Complex Tertiary Structure Governs RNA Editing of the *Drosophila Paralytic* Transcript

by

Leila E. Rieder

B.A., Reed College, Portland, OR, 2006

M.A., Brown University, Providence, RI, 2011

A dissertation submitted in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in The Division of Biology and Medicine, The Department of Molecular Biology,

Biochemistry, and Cellular Biology at Brown University

PROVIDENCE, RHODE ISLAND

May 2013

This dissertation by Leila E. Rieder is accepted in its present form by The Division of Biology and Medicine, The Department of Molecular Biology, Biochemistry, and Cellular Biology as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date_____ _____

Dr. Robert Reenan, Advisor

Recommended to the Graduate Council

Date_____ _____

Dr. Michael McKeown, Reader (Chairman)

Date_____ _____

Dr. Erica Larschan, Reader

Date_____ _____

Dr. Eric Morrow, Reader

Date_____ _____

Dr. Ronald Emeson, Outside Reader

Approved by the Graduate Council

Date_____ _____

Dr. Peter Weber, Dean of the Graduate School

Vitae

Leila Elizabeth Rieder

Brown University
Providence, RI 02912
518-428-2877 or leila.rieder@gmail.com

Education

Expected	**Doctor of Philosophy**, Dr. Robert Reenan, PI
May, 2013	Cellular Biology, Molecular Biology, and Biochemistry
	Brown University, Providence, Rhode Island 02912.
May, 2011	**Master of Arts**, Dr. Robert Reenan, PI
	Cellular Biology, Molecular Biology, and Biochemistry
	Brown University, Providence, Rhode Island 02912.
2007-2008	**Post-graduate classes**
	Harvard University Extension, Cambridge, Massachusetts 02138.
May, 2006	**Bachelor of Arts**, Biology
	Reed College, Portland, Oregon 97202.

Additional Research Experience

2008-2009	**PhD research**, Dr. Michael Sussman, PI,
	University of Wisconsin-Madison, 53706.
2006-2008	**Associate Scientist I**, Molecular Profiling, Biogen Idec, Inc.,
	Cambridge, Massachusetts 02142.
summers	**Research Assistant**, Marine Biological Laboratory,
2000-03, '05	Woods Hole, Massachusetts 02543.

Publications

LE Rieder, CJ Staber, B Hoopengardener, and RA Reenan. 2013. Tertiary structural elements determine extent and specificity of messenger RNA editing. In press, Nature Communications.

Savva, YA*, **LE Rieder***, and RA Reenan. 2012. The ADAR Protein Family. Genome Biology 13(12):252-32. (*Equal contributors)

Rieder, LE and RA Reenan. 2012. The Intricate Relationship between RNA Structure, Editing, and Splicing. Seminars in Cell and Developmental Biology 23(3): 281-8.

Allaire NE, **Rieder LE**, Bienkowska J, Carulli JP. 2008. Experimental Comparison and Cross-Validation of Affymetrix HT Plate and Cartridge Array Gene Expression Platforms. Genomics 92(5): 359-65.

Rieder, LE and AF Mensinger. 2001. Strategies for Increasing Growth of Juvenile Toadfish. Biological Bulletin 201(2): 283-5.

Grants, Awards, and Honors

Sept., 2011	**Poster prize**, Brown University Cellular Biology, Molecular Biology and Biochemistry Department annual retreat.
April, 2011	**Poster prize**, EMBO Workshop on chromatin structure, organization, and dynamics, Prague, CZ.
2009-2012	**National Science Foundation Graduate Research Fellowship Award**.
2008-2009	**Molecular Biosciences Training Grant**, University of Wisconsin-Madison Cellular Biology, Molecular Biology, and Biochemistry Department.
2005-2006	**Howard Hughes Undergraduate Research Grants**, Reed College.
2003-2006	**Commendation** for excellence in scholarship, Reed College.

Teaching Experience

2009-2013	**Mentor** to various Brown undergraduate students.
summer, 2012	**Teaching Assistant**, *Genome Explorers*, HHMI undergraduate summer course, Brown University.
2011	**Teaching certificate recipient**, Brown University Sheridan Center for Teaching and Learning.
2010	**Teaching Assistant**, *Plant Organism*, Brown University.
2006-2008	**Teaching Assistant**, Community Lab, Biogen Idec Inc., (middle and high school students).
2004-2006	**Science Center Tutor**, *Organic* and *Introductory Chemistry*, Reed College.

2004-2006	**Personal Tutor**, *Organic* and *Introductory Chemistry*, Reed College.
2003-2006	**Teaching Assistant**, *Introductory Chemistry* Lab, Reed College.

Professional Presentations

Jan., 2013	**Poster presenter**, Gordon Conference on RNA Editing, Galveston TX.
Jan., 2013	**Poster presenter**, Gordon-Kenan Research Seminar on RNA Editing, Galveston TX.
Sept., 2012	**Poster presenter**, EMBO conference, Nice France.
May, 2012	**Selected speaker**, RNA Society Annual Meeting, Ann Arbor IL.
Sept., 2012	**Selected speaker**, Brown University Molecular Biology, Cellular Biology and Biochemistry Department annual retreat, Providence RI.
April, 2011	**Poster presenter**, EMBO Workshop on chromatin structure, organization, and dynamics, Prague CZ. *Poster prize recipient.
April, 2011	**Co-chair**, RNA Processing session, EMBO Workshop on chromatin structure, organization, and dynamics, Prague CZ.
Feb., 2011	**Speaker**, Brown University Molecular Biology, Cellular Biology, and Biochemistry Department recruitment weekend, Providence RI.
Jan., 2011	**Poster presenter**, Gordon Conference on RNA Editing, Galveston TX.
Jan., 2011	**Poster presenter**, Gordon-Kenan Research Seminar on RNA Editing, Galveston TX.

Dedication

Both by Mark Nowogrodzki

RNA editing

In order to keep its cells in line,
Healthy, without aberrations,
A body constantly keeps a fine
Check on its genes' equation.

It monitors its DNA
And finds where corrections are due,
And edits its genes in its own way
Their balance to pursue.

That's what my granddaughter's quests embrace:
What are the body's controls?
Where and how does the edit take place,
And what are nature's goals?

She found substitutions in the RNA chain
Take place at particular spots,
In which the pattern of the genes' domain
Is changed into loops and knots.

And, though it may earn her a degree,
And maybe fame and glory,
The puzzle still appears to be
The same eternal story:

Are these peculiarities, strange and stark,
Genetic imperfections,
Or just a signal, to embark
On editing corrections?

This is the problem that ever imbued
The human intellect,
And that through history was pursued:
What's cause and what's effect?

The Fruit Fly

Sing praise for the humble fruit fly:
though with a tiny brain,
it teaches us the "what" and "why"
of nature's vast domain.

Whether Darwin or creation,
whether great design or chance,
it's many generations
our knowledge help advance.

Oh, what a humbling feeling
that reason does defy,
that nature's highest being
learns wisdom from a fly...

Acknowledgements

I owe so many thanks to the people who have helped me become a proud, confident scientist, and who have supported me through the dark times as well as the bright.

Thanks to:

... Dr. Allen Mensinger at the MBL and Dr. Keith Karoly at Reed College. The former shaped me into a young scientist and the latter turned me into a molecular biologist and geneticist.

... my unwavering support system in Madison, Wisconsin: Kelly April, Patricia Lindquist, Jessica Ciomperlik, and Rachel Nelson-Rodrigues. You kept me safe through the darkest time of my life, so that I could arrive whole and new on the other side. To Dr. Michael Sussman, a selfless PI who supported my transition to Brown without question.

... the University of Wisconsin-Madison CMB department, who wished me good luck without judgment, and the Brown University MCB department who welcomed me home.

... the friends I found at Brown when I arrived here lost and unsure. Together we suffered the ups and downs of graduate school. At each new challenge I drew strength from Rebecca Helm, Asli Sahin, Marcela Soruco, Selena Gell, and so many others.

... my lab mates throughout the years, without whom I would be lost. Dr. James Jepson, Dr. Yiannis Savva, Dr. Selena Gell, soon-to-be-Drs. Asli Sahin and Yao-Jen Chang. To the Brown undergraduate students, with whom I converse as colleagues and of whom I could not be more proud, Abby Kearson and Kasia Sierzputowska. To the Larschan lab members who provided my home away from home.

... Dr. Rob Reenan, who took a chance on me. I have never met a scientist more excited about research. To my academic committee and the professors, students, and staff in the Brown MCB department who collectively nurtured my professional scientific confidence.

... my family. My parents, Conly and Susan, who saw me through it all with just the right mix of caution, concern, and support. To my sister Ray, my personal inspiration and the bigger half of my heart. To Matt, whose strength of character never ceases to amaze me, and whose mere presence in my life is enough to shape me into a better person. To my grandpa, Mark Nowogrodzki, who catches all my typos and who effortlessly imbues meaning into poems that remind me to remember the beauty of science.

Contents

List of Tables

List of Figures

Chapter One

Introduction to RNA Editing

Parts of this chapter have been published as:

Rieder, LE, and RA Reenan. (2011) The intricate relationship between RNA structure, editing, and splicing. *Seminars in Cell and Developmental Biology.* May; 23(3):28-8.

I wrote the manuscript and crafted the figures. Dr. Robert Reenan provided editorial feedback.

and

Savva, YA*, **Rieder, LE***, and RA Reenan. (2012) The ADAR protein family. *Genome Biology.* Dec; 13(12):252-8. (*Denotes equal contributors)

Dr. Yiannis Savva and I share first authorship and jointly wrote the manuscript. I crafted the figures with help from Dr. Robert Reenan and Dr. James Jepson. Dr. Robert Reenan also provided editorial feedback.

1.1 ADARs and A-to-I RNA editing

The human genome is predicted to encode well over a million distinct, functional proteins [Nielsen et al., 2006], despite containing a mere 20,000-25,000 protein-coding genes [Nekrutenko, 2004]. Highly sophisticated post-transcriptional modifications of messenger RNAs lead to increased proteome diversity while limiting a genome's size and buffering it from mutation [Jin et al., 2007, Tian et al., 2008, Yang et al., 2008]. These enhancing processes include alternative splicing [Schmucker et al., 2000], alternative polyadenylation [Xing and Li, 2009], and RNA editing [Reenan et al., 2000, Pullirsch and Jantsch, 2010].

Unlike splicing, which can facilitate inclusion or exclusion of large segments of coding sequence, adenosine-to-inosine (A-to-I) RNA editing generally chemically alters only a few specific nucleotides within the coding regions of a transcript, providing a mechanism to subtly reshape the properties of resulting proteins. While editing causes the modification of transcripts at the level of single nucleotides, collectively these events are essential for viability in some organisms and correct neurological function in others [Higuchi et al., 2000, Nishikura et al., 2000, Tonkin and Bass, 2003, Palladino et al., 2000]. Although the global effect of loss of editing has been demonstrated in several model organisms, we have yet to fully uncover the functional importance of such seemingly small individual recoding events.

RNA editing was first discovered 25 years ago in *Xenopus* oocytes: injected antisense RNAs failed to form stable duplexes *in vivo*, suggesting an as-yet then uncharacterized unwinding activity [Rebagliati and Melton, 1987, Bass and Weintraub, 1987]. This activity was later attributed to the enzyme eventually named ADAR (Adenosine Deaminase acting on RNA) [Bass, 1997], and the unwinding was due to the conversion of select adenosines (A) into inosines (I) via hydrolytic deamination (Figure 1.1), destabilizing the duplexes as A-U Watson-Crick base pairs convert to I-U wobble pairs [Polson et al., 1991]. Because the properties of inosine mimic those of

guanosine (inosine will form two hydrogen bonds with cytosine, Figure 1.1), inosine is recognized as guanosine by the translational cellular machinery [Basillo et al., 1962]. Adenosine-to-inosine (A-to-I) RNA editing, therefore, effectively changes the primary sequence of RNA targets.

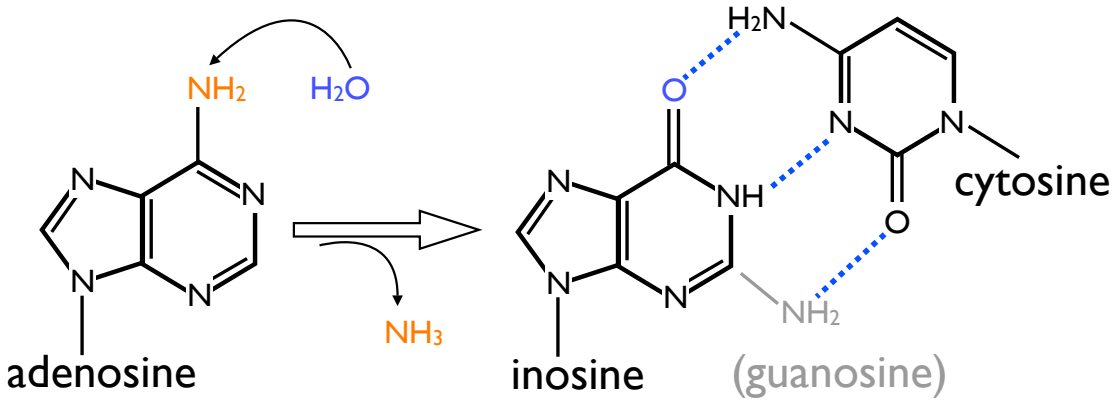


Figure 1.1: Hydrolytic deamination. A-to-I RNA editing involves the hydrolytic deamination of an adenosine into an inosine molecule. Inosine forms two hydrogen bonds with cytosine, resembling guanosine to the cellular machinery.

ADAR enzymes are highly conserved in metazoa [Bass, 2002], although the number of genes and isoforms varies between species. Mammalian genomes encode three ADARs: ADAR1, and ADAR2, which are catalytically active [Nishikura, 2010], and ADAR3, which is thought to be catalytically inactive [Chen et al., 2000]. The *C. elegans* genome encodes two genes, CeADR1 and CeADR2 [Tonkin et al., 2002], while only a single *adar* locus is present in the *Drosophila* genome [Palladino et al., 2000] (Figure 1.2A). The squid [Palavicini et al., 2009] and hydra (personal observation, R. A. Reenan) genomes each encode a single *adar* locus. Furthermore, ADAR genes are also present in the genomes of both sea urchin and sea anemone, suggesting an early origin of RNA editing enzymes in metazoan evolution [Jin et al., 2009]. In contrast, ADAR genes do not appear to be present in fungal, plant and yeast genomes [Jin et al., 2009].

Interestingly, although prokaryotic genomes do not contain ADAR genes,

they do encode a transfer RNA (tRNA) adenosine deaminase (TadA), which modifies specific tRNAs [Wolf et al., 2002]. Eukaryotic orthologs of this RNA editing enzyme, adenosine deaminases acting on tRNAs (ADATs), are also conserved in metazoa and catalyze the deamination of specific adenosines to inosines at or adjacent to the tRNA anticodon [Gerber and Keller, 2001]. Sequence homology between the catalytic domains of ADARs and ADATs suggests a model in which tRNA modifying enzymes are ancestral to ADARs (Figure 1.2B). In this model, a duplicate ADAT gene acquired one or more dsRNA binding domains that allowed the protein to recognize and bind dsRNAs. This novel gene is thought to have conferred selective advantage due to repair of detrimental genomic mutations by A-to-I modifications at the mRNA level [Bass, 2002].

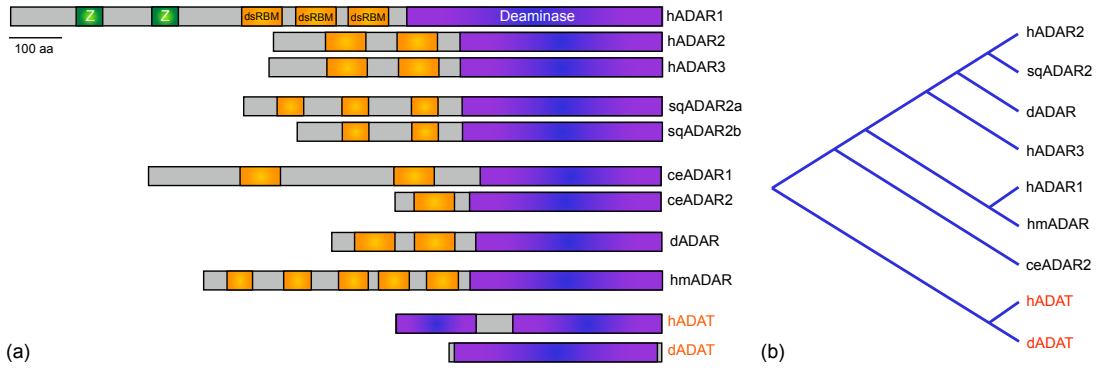


Figure 1.2: (a) Domain architecture of metazoan ADARs. The deaminase domain is depicted in blue, while the dsRBMs are shown in orange and Z-DNA binding domains, unique to human ADAR1, are presented in green. The human genome contains three ADAR genes (hADAR1-3). That of the squid, *Loligo pealeii*, contains an ADAR2-like gene (sqADAR2) that produces variants (a and b) through alternative splicing. *C. elegans* contains two genes (ceADAR1-2), while the genome of *D. melanogaster* encodes for only one (dADAR), an enzyme homologous to hADAR2. Although the dsRBM found in the *Hydra magnapapillata* genome are highly divergent, five such motifs are recognizable in hmADAR, the only identified gene in this family. Human and *Drosophila* ADAT architectures are included (red), as these enzymes are believe to be ancestral to present-day ADARs. (b) Cladogram based on ADAR catalytic domain sequences. MacVector was used to generate a relatedness tree based on the protein sequences of ADAR catalytic domains from different species. *C. elegans* ADAR2 is absent due to difficulty aligning the catalytic domain. Note that human and *Drosophila* ADATs (red) cluster as the outgroup. Reproduced from Savva et al. [2012].

In the decades of research following the initial discovery, RNA editing has

been proven to exert significant impact on proteomes and organisms. Mice deficient for ADAR1 have a heterozygous embryonic lethal phenotype [Nishikura et al., 2000]; those with complete lack of ADAR2 are seizure-prone and die soon after birth [Higuchi et al., 2000]. *Drosophila* null for dADAR, exhibit signs of severe neurological impairment, including locomotor uncoordination, temperature-sensitive paralysis and seizures [Palladino et al., 2000], while in *C. elegans*, both functional ADAR genes are required for normal chemotaxis [Tonkin et al., 2002]. Clearly the conserved process of A-to-I RNA editing is crucial for proper neurological function. Interestingly, genetically engineered *Drosophila* dADAR hypomorphs, which produce only 20% of dADAR protein compared to wild type animals, display alterations in complex behavior such as courtship, but not in general locomotor coordination [Jepson et al., 2011]. While RNA editing may serve different purposes in various organisms, ADARs are conserved among metazoa and share certain characteristics: between one and three tandem double-stranded RNA binding motifs (dsRBMs) and a C-terminal catalytic deamination domain.

1.1.1 ADARs: characteristic structural features

ADAR enzymes share a common domain architecture consisting of a variable number of N-terminal dsRNA binding domains (dsRBD) and a C-terminal catalytic deaminase domain [Bass, 2002] (Figure 1.2 A). Human ADARs (hADARs) possess two or three dsRBDs, while the *C. elegans* enzymes (ceADARs) contain one or two. The single *Drosophila* ADAR (dADAR) contains two dsRBDs, similar to the mammalian ADAR2. In squid two ADAR enzymes (sqADARs) are generated via splicing from a single *adar* locus: while isoform 2b contains two dsRBDs, the inclusion of an alternative exon leads to the generation of an ADAR enzyme (sqADAR2a) containing an additional dsRBD, which confers higher enzymatic activity *in vitro* [Palavicini et al., 2009]. Interestingly, recent evidence suggests that this extra dsRBD of sqADAR2a is required

for appropriate RNA editing in the high salt conditions of a marine environment [Palavicini et al., 2012], suggesting a link between the regulation of the RNA editing process and changes in the physical environment. In agreement with this, additional evidence suggests that RNA editing may be regulated by temperature [Garrett and Rosenthal, 2012, Savva et al., 2012] (please see **Chapter 3**). Finally, the *Hydra magnapapillata* ADAR enzyme (hmADAR) contains five identifiable dsRBDs (Figure 1.2 A) (personal observation, R. A. Reenan).

The crystal structure of the human ADAR2 deaminase domain (Figure 1.3) suggests that a catalytic center forms in the enzyme active site: a glutamic acid residue (E396) forms hydrogen bonds with a water molecule, while a histidine (H394) and two cysteine residues (C451 and C516) coordinate a zinc ion that activates the water molecule for nucleophilic attack (1.1) [Macbeth et al., 2005]. Buried in the catalytic core, an inositol hexakisphosphate (IP_6) molecule stabilizes multiple arginine and lysine residues and is also required for catalytic activity [Macbeth et al., 2005].

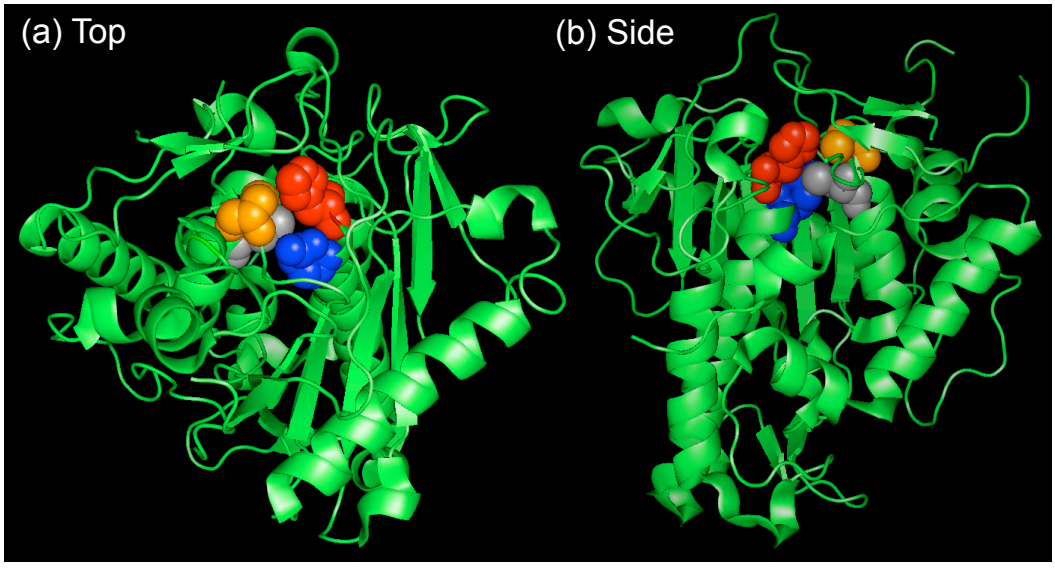


Figure 1.3: Crystal structure of the human ADAR2 deaminase domain shown fro the top (a) and side (b). The catalytic core is formed between H394 (red), E396 (blue), C451 (orange), and C516 (gray). The core also includes a water molecule, zinc ion, and IP_6 molecule (not shown). Reproduced from Savva et al. [2012].

1.1.2 ADARs: localization and function

The three mammalian *adar* genes give rise to four known isoforms: ADAR1p150, ADAR1p110, ADAR2, and ADAR3. ADAR1 variants and ADAR2 are expressed in many tissues, whereas the ADAR3 protein is only present in the brain [Chen et al., 2000, Melcher et al., 1996]. ADAR3 is thought to be catalytically inactive [Chen et al., 2000], but it is possible that it competes with ADARs 1 and 2 for RNA binding substrates, altering the editing profile through that mechanism. Alternative promoter usage within the ADAR1 transcript generates a full length (ADAR1p150) and a N-terminally truncated (ADAR1p110) isoform [Patterson and Samuel, 1995], both of which contain three dsRNA-binding domains and the deaminase domain.

The ADAR1 isoforms differ in their cellular distributions: ADAR1p110 exclusively localizes to the nucleus [Desterro et al., 2003, Sansam et al., 2003], while ADAR1p150 shuttles in and out of the nucleus, although it accumulates in the cytoplasm [Patterson and Samuel, 1995, Poulsen et al., 2001]. The cytoplasmic localization of ADAR1p150 suggests that a select class of dsRNAs could be targeted outside the nucleus as mature mRNAs. However, cytoplasmic localization of other ADARs is unusual.

Mammalian ADAR1p110 and ADAR2 are also present in the nucleolus and this localization is dependent on functional dsRBDs [Desterro et al., 2003, Sansam et al., 2003]. ADAR nucleolar localization may be mediated by dsRNAs generated from ribosomal RNAs (rRNAs) and small nucleolar RNAs (snoRNAs), both of which are concentrated within the nucleolus. However, to date there is no evidence that ADAR modification of rRNAs or snoRNAs occurs, thus, the significance of ADAR nucleolar localization remains enigmatic. Yet translocation of ADAR2 from the nucleolus to the nucleoplasm upon expression induction of a known editing target resulted broadly in higher editing levels in pre-mRNA substrates [Sansam et al., 2003]. Therefore, the nucleolar localization of ADARs may represent a mechanism

for regulating editing of pre-mRNAs through enzyme sequestration.

In *Drosophila*, Jepson et al. [2011] examined the complete expression pattern and cellular distribution of dADAR using a genetically engineered allele containing an HA epitope tag inserted into the endogenous locus. Visualization of dADAR-HA revealed that *Drosophila* ADAR is predominately expressed in the nervous system [Jepson et al., 2011]. Furthermore, both transgenic dADAR specifically expressed in the salivary glands of 3^{rd} instar larvae, as well as endogenous tagged dADAR, localize within the nucleus and accumulate within the nucleolus [Savva et al., 2012] (Figure 1.4), consistent with observations regarding mammalian ADARs.

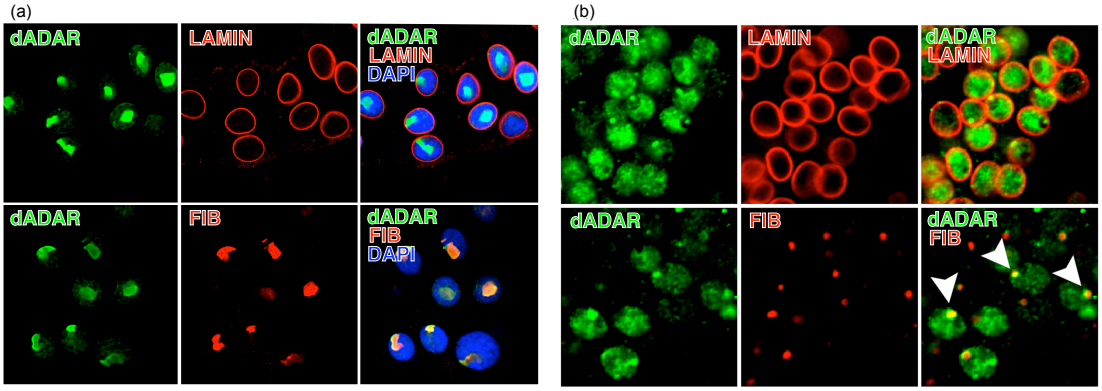


Figure 1.4: ADAR localization. (a) Transgenic HA-tagged ADAR (green) localizes within the nuclear envelope (lamin, red) and more specifically to the nucleolus (fibrillarin, red) in *Drosophila* salivary gland cells. (b) Endogenous HA-tagged ADAR (green) localizes to the *Drosophila* neuronal nucleus and colocalizes with the nucleolus, distinguished by the red fibrillarin signal (arrows). Reproduced from Savva et al. [2012].

1.1.3 Advantages of RNA editing

Adenosine-to-inosine RNA editing results in alteration of base pairing properties; inosine forms two hydrogen bonds with cytosine (Figure 1.1) and is recognized as guanosine by cellular machineries tasked with interpreting nucleic acid information, such as the translational system. Editing, therefore, effectively recodes RNA through A-to-G substitutions. These changes may have obvious impact in that editing has the

potential to recode stop codons (nonsense) into that of tryptophan (sense), resulting in read-through. Alternatively, editing can change key amino acid side chains (Figure 1.5) in conserved and functionally critical portions of the protein products.

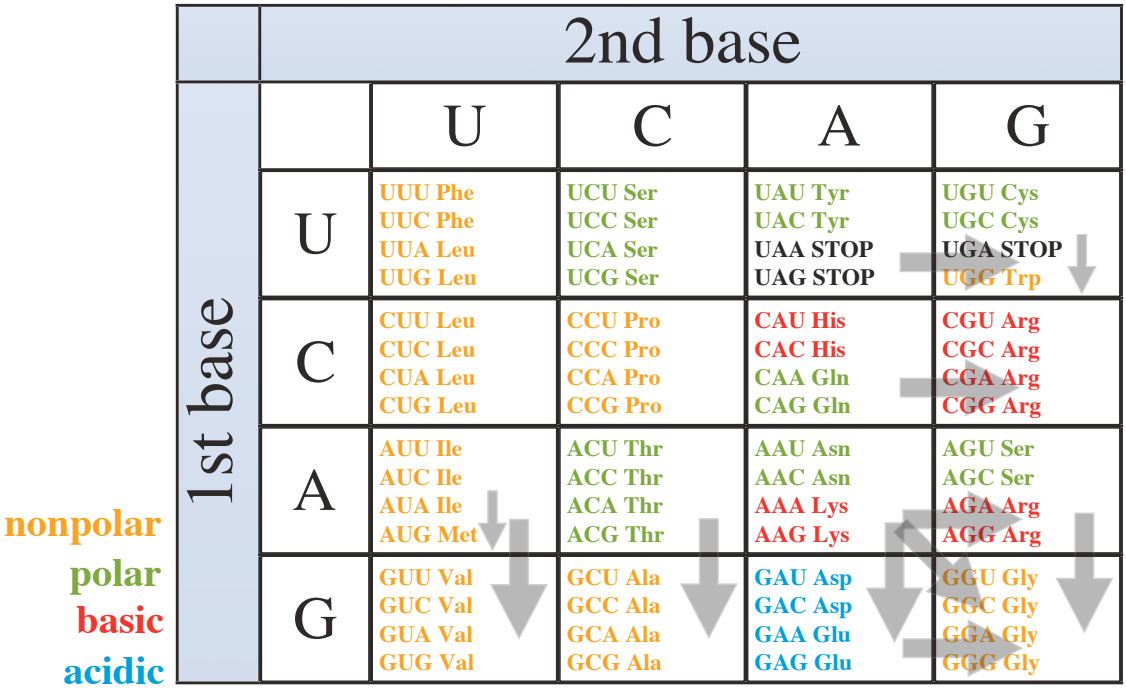


Figure 1.5: Adenosine-to-inosine RNA editing may recode transcripts. Recoding events may be silent (synonymous) or change amino acid residues (arrows; non-synonymous). Stop codons (black) may be edited to the tryptophan codon, resulting in read-through. Reproduced from Rieder and Reenan [2011].

What is the purpose of a system such as RNA editing versus simply fixing mutations at the DNA level? In fact, the effects of editing are more akin to genetic mutation than to alternative splicing, since the nature of editing alters the genetic code at the level of a single nucleic acid. Editing sites are sometimes hardwired as guanosine in closely related species, a phenomenon shown in both vertebrates [Li et al., 2009] and invertebrates [Yang et al., 2008, Tian et al., 2008], suggesting mutation at the DNA level is a viable alternative in select cases. Similar to alternative splicing, editing enables an organism (or even individual cells) to achieve genetic variation while safeguarding the genome from deleterious mutations [Gommans et al., 2009].

Additionally, post-transcriptional events such as splicing and editing diversify the protein repertoire [Hanrahan et al., 2000] allowing for flexibility in protein properties and possible adaptation of novel protein functions.

It is possible that cells such as neurons are particularly individualized through transcriptome diversification, at the levels of both transcription and transcript modification. ADARs are most highly expressed in the nervous system [Jepson et al., 2011] throughout development [Jacobs et al., 2009], although the presence of the protein does not necessarily cause high levels of editing. Rather, editing efficiency appears low in the embryonic mouse brain and gradually increases during development, implying that a regulatory network carefully controls editing during development despite ADAR concentration [Wahlstedt et al., 2009]. The observation that ADARs modify transcripts implicated in fast chemical and electrical neurotransmission [Hoopengardner et al., 2003] has led to suggestions that editing contributed to the evolution of complex metazoa [Jin et al., 2009] and to higher cognitive function in primates [Paz-Yaacov et al., 2010, Eisenberg et al., 2005]. Transcripts for proteins involved in synaptic signaling, including those of ion channels and synaptic release proteins, are heavily edited [Hoopengardner et al., 2003, Hanrahan et al., 2000], and even singular editing events can substantially impact channel properties [Bhalla et al., 2004].

For example, the human potassium channel $K_v1.1$ transcript is edited at a single adenosine. This one editing event causes a valine substitution for an isoleucine in the highly conserved ion-conducting pore, resulting in the effective removal of a single methyl group from the edited residue. Despite this subtle structural alteration, recoding of the transcript by RNA editing profoundly affects channel inactivation [Bhalla et al., 2004, Decher et al., 2010]. Similarly, the *Drosophila* Shaker potassium channel is edited at four locations, resulting in 16 possible channel isoforms. Each variant possesses subtly different biophysical properties. Furthermore, examination of multiply edited combinations revealed an intriguing degree of functional epistasis

between even distant editing sites [Ingleby et al., 2009].

It is evident that the complex interaction of multiple editing sites within a transcript confers the potential for functional fine-tuning. Editing can serve to increase proteomic diversity and produce variation within a cell population. While the significance of this process at the molecular level may appear subtle, the effect of single events at the organismal level can be quite striking.

1.1.4 Organismal impacts of individual RNA editing events

Recoding through RNA editing can produce variation in a protein population if some transcripts are edited while others reflect the literal genome. Yet certain transcripts are edited at a very high level, similarly impacting every resulting protein product. For example, the mouse GluR-B transcript, which encodes an AMPA receptor subunit, is edited to nearly 100% at one adenosine, replacing the genomically encoded glutamine with an arginine (Q/R site). Mice deficient for ADAR2 exhibit post-natal mortality, usually due to fatal seizure events, but substitution of the arginine form of GluR-B for the unedited form completely rescues this phenotype [Higuchi et al., 2000]. Editing of this transcript, therefore, is absolutely necessary for the organism's viability.

In addition to changing key residues (Figure 1.5) in essential proteins, RNA editing is known to contribute to cellular immunity. The presence of dsRNA in the cytoplasm, as with viral infection, triggers host defense responses including editing by mammalian cytoplasmic ADAR1. In many cases editing of a viral RNA genome results in loss or retardation of pathogenicity [Samuel, 2011]. Yet this process has also been co-opted by viruses as a method of regulating complex replication cycles. The hepatitis delta virus possesses a single-stranded RNA genome encoding a lone protein, HDAg. However the viral anti-genome folds into a duplex structure and directs the cellular RNA editing machinery to edit an amber stop codon in the HDAg transcript, resulting in read-through and production of a longer isoform of the viral

protein. The presence of HDAg-Long is necessary to switch the viral replication cycle from genome replication to packaging [Casey, 2011]. The replication of other viruses including HIV-1 [Doria et al., 2009] and the measles virus [Ward et al., 2011] is also enhanced by ADAR1 editing [Gelinas et al., 2011].

While the impact of editing events involved in temporal control of viral replication or recoding of nearly all GluR-B transcripts is apparent, other editing events may have subtler effects, such as those localized to introns or to the third nucleotides of codons (Figure 1.5). Considerable evidence exists, however, that even seemingly insignificant editing events may have substantial impact through effect on RNA structural stability, splicing, and even global cellular responses such as RNA interference (RNAi). It is clear that editing is not a purely random force in the transcriptome. As with splicing, the RNA itself encodes the highly conserved informational signals that direct ADARs.

1.2 RNA editing substrates

Early observations suggested that ADARs recognize only RNA substrates possessing duplex character with no activity detected on single-stranded RNA molecules [Bass and Weintraub, 1987]. This observation provoked two related questions: how do RNA substrates direct ADAR activity and how does the enzyme recognize target substrates? *In vitro* data revealed that specific deamination of single adenosines usually occurs on short imperfect duplex RNA, while long, highly stable, perfectly complementary duplexes are non-specifically, or promiscuously, modified at up to 50% of adenosine residues [Polson et al., 1991, Nishikura et al., 1991] (Figure 1.6 A).

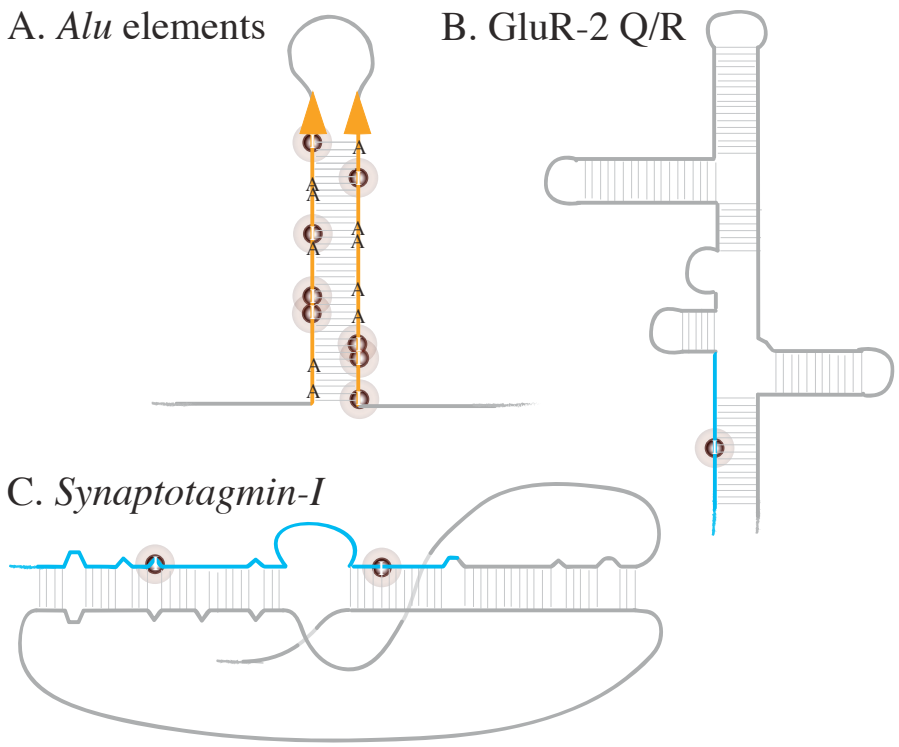


Figure 1.6: Promiscuous and specific RNA editing. Perfect duplex RNA structures are edited non-specifically (promiscuously), while short imperfect duplexes, with bulges, mismatches, and loops, are edited specifically. (A) Primate-specific *Alu* elements (orange) in reverse orientation may form long (approximately 300 basepairs) nearly perfect duplex editing substrates, which are edited by ADARs promiscuously. (B) The mammalian GluR-2 Q/R site is formed between an exon (blue) and downstream intron (gray). This imperfectly duplex substrate directs editing around a single nucleotide that results in a glutamine to arginine substitution in the AMPA receptor protein. (C) Part of the *Drosophila synaptotagmin-1* transcript forms a complex pseudoknot between an exon (blue) and following intron (gray) that directs precise deamination at two sites. Reproduced from Rieder and Reenan [2011].

