuel CE, Herbert A. 1997. A standardized nomenclature for adenosine deaminases that act on RNA. *RNA* **3**: 947–949.
- Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs FJ, Rechavi G, Li JB, 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.
- Behm M, Ohman M. 2016. RNA editing: A contributor to neuronal dynamics in the mammalian brain. *Trends Genet* **32**: 165–175.
- Berks M, Mapping CEG, Consortium S. 1995. The *C. elegans* genome sequencing project. *Genome Res* **5**: 99–104.
- Blango MG, Bass BL. 2016. Identification of the long, edited dsRNAome of LPS-stimulated immune cells. *Genome Res* **26**: 852–862.
- Blow M, Futreal PA, Wooster R, Stratton MR. 2004. A survey of RNA editing in human brain. *Genome Res* **14**: 2379–2387.
- Boguski MS, Lowe TM, Tolstoshev CM. 1993. dbEST—Database for “expressed sequence tags”. *Nat Genet* **4**: 332–333.
- Borozan I, Watt SN, Ferretti V. 2013. Evaluation of alignment algorithms for discovery and identification of pathogens using RNA-Seq. *PLoS ONE* **8**: e76935.
- C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282**: 2012–2018.
- Cattaneo R, Schmid A, Eschle D, Bacsko K, Meulen ter V, Billeter MA. 1988. Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell* **55**: 255–265.
- Cattenoz PB, Taft RJ, Westhof E, Mattick JS. 2013. Transcriptome-wide identification of A > I RNA editing sites by inosine specific cleavage. *RNA* **19**: 257–270.
- Chateigner-Boutin A-L, Small I. 2007. A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res* **35**: e114.
- Chen Y-C, Kao S-C, Chou H-C, Lin W-H, Wong F-H, Chow W-Y. 2008. A real-time PCR method for the quantitative analysis of RNA editing at specific sites. *Anal Biochem* **375**: 46–52.
- Chhangawala S, Rudy G, Mason CE, Rosenfeld JA. 2015. The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biol* **16**: 131.
- Clement NL, Snell Q, Clement MJ, Hollenhorst PC, Purwar J, Graves BJ, Cairns BR, Johnson WE. 2010. The GNUMAP algorithm: Unbiased probabilistic mapping of oligonucleotides from next-generation sequencing. *Bioinformatics* **26**: 38–45.
- Clutterbuck DR, Leroy A, O'Connell MA, Semple CAM. 2005. A bioinformatic screen for novel A-I RNA editing sites reveals recoding editing in BC10. *Bioinformatics* **21**: 2590–2595.
- Crowl JT, Gray EE, Pestal K, Volkman HE, Stetson DB. 2017. Intracellular nucleic acid detection in autoimmunity. *Annu Rev Immunol* **35**: 313–336.
- Cui P, Lin Q, Ding F, Xin C, Gong W, Zhang L, Geng J, Zhang B, Yu X, Yang J, et al. 2010. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* **96**: 259–265.
- Daugaard I, Hansen TB. 2017. Biogenesis and function of Ago-associated RNAs. *Trends Genet* **33**: 208–219.
- 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.
- Diroma MA, Ciaccia L, Pesole G, Picardi E. 2017. Elucidating the editome: Bioinformatics approaches for RNA editing detection. *Brief Bioinform* doi: 10.1093/bib/bbx129.
- Egebjerg J, Kukekov V, Heinemann SF. 1994. Intron sequence directs RNA editing of the glutamate receptor subunit GluR2 coding sequence. *Proc Natl Acad Sci* **91**: 10270–10274.
- Ehrenfeld E, Hunt T. 1971. Double-stranded poliovirus RNA inhibits initiation of protein synthesis by reticulocyte lysates. *Proc Natl Acad Sci* **68**: 1075–1078.
- Eigenbrod T, Keller P, Kaiser S, Rimbach K, Dalpke AH, Helm M. 2015. Recognition of specified RNA modifications by the innate immune system. *Meth Enzymol* **560**: 73–89.
- Eisenberg E. 2012. Bioinformatic approaches for identification of A-to-I editing sites. *Curr Top Microbiol Immunol* **353**: 145–162.
- Eisenberg E, Li JB, Levanon EY. 2010. Sequence based identification of RNA editing sites. *RNA Biol* **7**: 248–252.
- Elbarbary RA, Li W, Tian B, Maquat LE. 2013. STAU1 binding 3' UTR *IRAlus* complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes Dev* **27**: 1495–1510.
- Forni D, Mozzi A, Pontremoli C, Vertemara J, Pozzoli U, Biasin M, Bresolin N, Clerici M, Cagliani R, Sironi M. 2015. Diverse selective regimes shape genetic diversity at ADAR genes and at their coding targets. *RNA Biol* **12**: 149–161.
- Friedli M, Trono D. 2015. The developmental control of transposable elements and the evolution of higher species. *Annu Rev Cell Dev Biol* **31**: 429–451.
- Gallo A, Thomson E, Brindle J, O'Connell MA, Keegan LP. 2002. Microprocessing events in mRNAs identified by DHPLC analysis. *Nucleic Acids Res* **30**: 3945–3953.
- Garrett S, Rosenthal JJC. 2012. RNA editing underlies temperature adaptation in K⁺ channels from polar octopuses. *Science* **335**: 848–851.
- 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.
- Gilbert W. 1986. The RNA world. *Nature* **319**: 618.

- Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J, Hellsten U, Goodstein D, Couronne O, Tran-Gyamfi M, et al. 2004. The DNA sequence and biology of human chromosome 19. *Nature* **428**: 529–535.
- Hartner JC, Walkley CR, Lu J, Orkin SH. 2009. ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. *Nat Immunol* **10**: 109–115.
- Hellwig S, Bass BL. 2008. A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes. *Proc Natl Acad Sci* **105**: 12897–12902.
- Higuchi M, Single FN, Köhler M, Sommer B, Sprengel R, Seeburg PH. 1993. RNA editing of AMPA receptor subunit GluR-B: A base-paired intron–exon structure determines position and efficiency. *Cell* **75**: 1361–1370.
- Higuchi M, Maas S, Single FN, Hartner J, Rozov A, Burnashev N, Feldmeyer D, Sprengel R, Seeburg PH. 2000. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406**: 78–81.
- Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn K-A, Minoda A, Tolstorukov MY, Appert A, et al. 2014. Comparative analysis of metazoan chromatin organization. *Nature* **512**: 449–452.
- Hong C, Clement NL, Clement S, Hammoud SS, Carrell DT, Cairns BR, Snell Q, Clement MJ, Johnson WE. 2013. Probabilistic alignment leads to improved accuracy and read coverage for bisulfite sequencing data. *BMC Bioinformatics* **14**: 337.
- Hoopengardner B, Bhalla T, Staber C, Reenan R. 2003. Nervous system targets of RNA editing identified by comparative genomics. *Science* **301**: 832–836.
- Hurst SR, Hough RF, Aruscavage PJ, Bass BL. 1995. Deamination of mammalian glutamate receptor RNA by *Xenopus* dsRNA adenosine deaminase: Similarities to in vivo RNA editing. *RNA* **1**: 1051–1060.
- Jepson JEC, Reenan RA. 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.
- John D, Weirick T, Dimmeler S, Uchida S. 2017. RNAEditor: Easy detection of RNA editing events and the introduction of editing islands. *Brief Bioinform* **18**: 993–1001.
- Joyce GF. 2002. The antiquity of RNA-based evolution. *Nature* **418**: 214–221.
- Kaneko H, Dridi S, Tarallo V, Gelfand BD, Fowler BJ, Cho WG, Kleinman ME, Ponicsan SL, Hauswirth WW, Chiodo VA, et al. 2011. DICER1 deficit induces *Alu* RNA toxicity in age-related macular degeneration. *Nature* **471**: 325–330.
- Kikuno R, Nagase T, Waki M, Ohara O. 2002. HUGE: A database for human large proteins identified in the Kazusa cDNA sequencing project. *Nucleic Acids Res* **30**: 166–168.
- Kim U, Garner TL, Sanford T, Speicher D, Murray JM, Nishikura K. 1994. Purification and characterization of double-stranded RNA adenosine deaminase from bovine nuclear extracts. *J Biol Chem* **269**: 13480–13489.
- Kim DDY, Kim TTY, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A. 2004. Widespread RNA editing of embedded *Alu* elements in the human transcriptome. *Genome Res* **14**: 1719–1725.
- Kin T, Ono Y. 2007. Idiographica: A general-purpose web application to build idiograms on-demand for human, mouse and rat. *Bioinformatics* **23**: 2945–2946.