1.2.1 Promiscuous editing of perfectly duplex RNA substrates

Numerous efforts have recently attempted to identify the complete inosinomes [Wulff et al., 2011] of various organisms through deep sequence identification of the edited transcript population [Rosenberg et al., 2011, Bahn et al., 2011] or via other novel techniques [Li et al., 2009, Xia et al., 2005, Eisenberg, 2011]. Such endeavors reveal that a significant number of editing events occur in non-coding regions of the transcriptome–a phenomenon that spans metazoa, including primates, *C. elegans* [Wu et al., 2011, Morse et al., 2002] and *Drosophila* [Graveley et al., 2011]. The most

well studied example of non-coding ADAR activity is the promiscuous editing of primate-specific Alu repetitive elements, short interspersed elements of approximately 300 nucleotides, which inhabit up to 75% of human genes and comprise more than 10% of the human genome [Kim et al., 2004].

Given the prolificacy of Alu elements and their recent expansion in the human genome lineage, new insertions have frequently occurred in regions already occupied by other local Alu elements. The insertion of one element into the genome in reverse orientation to another nearby Alu can result in the formation of a highly complementary, stable, duplex structure in the transcript that may serve as an ADAR substrate (Figure 1.6 A). These elements, often found in gene-rich, more highly transcribed regions, lead to nearly perfect double-stranded RNA substrates *in vivo* and are heavily edited by ADAR [Athanasiadis et al., 2004, Paz-Yaacov et al., 2010, Levanon et al., 2004] (Figure 1.6 2A). In fact, due to Alu substrates, RNA editing in humans is at least an order of magnitude higher than that in mouse, *Drosophila*, or *C. elegans* [Eisenberg et al., 2005].

The purpose of promiscuous editing is unknown, although edited Alu-containing transcripts can be retained in the nucleus [Zhang and Carmichael, 2001]. Also, since inverted Alu elements so readily form double-stranded structures, they may also serve as Dicer substrates to generate Alu-specific small interfering RNAs that guide a subsequent RNA interference response. An Alu duplex in one transcript could mediate degradation of other Alu-containing transcripts via the RNA-Induced Silencing Complex (RISC). For example, a pair of inverted Alus in the 3' UTR of a GFP reporter mRNA strongly repress GFP expression *in vivo*, though a single element has no effect on the fluorescence signal [Chen et al., 2008]. Promiscuous editing of the Alu duplex would result in unwinding of the structure and loss of Dicer recognition, followed by alleviation of silencing. In this manner RNA editing could impact RNAi-based silencing and transcriptional regulation of many Alu-containing

genes. Indeed, RNAi is inhibited *in vitro* when target transcripts are first deaminated by ADAR2 [Scadden and Smith, 2001]. It is also interesting to note that aberrant chemotaxis phenotypes caused by *C. elegans* ADAR mutations [Tonkin et al., 2002] are rescued by deficiencies in various components of the RNAi pathway [Tonkin and Bass, 2003], suggesting that the modification activity of ADARs may unwind *in vivo* perfect dsRNA substrates capable of initiating an RNAi response through Dicer activity [Wu et al., 2011].

Promiscuous editing may also play a role in viral defense via interfacing with the RNAi pathway. Viruses with dsRNA genomes, for example, provide a perfect duplex substrate for promiscuous editing. These edited dsRNA viral genomes may be targeted for destruction: dsRNA with multiple I-U mismatches is cleaved by Tudor Staphylococcal Nuclease [Scadden, 2005], a component of the RISC. This targeting function appears to be conserved: mammalian ADAR1, ADAR2, and *Drosophila* dADAR are all capable of generating the preferred cleavage site when acting on dsRNA *in vitro* [Scadden, 2005]. Genome cleavage has an direct impact on viral pathogenicity.

1.2.2 Imperfectly duplex editing substrates direct specific editing events

Perfectly duplex substrates are edited promiscuously, while imperfect structures containing bulges, mismatches, and loops, are edited much more selectively. Specific editing targets are often identified due to A/G disparities between the cDNA sequence and that of the reference genome. To date, all confirmed ADAR substrates possess secondary structure around the region of editing; that is, the early-identified requirement of double-strandedness for editing has thus far not been violated. The secondary and tertiary structures of RNA around the editing site(s) in an imperfect RNA duplex serve to specifically direct editing of select adenosines. Also, it is rare

for inosine to replace adenosine in 100% of transcripts at these sites. Rather, levels of editing appear to be precisely modulated, can be evolutionarily conserved, and are dictated by RNA structure [Bratt and Ohman, 2003] (Figure 1.6 B, C).

Imperfectly duplex double-stranded editing substrates are sometimes generated within exonic sequences when the nascent transcript folds, creating simple hairpins [Bhalla et al., 2004, Hanrahan et al., 2000, Keegan et al., 2005, Ohlson et al., 2007]. Yet, more often intronic sequences with extensive complementarity to upstream or downstream exons containing the targeted adenosine(s) can pair with the exonic sequence to form structurally diverse RNA duplexes. These intronic *cis* elements, also known as editing site complementary sequences (ECS), can form simple exon-intron hairpin structures [Burns et al., 1997, Hanrahan et al., 2000, Higuchi et al., 1993, Lomeli et al., 1994, Wang et al., 2000] or more complex RNA secondary structures such as pseudoknots [Reenan, 2005] (Figure 1.7).

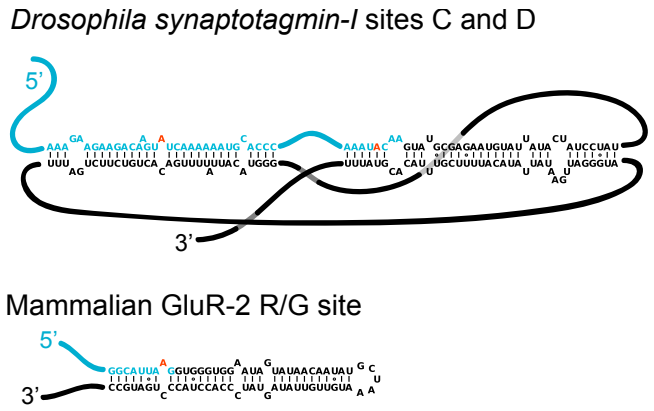


Figure 1.7: Diverse RNA structures that direct editing. The complex pseudoknot of *Drosophila synaptotagmin-1* is presented in contrast with the simple hairpin of mammalian GluR-2, both specific editing targets. Exons are represented in blue, introns in black. Adenosines targeted by ADAR are red. Reproduced from Savva et al. [2012].

The majority of ECS-mediated editing substrates usually contain mismatches, bulges, and loops. It is thought that the presence of such structural imperfections directs ADARs to specific locations on the RNA duplex without the requirement of primary sequence recognition. However, upon binding, ADARs do exhibit sequence

preferences for modifying select adenosines over others. *In vitro* studies using synthetic editing substrates revealed that ADAR enzymes preferentially target adenosines neighbored by a 5' uridine, while adenosines with a 5' guanosine neighbor are exceedingly rare [Lehmann and Bass, 2000, Polson and Bass, 1994]. In addition, adenosines found in mismatches with cytosines are edited more often when compared to other adenosines [Wong et al., 2001].

Specific RNA editing often leads to transcript recoding. Because inosine shares the base pairing properties of guanosine, the translational machinery interprets edited inosine as guanosine (Figure 1.1), altering the triplet codon, which can result in amino acid substitutions in protein products. Over half the triplet codons in the genetic code may be reassigned through RNA editing [Rieder and Reenan, 2011] (Figure 1.5). Due to the degeneracy of the genetic code, RNA editing can cause both silent and non-synonymous amino acid substitutions. However, statistically, RNA editing predominantly results in non-synonymous changes [Graveley et al., 2011] and thus favors the diversification of protein products. While the functional consequences of most specific recoding events are currently unknown, myriad studies in diverse model organisms indicate that specific RNA editing of certain mRNAs can result in profound changes in protein function [Bhalla et al., 2004, Dingledine et al., 1999, Ingleby et al., 2009, Rosenthal and Bezanilla, 2002, Seeburg et al., 2001].

As an example of RNA structure, the exonic GluR-2 Q/R site forms an imperfect RNA duplex by pairing with sequences in the downstream intron (Figure 1.6 B). Single nucleotide changes in the structure in either exonic or intronic sequence alter output editing levels of a mini-gene reporter at the Q/R site. This suggests that it is the precise RNA structure that finely tunes editing levels at this location [Higuchi et al., 1993]. As with the GluR-2 Q/R site, many target adenosines are encompassed in secondary structures composed of an exon and *cis*-acting elements present in a nearby intron. While most intronic sequence is under little selective pressure,

cis elements that participate in editing site duplex formation exhibit substantially higher sequence conservation compared to the surrounding intron. These observations allow putative identification of elements that participate in editing [Ingleby et al., 2009, Reenan et al., 2000].

Similarly, comparative genomics provides a useful tool to identify unknown editing sites based on conservation of nearby *cis* elements [Sixsmith and Reenan, 2007]. Prior to widespread use of deep sequencing technology comparative genomic approaches more than doubled the number of confirmed editing sites in *Drosophila* [Hoopengardner et al., 2003]. Although many *cis*-acting elements are localized in introns proximal to the edited exon, these sequences may also regulate editing from substantial distances.

Structures that direct editing may be relatively simple, such as an imperfectly base-paired hairpin [Stefl et al., 2010], but they may also be considerably more elaborate, as exemplified by the *Drosophila synaptotagmin-1* editing target (Figure 1.6 C). *Synaptotagmin-1* mRNA, which encodes the calcium sensor for fast neurotransmitter release at the synapse, forms arguably the most complex editing-directing RNA structure described in the literature. Editing sites C and D of the *synaptotagmin-1* transcript are encompassed in a long-range exon-intron structure of over 1500 nucleotides. This structure appears to form an intricate pseudoknot, the exact conformation of which serves to precisely regulate editing levels [Reenan, 2005].

1.2.3 Regulation of siRNA/miRNA biogenesis and function through RNA editing

Deamination of adenosines found in non-coding regions influences the biogenesis and target recognition of small interfering RNAs (siRNAs) involved in the RNA interference (RNAi) pathway [Nishikura, 2006]. The biogenesis of siRNAs requires processing of long dsRNA precursors into 21-23nt RNA duplexes, a task carried out

by the Dicer class of RNase III-like ribonuclease enzymes [Bernstein et al., 2001]. The mature siRNAs generated by Dicer are unwound for the incorporation of a single-stranded RNA (guide RNA) into the RNA-induced silencing complex (RISC). Ultimately, the targeting of RISC by the small guide RNA initiates transcriptional and post-transcriptional sequence-specific silencing [Hannon, 2002].

Growing evidence supports the interaction between the editing and RNAi pathways. Since the RNA editing and RNAi pathways both involve dsRNAs, editing may antagonize the RNAi pathway at two levels. First, adenosine deamination can alter the perfectly duplex structure of dsRNA substrates, leading to poor Dicer processing and thus decreased siRNA concentrations (Figure 1.8). Second, RNA editing of siRNAs may prevent base pairing of the RISC with the original mRNA targets, inhibiting the cleavage step of RNAi [Nishikura, 2006]. Similarly, editing may change the target of an siRNA by altering the primary sequence of the guide RNA. siRNA biogenesis is progressively inhibited with increased RNA editing *in vitro* [Scadden and Smith, 2001, Scadden and Smith, 2001]. However, the *in vivo* functional consequences of the antagonism between editing and RNAi pathways are still unclear.

Similar to the siRNA pathway, micro RNAs (miRNAs) are highly conserved, genomically encoded 20-25nt RNAs that mediate gene silencing and regulate diverse cellular processes including development, differentiation, and apoptosis [Bartel, 2004, Meister et al., 2004]. This class of small interfering RNAs is encoded within non-coding regions of the genome and form hairpin primary miRNA structures (pri-miRNAs), which are extensively processed to generate the mature miRNA. In their double-stranded state, these pri-miRNAs can serve as ADAR substrates (Figure 1.8).

Indeed, several miRNA precursors undergo A-to-I RNA editing at specific adenosines [Blow et al., 2006, Luciano et al., 2004]. Editing of these precursors can inhibit further cleavage and processing [Yang et al., 2006], as well as regulation of

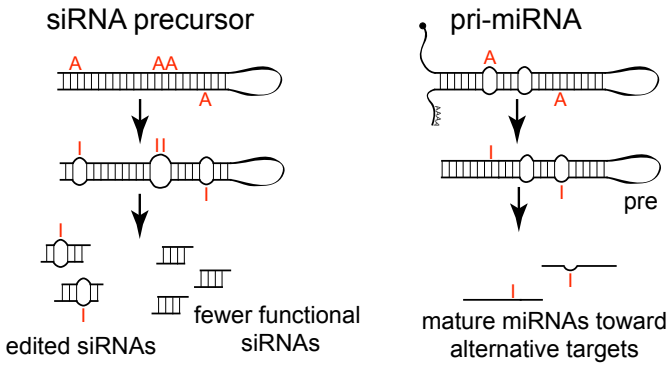


Figure 1.8: Editing of small RNAs. Editing can interfere with siRNA and micro RNA production and targeting. Perfectly duplex siRNA precursors are targets for hyper-editing by ADARs. Editing may result in improper Dicer processing and fewer functional siRNAs, or edited siRNAs. Primary (pri) microRNAs, imperfect duplexes, may be targets for specific editing, resulting in mature miRNAs toward alternative mRNA targets. Reproduced from Savva et al. [2012].

gene expression. A single edited adenosine within the 6nt targeting seed region of the miRNA is sufficient to redirect silencing to a new target [Kawahara et al., 2007] (Figure 1.8). In addition, A-to-I RNA editing is highly enriched at 3' UTRs within miRNA targets, suggesting a regulatory role for RNA editing in translation [Gu et al., 2012]. More importantly, RNA editing can inhibit or initiate miRNA/target interactions via modification at important positions where complementarity is required for appropriate miRNA targeting.

1.3 ADAR-substrate recognition

As discussed above, particular aspects of RNA structure serve to direct ADAR to edit specific adenosines. But how does the ADAR enzyme recognize substrates that vary in sequence almost as widely as the genome itself? No sequential or higher-order structure has yet been identified that is shared by all editing targets, which confounds the search for new inosinome candidates. Although ADARs recognize certain structures and precisely target adenosines in diverse transcripts, the only

feature common to all editing sites is a double stranded character.

The binding of an ADAR enzyme to an RNA substrate requires direct contact between the dsRBDs and dsRNA substrates [Valente and Nishikura, 2007]. Nuclear magnetic resonance and X-ray crystallography studies reveal that the binding of double stranded RNA binding proteins (dsRBPs) in general to dsRNA relies less on sequence than on structure [Bycroft et al., 1995, Nanduri et al., 1998, Ryter and Schultz, 1998]. Because dsRNA adopts an A-helical structure, which contains a deep and narrow major groove, dsRNA binding proteins are thought to recognize substrates through indirect readout and RNA structure [Draper, 1995].

Nuclear magnetic resonance data obtained from the dsRBMs of rat ADAR2 in solution with a simple hairpin substrate revealed unexpected structure-dependent discrimination by the individual motifs. In complex with the 71 nucleotide imperfect duplex encompassing the GluR-2 R/G editing site, dsRBM1 recognizes the conserved pentaloop at the apex of the hairpin structure, while dsRBM2 recognizes two bulged bases adjacent to the singular editing site [Stefl et al., 2006] (Figure 1.9 3A). However, these structural features are specific to the GluR-2 R/G site and are not common to all editing targets.

Consistent with the binding properties of other dsRBPs, ADARs can bind to any dsRNA without sequence specificity [Nishikura et al., 1991]. It has been widely assumed in the field that the dsRBDs of ADAR enzymes confer structural specificity during specific editing, determining which substrates are targeted and which adenosines are edited. The same group that reported structural recognition of the GluR-2 R/G site investigated this issue further and proposed a solution structure of the ADAR2 dsRBMs in complex with the GluR-2 R/G site that indicated sequence-specific recognition. Their data suggested that the ADAR2 dsRBMs contact the RNA minor groove, rather than the ribose backbone, allowing reading of the primary sequence [Stefl et al., 2010]. Two amino acid side chains contributed from each

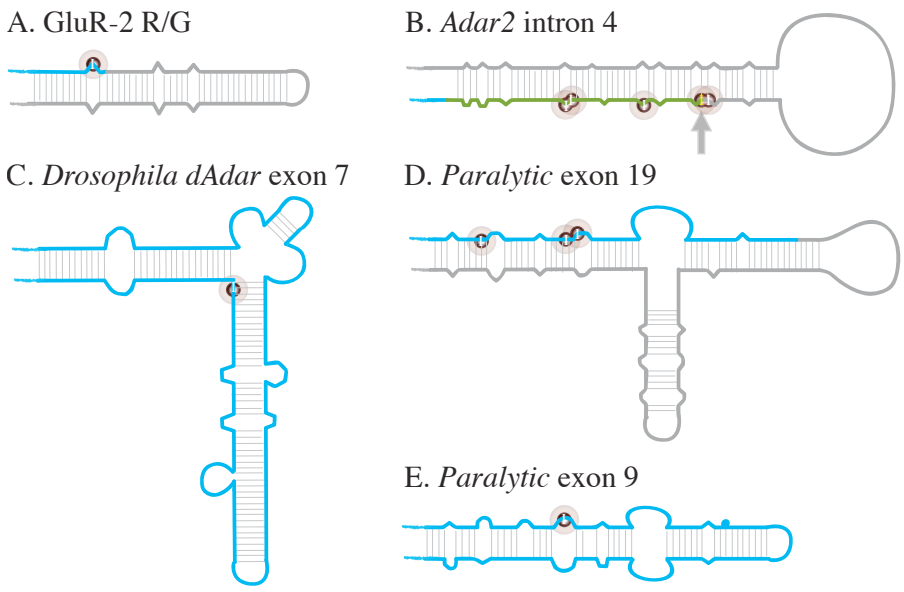


Figure 1.9: Proximity of editing sites to intron-exon junctions. Editing sites are often found near exon-intron (blue-gray) boundaries, but structures that direct editing can be formed from entirely exonic sequence (blue). (A) The mammalian GluR-2 transcript R/G site is located just one nucleotide from the intron-exon boundary. (B) The RNA structure that directs editing of mammalian ADAR2 is entirely intronic. Editing at one site (arrow) creates an alternative AG 3' splice acceptor site, leading to exonization of part of intron 4 (green). (C) *Drosophila dAdar* is edited at one adenosine in exon 7, substituting a glycine for the genomically-encoded serine. The *cis* elements that form this structure are included within the edited exon. (D) The *Drosophila* sodium channel transcript, *paralytic*, is edited at three adenosines in exon 19, located within 70 nucleotides of the downstream splice donor site. (E) *Paralytic* exon 9 is edited at a single site based on a structure formed from exonic sequence. Reproduced from Rieder and Reenan [2011].

dsRBM make sequence-specific contacts: each motif contacts a guanosine and an adenosine 9 or 10 nucleotides apart. *In vitro* studies revealed that the footprint of each dsRBM encompass 11-16 nucleotides of the 71 nucleotide hairpin GluR-2 substrate [Ohman et al., 2000].

However, it is not simply the dsRBMs of ADAR enzymes that provide substrate recognition; the C-terminal catalytic domain also contributes to editing specificity. Surprisingly, swapping the deamination domains of mammalian ADAR1 and ADAR2 reveals that these domains play the dominant role in determining editing specificity of the resulting chimeric proteins [Wong et al., 2001]. It is possible that the major function of the dsRBMs is to dimerize ADARs rather than contribute

to substrate specificity [Cho et al., 2003, Gallo et al., 2003], a view supported by the observation that ADAR1 and ADAR2 form homodimers independent of RNA substrate binding [Valente and Nishikura, 2007]. Both the amino terminal and dsRBM1 are required for dimerization of *Drosophila* dADAR on RNA, and point mutations in dsRBM1 are capable of abolishing both dimerization and editing activity. Nevertheless, while the amino terminus is required for activity and dimerization, dADAR peptides lacking this domain still recognize RNA substrates. This suggests that dADAR monomers bind individually and gain catalytic activity post-dimerization [Gallo et al., 2003]. Coupled editing sites are found on the same side of the RNA helix, indicating that only a single catalytic center remains active following dimerization [Enstero et al., 2009].

Yet recent *in vitro* work by Eggington et al. suggest that mammalian ADAR1 and ADAR2 are both able to edit when lacking N-terminal dsRBDs. Not only do these truncated enzymes effectively edit *in vitro*, but they are also able to selectively deaminate the same adenosines targeted by the full-length enzyme, although to different levels. The dsRBDs may function as RNA anchors [Stephens et al., 2004], but these data suggest that ADAR specificity is predominantly conferred, not by dsRBDs, but by the deaminase domain itself [Eggington et al., 2011]. However, experiments in an *in vivo* setting are required for the biological relevance of this observation to be understood.

Regardless of how ADARs recognize target substrates, editing seemingly must be accomplished before the splicing machinery recognizes local intron-exon junctions, removes introns, and abolishes the double-stranded structures that direct ADARs.

1.4 The complex relationship between RNA splicing and editing

Alternative splicing and RNA editing are both post-transcriptional modifications that diversify the resulting protein populations. These processes do not occur in isolation from each other; rather, editing and splicing are believed to be coupled, at the very least because both exonic and intronic sequences logically comprise most editing substrates in nascent transcripts. It is also thought that editing and splicing compete for access to the transcript and that one process contributes to the fine-tuning of the other. Further, recent high-throughput sequencing experiments revealed that co-transcriptional editing is widespread in *Drosophila* [Rodriguez et al., 2012].

1.4.1 Coupling of editing and splicing

Splicing is believed to occur co-transcriptionally [Goldstrohm et al., 2001] and recent evidence corroborates that spliceosomal assembly is required for proper transcription termination [Martins et al., 2011]. The RNA Polymerase II (Pol II) C-Terminal Domain (CTD) is thought to coordinate RNA processing events, including editing and splicing [Munoz et al., 2010]. Perhaps the most well known example of the interaction between editing and splicing is the editing by mammalian ADAR2 of its own transcript in intron 4 (Figure 1.9 3B). Editing of the *Adar2* RNA alters an AA into an AI (recognized as AG), creating a new 3' splice site (Figure 1.10). Edited forms of the transcript may then be alternatively spliced, and use of the new 3' acceptor causes a change to the open reading frame and a truncated protein [Rueter et al., 1999]. The Pol II CTD appears to coordinate editing and splicing of *Adar2* [Laurencikiene et al., 2006], suggesting that these post-transcriptional events may be coupled as a more general mechanism. The regulation of *adar2* transcript splicing via auto-editing represents a negative regulatory feedback mechanism to

modulate ADAR2 protein levels since inhibition of auto-editing in vivo increases ADAR2 expression and editing at many target adenosines in the transcriptome [Feng et al., 2006].

The transcript of the *Drosophila* homolog of mammalian ADAR2, dADAR, is also edited and spliced to create isoforms of different target specificity. Transcripts that contain alternative exon 3a due to recognition of a non-canonical 5' splice site also lack editing at the single target site in exon 7 [Keegan et al., 2005] (Figure 1.9 3C). The expression of dADAR isoforms with different specificity and activity is developmentally regulated and tissue-dependent [Marcucci et al., 2009], suggesting global editing is controlled in part by splicing of the *dAdar* transcript.

Adenosines targeted for editing are disproportionately localized near splice junctions in the pre-mRNA [Rieder and Reenan, 2011]. Therefore, during formation of a dsRNA ADAR substrate, intronic *cis*-acting sequences can form RNA duplexes encompassing splicing sites and potentially obscuring them from the splicing machinery (Figure 1.7). Furthermore, through modification of select adenosines, ADARs can create or eliminate splicing sites, broadly affecting later splicing of the transcript. Similar to the translational machinery, the spliceosome interprets inosine as guanosine, and therefore, a canonical GU 5' donor site and AG 3' acceptor site can be created via the deamination of AU (IU = GU) and AA (AI = AG), respectively [Valente and Nishikura, 2005]. Correspondingly, RNA editing can destroy a canonical AG 3' acceptor site (IG = GG) [Valente and Nishikura, 2005].

Yet even editing at sites that do not encode splicing signals can affect splicing. The mammalian GluR-2 transcript, for example, is edited at two non-synonymous sites: the Q/R and R/G sites, named for the residue changes resulting from editing. When the Q/R site, located in exon 11, as well as a downstream intronic hot spot, is edited, splicing is enhanced between exons 11 and 12 (Figure 1.10). In the same transcript, the R/G site is located just one nucleotide from the boundary between

exon 13 and the downstream intron (Figure 1.7). When this site is edited, splicing favors inclusion of exon 15 over that of exon 14 (Figure 1.10). The mechanism of how these particular editing sites affect splicing of the GluR-2 transcript is still unknown [Schoft et al., 2007].

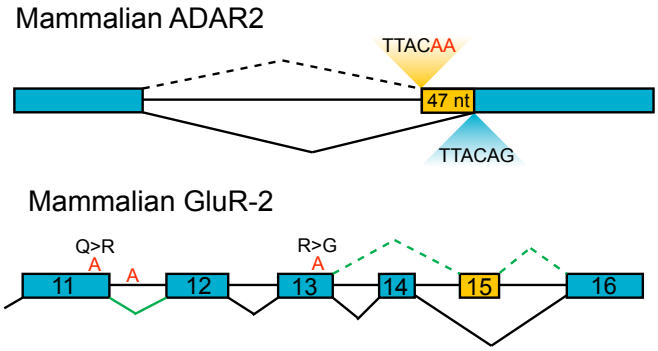


Figure 1.10: Editing affects splicing. Mammalian ADAR2 auto-edits its own transcript, creating a novel splice site (red), which results in the inclusion of 47 nucleotides (yellow) and a frameshift in the coding sequence. In the mammalian GluR-2 transcript, editing at both the Q/R site in exon 11 and an intronic hotspot (red) is required for efficient removal of the downstream intron. Editing of the R/G site (red) reduces efficacy of downstream splicing and favors an alternative final exon configuration (yellow). Reproduced from Savva et al. [2012].

Molecular data presented above indicates that editing and splicing are coordinated. Subcellular localization of ADARs also provides clues to the relationship between editing and splicing. Mouse ADAR2 is concentrated in the nucleolus [Sansam et al., 2003] and *Drosophila* dADAR is localized predominantly in the nucleus [Jepson et al., 2011]. Once a pre-mRNA has been spliced it is quickly shuttled from the nucleus to the cytoplasm for translation [Luo and Reed, 1999, Valencia et al., 2008]. Splicing and editing, therefore, are likely both compartmentalized in the nucleus.

1.4.2 Evidence that editing precedes splicing

Support for the coupling of editing and splicing comes from the nature of the structures that direct editing. Often these duplexes are formed between the region of an exon that

encompasses the editing site(s) and a nearby intron [Reenan, 2005, Hanrahan et al., 2000, Bratt and Ohman, 2003] (Figure 1.6 B,C; Figure 1.9 A, D). These structures can only form if editing precedes splicing: splicing results in loss of double-strandedness, prohibiting ADAR substrate recognition. Editing sites and the *cis* elements that direct editing tend to occur near exon/intron borders [Reenan, 2005, Rueter et al., 1999, Higuchi et al., 1993] (Figure 1.9 A, B, D).

Because RNA double-strandedness is required to direct ADAR, and editing sites are often located near splice junctions, splicing signals may be buried in secondary structures. However, these structures must be resolved in order for the intron-exon boundary to be accurately recognized by the splicing machinery. Conserved secondary structures are enriched at alternative splice sites [Shepard and Hertel, 2008], and encompassing the 3 or 5' splice site in local structure prevents spliceosomal recognition [Buratti and Baralle, 2004]. What does this mean for the splicing signals of constitutive exons, for example those represented in Figure 1.6 B and C, and Figure 1.9 A and D, which are involved in secondary structures directing editing? Resolution of these structures may be key.

Such a structure is predicted to direct editing at three sites in constitutive exon 19 of the *Drosophila* sodium channel gene *paralytic*. This structure also encompasses the downstream exon-intron boundary (Figure 1.9 D). A mutation in a dsRNA helicase, Maleless, appears to result in decreased recognition of the 5' splice site (obscured due to inclusion in secondary structure) and skipping of the edited exon–a splicing catastrophe that results in fewer functional sodium channels and whole organism temperature-sensitive paralysis [Reenan et al., 2000]. This evidence suggests that the helicase mediates interplay between editing and splicing: after deamination by dADAR the Maleless helicase resolves the secondary structure to expose the splicing signals. The mutant version of Maleless is either unable to resolve the structure or fails to dissociate from the region, resulting in exon skipping (see Figure 1.16). In

the mutant Maleless background correctly spliced transcripts are edited half as often as in a wild type helicase background. This finding suggests that the transcripts that do not fold into the structure recognized by dADAR are also those that are more easily identified by the splicing machinery. This hypothesis also suggests that the processes of splicing and editing compete for access to the nascent transcript.

While in some cases editing and splicing may be in direct competition, in others editing is known to direct splicing and therefore must occur first. The GluR-2 transcript is edited at two major recoding sites, R/G in exon 13 and Q/R in exon 11, and both are controlled by *cis* elements near the exon/intron borders (Figure 1.9 A and Figure 1.6 B, respectively). The Pol II CTD enhances editing at the R/G site by preventing premature splicing; yet editing of this site reduces splicing efficiency of the downstream intron and favors alternative splicing events [Schoft et al., 2007]. When the R/G region encompassing the 5' splice site was fused to the 3' splice site of an adenovirus reporter, editing and splicing were each found to interfere with the other *in vitro* [Bratt and Ohman, 2003]. Similar to the R/G site, editing at the Q/R site, which occurs at nearly 100%, is also facilitated by the CTD [Ryman et al., 2007]. Editing at both the Q/R site and an intronic hot spot is required for efficient splicing of the following intron (Figure 1.10). Further, when these sites are not edited the transcript is improperly spliced. Perhaps this serves as a quality control mechanism by ensuring that only edited transcripts mature and are exported to the cytoplasm for translation [Schoft et al., 2007].

Similar to the GluR-2 transcript, editing of *Alu* elements, though non-specific (Figure 1.6 A), directs future splicing events. Rather than favoring one discrete alternative exon over another, however, *Alu* editing can result in exonization of a previously non-coding sequence [Athanasiadis et al., 2004]. For example, the human nuclear prelamin A recognition factor transcript includes two *Alu*s in reverse orientation separated by 25 nucleotides, which may form a nearly perfect duplex.

Editing in the downstream *Alu* results in generation of a new 3' splice signal and alteration of splicing enhancers. In addition, a premature stop codon within the *Alu* is edited to that for tryptophan. Collectively these editing events can result in exonization of previously intronic sequence, a process that is regulated in a tissue-dependent manner [Lev-Maor et al., 2007].

While the significance of exonization by edited *Alu* elements is unmistakable, the relationship between splicing and editing may be considerably more opaque. Two *Drosophila* transcripts demonstrate correlation between splicing and editing across some distance. One encodes a calcium channel and the other a nicotinic acetylcholine receptor, and editing in each is directed by intronic *cis* signals. No correlation between editing and splicing is observed within the channel transcript, in which the edited exon is considerably upstream from the alternative splicing event. However, the alternate forms of the receptor transcript, in which the editing site is downstream from the alternative exons, display different editing efficiencies [Agrawal and Stormo, 2005]. These observations further suggest that editing precedes splicing and that this relationship may occur over considerable distances within the transcript.

Editing, therefore, must occur prior to splicing in most transcripts, either because the intron contains necessary *cis*-acting elements or because editing regulates subsequent splicing. Yet not every ADAR substrate contains an intronic component (Figure 1.9 C, E), and it is possible that select transcripts are deaminated post-splicing or even in isolation from other post-transcriptional processes.

1.4.3 Rare post-splicing editing events are possible

Several entirely exonic structures that direct editing have been identified. The human potassium channel gene $K_v1.1$ is intronless, yet edited via a 114 bp imperfectly duplex hairpin entirely contained within coding sequence [Bhalla et al., 2004]. One of the many editing sites in the *Drosophila* sodium channel transcript, *paralytic*, is directed

by an exonic hairpin [Hanrahan et al., 2000] (Figure 1.9 E). The *Drosophila dAdar* transcript itself undergoes editing based on a structure formed within exon 7 [Keegan et al., 2005] (Figure 1.9 C), which changes the subnuclear organization and specificity of the resulting enzyme for editing targets [Savva et al., 2012]. Auto-editing of the *dAdar* transcript decreases overall deaminase activity of the enzyme [Keegan et al., 2005], suggesting a feedback mechanism to control global editing levels.

In theory, intermolecular structures formed between two mature mRNAs could form double-stranded ADAR substrates. In such an example, one exonic sequence would act in *trans* as a regulatory element to direct editing of the other. This concept has even been explored *in vitro* and abstracted to the therapeutic realm: targeted ribo-oligonucleotides could direct ADAR to correct disease-related genomic mutations at the post-transcriptional level [Woolf et al., 1995], though controlling deamination of precise adenosines would be incredibly challenging. In addition, there is as yet no *in vivo* evidence of post-spliced intermolecular structures attracting ADAR, and no editing sites have been identified that lack a *cis* regulatory sequence.

Although it is possible that entirely exonic structures directing editing endure in cytoplasmic post-spliced mature mRNA, ADARs are most often found localized to the nucleus, as discussed above. However, some observations suggest ADARs may be present, at least transiently, in the cytoplasm. Mammalian ADAR1, for example, which has overlapping specificity with ADAR2 [Lehmann and Bass, 2000], shuttles between the nucleus and cytoplasm [Eckmann et al., 2001]. Inflammatory responses can cause increased ADAR1 expression in the cytoplasm [Liu et al., 1997] and the presence of extranuclear mammalian ADARs appears to be regulated by interferon responses [Yang et al., 2008, Patterson and Samuel, 1995]. Mammalian ADAR3, though not known to have catalytic activity, contains a nuclear localization signal recognized by nuclear importins. Certain isoforms of ADAR2 contain an ADAR3-like signal, suggesting ADAR3 may regulate sub-cellular localization of

more enzymatically-active ADARs by competing for importins [Maas and Gommans, 2009]. Sub-cellular localization of ADAR2 is also dependent on the phosphorylation-dependent peptidyl-prolyl *cis/trans* isomerase Pin1. In the absence of Pin1, ADAR2 mislocalizes to the cytoplasm. Cytoplasmic ADAR2 is targeted for degradation by WWP2, an E3 ubiquitin ligase [Marcucci et al., 2011], and so is unlikely to participate in editing of extranuclear targets.

Therefore, rare post-splicing or extranuclear editing may occur. If present, RNA editing activity in the cytoplasm could serve a distinct function. Splicing, while usually confined to the nucleus, also occurs in the cytoplasm of neuronal cell dendrites [Glanzer et al., 2005]. This suggests that pre-mRNA processing on-site may allow more rapid responses to synaptic signals. Cytoplasmic editing could serve a similar purpose, as editing targets are involved in chemical and electrical synaptic signaling. Post-transcriptional modification at the synapse would allow selection of the most appropriate protein isoforms and near-instant responses to changes in the dendritic environment. The relationship between cytoplasmic editing and splicing may occur much like nuclear events or be transcript-dependent.

1.5 Future directions

1.5.1 Frontiers in ADAR biology

Inappropriate RNA editing is associated with suicidal depression [Dracheva et al., 2008], cancer [Galeano et al., 2011], and neurodegenerative diseases [Blow et al., 2004] including amyotrophic lateral sclerosis [Kawahara et al., 2004, Takuma et al., 1999]. Thus a better understanding of the ADAR protein family functions will shed light into therapeutic strategies regarding these and other nervous system disorders.

ADARs modify specific adenosines to inosines in short imperfect dsRNA templates, a well-characterized role that usually leads to recoding of transcripts and

peptides (Figure 1.11 a). In contrast, ADAR modification of long, perfect duplexes results in the deamination of up to 50% of the adenosines [Cattaneo, 1994, Fischer et al., 2008, Peters et al., 2003] affecting siRNA biogenesis and activity (Figure 1.11 b). Although much progress has been achieved in ADAR biology over the last two decades, the consequences of such promiscuous RNA editing, especially *in vivo*, are not well understood.

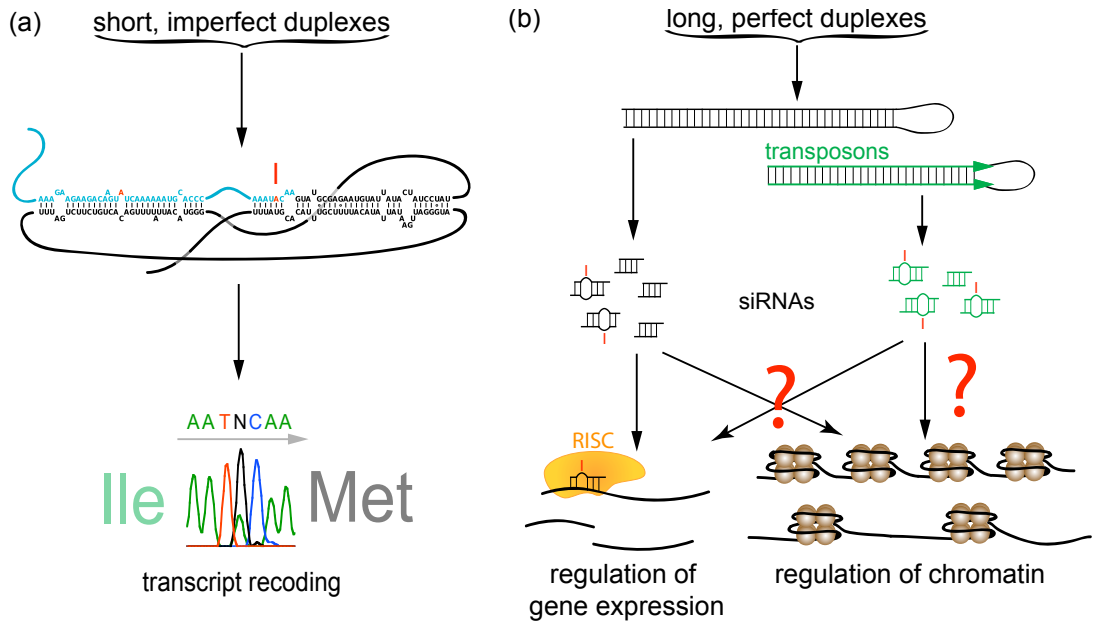


Figure 1.11: Consequences of specific versus non-specific editing. (a) Short, imperfect duplexes, such as *Drosophila synaptotagmin-1*, are specifically edited leading to transcript and peptide recoding. (b) Long, perfect dsRNA substrates, including those formed by nearby transposons in opposite orientation (green), are hyper-edited, leading to fewer or edited siRNAs. This may alter gene expression through RISC (yellow) targeting, but evidence also links the RNAi pathway to chromatin regulation. Reproduced from Savva et al. [2012].

Several lines of evidence indicate that promiscuous editing by ADARs antagonizes RNAi-mediated gene silencing [Nishikura, 2006, Wu et al., 2011]. Editing of endogenous siRNAs in *Drosophila* [Kawamura et al., 2008] suggests that ADARs interact with endogenous RNAi pathways, which are involved in somatic defense against transposable elements [Chung et al., 2008, Czech et al., 2008, Ghildiyal et al., 2008, Kawamura et al., 2008] and in the regulation of chromatin states [Fagegaltier

et al., 2009] (Figure 1.11 b). The consequences, however, of editing/RNAi interactions and the resulting effects on both gene expression and chromatin regulation remain enigmatic. Future studies should aim in deciphering the *in vivo* consequences of such interactions, as this avenue highlights a more global role for ADARs regarding the broad regulation of neuronal transcriptomes.

1.5.2 Directions in RNA post-transcriptional processing

The relationship between RNA editing and splicing is far from resolved and necessitates future study. Single nucleotide polymorphisms at the genetic level can result in differential splicing between individual humans, and may even impact gene expression [Lalonde et al., 2011]. Nucleotide changes in the transcript, for example, those caused by RNA editing, may have similar impacts on splicing and overall gene expression. In this way aberrant editing might lead to disease states. Over 50% of known disease-causing mutations affect splicing [Hammond and Wood, 2011], and the role of editing in these diseases has not been addressed.

The most widely used therapy to correct mis-splicing diseases is targeting antisense oligonucleotides to the pre-mRNA [Hammond and Wood, 2011], creating secondary structure around the splicing area of interest. Although these therapeutic oligonucleotides are chemically altered to resist nucleases and improve functionality and are not chemically equivalent to ribo-oligonucleotides, such therapies could also affect RNA editing by changing transcript structure, especially as editing sites tend to occur near intron-exon junctions (Figures 1.6 and 1.9). Off-target effects will have to be considered when designing antisense oligonucleotide therapies.

To more fully understand the relationship between splicing and editing, and effectively interpret the roles of such modifications in disease, the kinetic association between these processes should be explored to probe the model in which editing precedes splicing. For example, sequestering constitutive splice sites in synthetic RNA

secondary structure, as is observed for endogenous alternative splice sites [Shepard and Hertel, 2008], should allow more time for editing of nearby adenosines, resulting in increased editing or even editing of new target nucleotides. Decreasing RNA structure by reducing or altering intronic *cis* sequences predicted to obstruct splice sites should allow less time for ADAR to deaminate nearby, and so editing would be predicted to ablate or decrease. It would be most informative to conduct such experiments *in vivo* where the editing and splicing machinery exist in the correct stoichiometry and whole organismal effects may be observed. However, altering secondary structure around constitutive splice signals may also result in aberrant splicing [Goguel et al., 1993] and effects on behavior or viability of the animal.

Performing these experiments *in vivo* requires precision and accuracy to preserve the pre-mRNA sequence except where mutation is intentionally introduced. Although most often employed in yeast, homologous recombination techniques may be used to introduce genetic mutations into *Drosophila* [Staber et al., 2011] and have already been successfully employed to alter RNA editing in this organism [Jepson et al., 2011, Savva et al., 2012]. The homologous recombination approach has numerous benefits over traditional reporter transgenic constructs. For example, it allows manipulation of the endogenous gene, transcribed at the biologically relevant level rather than an exogenous driver generating overexpression of a reporter or minigene construct. This is especially important considering editing targets are most often involved in electrical and chemical neurotransmission, and altering gene dosage could have dramatic viability effects. In addition, manipulation of the endogenous gene allows for behavioral studies not possible using overexpression or transgene systems.

Investigating the interaction of editing and splicing *in vivo*, therefore, is disease-relevant and necessary for modeling the coordination between post-transcriptional processes. In addition, new editing sites identified through transcriptome

deep sequencing and other novel techniques continue to expand known inosinomes, and each new adenosine brings insight into possible structural or sequential patterns shared by ADAR target sites.

Further, ADAR itself may provide a therapeutic tool for genetic diseases. For example, fusing the ADAR catalytic domain to the bacteriophage λ N-peptide is sufficient to target the chimeric protein to dsRNA tagged with the boxB RNA recognition element. Swapping the bulged nucleotides in the target RNA structure alters editing specificity of the substrate (J. Rosenthal, personal communication). These results suggest that ADAR may provide a useful tool to combat genetic diseases at the RNA level.

1.6 Voltage-gated sodium channels

Complex neurological functions require precise control of neuronal properties such as membrane permeability to sodium ions. Voltage-gated sodium channels, responsible for the rising phase of the action potential in the membranes of neurons and other electrically excitable cells, must be flexible and sensitive enough to account for changing cellular and environmental conditions. Yet how sodium channel flexibility is achieved differs between organisms. For example, the human genome encodes at least nine major sodium channel genes (*SCN1A*-*SCN9A*), which may have subtly different properties [Mantegazza et al., 2010]. The *Drosophila* genome, however, contains just one sodium channel gene, *paralytic* (*para*), but the *para* transcript may be altered post-transcriptionally through alternative splicing and editing to produce over two million different channel isoforms [Graveley, 2001].

1.6.1 Sodium channel structure and function

It appears that vertebrate and invertebrate sodium channels are derived from a single ancestor, possibly a calcium channel, and have evolved independently in these two branches [Goldin, 2002]. Voltage-gated sodium channels consist of a highly processed α subunit of 260 kDa that is sufficient for functional sodium current, and auxiliary β subunits of approximately 36 kDa that are required for gating and normal kinetics [Goldin, 2002]. The β subunits may also be required for α subcellular targeting [Mantegazza et al., 2010]. The human genome contains nine sodium channel α genes and four β genes (*SCN1B-SCN4B*), while that of *Drosophila* contains only the single *para* functional α subunit and a potential β subunit, TipE [Feng et al., 1995, Hodges et al., 2002]. *TipE* is coexpressed with *para* in most cells and coimmunoprecipitates from *Xenopus* oocytes, suggesting physical association [Hodges et al., 2002].

Sodium channels are comprised of four homology domains (I-IV), each with six trans-membrane α helices (S1-S6; Figure 1.12). The channel is composed of a voltage-sensing domain, formed by S1-S4 and a pore-forming domain, consisting of S5 and S6. Voltage-gated sodium channels play a key role in neuronal excitability: in response to initial membrane depolarization, sodium channels open within a few hundred microseconds (activation), resulting in an inward Na+ current and further membrane depolarization. Channels subsequently inactivate in milliseconds, preventing additional ion flow. Inactivation, however, is often incomplete, leading to persistent Na+ current, which is known to be important to neuronal function [Mantegazza et al., 2010]. During channel gating, positively-charged residues on S4 move across the membrane in a "screw-helical" fashion, stabilized by negative charges in the S1 helix [Paldi and Gurevitz, 2010].

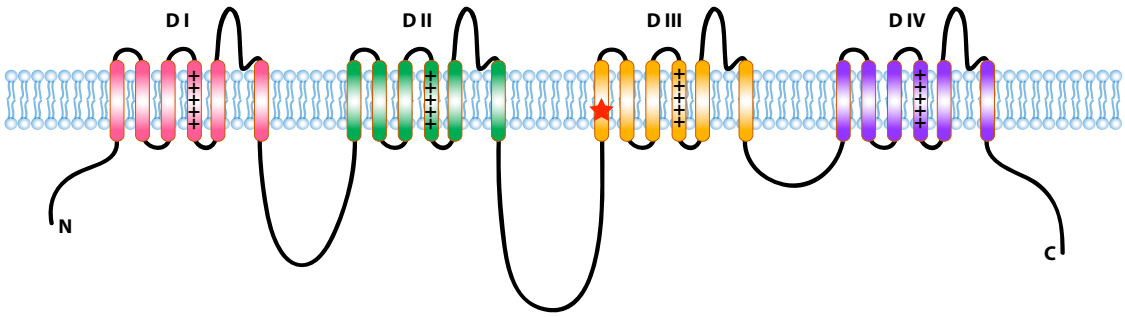


Figure 1.12: Voltage-gated sodium channel structure. The α sodium channel subunit, embedded in the cell membrane (blue), consist of four homology domains (DI-IV), each with six membrane-spanning helices (S1-S6). The fourth helices of each domain form the voltage sensor (+). The editing sites discussed in this dissertation (**Chapter 2**) are located on S1 of DIII (red star). α subunits are able to function autonomously, but their gating and localization may be affected by β subunits (not shown).

1.6.2 Sodium channelopathies

Perturbations in sodium channel structure or dosage can have substantial effects on channel function, neuronal activity, and resulting behavioral phenotypes. Mutations in sodium channel genes are implicated in the pathophysiology of human disorders as diverse as epilepsy, ataxia, migraine, neuropathic pain, autism, and neurodegenerative diseases including multiple sclerosis [Mantegazza et al., 2010, Meisler and Kearney, 2005]. Identified epilepsy mutations, such as those involved in generalized epilepsy with febrile seizures plus (GEFS+) and severe myoclonic epilepsy of infancy (SMEI), are found primarily in *SCN1A* and localize throughout the protein without any clear relationship to functional channel domains [Meisler and Kearney, 2005]. SMEI mutations, which mostly arise *de novo*, may be truncation or missense, while all identified GEFS+ mutations are missense in highly-conserved amino acids. Many studies suggest that most epilepsy-generating mutations lead to loss of channel function and haploinsufficiency [Mantegazza et al., 2010].

Many of the most widely-used antiepileptic drugs, such as phenytoin, carbamezepine, and lamotrigine, inhibit the function of voltage-gated sodium channels [Mantegazza et al., 2010]. While it appears curious that epilepsy, a disorder charac-

terized by brain hyperexcitability, should be caused by loss-of-function mutations in sodium channel genes, it appears that the main gene affected in epileptic patients, *SCN1A*, is most highly expressed in inhibitory neurons. This gene is also often affected in migraine patients, although mutations range from loss of function to gain of function [Mantegazza et al., 2010].

Sodium channel disorders may be aggravated or made apparent by increases in temperature. For example, epileptic attacks can occur in children due to increases in ambient or body temperature resulting from the immune response to infection or vaccination. Indeed, discovery of *paralytic* in 1971 was due to a temperature-sensitive mutation causing adult *Drosophila* paralysis [Suzuki et al., 1971], which suggested a nervous system defect. It was not until 1989 that *para* was determined to be the *Drosopihila* sodium channel homolog [Loughney et al., 1989, Ramaswami and Tanouye, 1989]. By then, nap^{ts}, another mutation believed to affect sodium channel abundance had been discovered [Wu et al., 1978]. It took until 2000 to discover that nap^{ts} is an allele of *maleless* [Reenan et al., 2000] (*mle*), an RNA helicase involved in *Drosophila* dosage compensation, that affects *paralytic* splicing and therefore channel abundance. This mutation is discussed below and is specifically addressed in **Chapter 4**. Additionally, human GEFS+ *SCN1A* epilepsy mutations knocked into *Drosophila paralytic* result in a temperature-induced seizure phenotype due to reduced inhibitory activity in the central nervous system [Sun et al., 2012].

1.6.3 *Paralytic* is the only sodium channel gene in *Drosophila*

Sodium channel mutations can lead to a variety of neurological syndromes and widespread behavioral phenotypes, but due to redundancy in the human genome molecular and phenotypic effects of isolated mutations are difficult to study. *Drosophila paralytic* presents an ideal model in which to investigate sodium channel processing and function, and is the subject of **Chapter 2**. Localization of the editing sites

studied in this chapter are represented in Figure 1.12 (protein) and Figure 1.13 (locus).

While mammalian genomes encode nine functional α sodium channel subunits, most invertebrate genomes, including that of *Drosophila*, contain only one, suggesting that invertebrates and vertebrates use different means of generating diversity in sodium channel function [Goldin, 2002]. Additionally, although some electrophysiological differences have been attributed to alternate mammalian channel spliceoforms, it is not clear how much functional diversity is achieved through alternative splicing, and no RNA editing has been identified in any mammalian sodium channel transcript. This is in sharp contrast to invertebrate sodium channels, which are heavily edited and alternatively spliced to generate protein diversity. Additionally, although *C. elegans* contains only 302 neurons, its genome encodes many more ion channel genes than the genome of *Drosophila*, an organism with about 250,000 neurons, again suggesting that *Drosophila* must achieve channel diversity through non-genomice molecular strategies [Littleton and Ganetzky, 2000]. Insect sodium channel spliceoforms generate a broad range of voltage-dependent activation and inactivation properties [Tan et al., 2002], and two A-to-I RNA editing sites in the German cockroach $BgNa_v$ sodium channel transcript lead to amino acid substitutions and altered gating properties [Song et al., 2004].

The *paralytic* locus, located on the X chromosome, is over 65 Kb long, and contains 26 constitutive exons [Loughney et al., 1989]. In addition to a myriad of mutually exclusive exons and alternative exons [Thackeray and Ganetzky, 1995], the transcript contains a dozen sites of RNA editing (Figure 1.13). Taken independently, these post-transcriptional modifications could generate over two million sodium channel isoforms [Hanrahan et al., 2000]. Evidence suggests only a fraction of these post-transcriptional modification combinations are utilized, and that the range and ratios of transcript isoforms change throughout development, among tissues,

and between sexes [Thackeray and Ganetzky, 1994, Ingleby et al., 2009]. Sodium channel spliceoforms have different characteristics in the membrane, such as varying voltage-dependent activation or inactivation [Olson et al., 2008, Lin et al., 2009], yet even minor sequence changes in the sodium channel transcript may translate to structural/functional changes, and provide flexibility in local or global neuronal membrane properties.

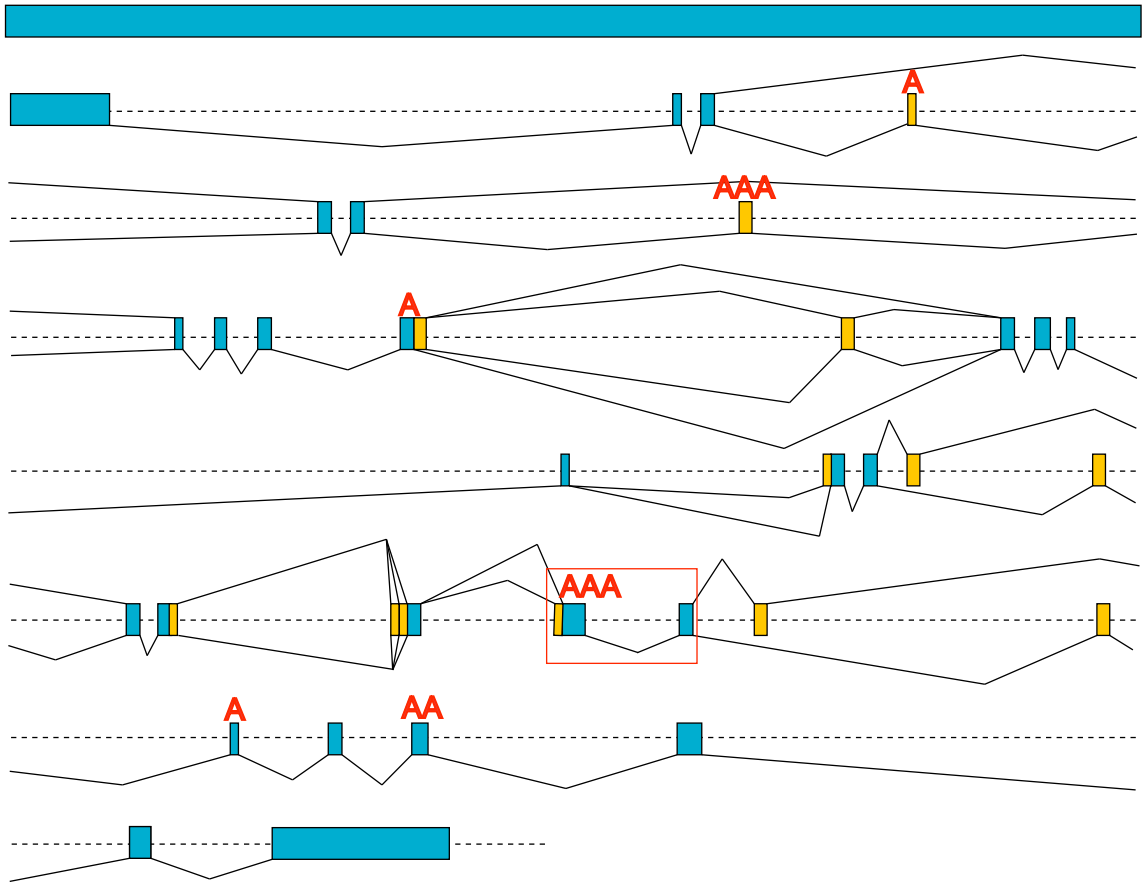


Figure 1.13: The *Drosophila paralytic* sodium channel locus. The entire *paralytic* locus is over 65 Kb. Constitutive exons are represented in blue, alternative exons in yellow, and introns as dotted lines. Editing sites are denoted by red "A"s. If all of the post-transcriptional modifications are considered separately, the single *paralytic* locus is able to give rise to over two million protein isoforms. The editing sites within the red box are the subject of **Chapter 2**.

1.6.4 Mechanisms of temperature-sensitive paralysis

As discussed above, sodium channel mutant phenotypes are exacerbated by increased temperature, which often leads to temperature-sensitive paralysis in affected *Drosophila*. Nelson and Wyman used *Drosophila paralytic* to propose a synergistic paralysis mechanism, suggesting that temperature-sensitive mutations in this locus uncover the inherent temperature sensitivity in all neurons. As temperature increases, neuronal repolarization occurs earlier due to an increase in the rates of activation and inactivation for both sodium and potassium currents [Huxley, 1959] (Figure 1.14 **A**). This "normal" effect of temperature on the action potential is augmented when there are fewer functional sodium channels, as with nap^{ts} (see below) and temperature-sensitive *para* mutations ($para^{ts}$, Figure 1.14 **B**), leading to paralysis when the sodium current fails to reach spiking threshold [Nelson and Wyman, 1990] (Figure 1.14 **C**).

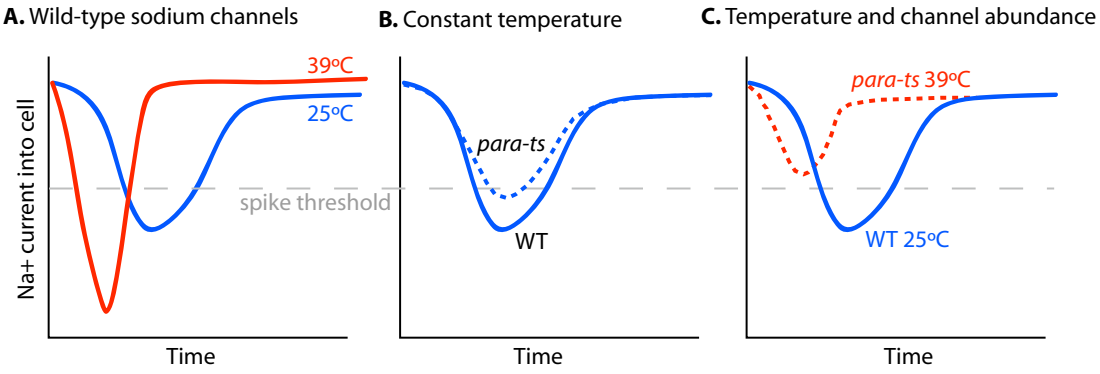


Figure 1.14: Sodium current in response to temperature and channel abundance. **A.** At elevated temperatures (red), sodium channels more rapidly activate and inactivate compared to at cooler temperatures (blue). **B.** $Para^{ts}$ and other mutations that affect channel abundance, reduce overall sodium current (dashed blue line). **C.** When *Drosophila* mutants with reduced sodium channel abundance are subjected to elevated ambient temperatures, a paralysis phenotype ensues due to the synergistic interaction between the normal effects of temperature on sodium channel current and reduced number of functional sodium channels because the sodium current (dashed red line) fails to reach spiking threshold (dashed gray line).

1.7 Splicing in *paralytic* is affected by a mutation in the Maleless RNA helicase

The nap^{ts} mutation in *Drosophila* was originally believed to be a mutation in an α or β sodium channel gene, but was finally shown in 2000 to be a mutation in the *maleless* gene [Reenan et al., 2000] encoding a member of the DEAH RNA helicase family (Mle) involved in dosage compensation. In Mle^{nap-ts} mutants, approximately 80% of *paralytic* transcripts are aberrantly spliced, always lacking constitutive exon 19 (harboring the three editing sites discussed in **Chapter 2**), and creating a widespread splicing catastrophe [Reenan et al., 2000]. At permissive temperatures, 20% of correctly spliced *para* transcript suffices for neuronal activity in mle^{nap-ts} mutants. However, at elevated temperatures (39°C) more sodium channel function is required for proper neuronal activity and 20% of the normal concentration of transcripts is insufficient. As even wild type flies paralyze at 43°C, mle^{nap-ts} and $para^{ts}$ mutations uncover normally temperature-sensitive traits in neurons [Nelson and Wyman, 1990](Figure 1.14).

Para null mutations, which are recessive lethal in a wild type background, become dominant lethal in a mle^{nap-ts} background, while $para^{ts}$ mutations, normally homozygous viable, become homozygous lethal. A duplication of wild type *para* in the genome rescues these effects, indicating that the observed phenotypes are due to *para* dosage [Ganetzky, 1984].

The mutation responsible for the mle^{nap-ts} "splicing catastrophe" phenotype is unknown. Candidates include a 7 bp intronic insertion and a single amino acid mutation (T415S) in the ATP hydrolysis domain of the protein [Smith, 2004]. Mle^{nap-ts} homozygous males are viable, while $mle^{-/-}$ males are inviable, indicating that the role of Mle in the Male-Specific Lethal complex required for dosage compensation is unperturbed by the *nap-ts* mutation. Polytene staining of nap^{ts} and wild type

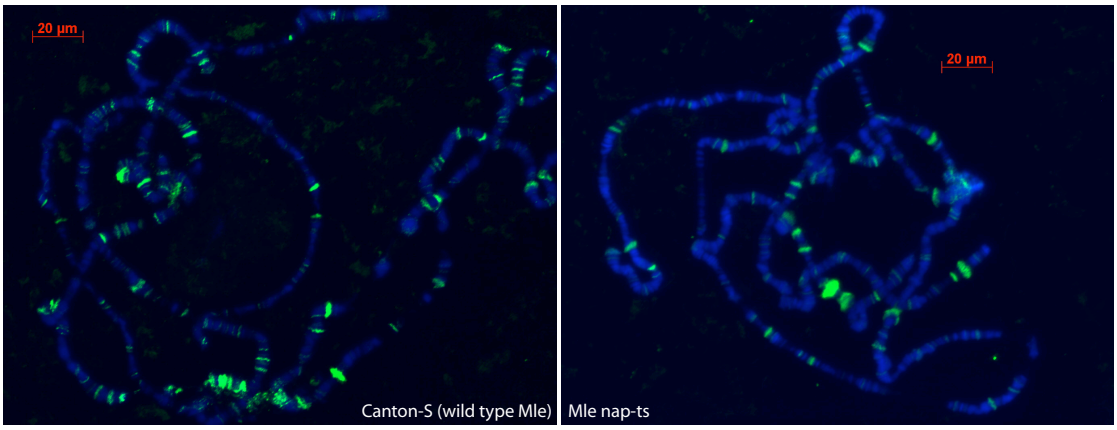


Figure 1.15: Third instar larvae polytene chromosome staining of the wild type Mle vs. the Mle^{nap-ts} mutant helicase. Female Canton-S polytenes are on the left, while female *mle*$^{nap-ts}$ chromosomes are on the right. Chromosomes are stained with DAPI (blue) while the helicase is stained with an anti-Mle antibody (green), courtesy of Mitzi Kuroda.

Mle reveal similar binding patterns in females, suggesting that the Mle^{nap-ts} helicase still localizes to active sites of transcription on polytenes (Figure 1.15). Curiously, homozygous *mle*$^{-/-}$ females are viable and do not have behavioral or molecular defects; a complete absence of the Mle protein does not result in exon skipping in the *para* transcript [Reenan et al., 2000].

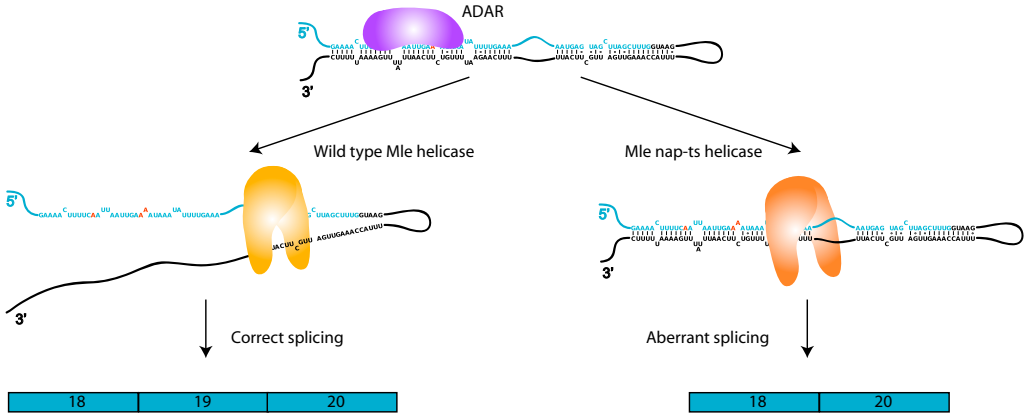


Figure 1.16: A model for the interaction of the Mle^{nap-ts} helicase with the *para* transcript. After dADAR (purple) edits three adenosines in *para* exon 19 (blue), a process that is dependent on the formation of RNA secondary structures, the wild type Mle helicase (yellow) resolves these structures prior to editing. The *mle*$^{nap-ts}$ mutation results in a recessive gain of function allele, giving rise to a mutant helicase (orange) that fails to unwind the RNA structure and/or obstructs the splicing machinery from removing the intron (black) leading to aberrant spliceoforms and temperature-senstive paralysis. Figure adapted from Reenan et al. [2000]

Therefore, the mle^{nap-ts} mutation results in a recessive gain of function allele. The current model for Mle^{nap-ts}-*para* interaction involves the wild type Mle helicase resolving *para* pre-mRNA secondary and tertiary structures after editing occurs but before splicing [Reenan et al., 2000] (Figure 1.16). In the complete absence of Mle, other RNA helicases participate in *para* unwinding, generating properly-spliced transcript isoforms. In an mle^{nap-ts} background, the mutant helicase not only fails to resolve the RNA structures surrounding *para* exon 19, but also blocks the splicing machinery in some way, leading to aberrant splicing and a temperature-sensitive paralysis phenotype.

Interestingly, the 20% of full-length *para* transcripts present in the mle^{nap-ts} mutant are edited at a significantly lower level than the full-length transcripts in $mle^{+/+}$ animals (37% vs. 73%) [Reenan et al., 2000]. This suggests that when the transcript folds into a structure that is not conducive to editing, the transcript is more likely to be correctly spliced even in a mle^{nap-ts} background. In other words, it may be the RNA structures responsible for directing editing that interfere with the Mle^{nap-ts} helicase. **Chapter 4** discusses this theory in greater detail.

1.8 Using Homologous recombination to engineer the *Drosophila* genome at the *paralytic* locus

Previously, a mini-gene reporter construct, encompassing *para* exon 18-20, including all intervening intronic sequences, was used to investigate the interaction between editing and splicing in the mle^{nap-ts} background [Smith, 2004]. While the wild type version of this construct is edited at the same three locations as endogenous *para* exon 19, it is to a much lower level. Additionally, while the mini-gene recapitulated the "splicing catastrophe" due to Mle^{nap-ts}, it was in the context of the endogenous wild type *para* gene. Finally, the mini-gene reporter system did not allow extrapolation

from genotype to phenotype.

Therefore, to introduce precise genetic mutations into the *Drosophila para* locus, we used homologous recombination (HR). HR is a technique widely used in mouse and yeast genetics, in which precisely engineered mutations are introduced into an endogenous genetic locus. The context of promoter, enhancer elements, chromatin state, and chromosomal location is preserved. Such a technique is especially important for a gene such as *para*, as sodium channel dosage affects organismal behavior.

HR was recently developed [Rong and Golic, 2000] and streamlined [Staber et al., 2011] in *Drosophila*, and research application for this versatile technique in the fly continues to expand. Once the transgenic animal has been created, the process (Figure 1.17) is predicted to take approximately three months. Briefly, a construct, contained on a pW25 vector with P-element insertion ends, is designed containing two "arms" homologous to the desired mutation site. The arms are flanked by restriction sites used for sub-cloning, flippase recognition targets (FRT), and I-SceI recognition sites for future genomic excision. Between the arms is a *mini-white* gene flanked by loxP sites, 6-frame stop codon sequences, and restriction cloning sites.

The construct is introduced into syncytial *Drosophila* embryos where it randomly integrates into the genome. Transgenic flies are mapped and stocked. Flies carrying the transgenic construct are crossed to animals expressing heat-shock flippase and heat-shock I-SceI endonuclease. When the progeny are heat shocked at 39°C, expression of the flippase and I-SceI enzymes excises the construct from the genomic integration location (Figure 1.17). In a small percentage of cases the linearized fragment will recombine into the genome at the endogenous location via cellular DNA repair mechanisms. Loss of the transgene is scored by white ommatidia, while re-integration is scored by red ommatidia, together generating a mosaic-eyed animal.

To ensure that the *mini-white* expression is due to integration after recom-

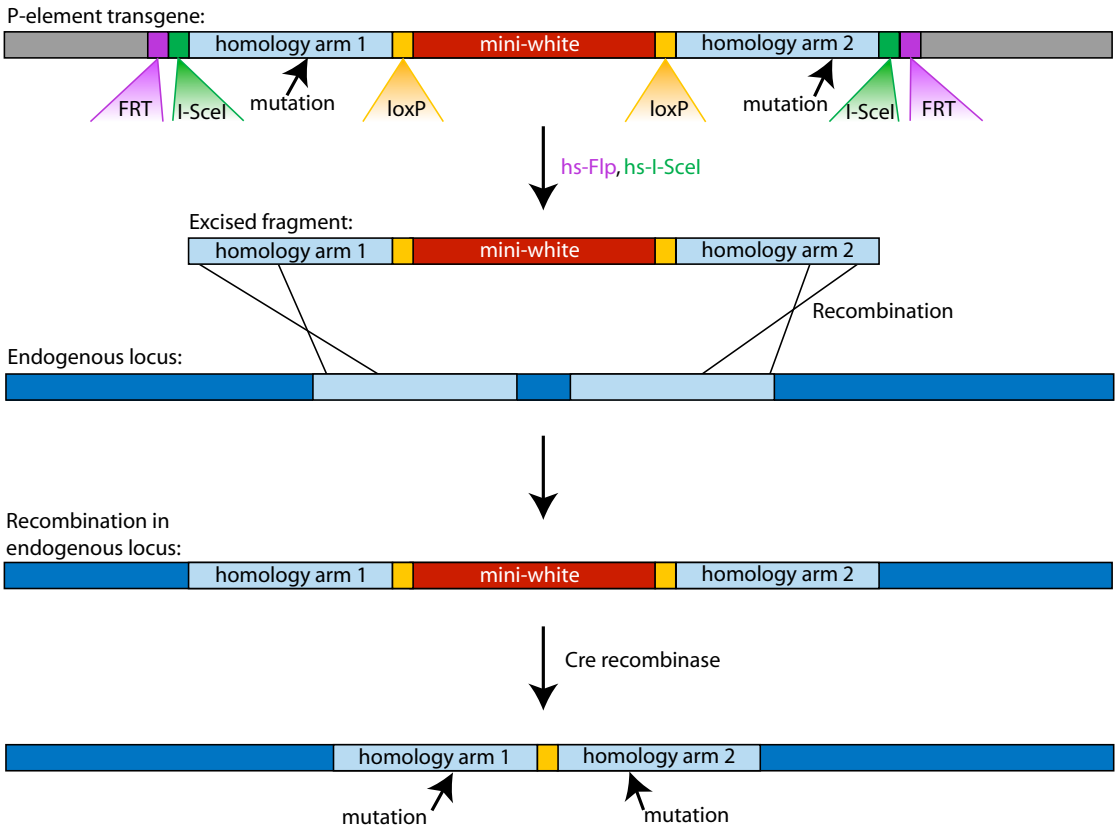


Figure 1.17: An overview of the homologous recombination process. A P-element transgene is engineered that contains two homology arms (light blue) flanked by restriction sites (not shown). Between the arms is a "*mini-white*" phenotypic marker (red) flanked by loxP sites (yellow). Arms are bracketed by FRT (flippase recognition target, purple) and I-SceI recognition sites (green). The desired mutations are introduced into the homology arms. After expression of heat-shock Flippase (Flp) and I-SceI recombinases, the construct is excised from the random genomic integration point and linearized. In a small percentage of cases, the fragment will recombine in the genome at the endogenous locus (blue). The addition of Cre recombinase, which specifically recognizes the loxP sequences, removes the *mini-white*, leaving behind a single loxP remnant of 76 bp, which is targeted to an intronic region of low sequence conservation.

bination and not because the construct failed to excise from the original genomic location, a flippase under control of an *eyeless* promoter (expressed only in the eye) is introduced. Progeny whose red eyes persist through the ey-FLP cross likely have successfully undergone homologous recombination at the endogenous locus. Finally, a Cre recombinase enzyme is introduced, which specifically recognizes the loxP sequences and excises the *mini-white* reporter (Figure 1.17), leaving behind a single loxP "remnant" of 76 basepairs in the genome. Loss of the red eye indicates successful Cre-mediated excision, and mutants are verified via PCR, restriction digestion, and Sanger sequencing.

The goal of HR is the integration of a desired mutation or combination of mutations into the endogenous locus in the genome. Control animals go through the same series of crosses as mutant lines and contain the wild type gene with the loxP insertion. For complete methodology see Staber et al. [2011]. Previously, assaying two mutations at once meant the design of a new HR construct, followed by injection and the complete HR procedure (Figure 1.17). However, I discovered that Cre recombinase swaps genomic material around loxP sites when two HR-generated alleles are placed *in trans*. Therefore, mutations that are on opposite sides of the loxP site (in different homology arms, Figure 1.17) may be easily combined through simple crossing methods. This technique, "Cre-mediated *trans* recombination," is outlined in **Appendix A**.

I used HR to introduce a series of precisely-designed mutations into the endogenous *Drosophila para* locus. These mutations were designed to probe the RNA structure(s) required to direct specific editing at three adenosines in exon 19 (**Chapter 2**). These structural mutations were also informative when I investigated the effect of temperature on RNA editing (**Chapter 3**), as the RNA structures ("thermistors") that control editing may be temperature-responsive. Finally, I assayed the effect of various engineered mutant alleles of *paralytic*, which are predicted to alter RNA

structure, on the interaction between the transcript, the Mle^{nap-ts} helicase, and RNA splicing (**Chapter 4**).

Chapter Two

Tertiary structural elements determine extent and specificity of messenger RNA editing

Parts of this chapter have been published as:

Rieder, LE, CJ Staber, B Hoopengardner, and RA Reenan. (2013) Tertiary structural elements determine extent and specificity of messenger RNA editing. *Nature Communications*, in press.

Dr. Rob Reenan and I conceived and designed experiments and I performed all experiments. Dr. Rob Reenan designed, constructed, and performed transgenesis of HR constructs. Ms. Cindi Staber assisted in HR construct transgenesis and mapping. Dr. Barry Hoopengardner cloned and sequenced species introns. Dr. Robert Reenan and I wrote the manuscript and I crafted the figures. All authors provided editorial feedback.

I would like to thank Dr. Yiannis Savva and Dr. James Jepson for the $dADAR^{hyp}$ and $dADAR^{5g1}$ *Drosophila* lines and for manuscript input, Mr. Matthew A. Reyna for assistance with figures, Ms. Rachel Whitaker for qRT-PCR guidance, Ms. Kasia Sierzputowska for Figure 2.6**e** and Dr. Erica Larschan for manuscript input. Dr. Barry Ganetzky supplied the $para^{lk5}$;;;para dup/Ci stock. This material is based upon work supported by a National Science Foundation Graduate Research Fellowship awarded to LER (No. DGE 0228243).