- Kiran AM, O'Mahony JJ, Sanjeev K, Baranov PV. 2013. Darned in 2013: Inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res* **41**: D258–D261.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Lässig C, Matheisl S, Sparrer KMJ, de Oliveira Mann CC, Moldt M, Patel JR, Goldeck M, Hartmann G, García-Sastre A, Hornung V, et al. 2015. ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA. *eLife* **4**: e10859.
- Lee J-H, Ang JK, Xiao X. 2013. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA* **19**: 725–732.
- Lehmann KA, Bass BL. 1999. The importance of internal loops within RNA substrates of ADAR1. *J Mol Biol* **291**: 1–13.
- Lehmann KA, Bass BL. 2000. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**: 12875–12884.
- Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Szybel D, et al. 2004. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* **22**: 1001–1005.
- Levanon EY, Hallegger M, Kinar Y, Shemesh R, Djinovic-Carugo K, Rechavi G, Jantsch MF, Eisenberg E. 2005. Evolutionarily conserved human targets of adenosine to inosine RNA editing. *Nucleic Acids Res* **33**: 1162–1168.
- Li JB, Levanon EY, Yoon J-K, Aach J, Xie B, Leproust E, Zhang K, Gao Y, Church GM. 2009. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**: 1210–1213.
- Li F, Zheng Q, Ryvkin P, Dragomir I, Desai Y, Aiyer S, Valladares O, Yang J, Bambina S, Sabin LR, et al. 2012. Global analysis of RNA secondary structure in two metazoans. *Cell Rep* **1**: 69–82.
- Li L, Qian G, Zuo Y, Yuan Y, Cheng Q, Guo T, Liu J, Liu C, Zhang L, Zheng H. 2016. Ubiquitin-dependent turnover of adenosine deaminase acting on RNA 1 (ADAR1) is required for efficient antiviral activity of Type I interferon. *J Biol Chem* **291**: 24974–24985.
- 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.
- Liscovitch-Brauer N, Alon S, Porath HT, Elstein B, Unger R, Ziv T, Admon A, Levanon EY, Rosenthal JJC, Eisenberg E. 2017. Trade-off between transcriptome plasticity and genome evolution in cephalopods. *Cell* **169**: 191–202.
- Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung M-S, Ercan S, Ikegami K, Jensen M, Kolasinska-Zwiercz P, et al. 2011. Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res* **21**: 227–236.
- Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP. 2011. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci* **108**: 11063–11068.
- Ma E, MacRae IJ, Kirsch JF, Doudna JA. 2008. Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol* **380**: 237–243.
- Mannion NM, Greenwood SM, Young R, Cox S, Brindle J, Read D, Nelläker 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.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**: 10–12.
- Melcher T, Maas S, Herb A, Sprengel R, Seeburg PH, Higuchi M. 1996. A mammalian RNA editing enzyme. *Nature* **379**: 460–464.
- Mills JD, Kawahara Y, Janitz M. 2013. Strand-specific RNA-seq provides greater resolution of transcriptome profiling. *Curr Genomics* **14**: 173–181.
- Minoche AE, Dohm JC, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* **12**: R112.
- Mohr S, Ghanem E, Smith W, Sheeter D, Qin Y, King O, Polioudakis D, Iyer VR, Hunicke-Smith S, Swamy S, et al. 2013. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**: 958–970.
- Morse DP, Bass BL. 1997. Detection of inosine in messenger RNA by inosine-specific cleavage. *Biochemistry* **36**: 8429–8434.