2.1 RNA editing of *paralytic*

The spatial folding of a given RNA *in vivo* poses an experimentally difficult but vital issue for key biological processes. Recognition of RNA substrates by binding proteins is believed to be largely dependent on secondary structure [Stefl et al., 2005], bestowed on these proteins by recognition motifs seeking RNA shapes rather than primary nucleotide sequence. Although recent evidence suggests that double-stranded (ds) RNA binding domains (RBDs) may be capable of primary sequence recognition via the shallow minor groove of duplex RNA [Masliah et al., 2012], three-dimensional RNA structure likely plays a major role in highly-specific protein-RNA recognition, supporting diverse cellular functions such as splicing, RNA interference, and translation. ADAR (adenosine deaminases that act on RNA) family members possess dsRBDs, recognize diverse RNA substrates, and catalyze the deamination of adenosines into inosines in a post transcriptional process known as A-to-I RNA editing [Savva et al., 2012]. Because the chemical properties of inosine mimic those of guanosine, cellular machines, including the ribosome, interpret inosine as guanosine. Therefore, RNA editing effectively causes an adenosine to guanosine change in the primary sequence of mRNAs, potentially recoding the resulting peptides.

A-to-I editing is thought to exist as a mechanism to expand protein repertoires within the nervous system [Stapleton et al., 2006, Hoopengardner et al., 2003, Rosenthal and Seeburg, 2012]. ADARs are strikingly conserved among metazoa [Nishikura, 2010, Savva et al., 2012] and engineered model organisms that lack ADAR activity display severe neurological phenotypes, including seizures and neurodegeneration [Higuchi et al., 2000, Palladino et al., 2000, Tonkin and Bass, 2003, Wang et al., 2000]. Furthermore, ADARs are localized to the neuronal nucleus [Savva et al., 2012] and nucleolus [Desterro et al., 2003], supporting the role of editing in nascent, immature RNA molecules. One exemplar *Drosophila* neuronal editing substrate is *paralytic* (*para*), the lone action-potential sodium channel gene in the fly genome [Ganetzky,

1984]. In contrast, nine voltage-gated sodium channel genes ($Na_v1.1$-$Na_v1.9$) are encoded in the mammalian genome [Meisler and Kearney, 2005]. Despite the fact that the *Drosophila* genome possesses only the *para* locus, *para* transcripts are extensively alternatively spliced and edited (Figure 2.1**a**). These post-transcriptional modifications, taken independently, result in over two million potential sodium channel isoforms [Hanrahan et al., 2000].

ADAR proteins typically contain at least one dsRBD, correlating well with the observation that edited adenosines are associated with dsRNA structures [Bass and Weintraub, 1987]. Nevertheless, very few structures that direct editing have been probed to reveal the extent to which three-dimensional properties contribute to editing specificity. Complicating this picture is the observation that the isolated ADAR catalytic domain, lacking dsRBDs, has significant activity on certain specific target dsRNAs [Eggington et al., 2011], and that mutations in the catalytic domain of human ADAR2 can result in altered editing site specificity [Kuttan and Bass, 2012]. At the RNA level, ADAR substrates range from perfectly complementary long RNA hairpins, which are promiscuously hyper-edited at up to 40% of adenosines [Wahlstedt and Ohman, 2011], to shorter (25 bp) imperfectly paired substrates with bulges and loops, which are specifically edited at one or a few adenosines [Nishikura et al., 1991].

The menagerie of specifically edited substrates includes simple short hairpins, such as the 71-nt stem-loop directing editing at the mammalian GluR-2 R/G site [Aruscavage and Bass, 2000], as well as comparatively enormous pseudoknotted structures, such as that found in the *Drosophila synaptotagmin-1* transcript, whose elements span thousands of nucleotides [Reenan, 2005]. Most RNA structures that recruit ADAR proteins involve exonic editing sites that interact via base pairing with complementary and conserved *cis* non-coding sequences from a nearby intron. Such imperfectly paired duplex structures direct modification to particular adenosines.

The diversity of ADAR targets provides a unique source for studying RNA-protein and RNA-RNA interactions.

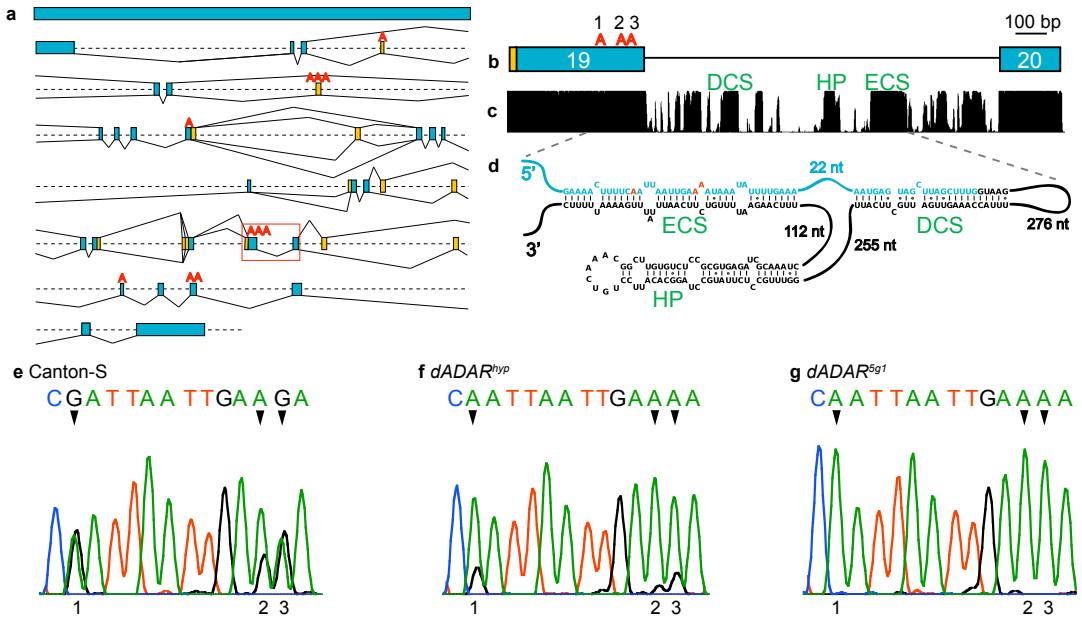


Figure 2.1: Organization of para exon 19 editing sites and ADAR dependence. **(a)** The *para* locus contains 26 constitutive exons (blue), 13 alternative exons (yellow) and 11 sites of RNA editing (red). The boxed region in **(a)** is enlarged in **(b)**, showing the three nearby editing sites in constitutive exon 19. **(c)** Sequence conservation of the region in **(b)** from 13 species of *Drosophilidae* found in the UCSC Genome Browser. Three regions of intronic homology, the ECS (editing site complementary sequence), DCS (donor site complementary sequence), and HP (hairpin) are putative *cis* sequences. **(d)** The indicated region of the *para* transcript from exon 19 (blue) and the downstream intron (black), folded in Sfold, show the conserved *cis* sequences denoted in **(c)** involved in double-stranded structures. Edited adenosines are shown in red. **(e)** Wild type (Canton-S) editing at the three sites in exon 19 is apparent as mixed A/G peaks in electropherograms from reverse-transcribed *para* mRNA. **(f)** Editing at these sites is sensitive to dADAR levels, as editing in a hypomorphic *dADAR* mutant ($dADAR^{hyp}$) are greatly reduced. **(g)** Editing is completely abolished in the $dADAR^{5g1}$ null mutant.

Containing eleven highly specific RNA editing sites (Figure 2.1**a**), *paralytic* mRNA provides an exceptional opportunity to study the enzyme-substrate interactions that direct RNA editing. However, previous *in vivo* studies of *paralytic* have only demonstrated an implied necessity for secondary structure [Reenan et al., 2000]. Here, we describe the *in vivo* molecular consequences of RNA structural perturbations, introduced via knock-in into the endogenous *para* locus in exon 19 and the downstream intron, using homologous recombination (HR) [Gong and Golic,

2003]. These mutations, designed to probe conserved *cis*-sequences in predicted RNA structures, establish the predicted necessity of imperfect double-strandedness around three nearby editing sites. We also demonstrate the presence of an accessory duplex, formed around the splicing donor site, which modulates an interaction between RNA splicing and the extent of editing. Finally, our data demonstrate the essential nature of a highly conserved tertiary pseudoknot that directs ADAR to only one of three nearby adenosines in the central duplex. Replacing the pseudoknot with a distinct yet functional natural example indicates the generality of the mechanism. Our data suggest that proximity of editing sites within a central duplex does not imply functional coupling. We show that extent and selectivity of editing can be determined by structural components assembled from non-local sequence elements.

2.2 Results

2.2.1 Evolutionary conservation and RNA structural predictions for an RNA editing site

Exon 19 in the *Drosophila paralytic* transcript possesses three edited adenosines within 12 nucleotides (Figure 2.1**b**), and is predicted to form dsRNA secondary structure (Figure 2.1**d**). An assay of *paralytic* RNA editing levels in wild type and mutant *Drosophila* can accurately be performed via RT-PCR and Sanger sequencing. Editing at the three adenosines (sites 1-3) in exon 19 is apparent as mixed A/G peaks in wild type electropherograms (Figure 2.1**e**). Editing at these adenosines is sensitive to levels of *Drosophila* ADAR (dADAR), as flies carrying a hypomorphic *dAdar* allele (approximately 80% dADAR reduction) [Jepson et al., 2011] show reduced editing levels (Figure 2.1**f**) and *dAdar* null males [Palladino et al., 2000] do not edit these sites (Figure 2.1**g**). In addition, the presence of editing at these three adenosines is conserved across *Drosophila* species [Reenan et al., 2000] (Figure 2.2).

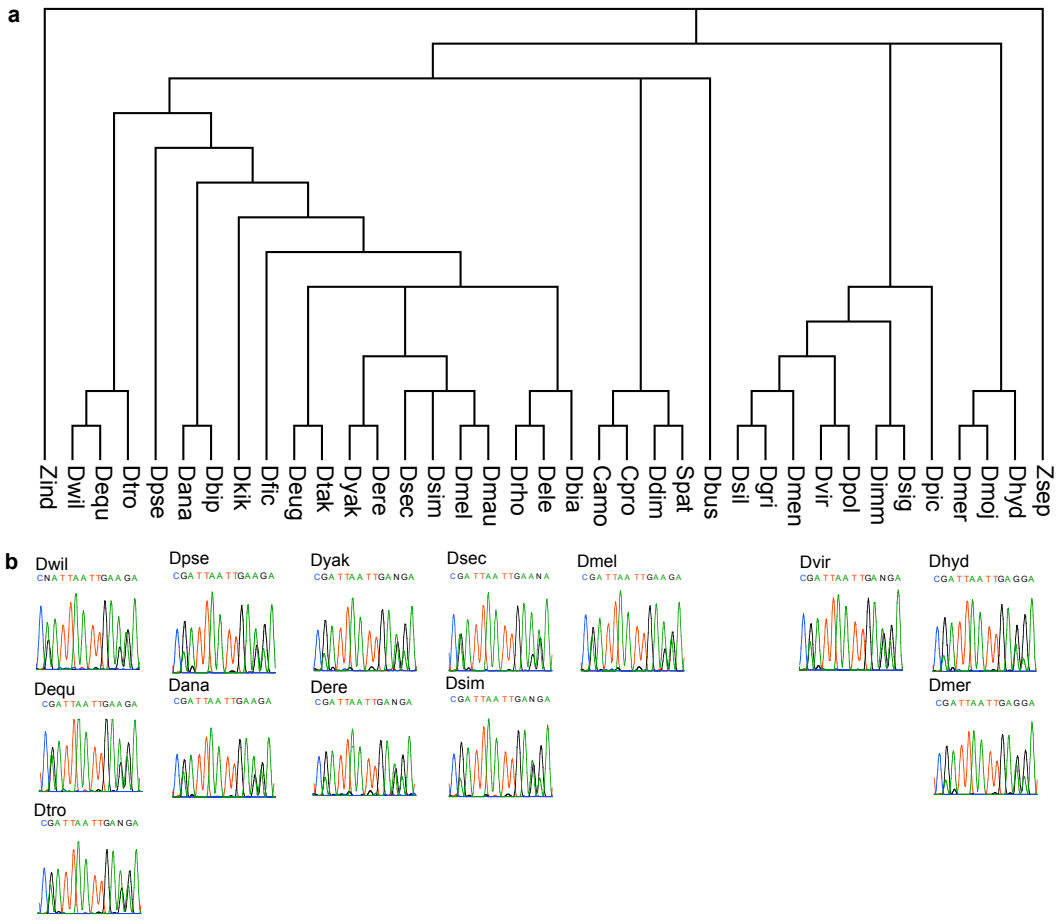


Figure 2.2: Paralytic exon 19 editing throughout *Drosophilidae*. **(a)** Cladogram of *Drosophilidae*. **(b)** Electropherograms of *para* exon 19 editing sites from species indicated in **(a)**. For full species names please see Figure 2.18.

While the previous study of these sites experimentally revealed the minimal region of primary sequence encompassing all intronic *cis*-elements (including an editing site complementary sequence or ECS) necessary for RNA editing, only two species were compared when making these predictions [Reenan et al., 2000]. We suspected that our analysis was far too limited to reveal more elusive conserved elements. Therefore, we obtained the genomic regions encompassing exon 19, intron 19, and exon 20 from 37 species of the genus *Drosophilidae* (Figure 2.3). Sequence alignments revealed several areas of extreme intron conservation in addition to the expected ECS element (Figure 2.1**c**, Figure 2.3). When we used the RNA folding algorithm Sfold [Ding et al., 2004] to predict dsRNA features of the *paralytic* transcript in this region, three striking features were predicted in the ensemble of all structures, and corresponded to regions of high sequence conservation (Figure 2.1**c**, **d**).

The expected central ECS-directed duplex feature places the three exonic editing sites within a 77 bp-long imperfect duplex formed between the exon and the region of the downstream intron with homology to the edited region (ECS; Figure 2.4**a**). In addition, a highly invariant region of the intron, which we call the donor-site complementary sequence (DCS; Figure 2.4**b**), is predicted to form a secondary duplex RNA structure encompassing the 5' splice donor site. Finally, a predicted intronic local hairpin (HP) of unknown function is highly conserved in all species surveyed (Figure 2.4**c**).

2.2.2 Knock-in mutagenesis demonstrates the requirement for dsRNA directing RNA editing *in vivo*

Traditional experimental manipulation of RNA editing structures are usually conducted *in vitro* [Daniel et al., 2012] with purified components or through the use of synthetic mini-gene reporters in tissue culture [Tian et al., 2011, Bhalla et al., 2004]. Many of these studies are capable of generally recapitulating editing and

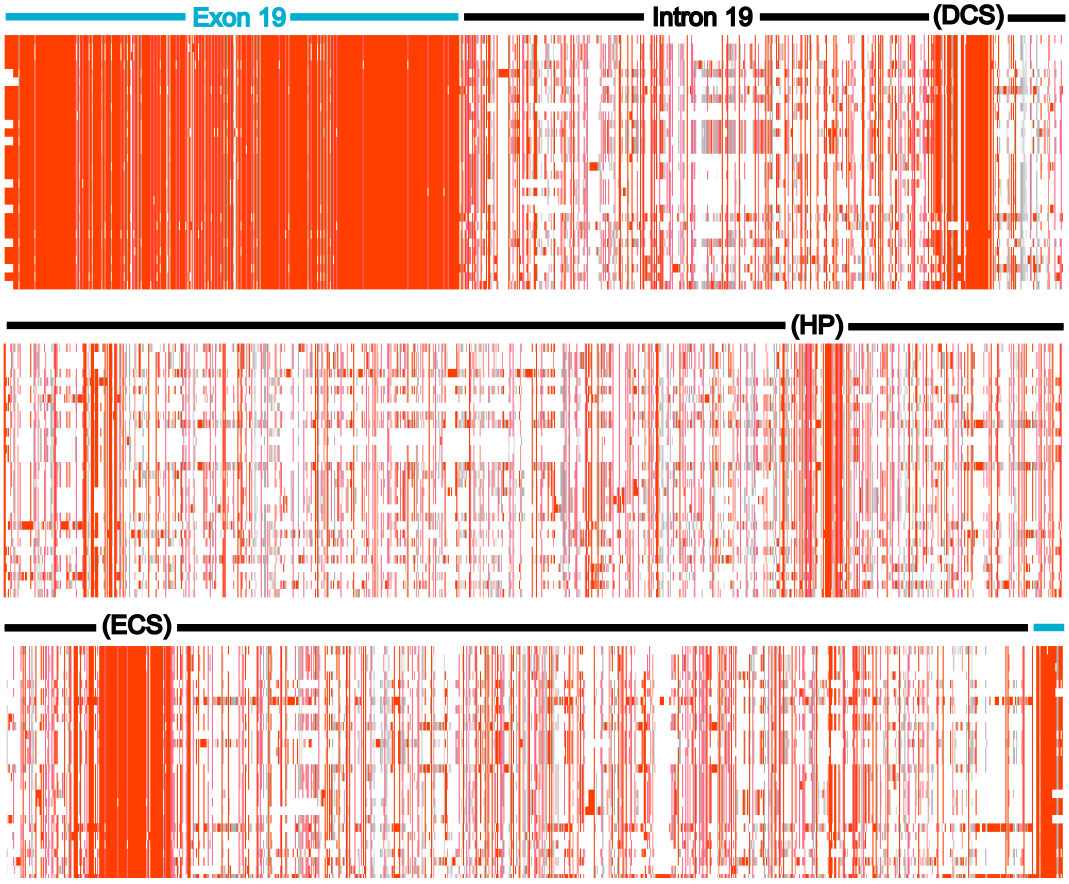


Figure 2.3: Genomic sequence conservation of *paralytic* intron 19 throughout *Drosophilidae*. Red indicates complete base conservation while pink represents near conservation, orange some conservation, and gray no conservation. White dashes indicate inserted nucleotides. Putative *cis* elements are denoted in parentheses. For complete ordered species list please see Figure 2.4**a**. Exonic sequence is denoted in blue while intronic is in black.

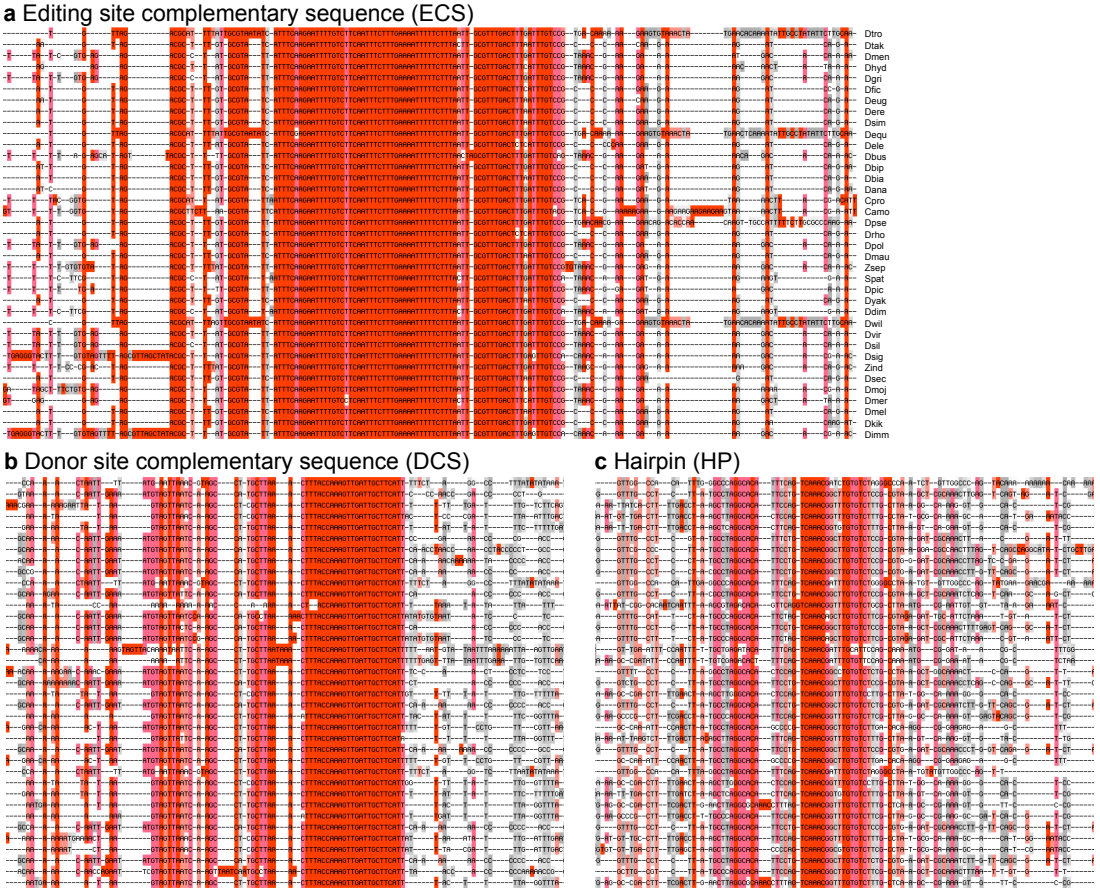


Figure 2.4: Nucleotide resolution of *paralytic* intronic *cis* element conservation. Color-coding as in Figure 2.3. **(a)** Editing site complementary sequence (ECS) and *Drosophilidae* species abbreviations. For full species names please see Figure 2.18. **(b)** Donor site complementary sequence (DCS). **(c)** Hairpin (HP).

delineating the most important *cis* elements. Model genetic systems have explored the effects of altered ADAR levels *in vivo*, but provide little detail about the fine structure of ADAR targets. We chose to use ends-out homologous recombination (HR) to study the *in vivo* details of RNA editing, where precise introduction of various specific mutations into the endogenous *Drosophila paralytic* locus [Staber et al., 2011] preserves endogenous gene regulation.

Briefly, the process of HR introduces mutations into the genome, as well as a 76 bp LoxP remnant directed to a region of the intron with low sequence conservation [Staber et al., 2011]. Specifically, we targeted the LoxP remnant to the non-conserved region of the intron between DCS and HP elements (Figure 2.5**a**). The initial targeting construct encompasses a selectable *mini-white* marker gene flanked by LoxP sites, unique restriction sites, and two homology arms into which the desired mutations are engineered (Figure 2.5**b**). After targeting, introduction of a Cre recombinase removes the *mini-white* gene between the LoxP sites, leaving a single LoxP remnant. A wild type construct is also targeted containing no mutations in the *para* locus, providing the appropriate control for all other mutants created through HR (Figure 2.5**c**, Figure 2.6**a**, Figure 2.9**a**).

We first assessed the role of the predicted ECS by precisely excising the 39 intronic nucleotides predicted to pair with the edited region. As expected, when the *para* ECS is deleted from intron 19, editing at exon 19 is completely abolished (Figure 2.5**d**) compared to the LoxP control (Figure 2.5**c**). This suggests that the ECS directs the formation of a central duplex RNA structure necessary for editing at all three adenosines. Although all three edited adenosines require the ECS, editing at only the first and third sites result in amino acid substitutions in the sodium channel peptide (Glu to Arg, Asn to Asp, respectively). We next mutationally hardwired the first and third edited adenosines to guanosines, mimicking complete ADAR modification at these sites (Figure 2.5**e**). This mutation (GAG) restores predicted base pairing at site

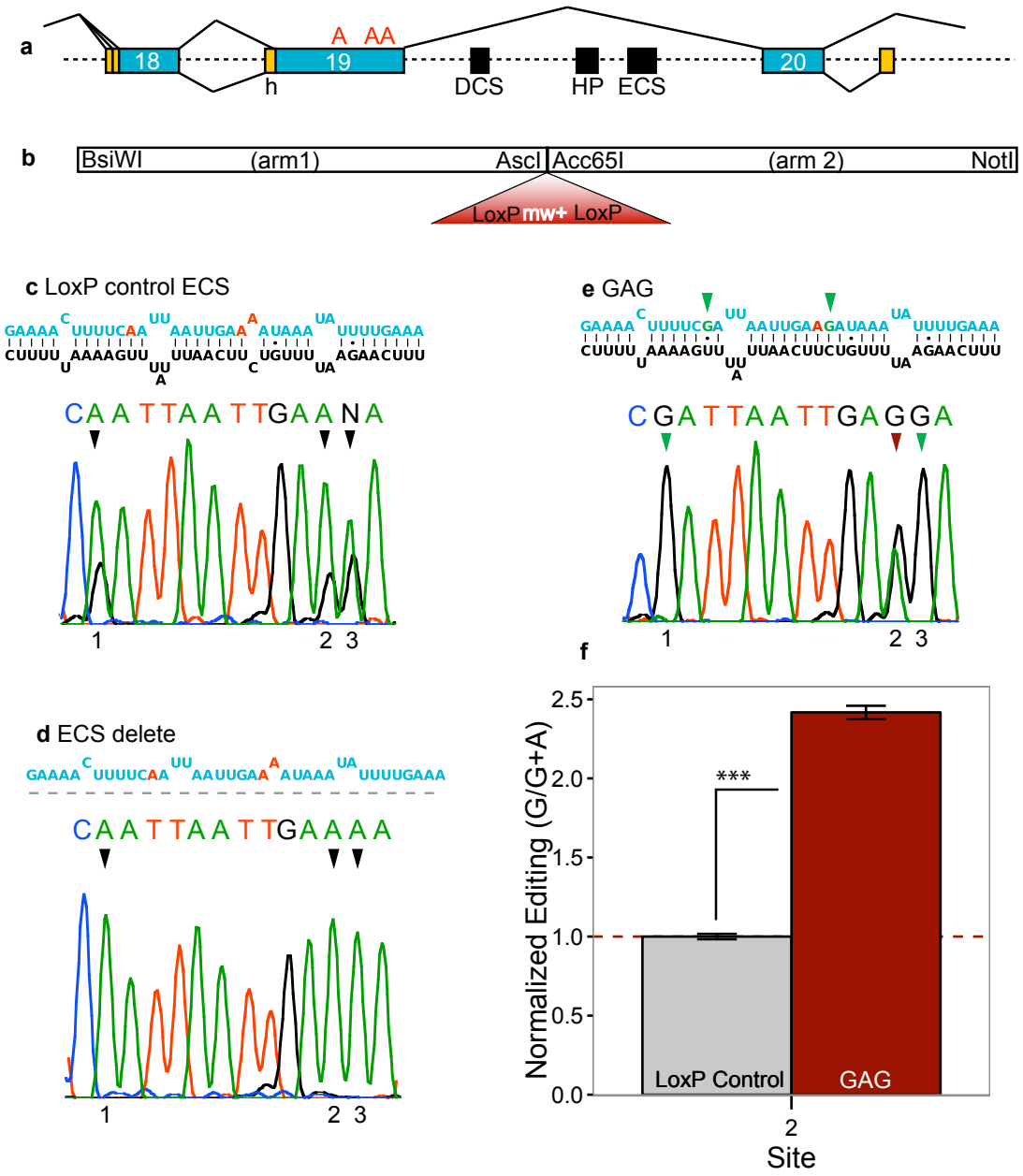


Figure 2.5: Editing mutations introduced by homologous recombination (HR) demonstrate necessity of central duplex. **(a)** The region of the *para* locus targeted by HR encompasses exon 19 and the putative cis elements (black). **(b)** The targeting construct for HR involves two arms (white), homologous to the endogenous para locus and bordered by unique restriction sites. A *mini-white* reporter gene flanked by LoxP sites (red) is situated between the arms and targeted to a region in the intron of low sequence conservation. **(c)** As a proper control, a construct containing wild type arms is targeted to create the LoxP control line, which is wild type except for the LoxP remnant in the intron. Editing at the three sites occurs in this control. **(d)** When the 39 nt ECS intronic *cis* sequence is deleted through HR (ECS delete), editing is completely abolished at all three adenosines. **(e)** The GAG mutation replaces the non-silent first (Glu > Arg) and third (Asn > Asp) adenosines with guanosines (green), resulting in sodium channel peptides representing only the completely edited form. **(f)** The GAG mutation (red) causes editing at the silent site 2 to increase compared to the LoxP control (gray). Bars represent standard error: $***$ $P < 0.0001$, please see Figure 2.17 for specific P values.

3 (versus an A-C mismatch, Figure 2.5**c**), resulting in a dramatic 2.5 fold increase in editing at synonymous site 2 (Figure 2.5**f**). These data are consistent with previous observations [Reenan et al., 2000], which suggest that editing of site 2 is coupled to editing at site 3. In addition, ADAR enzymes prefer editing adenosines neighbored by a 3 guanosine and a 5' adenosine [Lehmann and Bass, 2000]. Adenosines mismatched to a cytosine are also favored [Wong et al., 2001], suggesting that editing at site 3, an ideal dADAR target, precedes editing at site 2, which may occur as a sequential byproduct under little or no selective pressure. These results suggest that the GAG mutation improves the context of site 2 within the ECS/editing site duplex, resulting in increased editing levels.

2.2.3 Structural occlusion of splicing signals modulates RNA editing

A highly conserved intronic sequence (DCS) downstream of the 5' splice donor of intron 19 was predicted to generate a separate dsRNA structure encompassing the 5' splice donor in wild type *Drosophila* (Figure 2.6**a**). We therefore used HR to precisely delete the DCS (24 nt) (Figure 2.6**b**). In "DCS delete" alleles of *para*, editing is reduced uniformly but not eliminated at all three sites compared to the LoxP control (Figure 2.6**d**, Figure 2.13**b**).

We considered the possibility that the wild-type DCS duplex retards splicing, thereby enhancing editing, since the ECS is within the inron and more efficient splicing may disfavor editing by removing this crucial *cis* element. Thus, we designed a mutation (DCS zip) to extend the DCS duplex by 9 bp, which we hypothesized would energetically stabilize the DCS duplex. The insertion of the DCS zip mutation into the *para* intron (Figure 2.6**c**) resulted in substantially increased editing at all three sites (Figure 2.6**d**, Figure 2.13**c**). Our data support *in vitro* experiments documenting the interplay between editing and splicing [Rieder and Reenan, 2011].

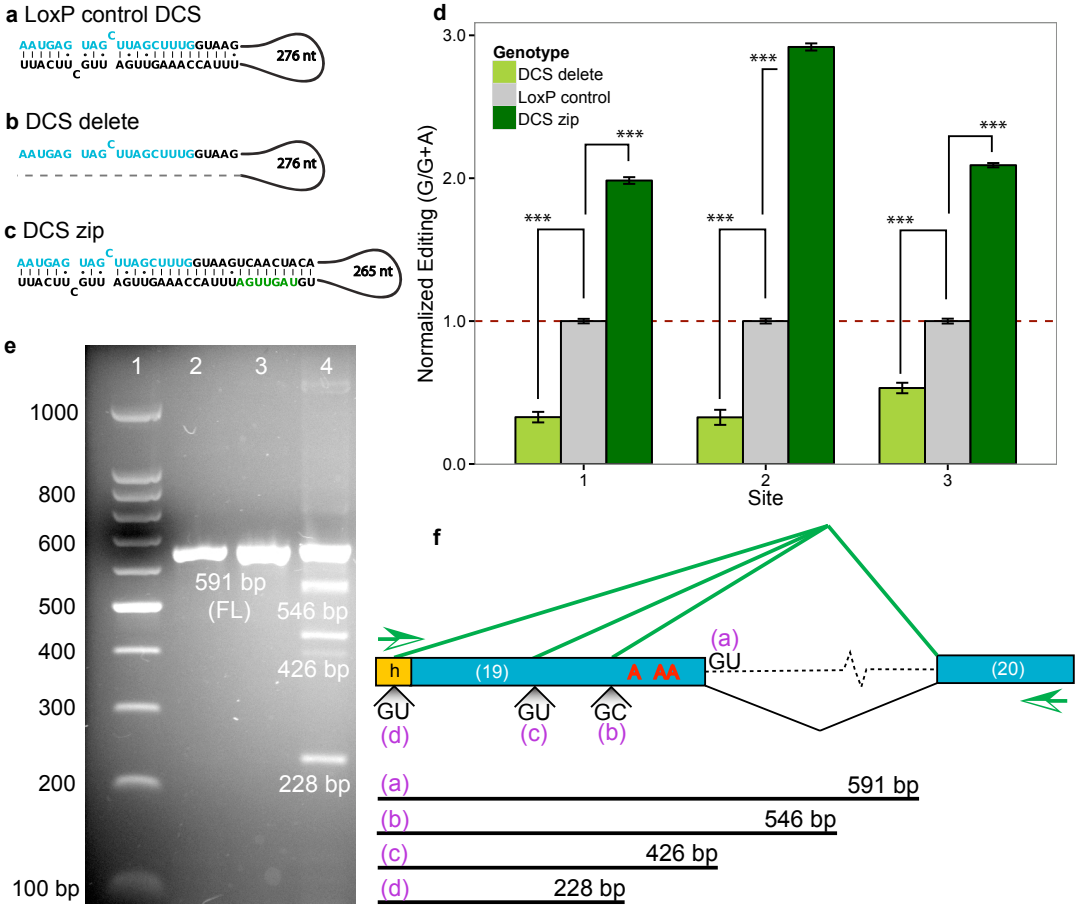


Figure 2.6: Mutations in the donor site complementary sequence (DCS) effect editing and splicing. **(a)** In the LoxP control, the wild type DCS intronic sequence is predicted to involve the donor site in secondary structure. **(b)** The "DCS delete" mutation excises the conserved 24 nt DCS *cis* sequence from the intron. **(c)** Seven intronic nucleotides (green) are added to the wild type DCS ("DCS zip"), and are predicted to result in a more stable secondary structure around the splice donor. **(d)** Data for each site are normalized to editing in the LoxP control. Editing at all three sites significantly decreases in the DCS delete *Drosophila* lines compared to the control, while editing significantly increases in animals homozygous for the DCS zip allele. Bars represent standard error ($***$ $P < 0.0001$, please see Figure 2.17 for specific P values). **(e)** PCR was performed on cDNA reverse-transcribed from *para* mRNA. Not all *para* transcripts from the DCS zip allele are full-length (FL). Lane 1, 100 bp ladder; lane 2, LoxP control; lane 3, DCS delete; lane 4, DCS zip. Sizes of bands are indicated on the 1.5% agarose gel. **(f)** Aberrant bands from **(e)** were isolated and sequenced to determine the activated cryptic donor sites within constitutive exon 19 (blue) and alternative exon h (yellow). Primers (green) in exons h and 20 produce the full-length PCR product (a) from the wild type splice donor. Aberrant shorter products (b-d) are formed due to activation of cryptic alternative donor sites in exons h and 19.

When the splicing machinery recognizes the donor site, the intron is removed and further processing, maturation, and export should decrease the likelihood that dADAR recognizes the transcript due to loss of double-strandedness. Further, we observed that para transcripts with a wild type intron or deleted for the DCS were properly spliced, while those from the DCS zip allele exhibited improper splicing, generating mRNA size variants (Figure 2.6**e**). Sequencing of aberrant products from the DCS zip animals revealed that they result from cryptic splice donor site activation from within constitutive exon 19 and the upstream alternative exon h (Figure 2.6**f**). All of the sequenced variants are predicted to result in nonsense mutations and, if translated, in substantially truncated non-functional sodium channel proteins. These observations strongly support the notion that the DCS *cis* element is involved in a dsRNA sequestering the splice donor, and that this structure modulates RNA editing levels in the central dsRNA duplex containing sites 1-3 by orchestrating the pace of intron removal by the splicing machinery.

We performed quantitative RT-PCR on RNA extracted from DCS zip and LoxP control animals using primers specific to the *para* spliceoforms detectable by conventional PCR (Figure 2.7**a**). This analysis revealed a significant decrease in the full length spliceoform in the DCS zip mutants compared to the LoxP control, while the presence of several aberrant spliceoforms was detectable and significantly increased in the mutant animals (Figure 2.7**b**). In addition, the DCS zip allele also gives rise to significantly more total *para* transcript than the LoxP control allele. These data suggest that transcription from the *para* locus is upregulated in DCS zip animals, perhaps in attempt to compensate for low levels of properly spliced para transcript and Para sodium channel protein.

The *para* locus encodes the only action potential-generating sodium channel gene in *Drosophila* [Ganetzky, 1984], and excitability of the fly nervous system is highly sensitive to *para* dosage [Stern et al., 1990]. Animals hemi- or homozygous

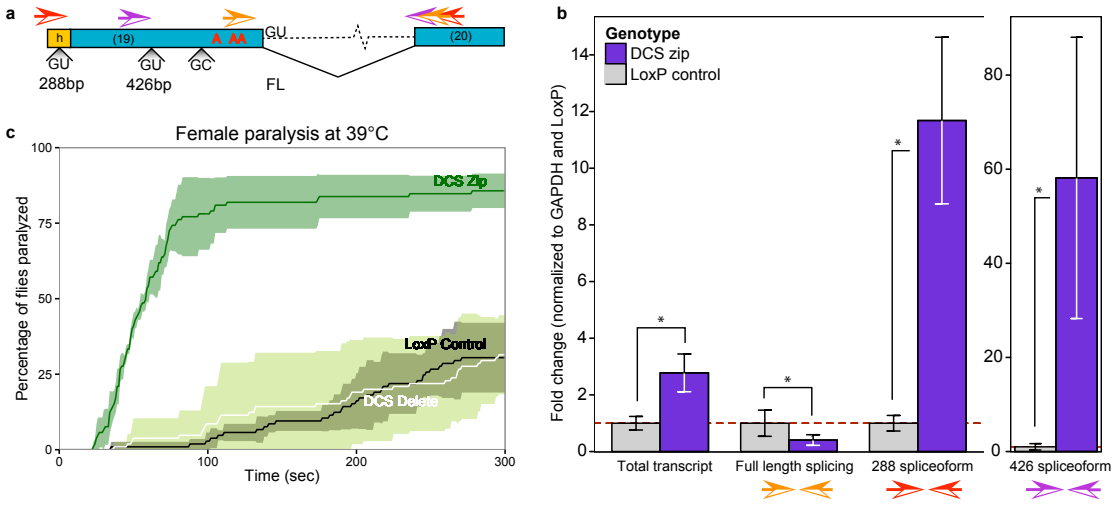


Figure 2.7: Consequences of aberrant splicing from the DCS zip allele **(a)** Location of qRT-PCR primers. Orange primers pick up full-length spliceoforms, while the red and purple primer pairs capture the 288bp and 426bp spliceoforms, respectively. **(b)** Quantitative RT-PCR shows total *para* transcript, detected by a primer pair spanning the 3 most exon-exon junction of the mature mRNA, is significantly increased from the DCS zip allele (purple) compared to the LoxP control (gray) (P=0.0228), while full- length transcript is significantly decreased (P=0.0493). The 288bp and 426bp aberrant spliceoforms are both significantly increased in the DCS zip animals (P=0.0404 and P=0.0037, respectively). Primers are as in **(a)**. The 546bp spliceoform detectable by conventional PCR (Figure 2.6**e**) is not detectable via qRT-PCR. Each bar represents three biological replicates and three technical replicates of each biological replicate. **(c)** Female flies were treated as males in Figure 2.8**a**. Flies homozygous for the DCS zip allele paralyze significantly faster than those of the LoxP control genotype ($P < 0.0001$). $N = 105$ flies per genotype. Ribbons represent standard deviations.

for temperature-sensitive (ts) *para* alleles, which reduce *para* expression, display ts paralytic phenotypes, while null alleles of *para* result in unconditional lethality. Therefore, we assessed the phenotypic consequences of altered *para* splicing in our DCS zip mutant. *Drosophila* hemi- or homozygous for the DCS zip mutation show a significant ts-paralytic phenotype (Figure 2.8**a**, Figure 2.7**c**), similar to other $para^{ts}$ alleles. To investigate the dosage of full-length functional *para* transcript originating from the DCS zip allele (Figure 2.6**c**), we crossed females homozygous for the DCS zip allele to males null for endogenous *para* on the X chromosome, but with a rescuing copy of *para* in a duplication translocated to the fourth chromosome. The lethality of female progeny from this cross demonstrates that the DCS zip allele is haploinsufficient. However, female siblings with an extra dose of the wild type *para* locus carried on the fourth chromosome duplication were rescued from this inviability (Figure 2.8**b**).

Interestingly, although males contain only one X chromosome, males hemizygous for the DCS zip allele are viable, likely due to the upregulation of genes on the X chromosome through the process of dosage compensation [Larschan et al., 2011]. In summary, we show that the extent of the DCS/donor site duplex strongly modulates the degree of RNA editing in a generic manner. The level of editing of all sites in the central duplex can be tuned up or down, and we propose that this is an effect modulated by the pace of splicing *in vivo*. While even subtle changes to intronic sequences predicted to strengthen this duplex enhance editing at the molecular level, they do so at the cost of incurring a substantial and presumably deleterious phenotypic consequence at the organismal level.

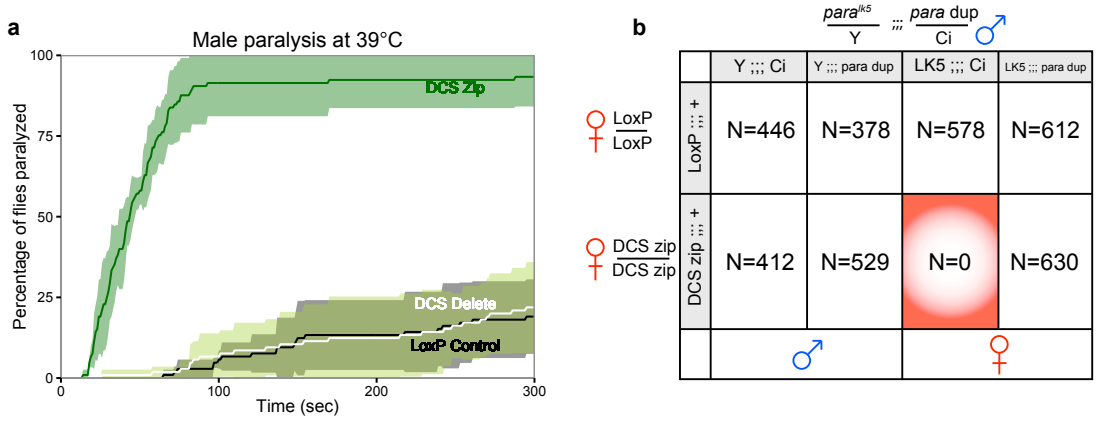


Figure 2.8: Phenotypic consequences of aberrant splicing from the DCS zip allele **(a)** Male flies of the indicated genotypes were placed in room temperature glass vials and then submerged in a circulating 39°C water bath. Time to first paralysis event was recorded over 5 minutes. Flies hemizygous for the DCS zip allele paralyze significantly faster than those with the LoxP control allele ($P < 0.0001$). $N = 105$ flies per genotype. Ribbons represent standard deviations. **(b)** Virgin females homozygous for the LoxP control or DCS zip alleles were crossed to males with a *para* deletion on the X chromosome ($para^{lk5}$) rescued by a wild type chromosomal duplication on the fourth chromosome (*para* dup). Allele contributions from each parent are shaded. No female progeny with only one copy of the DCS zip *para* allele were found (red shaded box), although this lethality is rescued by the wild type *para* duplication on the fourth chromosome. Lethality is not observed in male progeny hemizygous for the DCS zip allele, a phenomenon that is likely due to dosage compensation of the male X chromosome.

2.2.4 The hairpin structure is required for selective editing of one adenosine

The highly conserved intronic hairpin (HP) structure (Figure 2.9**a**) has no obvious relationship to either the ECS or DCS duplexes, and is without precedence in other known ADAR targets. Surprisingly, when the hairpin is deleted through HR editing is preferentially abolished at site 1, while editing at sites two and three remain near that of the LoxP control (Figure 2.9**g**, Figure 2.13**d**). Although the precise nucleotide sequence of the hairpin stem differs between species, comparative sequence analyses revealed extensive co-variation of stem sequences that maintain base pairing (Figure 2.10**a**, Figure 2.10**b**), suggesting that the structure (Figure 2.9**b**) rather than primary sequence is necessary for editing. In particular, two unpaired bulged regions of the HP stem are almost invariably pyrimidines, and the HP stem resembles a barbell. We therefore introduced the "HP zip" mutation, in which some of the bulged pyrimidines were paired, converting the HP stem into a more perfect duplex (Figure 2.9**c**). In this case, editing was severely reduced at the first adenosine, while editing at sites 2 and 3 occurred at control levels (Figure 2.9**g**, Figure 2.13**e**).

The HP stem is predicted to adopt a barbell structure due to the symmetric pyrimidine bulges (Figure 2.9**b**). While this structure is preserved in most species of *Drosophilidae*, as divergent as *D. melanogaster* (Figure 2.10**a**) and *D. melanica* (Figure 2.10**b**), *D. willistoni* subgroup members lack the predicted HP barbell due to stem mutations that pair the second group of pyrmidines (Figure 2.10**c**). Species within this subgroup therefore may provide an example of natural variation of the HP that resembles the HP zip synthetic allele (Figure 2.9**c**). We therefore compared editing at site 1 to editing at site 3 in species across *Drosophilidae* (Figure 2.2). Editing at site 1 compared to site 3 was significantly decreased in two closely related species, *D. willistoni* and *D. equinoxialis*, which lack the canonical HP barbell shape (Figure 2.10**c**), compared to *D. melanogaster* (Figure 2.10**d**). These data further

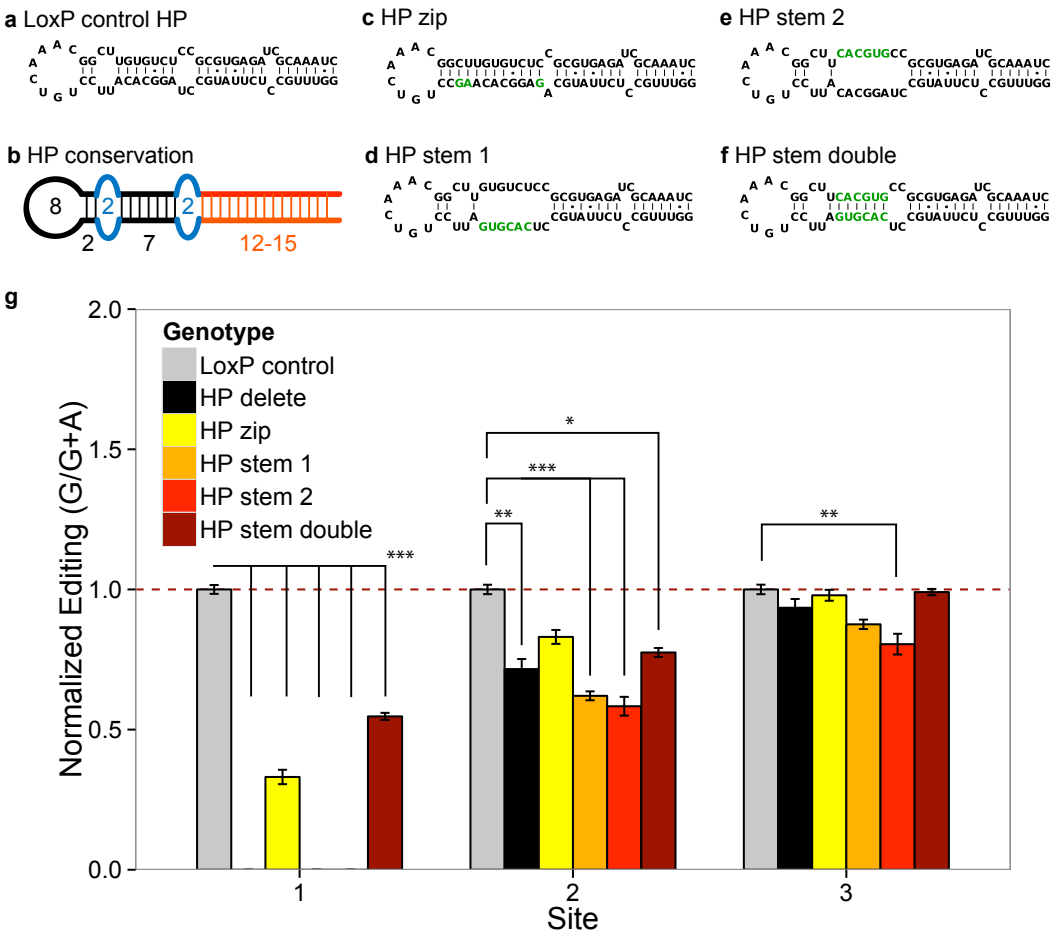


Figure 2.9: Editing at site 1, within the central duplex, requires the hairpin element. **(a)** The highly conserved predicted wild type hairpin (HP) element. **(b)** Conservation of the HP across *Drosophilidae*. Black regions of the structure are conserved, while red are variable. Blue represents conserved pyrimidines bulges. Numbers refer to the nucleotides in the loops or basepairs in the stem. **(c)** The HP zip mutation pairs some of the bulged, conserved pyrimidines (green) into the HP stem. **(d)** The HP stem 1 mutation introduces a destabilizing PacI site into one side of the stem, while **(e)** the HP stem 2 mutation introduces the PacI mutation from the opposite side. **(f)** Combining the HP stem 1 and 2 mutations (HP stem double) restores predicted base pairing in the HP stem. **(g)** Data for each site are normalized to editing in the LoxP control (gray). Deleting the HP (black) results in a selective loss of editing at site 1, while pairing the pyrimidine bulges (HP zip; yellow) results in a reduction of editing at site 1 compared to the LoxP control. Both the HP stem 1 and 2 mutations (orange and red, respectively) selectively abolish editing at site 1, while the restorative double mutation (HP stem double; brown) rescues editing at site 1. Bars represent standard error ($* \ P < 0.05$, $** \ P < 0.001$, $*** \ P < 0.0001$, please see Figure 2.17 for specific P values).

suggest that the HP stem shape serves to modulate RNA editing at site 1.

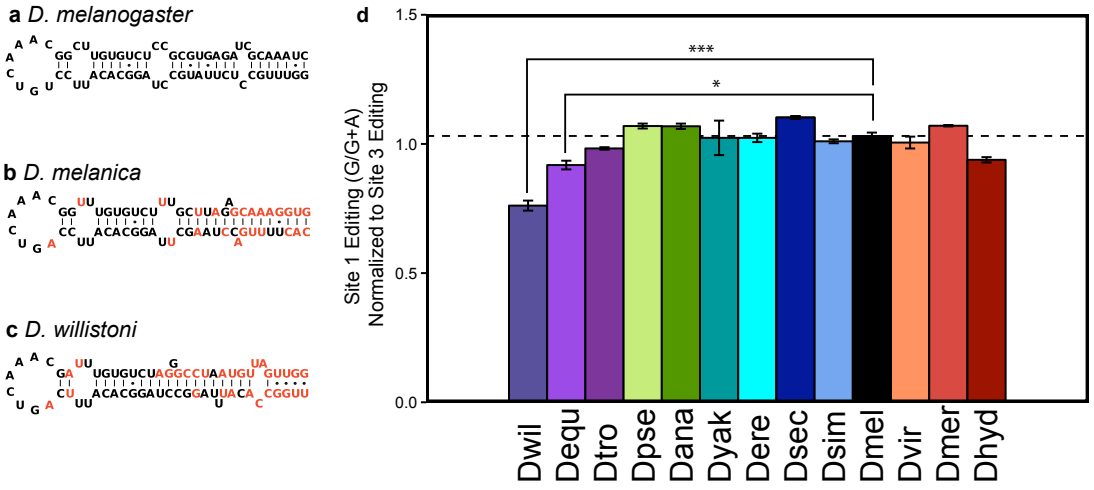


Figure 2.10: Hairpin conservation and editing consequences across *Drosophilidae*. **(a)** The predicted HP structure in *D. melanogaster*, compared to **(b)** that of *D. melanica*, which also contains the "barbell" shape due to bulged pyrimidines. In contrast, the HP of *D. willistoni* **(c)** does not have the predicted barbell. Nucleotides that differ from **(a)** are shown in red in **(b)** and **(c)**. **(d)** Editing of site 1 normalized to site 3 across species. The site 1 to site 3 ratio in *D. melanogaster* is represented by the black dotted line. This editing ratio is significantly decreased in two closely related species, *D. willistoni* and *D. equinoxialis*, compared to *D. melanogaster*. For species abbreviations please see Figure 2.18. Bars represent standard error ($* P < 0.05$, $*** P < 0.0001$, for exact P values please see Figure 2.18).

To test the notion that the duplex HP stem is necessary for editing at site 1, we introduced helix-disrupting mutations in the 5' or 3' side of the stem (Figure 2.9**d**, **e**). These mutations, "HP stem 1" and "HP stem 2," both selectively abolish editing at site 1, while restoring the base pairing by combining these two compensatory mutations ("HP stem double," Figure 2.9**f**) rescues editing at the first adenosine (Figure 2.9**g**, Figure 2.13**f-h**). Thus, editing at site 1, residing in a separate duplex distant from the HP, requires both the presence of the HP as well as the specific paired and unpaired bases within that structure. This is particularly interesting considering that two other adenosines only ten nucleotides away in the central duplex appear to be unaffected by the presence or integrity of the HP.

2.2.5 A long-range tertiary pseudoknot is mediated by the hairpin

A closer inspection of both the HP primary sequence as well as its potential to contribute to other structural elements was revealing. First, while the HP stem covaries between *Drosophila* species, the HP loop is completely invariant (Figure 2.9**b**, Figure 2.4**c**). Secondly, a short sequence just 3' to the ECS *cis* element is also completely conserved, was not predicted to base pair in standard folding predictions, and creates an exact docking site for the hairpin loop sequence–a potential tertiary pseudoknot (PK) interaction (Figure 2.12**a**). Such interactions are very difficult to predict with current RNA folding software programs except for in the smallest input sequences. Furthermore, there is no precedence for the absolute requirement of a long distance tertiary structural element orchestrating specific RNA editing in a different dsRNA helix.

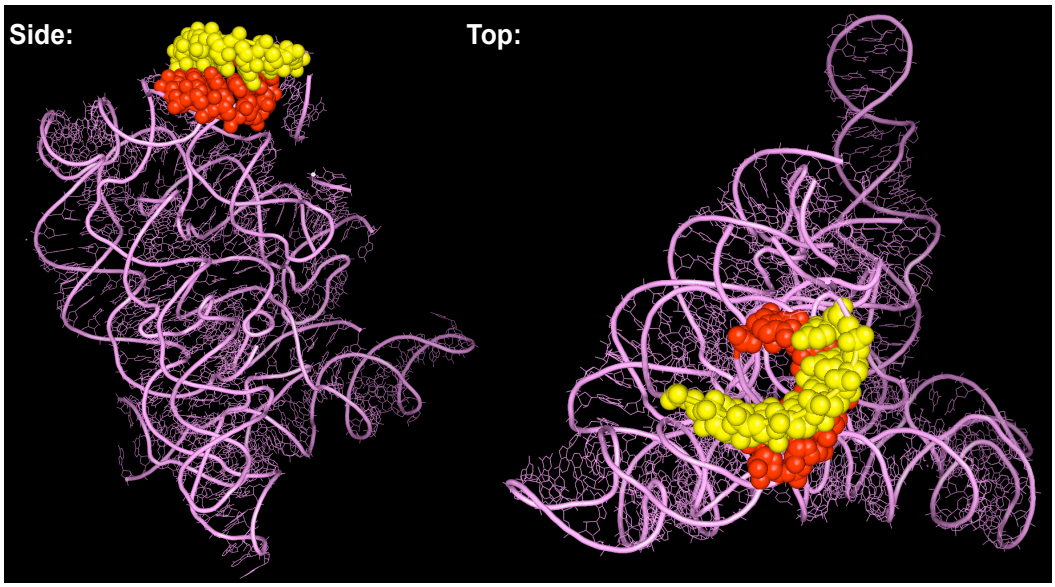


Figure 2.11: The group II self-splicing intron α/α' kissing loop interaction. Figures created with MacPyMOL freeware. Loop and dock nucleotides are colored yellow and red, respectively.

We were struck by the similarity of the predicted *para* PK to a structural ele-

ment of the group II self-splicing intron from the extremophilic bacterium *Oceanobacillus iheyensis* [Toor et al., 2008]. In this autocatalytic intron, the α/α kissing loop interaction is required to scaffold the formation of a five-way helical junction, ensuring the proper formation of a three-dimensional structure necessary for self-catalysis (Figure 2.11) [Harris-Kerr et al., 1993]. The primary sequence of the α/α kissing loop interaction differed from that of the *para* hairpin loop/dock interaction by only three nucleotides. We used HR to interconvert the hairpin loop sequence to the α sequence from the kissing loop interaction (Figure 2.12**b**) and the docking site into that of the α (Figure 2.12**c**). We found that these mutations, designed to disrupt the putative tertiary *para* PK, independently, selectively, and completely abolish editing at *para* site 1 (Figure 2.12**e**, Figure 2.13**i-j**). Combining the mutations in the double mutant, which recapitulates the naturally occurring α/α kissing loop interaction of the group II self-splicing intron within the *paralytic* intron (Figure 2.12**d**) restores editing at the site one adenosine (Figure 2.12**e**, Figure 2.13**k**). Thus, our data strongly suggest that not only does the tertiary pseudoknot exist *in vivo*, but it appears to use a structural variant similar to a natural example and serves to selectively direct the RNA editing enzyme to specifically deaminate only one of three nearby edited adenosines in the central duplex.

2.3 Discussion

2.3.1 Complex and novel structural requirements for RNA editing

The RNA informational complement of only four building blocks, amplified by simple base pairing rules, nevertheless allows for an almost infinite variety of secondary and tertiary structures, surfaces, and chemistries. *In silico* approaches such as comparative genomics and structural predictions are very useful in identifying important base

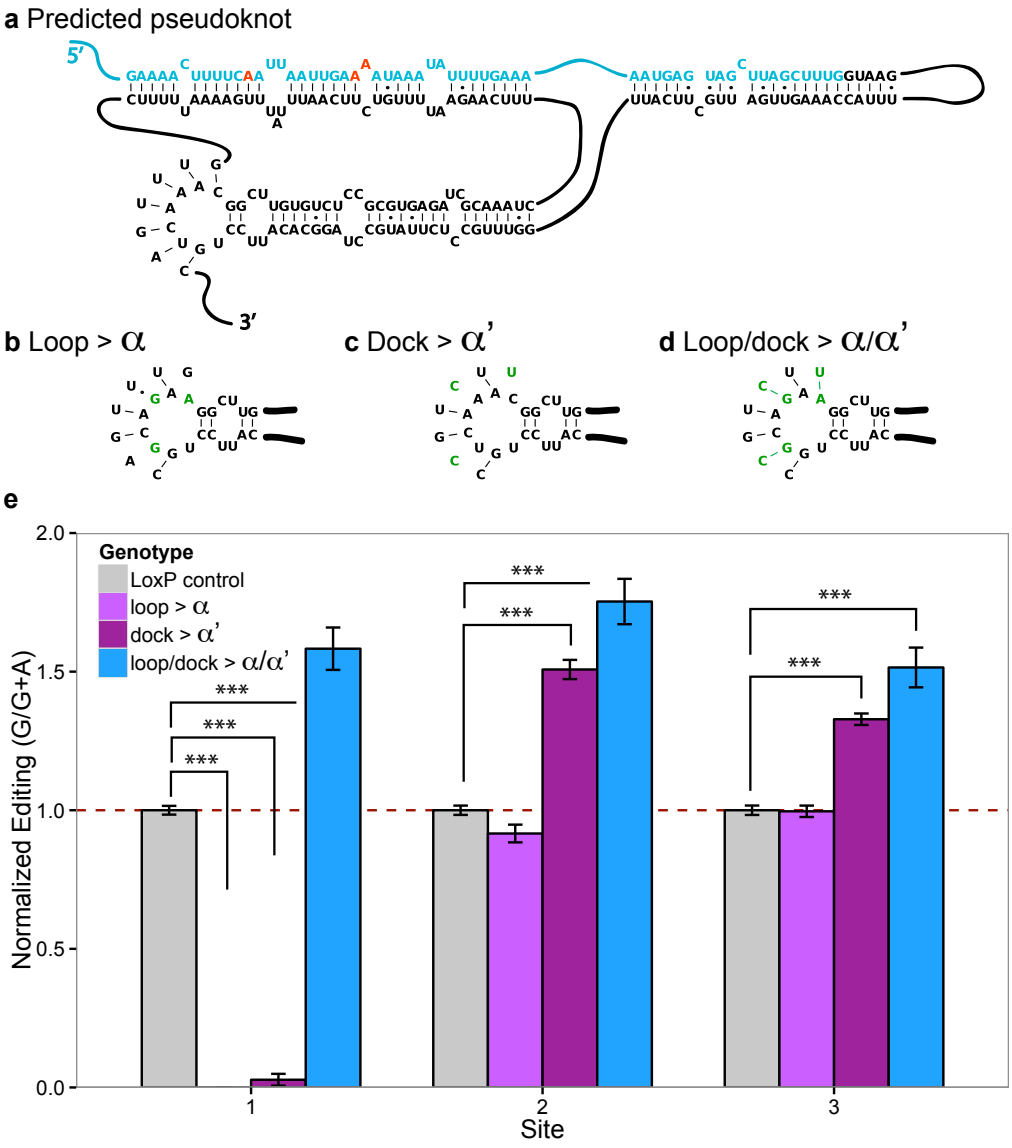


Figure 2.12: The hairpin is involved in a tertiary pseudoknot that directs selective editing in the central duplex, and can be functionally replaced. **(a)** The predicted intronic tertiary pseudoknot is composed of seven base pairs. Exon, blue; intron, black; editing sites, red. **(b)** The "Loop > α" mutation alters three nucleotides (green) in the HP loop so that the sequence recapitulates the α sequence from the group II self-splicing intron. **(c)** The "Dock > α' " mutation introduces three nucleotide changes (green) in the docking site of the predicted *para* pseudoknot, altering the sequence into that of the self-splicing intron α' sequence. **(d)** The double "Loop/dock > α/α' " mutation restores the predicted pseudoknot interaction using the α and α' sequences (green) from the group II self-splicing intron kissing loop interaction. **(e)** Editing is normalized to that in the LoxP control (gray). Both the "Loop > α" and the "Dock > α' " mutation (light and dark purple, respectively) abolish editing at site 1, while the double mutation, "Loop/dock > α/α' " (blue) rescues editing at this site. Bars represent standard error ($***$ $P < .0001$, please see Figure 2.17 for specific P values).

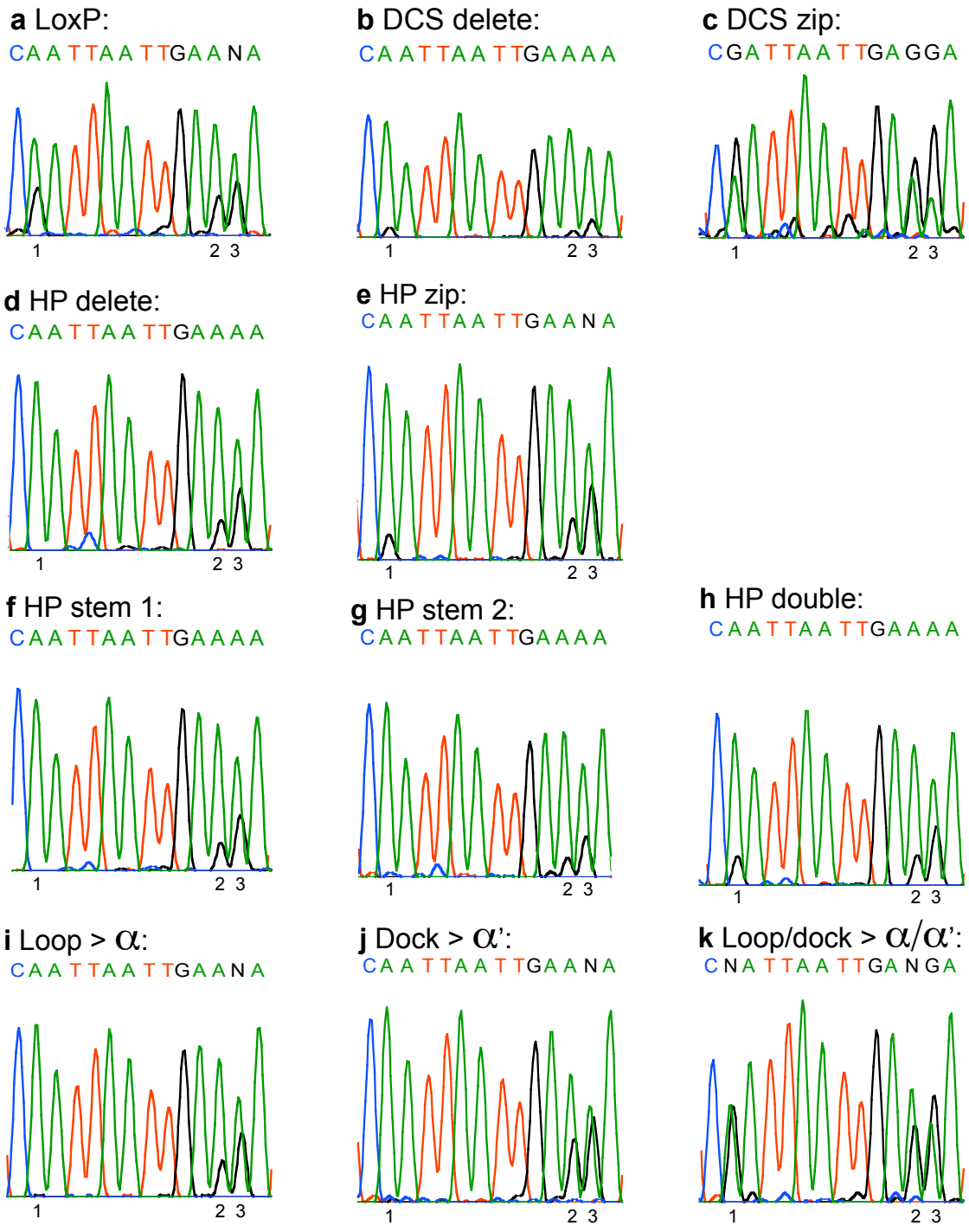


Figure 2.13: Representative electropherograms from all HR-generated mutant *Drosophila* lines. **(a)** LoxP control, reproduced from Figure 2.5**c**. **(b-k)** Mutations are as indicated.

pairing interactions in secondary structures. However, higher order RNA structures or subsidiary structural components and their spatial relationships may be difficult to discern. The enzyme-substrate requirement for RNA editing by ADARs has been pursued using both *in vitro* and *in vivo* systems. In general, most studies have focused on the necessity for relatively simple local secondary structural features of a more or less perfect central RNA duplex. In contrast, our studies reveal structural elements that are distant from the central duplex containing the edited adenosines, and play roles in both modulating and specifying sites of RNA editing *in vivo*.

2.3.2 Modulation of RNA editing by a structure separate from the central duplex

Many A-to-I editing sites are found proximal to intron-exon boundaries, and edited adenosine(s) and splicing signals are encompassed in the same dsRNA region [Rieder and Reenan, 2011]. In these cases, proximity of splicing boundaries and tempo of splicing are regulators of editing. The highly conserved subsidiary DCS/donor site RNA duplex is suggestive of a balance between splicing and editing; postponing splicing by occlusion of splicing signals would extend the time allotted for RNA folding, ADAR binding, or both, resulting in increased editing. Yet, efficient donor site recognition by the splicing machinery must eventually occur.

We previously reported that the "no action potential" (*nap*) mutation in the Maleless dsRNA helicase results in a splicing catastrophe around *para* exon 19 [Reenan et al., 2000], causing exon skipping due to avoidance of the 5' donor site for the edited exon. Therefore, precedence suggests that this constitutive splicing event requires unique processing. The deletion of the DCS decreases editing overall at sites in the central duplex and appears to have no deleterious effect on splicing. Thus, the main effect of the wild type DCS appears to be as an upward modulator of editing.

In contrast, the DCS zip mutation adds 7 intronic nucleotides 271 nucleotides

away from the splice donor (Figure 2.6**c**), designed to extend the DCS/donor site RNA helix. Even this modest non-coding change resulted in dramatic increase in RNA editing, at the cost of cryptic donor site activation within exon 19 and the upstream alternative exon h (Figure 2.6**f**). Due to this aberrant splicing, we hypothesize that fewer full-length sodium channel transcripts are produced (Figure 2.7**b**) resulting in a sodium channel hypomorphic allele. A distal intronic duplex in the mammalian Gabra-3 transcript was recently reported to stimulate specific editing, but this element does not include splicing signals [Daniel et al., 2012]. The authors hypothesize that the Gabra-3 element serves to increase ADAR concentration locally, which is quite different from our proposed mechanism. Interestingly, the predicted DCS of *Musca domestica* is extended in both directions and a fourth adenosine is edited within this structure (Figure 2.14**a**-**b**, **e**). The predicted *M. domestica* DCS, while more extensive than that in *Drosophilidae*, is predicted to contain a large central nucleotide bulge, which perhaps aides the splicing machinery in donor site recognition by destabilizing the DCS structure.

Sodium channel abundance affects *Drosophila* behavior, development, reproduction, and aging [Garber et al., 2012]. In both sexes, flies hemi or homozygous for the DCS zip allele are sensitive to temperature (Figure 2.8**a**, Figure 2.7**c**), displaying a paralysis phenotype consistent with that of other temperature-sensitive *para* alleles that are due to loss of *para* activity [Ganetzky, 1984]. Thus, while substantial increases in editing occur in the DCS zip mutant, this increase comes at the expense of splicing defects and likely explains why the DCS is not larger in size and effect than that observed. Splicing and editing occur co-transcriptionally [Rodriguez et al., 2012], and recent evidence points toward a role for RNA structure in splicing [McManus and Graveley, 2011, Zhang et al., 2011]. Our data suggest that an RNA duplex sequestering splicing signals, distant from the central duplex, is finely-tuned in size to increase RNA editing levels by effecting splicing, while simultaneously not exceeding

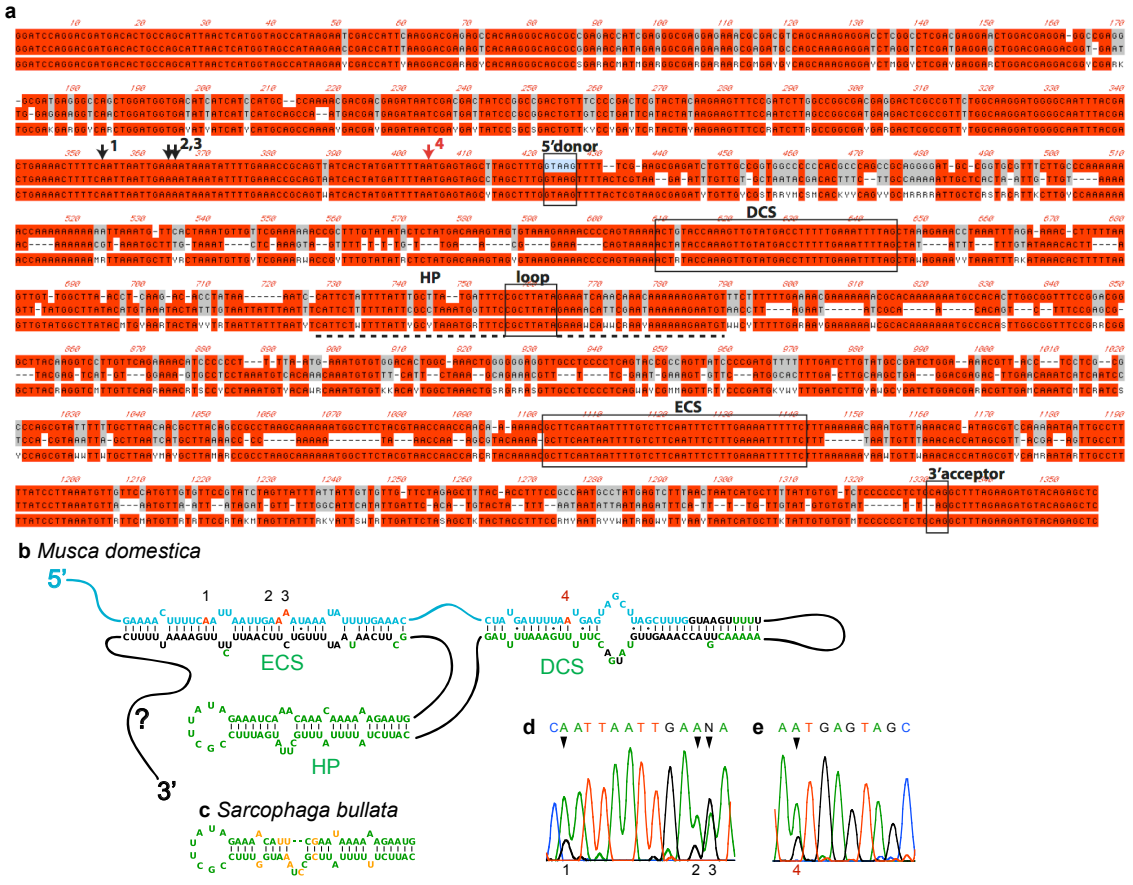


Figure 2.14: Editing and structure in outgroup species. **(a)** Genomic sequence comparison between *Musca domestica* (top) and *Sarcophaga bullata* (middle). Consensus sequence is on the bottom. Red nucleotides represent conservation, while gray denotes disparity. The relevant *cis* elements and four editing sites are noted. **(b)** Predicted *para* structure of *Musca domestica*. Four editing sites are shown in red, while intronic nucleotides that differ from the sequence of *D. melanogaster* are in green. Although the HP structure resembles that in *Drosophilidae*, the HP sequence is completely different. **(c)** The predicted HP from *Sarcophaga bullata*. Nucleotides different from *Musca domestica* **(b)** are show in orange. **(d)** Electropherograms representing editing of adenosines 1-3 and **(e)** 4 in *M. domestica*, which is not edited in *Drosophilidae*.

a size beyond which splicing is compromised (Figure 2.16**a**).

Suggestively, editing levels are increased in *Drosophila* carrying a mutant strain of "slow" RNA polymerase II (pol II) compared to Canton-S wild type animals with a "normal" pol II (Figure 2.16**b-c**). The C4 point mutation in the largest polymerase subunit confers on the enzyme a slower elongation rate, and is known to cause alternative splicing on endogenous genes in both human cells and *Drosophila* [de la Mata et al., 2003]. Because splicing is believed to occur co-transcriptionally, it is possible (but speculative) that splicing is also slowed when transcription rate is retarded. With this speculation, the data presented in Figure 2.16**b** is consistent with our model (Figure 2.16**a**) in which slower splicing (due to slowed pol II) results in an increased level of editing at all three *para* sites. However, the slow pol II and Canton-S lines do not share a genetic background and therefore these data are simply suggestive.

Additionally, we assayed behavioral phenotypes of the editing extremes: the GAG mutant, in which all *para* transcripts are 100% edited at sites 1 and 3, resulting in Para channels with Arg and Asp residues from the edited codons, and the ECS delete mutant, in which editing is abolished at sites 1-3 leading to 100% Glu and Asn residues in the peptide. The sodium channel α subunit consists of four similar domains, each consisting of six transmembrane helices, and the exon 19 editing sites map to the first transmembrane helix of domain III. The editing extreme mutations conferred no detectable activity or mating phenotype (Figure 2.15), suggesting, as is consistent with theories of mRNA editing, that these changes result in subtle alterations in channel function.