- Morse DP, Bass BL. 1999. Long RNA hairpins that contain inosine are present in *Caenorhabditis elegans* poly(A)⁺ RNA. *Proc Natl Acad Sci* **96**: 6048–6053.
- Morse DP, Aruscavage PJ, Bass BL. 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* **99**: 7906–7911.
- Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Neeman Y, Levanon EY, Jantsch MF, Eisenberg E. 2006. RNA editing level in the mouse is determined by the genomic repeat repertoire. *RNA* **12**: 1802–1809.
- Nishikura K. 2010. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* **79**: 321–349.
- Nishikura K. 2016. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **17**: 83–96.
- Nottingham RM, Wu DC, Qin Y, Yao J, Hunnicke-Smith S, Lambowitz AM. 2016. RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA* **22**: 597–613.
- O'Connell MA, Mannion NM, Keegan LP. 2015. The epitranscriptome and innate immunity. *PLoS Genet* **11**: e1005687.
- Ohlson J, Enstero M, Sjöberg B-M, Ohman M. 2005. A method to find tissue-specific novel sites of selective adenosine deamination. *Nucleic Acids Res* **33**: e167.
- Ohman M, Källman AM, Bass BL. 2000. In vitro analysis of the binding of ADAR2 to the pre-mRNA encoding the GluR-B R/G site. *RNA* **6**: 687–697.
- O'Neil D, Glowatz H, Schlumpberger M. 2013. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr Protoc Mol Biol* doi: 10.1002/0471142727.mb0419s103.
- Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, et al. 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* **36**: 40–45.
- Parker GS, Maity TS, Bass BL. 2008. dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. *J Mol Biol* **384**: 967–979.
- Patterson JB, Samuel CE. 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.
- Paul MS, Bass BL. 1998. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J* **17**: 1120–1127.
- 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.
- Pfaller CK, Mastorakos GM, Matchett WE, Ma X, Samuel CE, Cattaneo R. 2015. Measles virus defective interfering RNAs are generated frequently and early in the absence of C protein and can be destabilized by adenosine deaminase acting on RNA-1-like hypermutations. *J Virol* **89**: 7735–7747.
- Picardi E, Pesole G. 2013. REDIttools: High-throughput RNA editing detection made easy. *Bioinformatics* **29**: 1813–1814.
- Picardi E, D'Erchia AM, Giudice Lo C, Pesole G. 2017. REDiportal: A comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res* **45**: D750–D757.
- Polson AG, Bass BL. 1994. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J* **13**: 5701–5711.
- Polson AG, Crain PF, Pomerantz SC, McCloskey JA, Bass BL. 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.
- Polson AG, Bass BL, Casey JL. 1996. RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature* **380**: 454–456.
- Porath HT, Carmi S, Levanon EY. 2014. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat Commun* **5**: 4726.
- Porath HT, Knisbacher BA, Eisenberg E, Levanon EY. 2017. Massive A-to-I RNA editing is common across the metazoa and correlates with dsRNA abundance. *Genome Biol* **18**: 185.
- Prachumwat A, DeVincentis L, Palopoli MF. 2004. Intron size correlates positively with recombination rate in *Caenorhabditis elegans*. *Genetics* **166**: 1585–1590.
- Ramaswami G, Li JB. 2014. RADAR: A rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res* **42**: D109–D113.
- Ramaswami G, Li JB. 2016. Identification of human RNA editing sites: A historical perspective. *Methods* **107**: 42–47.
- Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'Connell MA, Li JB. 2013. Identifying RNA editing sites using RNA sequencing data alone. *Nat Meth* **10**: 128–132.
- Rebagliati MR, Melton DA. 1987. Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. *Cell* **48**: 599–605.
- Reich DP, Tyc KM, Bass BL. 2018. *C. elegans* ADARs antagonize silencing of cellular dsRNAs by the antiviral RNAi pathway. *Genes Dev* **32**: 271–282.
- Rich A. 1962. On the problems of evolution and biochemical information transfer. In *Horizons in biochemistry* (ed. Kasha M, Pullman B), pp. 103–126, Academic, New York.
- Rieder LE, Staber CJ, Hoopengardner B, Reenan RA. 2013. Tertiary structural elements determine the extent and specificity of messenger RNA editing. *Nat Commun* **4**: 2232.
- Robertson MP, Joyce GF. 2012. The origins of the RNA world. *Cold Spring Harb Perspect Biol* **4**: a003608.
- Rosenthal JJC. 2015. The emerging role of RNA editing in plasticity. *J Exp Biol* **218**: 1812–1821.
- Rueter SM, Dawson TR, Emeson RB. 1999. Regulation of alternative splicing by RNA editing. *Nature* **399**: 75–80.
- Ruffalo M, LaFramboise T, Koyutuerk M. 2011. Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics* **27**: 2790–2796.
- Rybak-Wolf A, Jens M, Murakawa Y, Herzog M, Landthaler M, Rajewsky N. 2014. A variety of dicer substrates in human and *C. elegans*. *Cell* **159**: 1153–1167.
- Sakurai M, Yano T, Kawabata H, Ueda H, Suzuki T. 2010. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. *Nat Chem Biol* **6**: 733–740.
- Sakurai M, Ueda H, Yano T, Okada S, Terajima H, Mitsuyama T, Toyoda A, Fujiyama A, Kawabata H, Suzuki T. 2014. A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res* **24**: 522–534.
- Sakurai M, Shiromoto Y, Ota H, Song C, Kossenkova AV, Wickramasinghe J, Showe LC, Skordalakes E, Tang H-Y, Speicher DW, et al. 2017. ADAR1 controls apoptosis of stressed cells by inhibiting Staufen1-mediated mRNA decay. *Nat Struct Mol Biol* **24**: 534–543.
- Saldi TK, Ash PE, Wilson G, Gonzales P, Garrido-Lecca A, Roberts CM, Dostal V, Gendron TF, Stein LD, Blumenthal T, et al. 2014. TDP-1, the *Caenorhabditis elegans* ortholog of TDP-43, limits the accumulation of double-stranded RNA. *EMBO J* **33**: 2947–2966.
- Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. 2012. The birth of the Epitranscriptome: Deciphering the function of RNA modifications. *Genome Biol* **13**: 175.
- Samuel CE. 2011. Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology* **411**: 180–193.
- Savva YA, Jepson JEC, Chang Y-J, Whitaker R, Jones BC, St Laurent G, Tackett MR, Kapranov P, Jiang N, Du G, et al. 2013. RNA editing regulates transposon-mediated heterochromatic gene silencing. *Nat Commun* **4**: 2745.
- Schlee M. 2013. Master sensors of pathogenic RNA—RIG-I like receptors. *Immunobiol* **218**: 1322–1335.