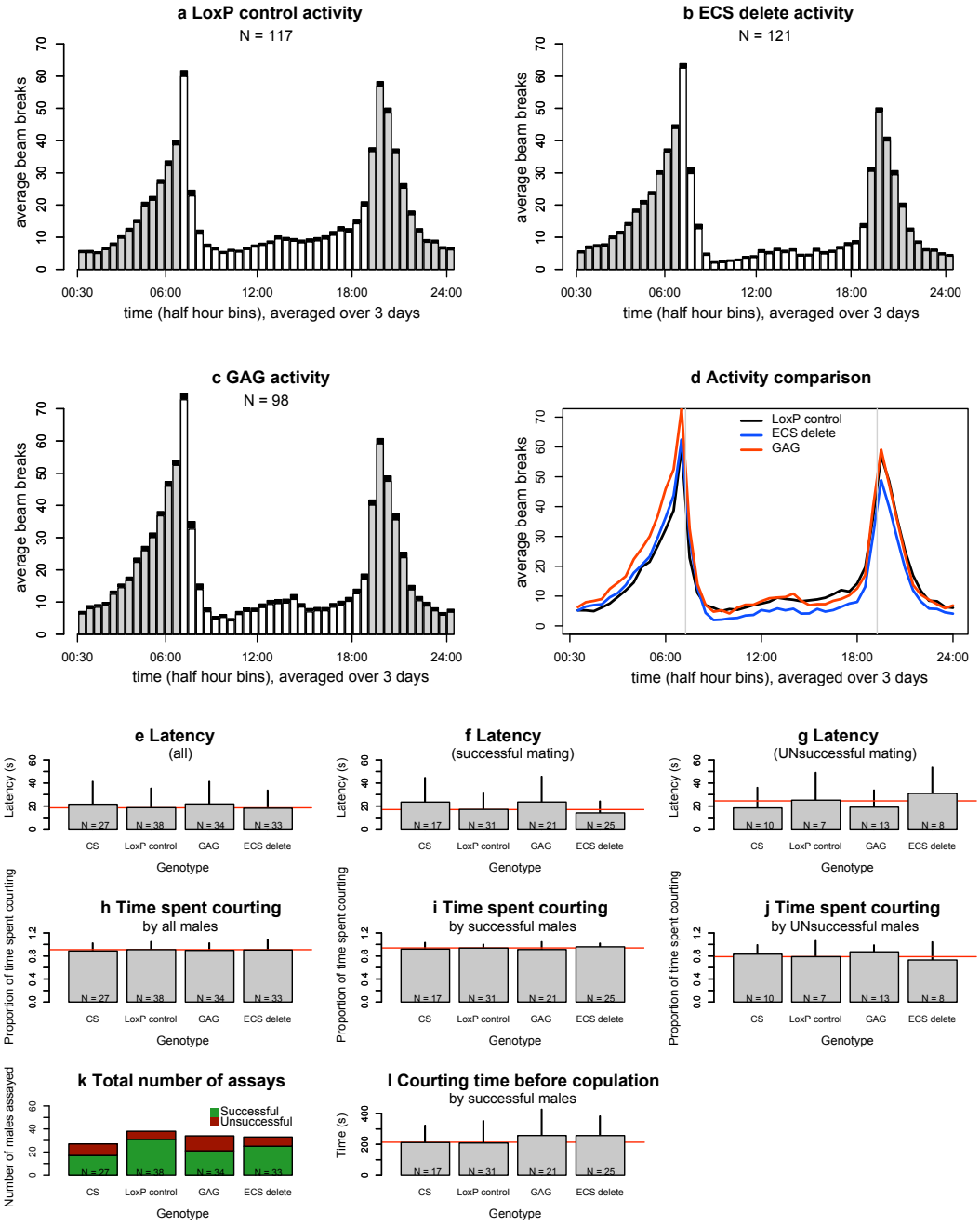


Figure 2.15: Diurnal and mating behavior in editing extremes. **(a)** Circadian activity (beam breaks) of LoxP control males ($N = 117$), averaged over 3 days into half-hour bins. White bars represent lights-on, while gray bars represent lights-off. Black shadows represent standard deviation. **(b)** Activity of ECS delete males ($N = 121$). **(c)** Activity of GAG males ($N = 98$). **(d)** Comparison of activity between LoxP control males (black), ECS delete males (blue) and GAG males (red). Gray lines represent lights-on and lights-off, as in **(a-c)**. **(e-g)** Mating latency (time before courtship initiation) of all males **(e)**, successful males **(f)**, and unsuccessful males **(g)**. Numbers are indicated within bars and vertical lines represent standard deviations. The red line aligns with LoxP control average in each graph. **(h-j)** Total time spent courting by all males **(h)**, successful males **(i)**, and unsuccessful males **(j)**. **(k)** Total number of assays by male genotype. Green represents successful males, while red represents unsuccessful males. **(l)** Courting time before copulation of successful males.

2.3.3 RNA editing specificity conferred by a tertiary pseudoknot

Complex tertiary interactions that extend base pairing to extrahelical and non-canonical base pairs in dsRNA structures are exemplified by the pseudoknot. PK structures are essential to diverse biological processes. For example, in protein synthesis, the central PK of the 30S ribosomal subunit serves to scaffold the proper arrangement of three major structural domains [Brink et al., 1993]. Likewise, telomerase function is dependent on a conserved core PK within the telomerase RNA linked to the RNA template strand for DNA synthesis [Blackburn and Collins, 2011]. In contrast to these structural PKs, there are fewer and less highly conserved PKs reported in mRNA. However, PKs figure prominently in viral mRNAs in the process of ribosomal frameshifting, a process vital to viral life cycles [Brierley et al., 2007].

ECS elements underlie the central duplex containing sites of ADAR action and, while they can be found at a significant distance from editing sites, are generally thought to contain all of the information for ADAR specificity. For example, the *para* ECS appears to be necessary and sufficient for editing at sites 2 and 3 (Figure 2.16**d**). However, the PK interaction we report here is unique in its characteristics as a tertiary structure directing specific RNA editing. Length and helix discontinuities (loops and bulges) of RNA central duplexes limits ADAR activity to particular adenosines [Ohman et al., 2000]. We propose that the exon/ECS duplex is inconsistent in either size or structure with a standard binding of dADAR to the duplex for editing at site 1, and necessitated the evolution of a more complex structure to orchestrate binding of dADAR and engage the catalytic domain on the site 1 adenosine (Figure 2.16**e**).

The *paralytic* pseudoknot interaction reported herein encompasses seven Watson-Crick base-pairing interactions. Although the hairpin loop and docking site are absolutely conserved in all the *Drosophilidae* species we investigated (Figure 2.9**b**, Figure 2.4**c**), replacing the pseudoknot sequences with a naturally occurring motif

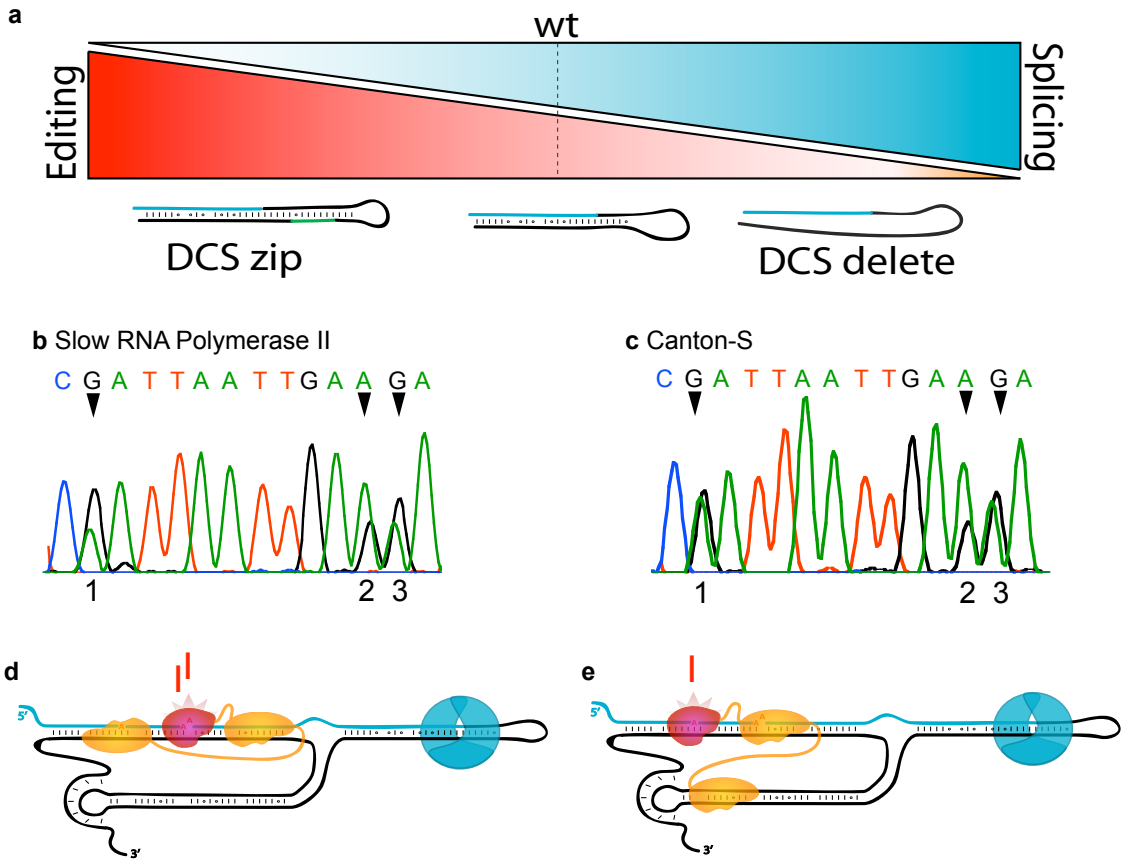


Figure 2.16: A model for *paralytic* pre-mRNA editing and splicing. **(a)** Editing (red) and splicing (blue) are inversely correlated. When splicing is very slow or inefficient, editing is high, for example, in the *para* "DCS zip" allele. If splicing is rapid or highly efficient, the *paralytic* structure is quickly resolved, resulting in lower levels of editing, for example, transcripts from the *para* "DCS delete" allele. **(b)** Editing in "slow" RNA Polymerase II mutant *Drosophila*, compared to editing in (**c**) Canton-S with a "normal" Polymerase. (**d**) dADAR dsRNA binding domains (orange) bind to only the ECS to edit sites 2 and 3 into inosine (red). This occurs before the splicing machinery (blue) recognizes the splice donor. **(e)** dADAR binds to the ECS and HP duplexes to correctly position the catalytic domain (red) to edit *para* site 1.

(the kissing loop sequences from the group II self-splicing intron) not only completely supports editing *in vivo*, but increases editing at all three edited adenosines (Figure 2.12**e**). While the *Drosophila* PK sequences are invariant, the α/α' sequences from group II introns do co-vary, preserving structure while altering primary sequence [Michel et al., 1989]. Indeed, seven base-pairs is thought to be a minimum required for rapid nucleotide annealing [Cisse et al., 2012], although certain kissing loops comprising as little as two base pairs can have surprising stability [Chen and Garcia, 2012].

Musca domestica and *Sarcophaga bullata* show RNA editing at sites 1-3 (Figure 2.14**a**, **d**), and comparative genomics reveal a conserved hairpin, located between the orthologous DCS and ECS elements, with an invariant 8-nucleotide loop (Figure 2.14**a-c**). Curiously, the loop sequence differs from that seen in *Drosophilidae*, and no obvious conserved docking site was observed. Nevertheless, we believe that the HP in these species directs site 1 editing, probably through a long-range tertiary interaction that remains to be determined.

2.3.4 Broader implications for RNA editing

Recent effort in the editing field has focused on identification of new editing sites in multiple organisms via RNAseq, thereby compiling comprehensive inosinomes [Wulff et al., 2011]. Yet these studies are often controversial [Li et al., 2011], and pose substantial analytical challenges due to false positives [Bass et al., 2012]. More broadly, our observations suggest that proximity of editing sites does not imply functional coupling. Future efforts to find polymorphisms affecting editing levels in human genes should not assume that standard ECS elements are sufficient to determine editing extent and specificity.

Editing is implicated in an expanding list of diseases, including breast cancer [Shah et al., 2009], suicidal depression [Lyddon et al., 2012] and ALS. Substantial

evidence points toward ADAR2 mis-regulation in motor neurons of ALS patients [Hideyama et al., 2012], and recent identification of a hexanucleotide repeat associated with ALS [Renton et al., 2011] has strengthened the hypothesis that aberrant RNA processing is involved in ALS, possibly through the seclusion of RNA binding proteins such as ADAR, with irregular transcripts in inclusion bodies [Baloh, 2012].

Additional insights into the secondary and tertiary RNA structures that direct editing may allow for prediction of potential editing substrates, and help to rule out false positives identified via deep sequencing methods. In addition, artificial editing substrates may be designed, enabling the co-option of endogenous ADAR enzymes as tools in specific RNA therapies. For example, a 937 bp antisense sequence designed to trigger RNA editing and degradation of the HIV *env* transcript, is already in clinical trials [Burnett et al., 2012]. Further, the adenosine preference of the human ADAR protein is mutable [Kuttan and Bass, 2012], and the enzyme itself therefore presents a possible target for genetic disease therapy.

2.4 Methods

Sequence alignments

Sequence alignments were performed using the ClustalW multiple sequence alignment package included with MacVector^{tm} 7.2 (Accelrys Inc.). Open and extend gap penalties were set to 1.0. The 37 species alignment was then used for the cladogram in the supplemental material. The cladogram was generated using the Neighbor-joining method, bootstrap (1000 reps).

RNA structural predictions

Structural predictions for regions encompassing all elements and the HP element alone were performed using the SFOLD web-based algorithm (http://sfold.wadsworth.org/cgi-bin/index.pl) using the SRNA package to generate general features and output for

Genotype	Site	Raw editing level (SE)	*P* value vs. LoxP	Significance code
LoxP control	1	0.314 (0.0157)	NA	
GAG	1	1	NA	
DCS delete	1	0.102 (0.0115)	< 0.0001	***
DCS zip	1	0.623 (0.0073)	< 0.0001	***
HP delete	1	0	< 0.0001	***
HP zip	1	0.103 (0.0080)	< 0.0001	***
HP stem 1	1	0	< 0.0001	***
HP stem 2	1	0	< 0.0001	***
HP stem dbl	1	0.171 (0.0039)	< 0.0001	***
Loop > a	1	0	< 0.0001	***
Dock > a'	1	0.008 (0.0066)	< 0.0001	***
Loop/Dock > a/a'	1	0.497 (0.0240)	< 0.0001	***
LoxP control	2	0.234 (0.0168)	NA	
GAG	2	0.566 (0.0097)	< 0.0001	***
DCS delete	2	0.076 (0.0123)	< 0.0001	***
DCS zip	2	0.684 (0.0058)	< 0.0001	***
HP delete	2	0.167 (0.0084)	0.0003	**
HP zip	2	0.194 (0.0058)	0.0644	
HP stem 1	2	0.145 (0.0037)	< 0.0001	***
HP stem 2	2	0.136 (0.0078)	< 0.0001	***
HP stem dbl	2	0.181 (0.0037)	0.0056	*
Loop > a	2	0.214 (0.0074)	0.8318	
Dock > a'	2	0.353 (0.0081)	< 0.0001	***
Loop/Dock > a/a'	2	0.410 (0.0191)	< 0.0001	***
LoxP control	3	0.387 (0.0170)	NA	
GAG	3	1	NA	
DCS delete	3	0.205 (0.0141)	< 0.0001	***
DCS zip	3	0.809 (0.0058)	< 0.0001	***
HP delete	3	0.362 (0.0120)	0.6976	
HP zip	3	0.382 (0.0075)	1	
HP stem 1	3	0.339 (0.0064)	0.0728	
HP stem 2	3	0.311 (0.0143)	0.0009	**
HP stem dbl	3	0.383 (0.0043)	1	
Loop > a	3	0.385 (0.0079)	1	
Dock > a'	3	0.514 (0.0080)	< 0.0001	***
Loop/Dock > a/a'	3	0.586 (0.0277)	< 0.0001	***

Figure 2.17: Specific P values. P values are the result of one-way ANOVAs with Dunnett post-hoc tests (alpha = 0.05). Testing is always against the LoxP control. Significance codes: $*$ $P < 0.05$, $**$ $P < 0.001$, $***$ P <0.0001.

Flybase abbreviation	*Drosophila* species	*P* value vs. Dmel	Significance code
Zsep	*Zaprionus sepsoides*		
Dhyd	***Drosophila hydei***	0.0568	
Dmoj	*Drosophila mojavensis*		
Dmer	***Drosophila mercatorum mercatorum***	0.8212	
Dpic	*Drosophila picticornis*		
Dsig	*Drosophila similis grenadensis*		
Dimm	*Drosophila immigrans*		
Dpol	*Drosophila polychaeta*		
Dvir	***Drosophila virilis***	0.986	
Dmen	*Drosophila melanica*		
Dgri	*Drosophila grimshawi*		
Dsil	*Drosophila silvestris*		
Dbus	*Drosophila busckii*		
Spat	*Scaptodrosophila pattersoni*		
Ddim	*Drosophila dimorpha*		
Cpro	*Chymomyza procnemis*		
Camo	*Chymomyza amoena*		
Dbia	*Drosophila biauraria*		
Dele	*Drosophila elegans*		
Drho	*Drosophila rhopaloa*		
Dmau	*Drosophila mauritiana*		
Dmel	***Drosophila melanogaster***	NA	
Dsim	***Drosophila simulans***	0.9973	
Dsec	***Drosophila sechellia***	0.2119	
Dere	***Drosophila erecta***	1	
Dyak	***Drosophila yakuba***	1	
Dtak	*Drosophila takahashii*		
Deug	*Drosophila eugracilis*		
Dfic	*Drosophila ficusphila*		
Dkik	*Drosophila kikkawai*		
Dbip	*Drosophila bipectinata*		
Dana	***Drosophila ananassae***	0.8583	
Dpse	***Drosophila pseudoobscura pseudoobscura***	0.8429	
Dtro	***Drosophila tropicalis***	0.6376	
Dequ	***Drosophila equinoxialis***	0.0131	*
Dwil	***Drosophila willistoni***	<0.0001	***
Zind	*Zaprionus indianus*		

Figure 2.18: *Drosophilidae* species abbreviations and P values. *Drosophilidae* species abbreviations and species, as given in Flybase. P values for the 13 species represented in Figure 2.2**b** and Figure 2.10**d** (bold) are the result of one-way ANOVAs with Dunnett post-hoc tests (alpha = 0.05). Comparison is to *D. melanogaster*. Significance codes: $* \ P < 0.05$, $** \ P < 0.001$, $*** \ P < 0.0001$.

statistical RNA folding.

***Drosophila* stocks**

Stocks were maintained at 25°C under 12-h light/dark cycles on standard cornmeal molasses food. The slow RNA polymerase mutant *Drosophila* line "RpII215^{C4}" was obtained from Bloomington Stock Center (No. 3663).

Ends-out homologous recombination of the *paralytic* locus

We performed ends-out homologous recombination after Staber et al. [2011]. This technique involves two 2.5 Kb arms homologous to the *para* locus and flanked by recognition sites for the Flp and I-Sce I endonucleases. We cloned and sequenced homology arms in pTOPO (Life Technologies) and then ligated them into a p[w25.2] targeting vector. We achieved mutations in the desired region using the Quik Change II XL Site-directed Mutagenesis Kit (Agilent Technologies). The mini-gene *white*$^{+}$, situated between the arms and flanked by LoxP sites, acts as a selectable eye color marker and is removed later with Cre-recombinase. The vector was introduced into the *Drosophila* genome by injection (Genetic Services, Inc.).

We targeted the endogenous *para* locus so as to generate multiple independent events. We removed the *white*$^{+}$ marker from each line by crossing in Cre-recombinase and isolating individuals carrying targeted alleles with a single LoxP remnant. We then validated targeted alleles via Sanger sequencing (University of Wisconsin Biotechnology Center).

RNA editing analyses

For each mutation analyzed we performed six analyses, two PCRs each from three independent *Drosophila* HR lines. Species-specific editing data were obtained from 3 PCR replicates derived from a single RNA sample. For each sample we extracted RNA from male heads ($N = 15 - 20$) using Tri Reagent (Molecular Research Center, Inc.). We amplified cDNAs via RT-PCR using a *para*-specific primer (Table 2.1), performed PCR and electrophoresed samples on agarose gels. Species-specific editing

data were obtained by designing PCR and RT primers to regions of homologous sequence (Table 2.1). We cleaned the products using Wizard Gel and PCR Cleanup Kits (Promega) and detected editing by Sanger sequencing (University of Wisconsin Biotechnology Center). We determined editing ratios by measuring the area under select A and G nucleotide traces in Adobe Photoshop.

Table 2.1: Specific *paralytic* primer sequences.

Step	Forward (**F**) and Reverse (**R**) Primers
Reverse transcription	CCTGTTCGTTGGTTAAGCACTTTGC
PCR (*D. melanogaster*)	**F:** CAGATGATTGGCAACTCAATTAACCACC
	R: CGCGACCATGTTCGATATGGCAGGTAG
PCR (species)	**F:** CCATAAGAATCGACCATTCAAGGACG
	R: CTTTGAAGCCGAGCGCCAACCACTTG
Sanger sequencing	CTTTGAAGCCGAGCGCCAACCACTTG
qRT-PCR (total transcript)	**F:** CAACGACGTCTACAACTTCAAGACC
	R: CTATTACCGCAGGTCATGTGGTAG
qRT-PCR (full-length transcript)	**F:** GGATGGGGCAATTTACGACTGAAAAC
	R: GTGGCAGATGTACATCTTCTAATGC
qRT-PCR (426 bp spliceoform)	**F:** GGTCTCGACGAGGAACTGGAC
	R: CCTGCAGTATGGGTCTTTGTGGCAG
qRT-PCR (288 bp spliceoform)	**F:** CTCAATTAACCACCAAGACAATAG
	R: CTTTGAAGCCGAGCGCCAACCACTTG
qRT-PCR(GAPDH)	**F:** GACGAAATCAAGGCTAAGGTCG
	R: AATGGGTGTCGCTGAAGAAGTC

Identification of cryptic splice site activation

We cleaned PCR products, as above, and sequenced individual products. Please see Table 2.1 for specific primer sequences.

Temperature-sensitivity paralysis assay

We raised *Drosophila* at room temperature (18°C) and tested all animals within 24 hours after ecclosion. We mouth pipetted single flies into room temperature glass vials, which we then submerged in a circulating 39°C water bath. We measured time from submersion until first detectable paralytic event, defined as 5 seconds lying on the back or (rarely) side. We tested 105 flies per genotype (35 each from 3 independent HR lines) per sex.

Quantitative RT-PCR assay

We isolated mRNA from 40 male *Drosophila* heads per biological replicate using oligo $d(T)_{25}$ magnetic beads (New England BioLabs). We normalized and reverse-transcribed the mRNA using the iScript cDNA Synthesis Kit (Bio-Rad). We then performed quantitative RT-PCR using the SYBR Green PCR Master Mix (Applied Biosystems) and ran plates on the 7500 Fast Real-Time PCR System machine (Applied Biosystems). We performed three technical replicates per biological replicate, and three biological replicates for both the LoxP control and DCS zip lines. We used primers specific to the mature GAPDH transcript to normalize the data from replicates (Table 2.1). To detect total *para* transcript, we designed primers specific to the 3 most exon-exon junction of the mature *para* mRNA. To detect spliceoforms we designed primer pairs within exons 19 and 20 (see Figure 2.7**a**, Table 2.1). We performed melting curves on each PCR product to ensure primer specificity.

Statistical analyses

For editing analyses, we performed one-way ANOVAs (alpha = .05) followed by Dunnett post hoc tests. $* \ P < 0.05$, $** \ P < 0.001$, $*** \ P < 0.0001$. All experimental lines were compared to LoxP control lines. For specific P values please see Figure 2.17. For paralysis analyses we performed logrank tests on the raw data. We performed one-way ANOVAs on qRT-PCR data after Rieu and Powers [Rieu and Powers, 2009].

Chapter Three

Temperature-dependent modulation of editing via RNA "thermistors"

The work for this chapter was carried out in collaboration with Dr. Yiannis Savva (Brown), Mr. Matthew Reyna (RPI).

I wrote the manuscript, crafted the figures, and carried out data analysis. Dr. Yiannis Savva and I shared data collection responsibilities. Mr. Matthew Reyna scripted a program to compute editing ratios. Editorial feedback was contributed by all authors.

Thanks to Mr. Yao-jen Chang and Ms. Patricia Santos for western help, the laboratory of Dr. Stephen Helfand for incubator use, and Dr. James E.C. Jepson for tagged endogenous dADAR mutants.

3.1 RNA as *cis*-acting "molecular sensor"

The "RNA World hypothesis," first suggested by Walter Gilbert in 1986, proposes that RNA is the ancient molecule [Gilbert, 1986] capable of self-replication, chemical reaction catalysis (e.g. "ribozymes"), and environmental sensing. While DNA is limited to double-stranded helical shapes, the repertoire of possible RNA secondary and tertiary structures appears limitlesss. Wan et al. recently measured RNA folding energies genome-wide at single-nucleotide resolution, revealing that classes of RNAs fall into categories of thermodynamic stability [Wan et al., 2012]. These findings reveal, unsurprisingly, that RNA secondary structure is correlated to its function, and that both of these properties–structure and stability–contribute to transcriptional regulation [Bonetti and Carninci, 2012]. RNA structures are dynamic, and, as detailed below, they are incredibly responsive to input in the form of molecular and environmental factors. It is the mutable property of RNA structure that allows RNA to act as sensor and elicit rapid responses on the cellular level [Wan et al., 2011].

RNA structures are involved in transcription, splicing, cellular localization, silencing, translation, and RNA turnover [Wan et al., 2011]. Therefore, it is not surprising that RNAs also play a role in transcriptional regulation in response to environmental signals. Although the process of environmental sensing is often carried out by proteins, a range of RNA molecules have recently been discovered that recognize ligands or other signals and control post-transcriptional gene expression. Amazingly, these RNA sensors use only the four ribonucleotide building blocks, in contrast to the twenty amino acids available to proteins. In addition, RNA sensors are often highly specific, a property that has garnered much interest from the medical field.

For example, RNA aptamers, single-stranded nucleic acids that bind to molecular ligands and inhibit their functions (Figure 3.1 **A**), are easily selected in an *in vitro* process called SELEX (Systematic Evolution of Ligands by EXponential enrichment) [Tuerk and Gold, 1990]. The first aptamer therapeutic, Pfizer's Macugen

to treat age-related macular degeneration, was approved by the FDA in 2005. However, it remains the only aptamer therapeutic on the market [Sundaram et al., 2013].

3.1.1 Riboswitches

A sub-class of RNA aptamers,"riboswitches," were discovered in 2002 in bacteria [Mironov et al., 2002, Nahvi et al., 2002, Winkler et al., 2002]. A riboswitch, which typically consists of an aptamer domain and an expression domain (Figure 3.1 **B**), senses cellular metabolites to modulate gene expression through changes in RNA structure. For example, the vitamins B_1, B_2, and B_{12} directly interact with mRNAs to control operons involved in metabolite biosynthesis and transport [Serganov and Nudler, 2013]. The *btuB* gene in *E. coli* and the *cob* operon in *S. typhimurium* both bind coenzyme B_{12}, precisely regulating internal cellular concentrations of this metabolite. More than 200 similar variants of this riboswitch have been found in diverse bacteria [Mandal and Breaker, 2004].

Although widespread in prokaryotes, only a single family of riboswitch has been found in eukaryotes [Wan et al., 2011]. Instead of exerting transcriptional control, eukaryotic thymidine pyrophosphate (TPP) riboswitches regulate alternative splicing. Riboswitch structures form in some eukaryotic mRNAs, involving an intergenic region or 3' UTR sequence that base pairs around a splice junction. In the absence of the ligand, TPP, normal splicing occurs. However, in the presence of TPP, the RNA structure is altered such that the alternative exon is included, leading to translation termination or aberrant peptide synthesis [Serganov and Nudler, 2013]. TPP-regulated riboswitches are found in bacteria, plants, and fungi [Sudarsan et al., 2003], suggesting they represent an ancestral form of genetic control. Therefore, the ability of RNAs to sense environmental signals and effect change at the level of post-transcriptional gene expression is found in diverse life forms.

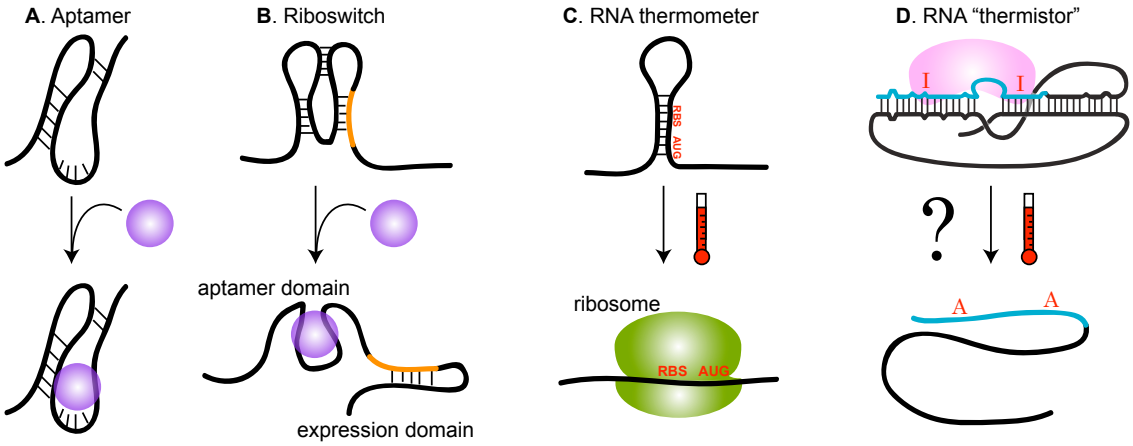


Figure 3.1: RNA sensor families. **A.** Aptamers, which may be DNA or RNA, bind small molecule ligands (purple) and inhibit their fuction. **B.** Bacterial RNA riboswitches contain an aptamer domain and an expression domain. When the aptamer domain binds a ligand (purple, often a cellular metabolite), the RNA changes shape and affects subsequent translation. Certain sequences of the RNA (orange) have alternative *cis* binding partners. **C.** RNA thermometers, most often found in bacteria, structurally sequester the ribosome binding site (RBS) and start codon (red). When the structure melts, the ribosome (green) may bind and translate the RNA. RNA thermometers most often regulate heat- and cold-shock genes and virulence factors. **D.** We suggest that there exist RNA "thermistors" structures, sometimes complex, required for ADAR (pink) recognition and A-to-I editing, that are sensitive to temperature. Thus a thermistor could affect RNA editing levels in response to temperature fluctuations.

3.1.2 RNA thermometers: temperature, RNA structure, and gene regulation

Yet RNA sensing is not limited to molecular ligands; RNA structure is fundamentally sensitive to temperature. The differential stability of RNA molecules in the transcriptome correlates to the diverse cellular roles for RNA structures. For example, Wan et al. detected specific "signatures" of yeast RNA melting temperatures that detected RNA properties, including the polarity of mRNA open reading frames, regulatory motifs in 3' UTRs and noncoding versus coding RNAs [Wan et al., 2012]. Computational programs are reasonably accurate when predicting secondary structure of short (<700 bp) input RNA sequences [Wan et al., 2011]. Yet the probability of finding this, the most thermodynamically-stable (minimum free energy) structure, drops as temperature increases [Huynen et al., 1997]. At higher temperatures RNA molecules dynamically adopt various probable secondary and tertiary structures.

RNA thermometers, as their name suggests, are riboswitches that are sensitive to temperature (Figure 3.1 **C**). RNA thermometers are found in bacteria, which employ these temperature-senstive regulatory elements to control transcription of heat-shock, cold-shock, and some virulence genes (thought to be triggered by entry into a warm-blooded host) [Narberhaus et al., 2006]. All known bacterial RNA thermometers sequester the ribosome binding site (Shine-Dalgarno sequence) and start codon at certain temperatures to inhibit translation. Simple synthetic RNA thermometers have been created by structural occlusion of the ribosome binding site, indicating that they most likely operate though a basic melting mechanism [Neupert et al., 2008].

Although there is no direct evidence of RNA thermometers in eukaryotes, several avenues of evidence suggest that some eukaryotic RNA structures respond to temperature to control translation. For example, many eukaryotic mRNAs contain complex IRES (internal ribosome entry sites) sequences that are believed to form

secondary or tertiary structures [Perard et al., 2013, Filbin et al., 2013] that may be responsive to temperature. In addition, eukaryotic cells also contain a heat-shock response pathway, suggesting a mechanism of genetic control by temperature change [Narberhaus et al., 2006]. With the addition of intronic sequences and the compartmental separation of transcription and translation, eukaryotic RNA thermometers could be considerably less simple and conserved than those found in bacteria, confounding detection. Temperature-sensitive structures found in eukaryotic mRNA could theoretically act anywhere in the transcript to alter splicing, translation, transport, degradation, or protein binding.

3.1.3 RNA "thermistors," putative thermo-sensitive effectors of RNA editing

A post-transcriptional modification known to involve diverse secondary and tertiary RNA structures is adenosine-to-inosine RNA editing (please see **Chapters 1** and **2**). We reasoned that there might be a class of eukaryotic RNA thermometer-like structures, "thermistors," that, instead of controlling post-transcriptional gene expression, rather exert subtle pressure on RNA editing levels (Figure 3.1 **D**). In this way the transcriptome and proteome could be responsive to temperature via RNA structure.

Certain evidence exists in the literature suggesting that RNA editing is temperature-sensitive. Savva et al. demonstrated that dADAR auto-editing of the *dAdar* transcript decreases as temperature increases. As the ratio of edited to unedited dADAR isoforms fine-tunes the global inosinome as well as complex organismal behavior, these data point toward an intriguing role of thermal control in RNA editing [Savva et al., 2012]. Additionally, not only is RNA editing in octopus delayed rectifier K^+ channel transcripts responsive to ambient water temperature, but this response appears to be adaptive to the organism [Garrett and Rosenthal, 2012].

To search for RNA thermistors we chose to use *Drosophila*, a poikilotherm with known robust RNA editing machinery and a well-characterized inosinome [Savva et al., 2012]. Although poikilotherms are at the mercy of their environment as they cannot regulate their own body temperatures, they have evolved physiological processes that allow them to function within the natural temperature range of their habitats [Economos and Lints, 1986]. It is possible that differential RNA editing is one process that allows poikilothermic animals such as *Drosophila* to function at varying environmental temperature.

The above observations led us to undertake an exhaustive survey of 55 *Drosophila* editing sites across a 20°C biologically-relevant temperature range (10°C-30°C). Global editing patterns suggest that editing decreases at higher temperatures, but examination of individual editing sites reveal a broad range of temperature-responsive editing patterns. To test the involvement of RNA structures, we repeated the experiment using *Drosophila* with HR-generated structural mutations in the *paralytic* editing substrate, revealing slightly altered temperature sensitivity due to changing structure. Our results suggest that RNA editing is sensitive to temperature, and that this response is effected by RNA "thermistors," temperature-sensitive secondary and tertiary structures that direct editing.

3.2 Results

In order to undertake this ambitious experiment, we first needed a better method of RNA editing analysis. Previous methods involved capturing a screen shot of a mixed peak chromatogram, importing it into Adobe Photoshop, and tracing each peak in order to calculate the area under each curve (in pixels). These numbers were then recorded and stored in a spreadsheet for further analysis. This method, while accurate [Rinkevich et al., 2012], is incredibly laborious and time consuming. We

therefore partnered with a mathematician, Matthew Reyna (Rensselaer Polytechnic Institute), who designed a program to automate the process. This program, Mixed Peak Image Analysis, requires the user to identify the mixed peak and returns the numerical editing ratio as well as intermediate data from each stage of the procedure (Figure 3.2). The program allows the simultaneous processing of many samples, is incredibly time-efficient, and is an accurate substitution for by-hand data processing (Figure 3.3), which may introduce user bias.

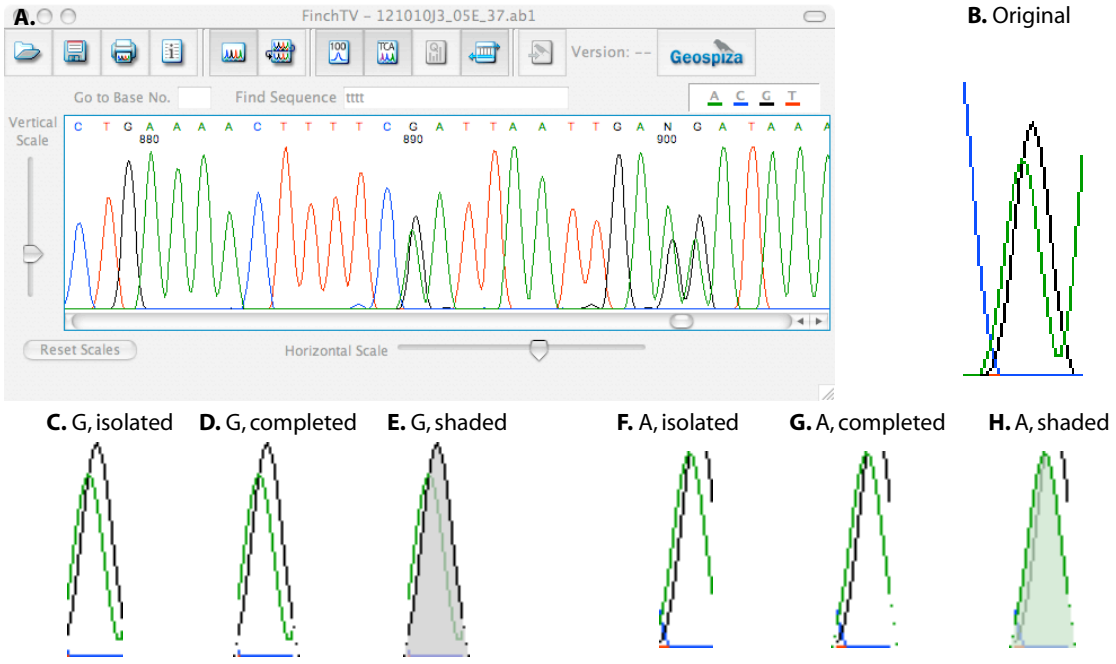


Figure 3.2: Progression of the Mixed Peak Image Analysis program. The program requires the user to select mixed peaks from a chromatogram (e.g. position 890 in **A.**), and save the original (**B.**) in an input folder. The program then performs the isolation (**C., F.**), completion (**D., G.**), and area calculation (**E., H.**) steps for the guanosine (green) and adenosine (black) traces. It then returns a .txt file containing calculated editing ratios.

We grew flies at 25°C before transferring newly-ecclosed animals to incubators held at 10, 20, and 30°C. At 20°C *Drosophila* eggs are the most viable [Economos and Lints, 1986], therefore the 20°C temperature range chosen is biologically-relevant. After 72 hours of acclimation the animals were frozen at -80°C and then processed as described below. The Mixed Peak Image Analysis program was used for all samples.

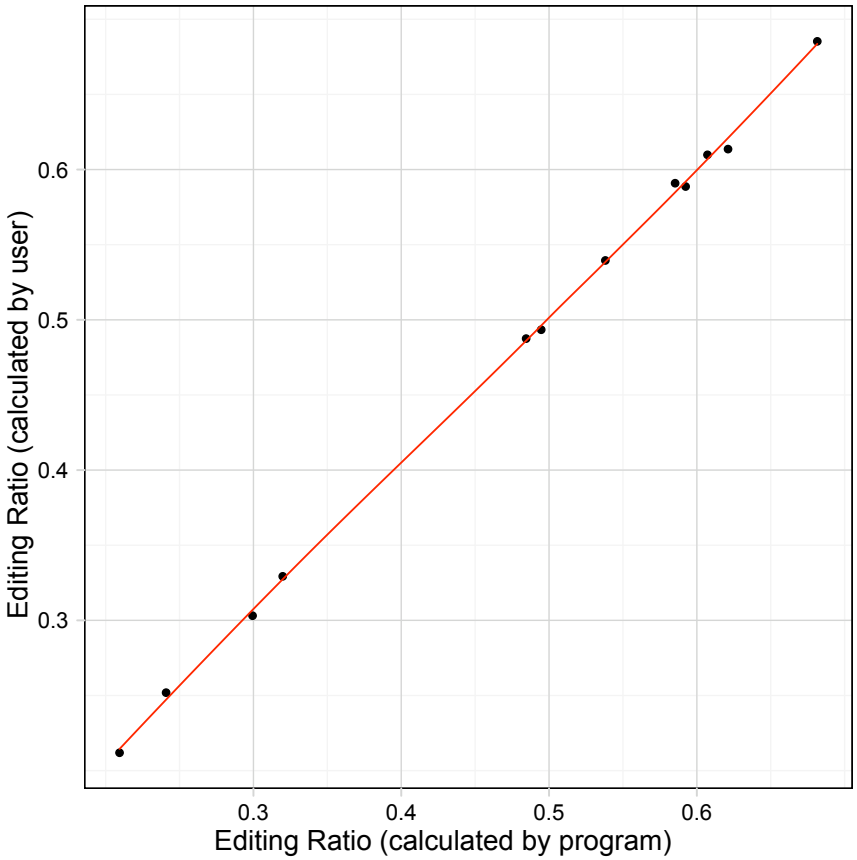


Figure 3.3: Accuracy of the Mixed Peak Image Analysis program. The program was compared to the standard method of editing calculation, which requires the user to import .tiff files into Adobe Photoshop, trace curves, and record and calculate peak area ratios by hand. The same peaks were analyzed by the program and by hand. Editing ratios are G/G+A. The red line represents best fit, $y = 0.9996x$, $R^2 = 0.9987$.

3.2.1 Temperature changes affect editing of specific transcripts

To investigate the effect of temperature on editing in specific transcripts we limited our scope to *Drosophila melanogaster*, in which the inosonome has been well-characterized [Savva et al., 2012]. We investigated 55 editing sites in sixteen different transcripts. The results from each individual site are compiled in **Appendix B**.

Temperature-sensitive editing at individual sites illustrates the full range of possible patterns. For example, editing in the Dopamine/Ecdysteroid receptor transcript sites 1-3 are all very sensitive to temperature: as temperature increases, editing decreases substantially (Figure 3.4 **A**). However, editing at *synaptotagmin-1* sites 2-4 is largely temperature resistant (Figure 3.4 **B**). *Paralytic* sites 1-3 are resistant to temperature between 10 and 20°C, but sensitive at 30°C (Figure 3.4 **C**), while editing of site 6 in *shab* is sensitive to temperature, but increases at higher temperatures (Figure 3.4 **D**). Notably, *shab* site 7, located adjacent to site 6, is edited at 100% and is temperature-insensitive within the 20°C range studied.

3.2.2 Conservation of editing responsiveness to temperature in *Drosophilidae*

In order to determine whether the observed temperature-dependent editing responses are conserved in other species of *Drosophila*, we studied editing of several transcripts in five closely-related *Drosophilidae* species (Figure 3.5).

Interestingly, we observed that certain sites, such as *complexin* site 2 and *synaptotagmin-1* site 4, are very tightly regulated not only across temperatures, but also between the species investigated (Figure 3.6). Generally, if an editing site is responsive to changes in temperature in one species, it is also responsive in other species (for example, *dAdar* auto-editing), although the temperature-dependent

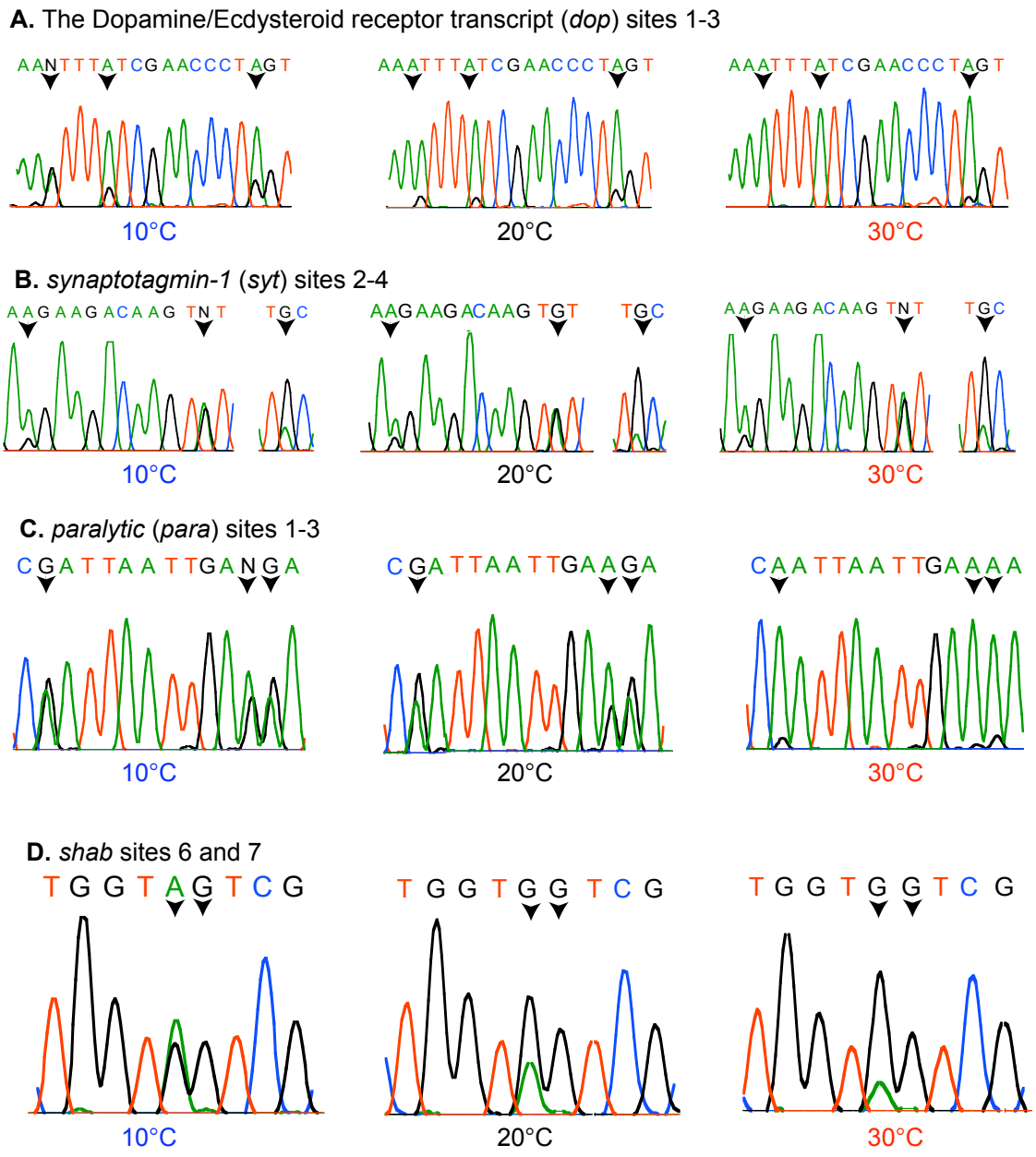


Figure 3.4: Temperature affects editing in specific transcripts. **A.** Editing at sites 1-3 in the Dopamine/Ecdysteroid receptor transcript decreases with increasing temperature. **B.** Editing of *synaptotagmin-1* sites 2-4 are temperature-insensitive between 10 and 30°C. **C.** Editing at *paralytic* sites 1-3 is stable between 10 and 20°C, yet decreases dramatically at 30°C. **D.** Editing at *shab* site 6 is one of the few sites surveyed that increases with rising temperature. In contrast, adjacent site 7 is edited at 100% regardless of temperature. In all figures, black carrots denote editing sites.

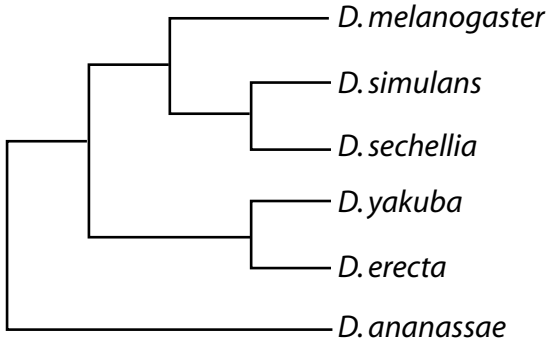


Figure 3.5: Phylogeny of *Drosophila* species used in this study. *D. ananassae* represents an outgroup and was not included in the investigation. Based on the *Drosophilidae* phylogeny from the UCSC Genome Browser [Meyer et al., 2012].

editing trend is not always in the same direction (for example, *uncoordinated-13*).

It is possible that the observed temperature-responsiveness of editing is a result of altered RNA structures, perhaps due to single nucleotide changes or polymorphisms within the genes of the *Drosophila* species studied here. As most structures that direct editing are formed between exonic and intronic sequences, changes in primary sequence often occur in intronic *cis* elements, leading to alterations in RNA structure and a corresponding change in editing. However, it is notable that both *dAdar* [Keegan et al., 2005] and *unc-13* are editing sites directed by entirely exonic secondary structure, whereas *complexin* and *synaptotagmin-1* [Reenan, 2005] sites are directed by structures comprised of paired exon and intron sequences. One would therefore expect the structures of *adar* and *unc-13* to be under higher sequence, and therefore structural, conservation, leading to highly similar editing at these sites across *Drosophilidae*. While editing in the *dAdar* transcript satisfies this prediction, editing in *unc-13* is highly variable (Figure 3.6), suggesting that either the *unc-13* RNA structure or the specificity of editing machinery has evolved in the species studied, leading to the observed variability.

These observations suggest broadly that when editing is tightly controlled across temperature it is also conserved between species. This further suggests that

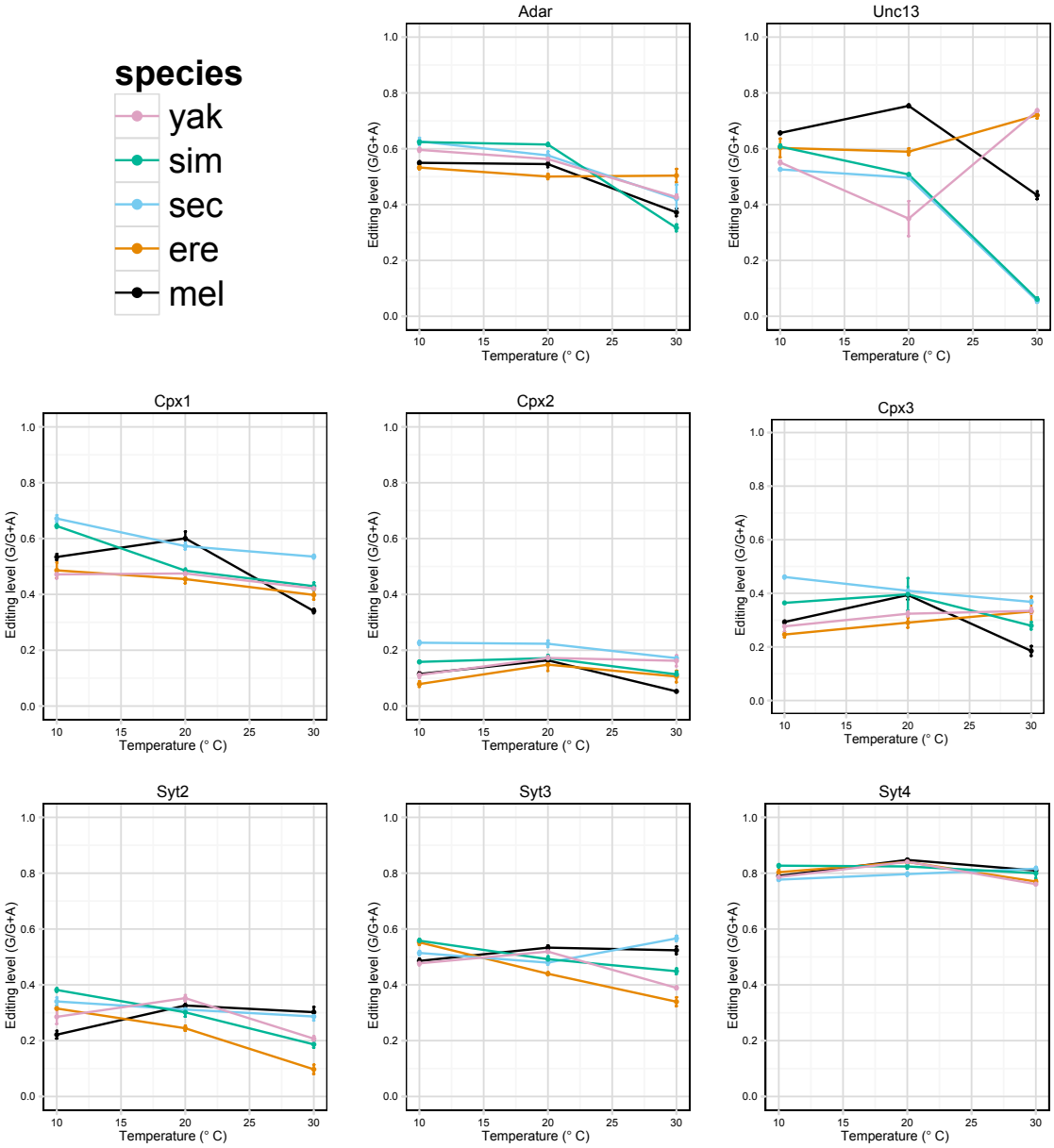


Figure 3.6: A comparison of temperature responsiveness of select editing sites across *Drosophilidae*. Species color-coding is constant in all graphs. Included are *dAdar* (Adar), *uncoordinated-13* (Unc13), *complexin* (Cpx) sites 1-3, and *synaptotagmin-1* (Syt) sites 2-4. Bars represent standard error. Pink: *D. yakuba* (yak), green: *D. simulans* (sim), blue: *D. sechellia* (sec), orange: *D. erecta* (ere), black: *D. melanogaster* (mel).

the RNA structures that direct editing at certain sites are highly thermodynamically stable and are also preserved during evolution. Yet structures that are inherently more temperature labile, allowing editing levels to fluctuate with temperature, are less conserved.

3.2.3 The effect of temperature on global RNA editing trends

We investigated the global effect of temperature on RNA editing in *D. melanogaster*. Figure 3.7 independently plots all sites studied by rank at different temperatures. When plotted thusly, it is apparent that while global editing profiles at 10 and 20°C appear very similar, global editing at 30°C is decreased, suggesting that, at least on a global scale, editing decreases as temperature increases.

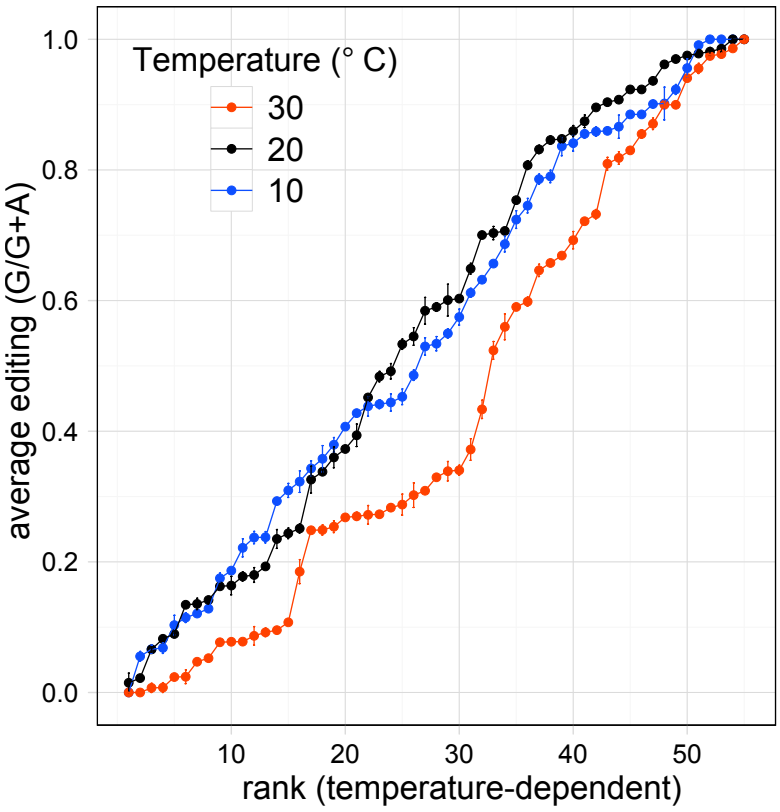


Figure 3.7: Global effects of temperature on editing levels, independently ranked. Each site is ranked separately at each temperature. Bars represent standard error.

When the same data is instead presented as in Figure 3.8, in which all sites are ranked only at 20°C, the independent axis then represents individual editing sites. From this presentation of the data it is apparent that, while the global trend may suggest decreased editing at 30°C, each individual site behaves very differently with respect to its temperature profile.

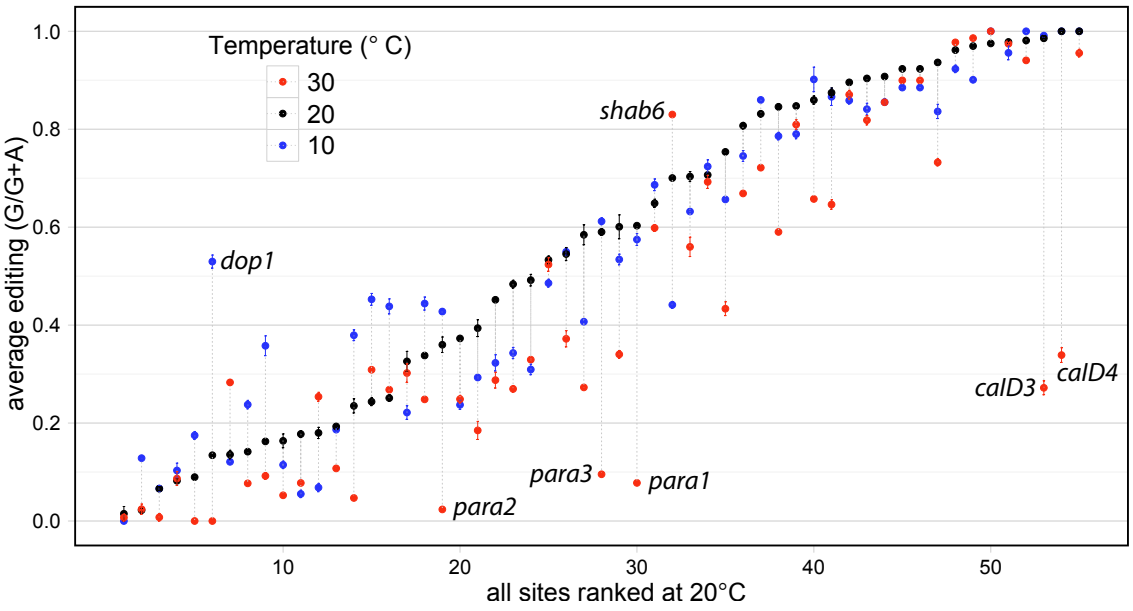


Figure 3.8: Global effects of temperature on editing levels, ranked at 20°C. Sites are ranked by editing level at 20°C and this rank is preserved for editing at 10°C and 30°C. The temperature range at each site is represented by vertical gray dotted lines. Notable sites are identified. Bars represent standard error.

3.2.4 HR-generated RNA structural mutants

Temperature, therefore, impacts editing at individual sites as well as globally. We hypothesized that these editing changes are brought about by the melting of RNA structures that direct editing. At higher temperatures these specific structures occupy a smaller proportion of the structural space, leading to a decrease in editing. Similarly, at lower temperatures editing would be predicted to increase as these structures become more stable.

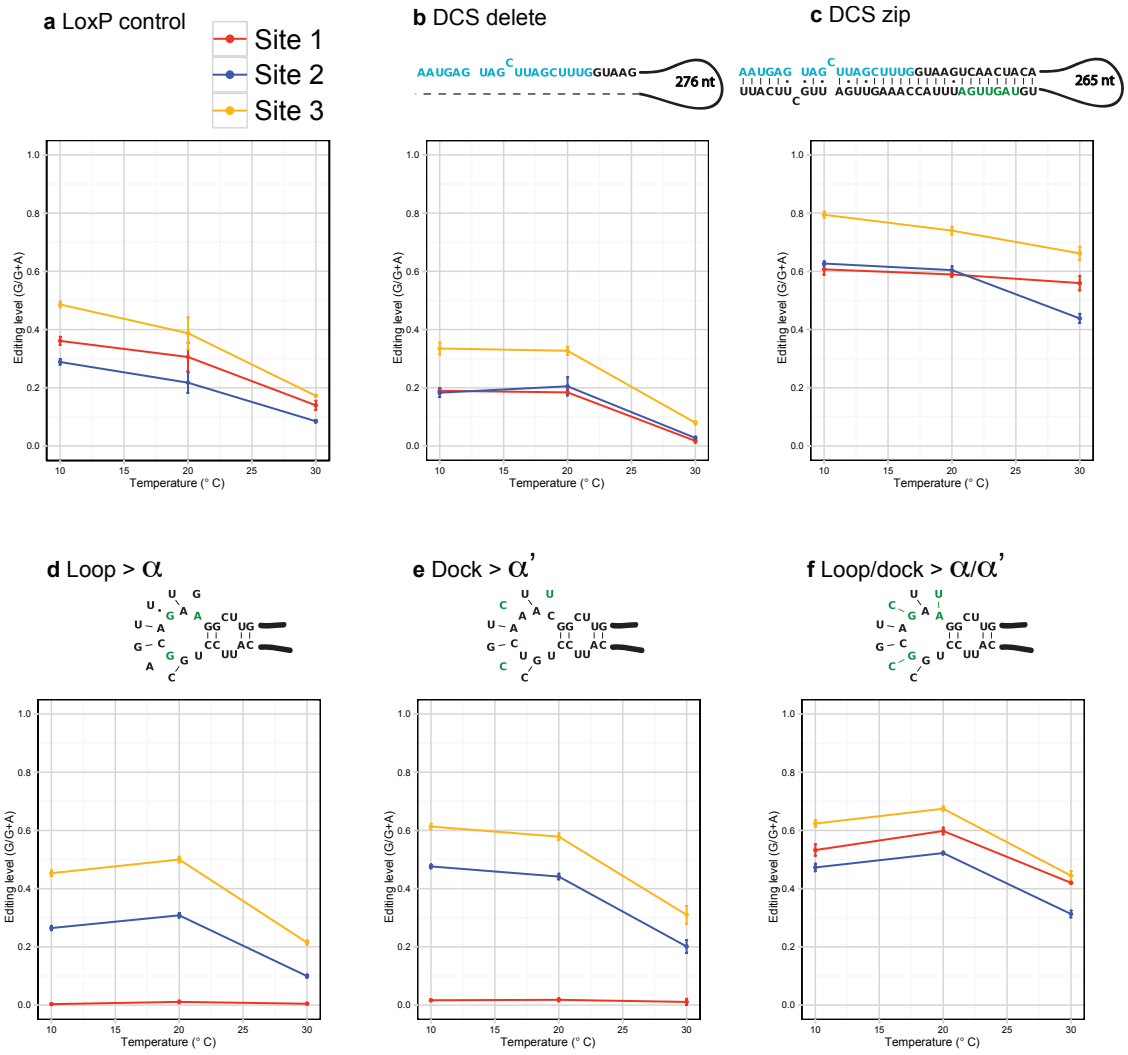


Figure 3.9: Temperature effects of HR-generated mutations in *paralytic*. The wild type *paralytic* transcript folds into a complex pseudoknotted structure to direct editing at three sites (see **Chapter 2**). (**a**) In the loxP control animal, all three sites are decreased at 30°C, similar to the temperature-responsiveness of wild type *paralytic* (Figure 3.4 **C**). (**b**) The "DCS delete" mutation maintains a similar pattern, although all three sites are decreased overall, as is seen with this mutation in Figure 2.6 **d**. (**c**) Similarly, the "DCS zip" mutation, which increases secondary structure around the splice donor and is known to universally increase editing at all three sites (Figure 2.6 **d**) displays a temperature-dependent pattern similar to the loxP control, although curiously site 1 appears temperature-resistant in this mutant. The "Loop > α" (**d**) and "Dock > α' " mutations, which selectively abolish editing at site 1 (Figure 2.12 **e**), are temperature-responsive at sites 2 and 3 similar to the loxP control (**a**). Similarly, the "Loop/Dock > α/α'," which rescues editing at site 1, appears to have a similar temperature profile to that seen in the loxP control. In all graphs, site 1 is red, site 2 is blue, and site 3 is orange. Bars represent standard error. Relevant HR-introduced mutations are shown above graphs. Deleted nucleotides are noted by dashed gray lines, while altered nucleotides are shown in green. Blue represents exon, while black denotes intronic sequence.

We tested the editing response of several HR-generated RNA structural mutations in the *para* transcript (see **Chapter 2**) at varying temperatures, compared to the loxP control (Figure 3.9 **a**). For example, the "DCS delete" mutation, which is predicted to decrease secondary structure, decreases editing at all three *para* editing sites, while the "DCS zip" mutation, which increases secondary structure, universally increases editing (Figure 2.6 **d**). The temperature-responsive editing profile of these structural mutations appears to be a product of both temperature and structure: the DCS zip mutation displays decreased editing at all three sites in response to elevated temperature (Figure 3.9 **b**), yet editing is still decreased from the loxP control (Figure 3.9 **a**). Similarly, the DCS zip mutation also shows decreased editing at 30°C, yet editing at all three sites is higher than in the loxP control (Figure 3.9 **c**).

In addition, we tested the temperature-responsiveness of the *paralytic* pseudoknot mutations (Figure 2.12). The individual pseudoknot mutations, "Loop > α" and "Dock > α'," both selectively abolish editing at site 1, while preserving editing at sites 2 and 3. The double mutation, "Loop/Dock > α/α'," which is predicted to restore the tertiary psedudoknot structure, rescues editing at site 1. These mutations reveal temperature sensitivity similar to that seen in the loxP control (Figure 3.9 **d**-**f**), suggesting again that for these particular editing sites, editing levels are affected by both RNA structure and temperature. In this case the effect appears additive.

3.3 Discussion

RNA structures possess the ability to bind molecular ligands with high specificity, regulating diverse cellular tasks such as transcriptional regulation and protein synthesis. Structures are inexorably linked to RNA splicing. For example, the *Drosophila Down Syndrome Cell Adhesion Molecule* (*Dscam*) gene contains 48 mutually exclusive exons in cluster 6. Exon selection is regulated by competing RNA secondary structures,

and the thermodynamic strength of the structure governing exon selection is directly correlated with frequency of exon inclusion [May et al., 2011]. RNA structure also mediates splicing in *Saccharomyces cerevisiae*, and alternative 3' splice site selection, governed by RNA secondary structure, is at least partially regulated by temperature [Plass et al., 2012].

RNA thermometers, involved in transcriptional regulation in prokaryotes, are now not the only RNA structures that are temperature-responsive, though they may easily be the best characterized. RNA editing, another post-transcriptional process dictated by RNA secondary and tertiary structures (please see **Chapter 2**), may also be sensitive to temperature changes, especially in a poikilothermic animal such as *Drosophila*. We therefore investigated the temperature-responsiveness of 55 well-characterized *Drosophila* editing sites [Savva et al., 2012] across a biologically-relevant 20°C range [Dillon et al., 2009].

Our analysis revealed that editing is globally decreased at 30°C (Figure 3.7), suggesting that this temperature may affect the RNA structures involved in editing. As *D. melanogaster* show fairly constant viability and fertility between approximately 12 and 30°C [Hoffman, 2010], we therefore suggest that this trend may represent an adaptive cellular response rather than a breakdown in editing efficiency. Indeed, the sensitivity of certain editing sites to elevated temperature is in the opposite direction of this global trend (Figure 3.4 **D**).

Figure 3.8 suggests that each individual site behaves differently in response to temperature changes, an observation consistent with the data presented in Figure 3.4. Also congruent is the previous observation that certain sites are sensitive to dADAR isoform and concentration of dADAR protein [Jepson et al., 2011]. In addition, all known editing sites are encompassed in duplex RNA structure, although other than double-strandedness, these sites have no identified common structural property. This suggests that sites may behave differently because variable RNA structures act as

mechanistically distinct "thermistors," depending on the function of the editing site residue.

This hypothesis is supported by the observation that neighboring sites within a transcript, which are predicted to be involved in the same governing RNA structure, often show similar temperature profiles (Figure 3.4 **A-C**). However, there are other sites (Figure 3.4 **D**) in which adjacent sites behave differently with respect to temperature. This could indicate that these sites are "more important" than labile sites. Indeed, the top 25% of sites ranked in Figure 3.8 are largely insensitive to temperature (with a few noted exceptions). This suggests that sites that are edited close to 100% are edited at this level regardless of temperature, perhaps because these structures are the most thermodynamically stable and possible alternative structures are much less likely to form within the investigated temperature range. In support, Tian et al. observed that while editing of constructs representing the wild type Gabra3 I/M site is insensitive to temperature, synthetic constructs containing structural mutations are temperature-sensitive, suggesting an evolutionary advantage to temperature-insensitivity at this particular site [Tian et al., 2011].

The species represented in the above analysis, although closely related (Figure 3.5), represent diverse geographical (and temperature) ranges. While *D. melanogaster* and *D. simulans* are cosmopolitan species, *D. yakuba* is found in savannah climes, *D. erecta* is found in west central Africa, and *D. sechellia* was previously confined to the Seychelles islands [Lachaise et al., 1988]. Our investigation into temperature sensitivity of editing across species was brief. Nevertheless, it appears that when editing level is tightly regulated across temperature, it is similarly regulated across species, suggesting conservation of stable RNA structures. However, editing sites that are more temperature-sensitive are also more variable across species (Figure 3.6), suggesting that labile RNA structures are under less selective pressure. However, drawing broad conclusions based on only eight editing sites in four transcripts is

unwise.

To test our hypothesis that the responsiveness of RNA editing to temperature is mediated by RNA structure, we tested several HR-generated *paralytic* RNA structural mutations at different temperatures. These mutations are described in detail in **Chapter 2** and represent bidirectional changes to both a secondary (DCS) and tertiary (pseudoknot) structural element. The *paralytic* sites investigated in this study are responsive to temperature primarily at 30°C, at which they drastically decrease (Figure 3.4 **C**). Interestingly, in this particular transcript, editing level appears to be responsive to both structure and temperature in an additive manner (Figure 3.9), suggesting levels of regulation at both the molecular and abiotic levels. However, this observation may be specific to these particular editing sites in *paralytic* and is not necessarily representative of global trends.

3.4 Methods

Drosophila stocks

We raised wild type (Canton-S) *Drosophila* at 25°C in humidity-controlled incubators with 12-hour light/dark cycles. We collected animals daily as they ecclosed and shifted newly emerged animals to one of three temperatures: 10, 20 and 30°C, using temperature, humidity, and light-controlled (12-hour light/dark) incubators. From experience we know that flies can live within this temperature range for 72 hours, although they are most often subjected to temperatures between 20 and 30°C. We allowed these animals to remain in the incubators for 72 hours, long enough for the RNA editing machinery to adjust to the temperature, and then immediately froze flies from this population for long-term storage at -80°C.

Additionally, we raised and collected *D. erecta*, *D. sechellia*, *D. simulans*, and *D. yakuba* in the same manner as *D. melanogaster* above. We did the same with

mutant *D. melanogaster* in which the endogenous *dAdar* gene had been HA-tagged using homologous recombination. *Drosophila* with engineered mutations in *paralytic* were generated through homologous recombination (please see **Chapter 1**). The *paralytic* mutations are extensively described in **Chapter 2**.

RNA editing analyses

For each unique species/genotype/temperature sample, we extracted RNA from male fly heads ($N = 15 - 20$) using Tri Reagent (Molecular Research Center, Inc.). We chose only males to control for sex-related editing changes. For each sample, three separately-extracted RNA biological replicates were used. We amplified cDNAs via RT-PCR using random primers, performed PCR using target-specific primers (2 per technical PCR replicates per RNA sample) and electrophoresed samples on agarose gels. We cleaned the PCR products using Wizard Gel and PCR Cleanup Kits (Promega) and detected editing by Sanger sequencing (University of Wisconsin Biotechnology Center). For species-specific editing, we designed primers to regions of high sequence conservation.

Analysis of editing using Mixed Peak Imaging program

Our collaborator, Matthew Reyna (RPI) wrote a program to calculate the ratio of mixed peak areas. This program represents a rapid method of editing calculation. It requires the user to identify mixed peaks, which the program analyzes to give editing ratios. The program also returns data on each stage of the calculation (Figure 3.2).

Briefly, the algorithm can be broken down into five main tasks, where only a single task is provided by the user. The computer performs the final four tasks and provides feedback for each task and warnings when a task may provide suspect results. After the user identifies a mixed peak for analysis, the program isolates each curve by discarding information corresponding to or influenced by adjacent peaks. The program then completes each curve by performing a linear interpolation and extrapolation to recover missing parts of each curve. The program finally computes

the area under each curve and calculates the mixed peak ratio of the curves, a number between 0 and 1, which corresponds to the editing level (G/G+A).

Chapter Four

Mle^{nap-ts}, an RNA helicase, interacts with the *paralytic* transcript

The TALEN work for this chapter was carried out in collaboration with Ms. Kasia Sierzputowska, an undergraduate student at Brown University.

Ms. Sierzputowska and I crafted the figures with help from Ms. Marianna Neubauer. I wrote the manuscript and Ms. Sierzputowska provided editorial feedback.

Thanks to Dr. James Gagnon (Harvard University) for TALEN design advice. Thanks to Dr. Koen Venken (Baylor College of Medicine, Department of Biochemistry and Molecular Biology) for the gift of a 3X-P3- and EGFP-containing plasmid. This research was funded in part through a Brown Research Grant to Ms. Kasia Sierzputowska.

In theory, RNA editing, which often relies on connectivity between an exonic and intronic sequence, is linked to splicing (see **Chapter 1**). However, until 2000 this relationship was only hypothesized [Seeburg, 2000]. Reenan et al. discovered that the nap^{ts} (*no action potential-temperature sensitive*) mutation in the Maleless (Mle) RNA helicase causes a "splicing catastrophe" in the *paralytic* (*para*) transcript [Reenan et al., 2000]. This evidence suggests that the double-stranded RNA structures required for RNA editing must be resolved prior to efficient splicing [Seeburg, 2000]. Rent high-throughput sequencing efforts reveal that co-transcriptional RNA editing is widespread in *Drosophila* [Rodriguez et al., 2012], suggesting that defining the relationship between editing and splicing in *para* may reveal trends of a broader post-translational relationship.

4.1 The DExD/H box RNA helicase family

Maleless, along with its mammalian homolog, RNA helicase A (RHA), is a member of the DExD/H box RNA helicase family. Members of this enzyme family share a conserved core containing motifs involved in ATP hydrolysis and RNA binding, yet their N- and C-terminal domains vary widely, suggesting diverse functionality.

DExD/H helicases are associated with almost all aspects of RNA metabolism, including synthesis, pre-mRNA processing, turnover, and folding [Fuller-Pace, 2000]. Evidence links RHA directly to RNA polymerase II [Nakajima et al., 1997, Anderson et al., 1998], suggesting that RHA specifically is involved in transcription. Interestingly, the ATPase/helicase activity of these enzymes does not seem to be involved in their function in transcription, and so they may play multiple roles in the cell [Fuller-Pace, 2000]. A dual role for Mle was uncovered due to the nap^{ts} mutation.

4.1.1 Dual functionality of the *Drosophila* Maleless helicase

While Mle appears to be a general dsRNA helicase, it is best understood as an important player in the *Drosophila* dosage compensation pathway. Mle is one of five integral protein components of the Male Specific Lethal (MSL) complex, which is present only in males and targets the X-chromosome for transcriptional upregulation. The MSL complex in males lacking this helicase cannot properly target the X chromosome for upregulation, leading to lethality [Kuroda et al., 1991]. Mle is also expressed in females, hinting at functionality outside of dosage compensation, yet females lacking Mle are viable and without discernible phenotype.

Further evidence for the dual functionality of Mle was provided when Reenan et al. discovered that the Mle^{nap-ts} mutant form of the helicase causes aberrant *para* transcript splicing: constitutive exon 19 and often additional upstream exons are spliced out of the mature transcript (Figure 1.16 **a**), leading to a temperature-sensitive paralytic phenotype. While the helicase/ATPase activity of Mle is required for its function in dosage compensation [Lee et al., 1997, Morra et al., 2008], mutations in these domains do not appear to affect *para* splicing. Conversely, the *nap^{ts}* phenotype, which is likely due to a transversion C to G mutation in the Walker A NTP-binding motif (GxGKTT to GxGKTS) [Zhang and Grosse, 2004], does not result in male-specific lethality or affect Mle targeting in females (Figure 1.15) or MSL complex targeting in males.

4.1.2 A model for Mle^{nap-ts}-*para* interaction

The current hypothesis for Mle^{nap-ts}-*para* interaction involves *nap-ts* as a recessive gain of function allele. It was hypothesized that the mutant helicase either cannot dissociate from the secondary structure around *para* exon 19 or structurally blocks the splicing machinery from recognizing the splice donor [Reenan et al., 2000] (Figure 1.16). We reasoned that mutations in the *para* pre-mRNA secondary and tertiary

structures could rescue or antagonize the splicing and behavioral defects resulting from Mle^{nap-ts}.