- Schmedt C, Green SR, Manche L, Taylor DR, Ma Y, Mathews MB. 1995. Functional characterization of the RNA-binding domain and motif of the double-stranded RNA-dependent protein kinase DAI (PKR). *J Mol Biol* **249**: 29–44.
- Schönborn J, Oberstaß J, Breyel E, Tittgen J, Schumacher J, Lukacs N. 1991. Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. *Nucleic Acids Res* **19**: 2993–3000.
- Sijen T, Plasterk RHA. 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**: 310–314.
- Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: Key considerations in genomic analyses. *Nat Rev Genet* **15**: 121–132.
- Sinha NK, Trettin KD, Aruscavage PJ, Bass BL. 2015. *Drosophila* Dicer-2 cleavage is mediated by helicase- and dsRNA termini-dependent states that are modulated by Loquacious-PD. *Mol Cell* **58**: 406–417.
- 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, Köhler M, Sprengel R, Seeburg PH. 1991. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**: 11–19.
- St Laurent G, Tackett MR, Nechkin S, Shtokalo D, Antonets D, Savva YA, Maloney R, Kapranov P, Lawrence CE, Reenan RA. 2013. Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in *Drosophila*. *Nat Struct Mol Biol* **20**: 1333–1339.
- Steff R, oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB, 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.
- Suzuki T, Ueda H, Okada S, Sakurai M. 2015. Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. *Nat Protoc* **10**: 715–732.
- Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, 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 PC, Diegelman-Parente A, Mathews MB. 2004. The double-stranded-RNA-binding motif: Interference and much more. *Nat Rev Mol Cell Biol* **5**: 1013–1023.
- van Gurp TP, McIntyre LM, Verhoeven KJF. 2013. Consistent errors in first strand cDNA due to random hexamer mispriming. *PLoS ONE* **8**: e85583.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
- Wagner RW, Nishikura K. 1988. Cell cycle expression of RNA duplex unwindase activity in mammalian cells. *Mol Cell Biol* **8**: 770–777.
- Wagner RW, Smith JE, Cooperman BS, 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* **86**: 2647–2651.
- Walkley CR, Li JB. 2017. Rewriting the transcriptome: Adenosine-to-inosine RNA editing by ADARs. *Genome Biol* **18**: 205.
- Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57–63.
- Wang IX, So E, Devlin JL, Zhao Y, Wu M, Cheung VG. 2013. ADAR regulates RNA editing, transcript stability, and gene expression. *Cell Rep* **5**: 849–860.
- Wang Z, Lian J, Li Q, Zhang P, Zhou Y, Zhan X, Zhang G. 2016. RES-Scanner: A software package for genome-wide identification of RNA-editing sites. *Gigascience* **5**: 37.
- 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.
- Wheeler EC, Washburn MC, Major F, Rusch DB, Hundley HA. 2015. Noncoding regions of *C. elegans* mRNA undergo selective adenosine to inosine deamination and contain a small number of editing sites per transcript. *RNA Biol* **12**: 162–174.
- 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.
- Wolfe D, Dudek S, Ritchie MD, Pendergrass SA. 2013. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min* **6**: 18.
- Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**: 873–881.
- 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.
- Xiang J-F, Yang Q, Liu C-X, Wu M, Chen L-L, Yang L. 2018. N⁶-Methyladenosines modulate A-to-I RNA editing. *Mol Cell* **69**: 126–135.e6.
- Yasukawa K, Konishi A, Inouye K. 2010. Effects of organic solvents on the reverse transcription reaction catalyzed by reverse transcriptases from avian myeloblastosis virus and Moloney murine leukemia virus. *Biosci Biotechnol Biochem* **74**: 1925–1930.
- Yoo J-S, Kato H, Fujita T. 2014. Sensing viral invasion by RIG-I like receptors. *Curr Opin Microbiol* **20**: 131–138.
- Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. 2014. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**: 419.
- Zhao H-Q, Zhang P, Gao H, He X, Dou Y, Huang AY, Liu X-M, Ye AY, Dong M-Q, Wei L. 2015. Profiling the RNA editomes of wild-type *C. elegans* and ADAR mutants. *Genome Res* **25**: 66–75.
- Zhang Q, Xiao X. 2015. Genome sequence-independent identification of RNA editing sites. *Nat Meth* **12**: 347–350.
- Zhu S, Xiang J-F, Chen T, Chen L-L, Yang L. 2013. Prediction of constitutive A-to-I editing sites from human transcriptomes in the absence of genomic sequences. *BMC Genomics* **14**: 206.

Cold Spring Harbor Perspectives in Biology

Mapping the dsRNA World

Daniel P. Reich and Brenda L. Bass

Cold Spring Harb Perspect Biol 2019; doi: 10.1101/cshperspect.a035352

Subject Collection [RNA Worlds](#)

Alternate RNA Structures

Marie Teng-Pei Wu and Victoria D'Souza

Approaches for Understanding the Mechanisms of Long Noncoding RNA Regulation of Gene Expression

Patrick McDonel and Mitchell Guttman

Principles and Practices of Hybridization Capture Experiments to Study Long Noncoding RNAs That Act on Chromatin

Matthew D. Simon and Martin Machyna

Linking RNA Sequence, Structure, and Function on Massively Parallel High-Throughput Sequencers

Sarah K. Denny and William J. Greenleaf

Extensions, Extra Factors, and Extreme Complexity: Ribosomal Structures Provide Insights into Eukaryotic Translation

Melanie Weisser and Nenad Ban

Nascent RNA and the Coordination of Splicing with Transcription

Karla M. Neugebauer

Combining Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) Spectroscopy for Integrative Structural Biology of Protein–RNA Complexes

Alexander Leitner, Georg Dorn and Frédéric H.-T. Allain

Structural Biology of Telomerase

Yaqiang Wang, Lukas Susac and Juli Feigon

Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes

Berthold Kastner, Cindy L. Will, Holger Stark, et al.

What Are 3' UTRs Doing?

Christine Mayr

Single-Molecule Analysis of Reverse Transcriptase Enzymes

Linnea I. Jansson and Michael D. Stone

CRISPR Tools for Systematic Studies of RNA Regulation

Jesse Engreitz, Omar Abudayyeh, Jonathan Gootenberg, et al.

Relating Structure and Dynamics in RNA Biology

Kevin P. Larsen, Junhong Choi, Arjun Prabhakar, et al.

Beyond DNA and RNA: The Expanding Toolbox of Synthetic Genetics

Alexander I. Taylor, Gillian Houlihan and Philipp Holliger

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>

**All Modifications and
Oligo Types Synthesized**

Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

Oligo Modifications?

Your wish is our command.

Discovering and Mapping the Modified Nucleotides That Comprise the Epitranscriptome of mRNA

Bastian Linder and Samie R. Jaffrey

Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast

Clemens Plaschka, Andrew J. Newman and Kiyoshi Nagai

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>

All Modifications and Oligo Types Synthesized

Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

Oligo Modifications?
Your wish is our command.