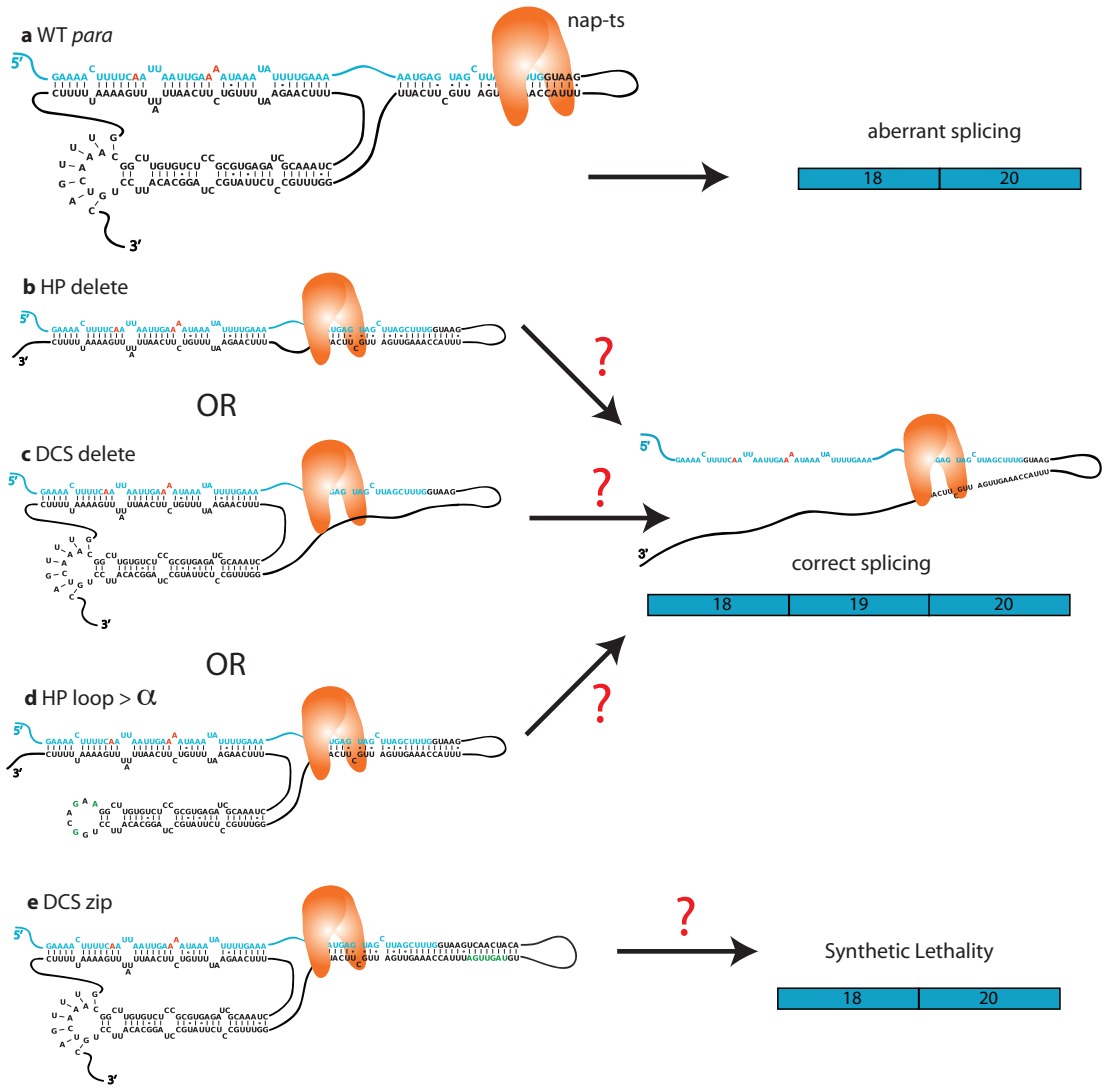


Figure 4.1: Model for Mle^{nap-ts} interaction with *para* structural mutants. (**a**) Mle^{nap-ts} (orange) causes a "splicing catastrophe" in wild type *para* background, causing exon skipping. (**b-d**) It is possible that *para* structural mutants that decrease secondary structure, such as the "HP delete" (**b**) or "DCS delete" (**c**), or those that alter tertiary structure, including the HP "Loop > α" (**d**), will rescue the *nap^{ts}* splicing phenotype because the mutant helicase can now resolve the RNA secondary/tertiary structures prior to splicing. (**e**) The *mle^{nap-ts}* mutation may be functionally lethal in the context of "DCS zip" allele of *para*, as increased secondary structure will further interfere with proper splicing, leading to even fewer full-length Para sodium channels.

For example, structural mutations that decrease secondary structure, such as the *para* HP, DCS, or ECS delete mutations, or those that alter tertiary structure,

including the HP "Loop > α" and "Dock > α' " mutations, might not require the Mle^{nap-ts} helicase for structural resolution before splicing (Figure 4.1 **b-d**). Alternatively, the structural changes around *para* exon 19 might not attract the mutant helicase, leading to structural resolution by other mechanisms and a rescue of the molecular and phenotypic defects.

In contrast, the *para* DCS zip allele increases secondary structure around the splice donor and even in a wild type Mle background this allele generates aberrant splicing and a temperature-senstive paralysis phenotype (Figure 2.8). We reasoned that the homo- or hemizygous DCS zip allele in combination with homozygous *mle*$^{nap-ts}$ might result in synthetic lethality (Figure 4.1 **e**). Animals homozygous for both the DCS zip and *mle*$^{nap-ts}$ alleles are hypothesized to produce even fewer full-length sodium channel transcripts compared to either single mutant.

4.2 Results and Discussion

4.2.1 The Mle^{nap-ts} helicase is rescued by the presence of the loxP in the *paralytic* pre-mRNA

We therefore combined our *para* structural mutations generated through homologous recombination (**Chapter 2**) with the *mle*$^{nap-ts}$ mutation in flies. Surprisingly, we discovered that every HR-generated *para* mutation, including the otherwise wild type loxP control, resulted in a rescue of the *nap*^{ts} splicing phenotype (Figure 4.2).

Surprised by this observation, we turned to the loxP remnant, a 76 basepair intronic insertion that remains in the genome after the HR process [Staber et al., 2011]. The remnant includes not only the loxP sequence (34 bp), but also flanking 6-frame stop codons (14 bp each), and AscI (8 bp) and Acc65I (6 bp) restriction digestion sites required for arm subcloning (Figure 4.6). The core of the remnant is therefore palindromic and is predicted to form a perfect hairpin structure in the

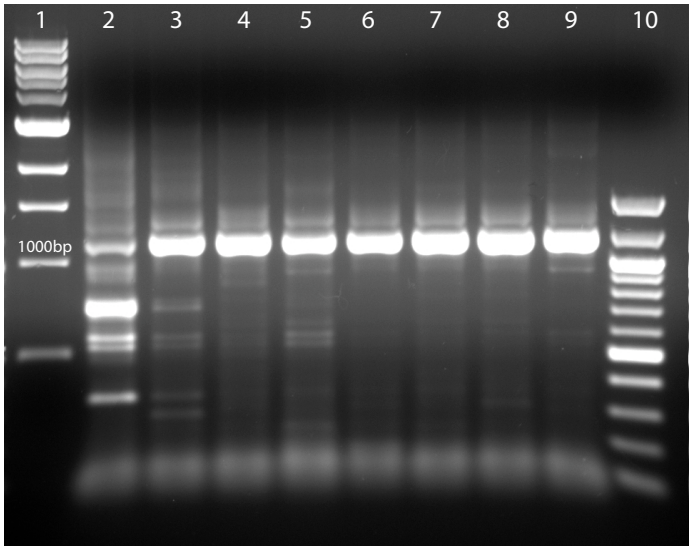


Figure 4.2: The loxP suppresses the skipping phenotype in a mle^{nap-ts} background. Lane 1: 1 Kb ladder; Lane 2: mle^{nap-ts}; Lane 3: *para-loxP* ; mle^{nap-ts}; Lane 4: *para-loxP*; Lane 5: *para-ECS delete* ; mle^{nap-ts}; Lane 6: *para-ECS delete*; Lane 7: *para-DCS delete* ; mle^{nap-ts}; Lane 8: *para-DCS delete*; Lane 9: Canton-S; Lane 10: 100 bp ladder. Primers are given in Table 4.2

para intron between the DCS and Hairpin (Figure 4.3). We hypothesized that the addition of the loxP hairpin results in the rescue of the splicing catastrophe in the presence of Mle^{nap-ts}.

In an effort to disrupt the loxP structure (Figure 4.3), we first repeatedly targeted the *para* loxP with constitutively-expressed Cre recombinase. However, it became apparent that Cre preserves the integrity of the loxP to a striking degree. This observation led us to search for alternative molecular strategies for targeted loxP disruption.

4.2.2 Transcription Activator-Like Effector Nucleases (TALENs)

TALENs (Transcription Activator-Like Effector Nucleases) are a novel genome editing technology that presents an ideal tool for loxP disruption. Recently, two back-to-back *Science* papers broke the cipher that governs DNA sequence recognition by TAL

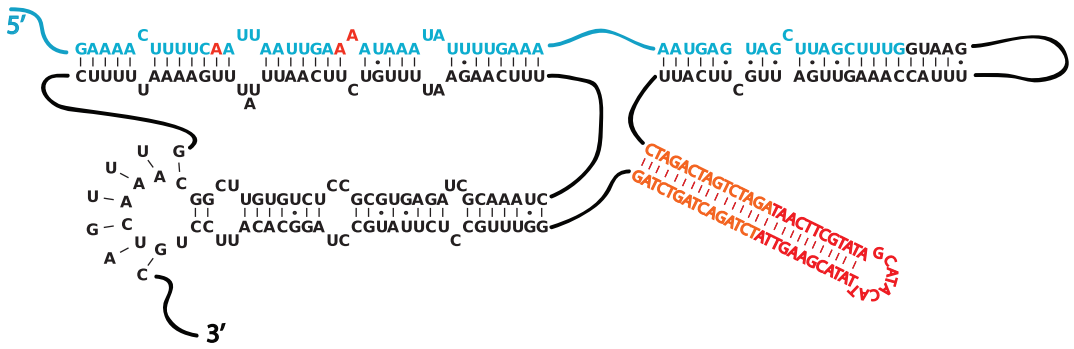


Figure 4.3: The loxP remnant may form a perfect duplex in the *paralytic* transcript. Exon 19 is depicted in blue with three edited adenosines (red). The downstream intron with conserved *cis* elements is shown in black. Structural elements are consistent with those discussed in **Chapter 2**. The loxP remnant, including two 6-frame stop codons (orange) and the loxP site (red) may form a perfect duplex in the pre-mRNA structure.

(Transcription Activator-Like) Effector proteins [Moscou and Bogdanove, 2009, Boch et al., 2009]. These transcription factors, which are injected into plant cells by bacterial pathogens in order to manipulate cellular processes, contain highly-conserved repeat arrays. Two hyper-variable residues in each repeat, repeat variable diresidues (RVDs), are responsible for DNA sequence recognition (Figure 4.4).

Soon after this discovery, TAL DNA recognition domains were fused to the Fok1 nuclease, creating TALENs (TALE Nucleases, Figure 4.4) that could be employed for genome cleavage [Joung and Sander, 2013]. Unlike zinc-finger nucleases, TALENs are highly sequence specific, have few off-target effects, and are very inexpensive to design and synthesize. Further, TALENs are significantly more mutagenic than zinc-finger nucleases [Chen et al., 2013], predicting their widespread use in biotechnology. Relatively few rules govern the choice of target DNA sequence [Cermak et al., 2011], and kits are widely available to facilitate assembly.

TALENs have been employed in most model systems, including mouse [Sung et al., 2013], human cell lines [Kim et al., 2013, Ding et al., 2013], *Xenopus laevis* [Sakuma et al., 2013], and *Danio rerio* [Bedell et al., 2012, Zu et al., 2013], as well as in non-model animals such as cow and pig [Carlson et al., 2012] and medaka

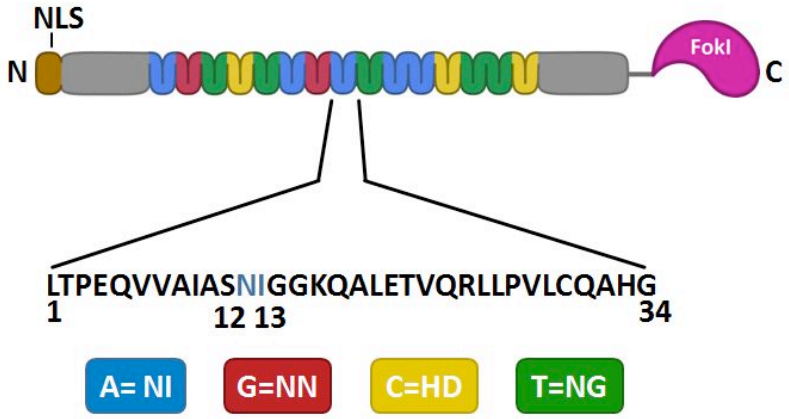


Figure 4.4: Structure of a Transcription Activator-Like Effector Nuclease. A TALEN consists of an N-terminal nuclear localization signal (NLS), a variable number of repeats, and a C-terminal Fok1 nuclease (pink). Residues 12 and 13 from each repeat represent the repeat-variable di-residue (RVD), and via the RVD each 34 amino acid repeat targets a single nucleotide. The four most common RVDs each target a separate nucleotide in the DNA. Image courtesy of Marianna Neubauer and Kasia Sierzputowska.

[Ansai et al., 2013], suggesting that TALENs are useful tools in a variety of biological model systems. The Jiao group efficiently targeted the *Drosophila yellow* gene for disruption using transiently-expressed TALENs [Liu et al., 2012], suggesting TALENs are also operative and useful in the fly. However, to date, this study remains the sole experimental publication employing TALENs in *Drosophila*, although a second study evaluated TALEN use in *Drosophila* as a proof of utility [Sakuma et al., 2013].

We chose to synthesize TALENs targeted to the loxP remnant. Because the remnant is palindromic, we were able to design a single TALEN target sequence that provides correct target and spacer lengths, allowing Fok1 homodimerization and targeted cleavage within the loxP (Figure 4.5).

4.2.3 TALEN design

To date, all studies using TALENs have injected TALEN-encoding mRNA into cells (or syncytial embryos, in the case of *Drosophila*), allowing for transient TALEN

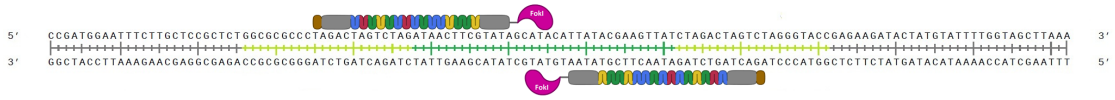


Figure 4.5: Targeting the loxP for disruption using TALENs. Because the loxP remnant (yellow/green) is palindromic, we targeted a single TALEN to this region, creating double-stranded breaks in the loxP when the Fok1 dimerizes. Image courtesy of Marianna Neubauer and Kasia Sierzputowska.

expression. For several reasons we choose to express the loxP-targeting TALEN ("loxP-TALEN") directly from the *Drosophila* genome. First, traditional elegant *Drosophila* genetics often involves crossing flies to generate appropriate *in vivo* combinations of alleles. In keeping with the litany of genetic tools used in the fly, a genomically-expressed TALEN could easily be crossed in and out of a population. Secondly, stable genomic expression of the loxP-TALEN would allow researchers using homologous recombination in the fly to cross in the TALEN, if loxP disruption is desired. In this way, the loxP-TALEN would become a tool extending the use of HR in *Drosophila*. This would also mean that laboratories unequipped for embryo injection could keep a stable stock of loxP-TALEN animals without the need to constantly inject loxP-TALEN mRNA into different *Drosophila* lines. Finally, the stable genomic expression of a TALEN from a genome has yet to be characterized. The loxP, as a foreign DNA remnant, presents a unique opportunity to safely express a TALEN from the *Drosophila* genome, allowing us to characterize any unintended consequences from genomic expression without targeting endogenous sequence.

When designing our TALEN strategy we decided to take advantage of the HR process. To use the loxP-TALEN in "post-Cre" individuals, whose genomes contain a single loxP, would require genotyping. Instead, we chose to use "pre-Cre" animals, whose genomes still contain the *mini-white* phenotypic marker flanked by two identical loxP sites (Figure 1.17). In this experimental design the loxP-TALEN is predicted to target *both* loxP sites, removing the *mini-white* marker and resulting in an easily-visible mosaic-eyed animal while simultaneously disrupting the loxP

integrity. In addition, TALENs have yet to be used to target an entire gene for removal rather than disruption.

4.2.4 TALEN vector construction

Because of our above experimental design, we constructed a vector using the pUAST backbone. The vector contains the TALEN gene, assembled using the AddGene Golden Gate TALEN Assembly Kit [Cermak et al., 2011], an EGFP phenotypic marker, and P-element ends for genomic insertion (Figure 4.7). First the pUAST vector was modified from containing the *mini-white* phenotypic marker to containing the EGFP marker under control of an eye-specific 3X-P3 promoter. The 3X-P3 and EGFP sequences were kindly donated by Dr. Koen Venken (Baylor College of Medicine, Department of Biochemistry and Molecular Biology). These sequences were amplified from the donated plasmid using the primers indicated in Table 4.1 and inserted into the pUAST backbone in place of the *mini-white* using PciI.

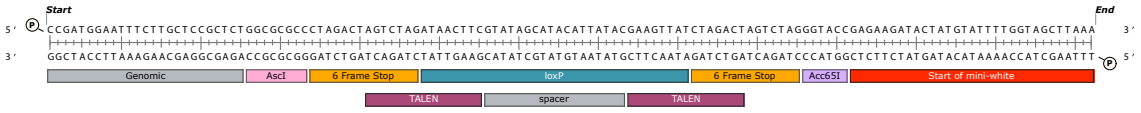


Figure 4.6: The loxP remnant and TALEN targeting sequence. The loxP remnant consists of the loxP sequence (green, 34 bp), flanking 6-frame stop codons (orange, 14 bp each), and AscI (pink, 8 bp) and Acc65I (lavender, 6 bp) sites. The targeted TALEN sequence (purple) spans the 6-frame stop and loxP sequences. The remnant upstream of the *mini-white* gene (red) is depicted.

The repetitive domain of the TALEN was designed to target the loxP remnant (Figure 4.6) The TAL repeats were assembled in the "pTAL2" backbone vector included in the AddGene Golden Gate TALEN Kit. The sequence encoding the Fok1 nuclease was captured from the provided TAL "pTAL3-His" plasmid using primers indicated in Table 4.1. The Fok1 was cloned into pTAL2 after the C-terminal TAL domain using BbvCI. The entire TALEN gene was then inserted into the EGFP pUAST vector using EcoRI. The TALEN gene was inserted under control of a UAS-

HSP 70 promoter and so will express constitutively in the presence of a Tublin-GAL4 driver.

Figure 4.7: Composition of the TALEN plasmid. The pUAST backbone encodes P-element ends (peach) for genomic insertion, and an EGFP marker (green) under control of the 3X-P3 eye-specific promoter (blue). The complete TALEN sequence (orange, yellow, and purple) is under control of a UAS-HSP 70 promoter (green and white). The EGFP was cloned into the pUAST backbone using PciI, while the TALEN sequence was added using EcoRI after BbvCI insertion of the Fok1 sequence.

4.3 Future directions

Once the loxP-TALEN EGFP-pUAST plasmid validation is complete it will be introduced into the *Drosophila* genome by injection (Genetic Services, Inc.). Progeny of injected flies will be scored for EGFP eyes, indicating successful random P-element

Table 4.1: Primers used in the synthesis of the loxP-TALEN EGFP vector. Underlines represent restriction sites added by primers.

	Forward (F) and Reverse (R) Primers	R. Enzyme
Fok1	**F:** CAAACACCGGATCAGGCGTCTTTG	BbvCI
	R: GGCTGAGGATTCAATCTTAAGAAAC	
EGFP	**F:** TACATGTCAAACATGAGAATTGGTCGAG	PciI
	R: TACATGTTATTGATCATAATCAGCCATACC	

genomic insertion. These animals will be mapped and stocked and used in the TALEN crossing scheme proposed in Figure 4.8. This scheme will introduce the tubulin-GAL4 driver, the UAS-TALEN, and loxP-containing *para* alleles (from "pre-Cre" flies) into the same animal. These flies will express the EGFP phenotypic marker in all eye ommatidia, while the action of the loxP-TALEN will excise the *mini-white* marker from some ommatidia, resulting in eyes that are mosaic for *mini-white* expression.

As loxP-TALEN expression will be driven by a ubiquitous tubulin driver, it will act on the genome not only in the eye, but also in the germline. F_2 progeny from the crossing scheme are predicted to have either red eyes, indicating that the TALEN did not act on the dual loxP sequences to excise the *mini-white* marker, or white eyes, indicating the intended action of the TALEN (Figure 4.8).

We hypothesize that TALENs may be effectively expressed from the *Drosophila* genome. This is the first study in which TALENs are not transiently expressed through mRNA injection, and so will provide an alternative method for TALEN introduction into an important model organism. Transient TALEN expression results in various imprecise indels and nucleotide additions to the targeted sequence [Joung and Sander, 2013]. It appears that TALENs continue to bind, cut, and rebind until all or most of their target sequence is deleted (Dr. James Gagnon, personal communication), resulting in somatic mosaic individuals. However, our strategy of driving TALEN expression from the genome followed by isolation of F_2 progeny (Figure 4.8) will result in animals that each represent a single allele generated from TALEN targeting. Therefore, characterizing TALEN action will be as simple as genotyping F_2 flies.

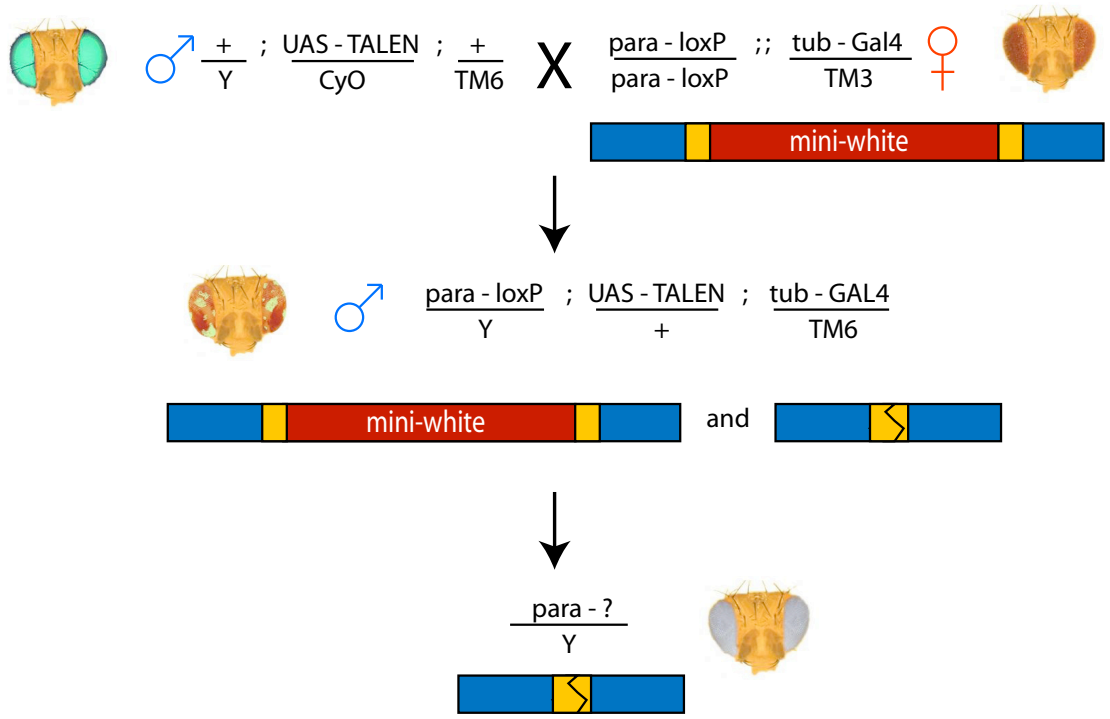


Figure 4.8: Proposed TALEN crossing scheme. "Pre-Cre" flies will be used because the *para* locus (blue) contains a *mini-white* gene (red) flanked by target loxP sites (orange). We can then score for the loss of red pigment, indicative of TALEN action. The scheme depicts a TALEN transgene on chromosome 2, followed by the EGFP adult eye marker, but TALENs on the X or on the 3rd could be used in similar crosses. A ubiquitous driver, such as tubulin-GAL4 on chromosome 3, drives TALEN expression in all cells throughout development, resulting in animals mosaic for *mini-white* expression. X-chromosomes from the F_1 generation are isolated by crossing to an Fm7 balancer line (not shown), and F_2 progeny lacking the red eye color will then be balanced and stocked for analysis. Image adapted from Kasia Sierzputowska.

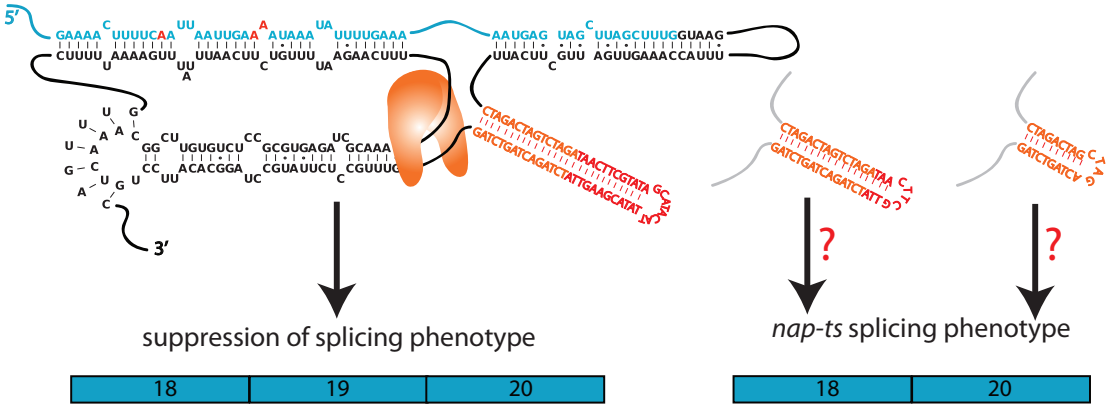


Figure 4.9: TALEN targeting of the loxP will generate animals with various degrees of loxP integrity. The intact loxP suppresses the splicing phenotype seen in Mle^{nap-ts} animals, while various loxP remnant deletions will allow us to map how much of the loxP is required for phenotype suppression.

This strategy will also generate animals whose *para* loci represent various degrees of loxP integrity (Figure 4.9). This will provide two benefits. First, it will allow us to map the loxP location that interferes with the Mle^{nap-ts} mutant helicase, possibly through direct binding. Secondly, we may target the loxP in animals containing *para* intronic mutations generated through HR. We may select loxP deletions that we know from control individuals do not suppress the splicing phenotype. We may then assay the interaction of the mutant helicase with the HR-generated structural mutants, without the interference of the intact loxP. In these experiments, we hypothesize that mutations that disrupt the wild type *para* structure (see **Chapter 2**) will rescue proper *para* splicing, while the DCS zip mutation, which results in an aberrant splicing pattern even in the presence of wild type Mle, will be synthetically lethal in the presence of Mle^{nap-ts} (Figure 4.1).

4.4 Methods

RNA splicing analysis

RNA was isolated and reverse-transcribed as in **Chapter 2**. cDNA was amplified via PCR using primers specified in Table 4.2

Table 4.2: Primers used in the mle^{nap-ts} splicing analysis.

Location	Forward (F) and Reverse (R) primers
Constit. exon 16	**F:**CCTTCTTCTTGGCCACCGTTGTCATCGGC
Alt. exon "N"	**R:** CCTCATGCCCTGCATACGGGACATGGCAC

TALEN construction

The TALEN was constructed using the Golden Gate TALEN Kit available through AddGene. The repetitive region was designed to target the sequence AGTCTA-GATAACTTC, as this sequence satisfies targeting guidelines [Cermak et al., 2011], and is present as a palindrome in the loxP remnant (see Figure 4.5), allowing us to synthesize only a single TALEN (Figure 4.6).

Due to kit design, backbone plasmid pTAL3-His contained the required Fok1 sequence and no convenient subcloning restriction sites, while backbone plasmid pTAL2 contained convenient EcoRI subcloning sites but no Fok1 sequence [Cermak et al., 2011]. We therefore amplified the Fok1 sequence from pTAL3-His using the primers indicated in Table 4.1. After the TAL repeats were assembled in the pTAL2 vector, the Fok1 sequence was added using BbvCI and T4 DNA ligase (Promega). The successful addition of Fok1 was verified by PCR and Sanger sequencing.

EGFP-pUAST construction

We replaced the majority of the *mini-white* gene in the pUAST P-element vector with an EGFP gene under control of a 3X-P3 promoter. The plasmid supplying the 3X-P3 and EGFP sequences was kindly donated by Dr. Koen Venken (Baylor College of Medicine, Department of Biochemistry and Molecular Biology). The relevant fragment was amplified from the supplied vector using the primers indicated in Table 4.1. The 3X-P3 and EGFP sequences were subcloned into the pUAST backbone using PciI and T4 DNA ligase. The addition of the EGFP was verified by PCR and Sanger sequencing.

Vector assembly

The complete TALEN fragment was subcloned out of pTAL2 and into EGFP-pUAST using EcoRI and T4 DNA ligase.

Bacterial strains and transformation

Vectors were grown in One Shot TOP10 Competent Cells (Invitrogen). Cells were transformed according to the protocol supplied with the cells. Colonies were grown on LB agar plates containing the appropriate antibiotic at each stage of vector construction [Cermak et al., 2011]. Several colonies were picked and grown in 3 mL of liquid LB medium containing the appropriate antibiotic. Plasmids were isolated from cultures with the PureYield Plasmid Miniprep system (Promega).

Chapter Five

Conclusions and Future Directions

5.0.1 RNA editing by ADARs

Post-transcriptional modifications, such as adenosine-to-inosine RNA editing and alternative splicing, diversify the proteome while limiting the necessary size of the genome. Although splicing globally rearranges existing information within the transcript, the conserved process of RNA editing recodes the message through single nucleotide changes, often at very specific transcript locations. Because inosine is interpreted as guanosine by the cellular machineries, editing effectively results in the substitution of a guanosine for an adenosine in the primary RNA sequence. A highly conserved group of enzymes, the adenosine deaminase acting on RNA (ADAR) family, mediates this reaction and is highly expressed in the metazoan nervous system [Savva et al., 2012]. Mutation or deletion of ADAR genes results in striking phenotypes in multiple model organisms, including seizure episodes, extreme uncoordination, and neurodegeneration.

Neuronal transcripts are specifically targeted for editing. These include mRNAs of proteins involved in electrical and chemical neurotransmission, including pre-synaptic release machineries, and voltage- and ligand-gated ion channels. One such transcript is *paralytic*, the only functional voltage-gated sodium channel gene in the *Drosophila* genome [Ganetzky, 1984]. While mammalian genomes contain nine sodium channel genes, insects achieve channel diversity through post-transcriptional modification of the single channel transcript [Goldin, 2002].

5.0.2 RNA editing of the *paralytic* transcript

We therefore investigated RNA editing in *Drosophila paralytic* to probe the extent of the RNA structures directing editing. Using homologous recombination, we surgically altered conserved *cis* elements associated with a cluster of ADAR modification sites within the endogenous *para* gene. In addition to the local requirement for a central imperfect RNA duplex containing the modified adenosines, we demonstrated that

a secondary RNA duplex containing the splice donor significantly modulates RNA editing. When this structure is destroyed, editing decreases, perhaps due to more rapid intron removal, while a subtle non-coding mutation, extending base pairing of this accessory helix, is predicted to stabilize the secondary structure and leads to higher levels of editing. This mutation also confers significant temperature-sensitive phenotypic consequences via effects on splicing, suggesting a complex relationship between these post-transcriptional processes.

Through mutation/counter-mutation, we also uncovered a highly conserved intronic long-range tertiary pseudoknot that is absolutely required for deamination of one particular adenosine in the central duplex. Replacement of this tertiary pseudoknot with sequences from a Group II self-splicing RNA kissing loop results in a functional structure that efficiently directs editing. Our results demonstrate that complex RNA tertiary structures, which may be difficult to predict computationally, form *in vivo* and can regulate RNA editing events. These structures may assemble at a considerable distance from the edited adenosines [Rieder et al., 2013].

5.0.3 The complex relationship between editing and splicing

Precise control of editing is dictated by duplex structures in the RNA, often formed between the exonic region surrounding the editing site and *cis* regulatory elements localized to a nearby intron, suggesting that editing must precede splicing [Rieder and Reenan, 2011], and additional recent evidence suggests widespread co-transcriptional editing [Rodriguez et al., 2012]. However, the precise relationship between the post- or co-transcriptional processes of splicing and editing remains unclear.

Paralytic again presents an ideal system in which to investigate this complex relationship, as the nap^{ts} mutation in the Maleless dsRNA helicase generates a "spicing catastrophe" around edited *para* intron 19 [Reenan et al., 2000]. It is possible that release of the complex secondary and tertiary RNA structures directing editing would

rescue the mle^{nap-ts} molecular and temperature-senstive phenotypes. Surprisingly, we discovered that the loxP remnant alone, which is predicted to form a perfect duplex in the RNA, rescues this splicing phenotype. We are therefore currently targeting a TALEN to the remnant with the goals of both loxP disruption, as well as development of the TALEN technology in *Drosophila*. This technology will hopefully generate animals with degrees of loxP disruption, allowing us to determine the extent of loxP structural integrity that interacts with the Mle^{nap-ts} mutant helicase and restores the splicing catastrophe. Moreover, the loxP-targeting TALEN may be used in combination with HR-generated alleles at any locus, providing an accessory tool to homologous recombination in *Drosophila*.

5.0.4 Temperature effects on RNA editing

RNA structures fulfill numerous cellular roles at all levels of gene expression. RNA thermometers, for example, are prokaryotic structural elements that control translation by sequestering the ribosome binding site and start codon within secondary structure [Narberhaus et al., 2006]. In response to temperature changes, these structures rearrange to allow translation. The RNA structure therefore acts as a molecular sensor, eliciting cellular responses in response to abiotic factors.

We propose that RNA structures involved in editing act as "thermistors," modulating editing levels in response to temperature. Indeed, global editing is decreased at 30°C compared to 10 and 20°C, although individual editing sites behave differently in response to temperature shifts. This likely indicates a general melting of RNA structures at elevated temperature, but also reflects the range of RNA structures known to direct specific editing. While the response of editing to temperature is known to be adaptive in some cold blooded species [Garrett and Rosenthal, 2012], it is not known if the temperature-sensitivity of editing in poikilothermic *Drosophila* is adaptive or if it instead represents a breakdown of the editing system, either the

signals in the form of RNA structure or the dADAR editing machinery itself.

5.0.5 Sodium channel epilepsy mutations

Curiously, epilepsy-causing mutations in the human sodium channel gene *SCN1A* occur throughout the channel and confer varying degrees of symptom severity [Meisler and Kearney, 2005], while the only *Drosophila* homolog, *paralytic*, is edited throughout the transcript to achieve channel diversity. A GEFS+ (Generalized Epilepsy with Febrile Seizures Plus) mutation found in human patients (K1270T), when introduced into *Drosophila paralytic*, confers a temperature-sensitive seizure phenotype [Sun et al., 2012]. This particular GEFS+ mutation occurs on S2 of DIII (red on yellow helix in Figure 5.1) in the sodium channel protein, and is predicted to be in close steric proximity to the residue altered by the unique *Musca domestica* editing site (M1230V, bottom green on white helix in Figure 5.1).

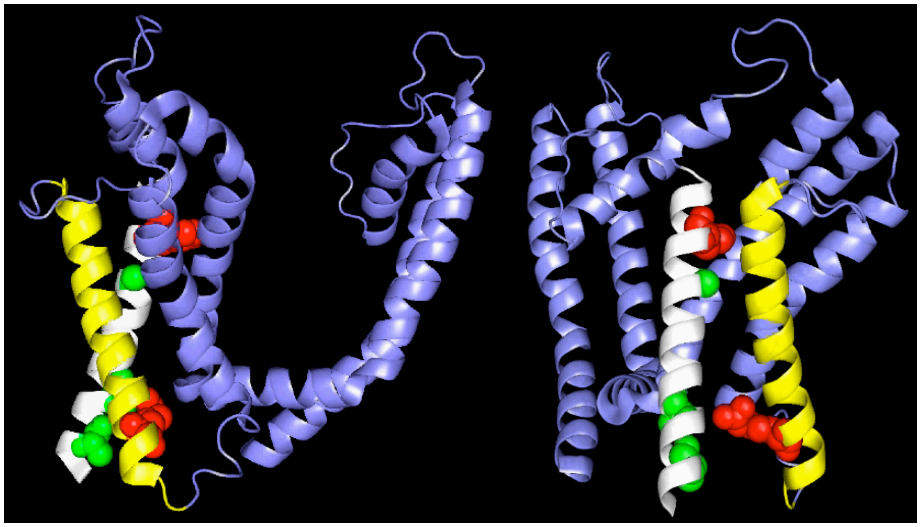


Figure 5.1: Steric relationship between residues affected by RNA editing and epilepsy mutations in the sodium channel. Two views of the six trans-membrane helices of Para DIII are depicted, S1 in white, S2 in yellow, and S3-S4 in blue. Residues affected by RNA editing in *para* exon 19 are shown in green. The top sites are found in *Drosophilidae* (Q1213R, N1217D), while the bottom (M1230V) is unique to *M. domestica* (Figure 2.14). An amino acid substitution found in GEFS+ patients (K1270T, red) is localized to S2. In addition, an amino acid substitution (red) leading to a more severe form of epilepsy, SMEI (Severe Myoclonic Epilepsy of Infancy, or Dravet Syndrome), localizes to S1 close to the residues affected by editing in *Drosophilidae*. Figure courtesy of Dr. Robert Reenan.

The residues (Q1213R, N1217D) affected by the *Drosophila paralytic* editing sites studied in **Chapter 2** localize to S1 of DIII (upper green on white helix in Figure 5.1). While an amino acid substitution (W1204R) leading to a more severe epilepsy presentation, SMEI (Severe Myoclonic Epilepsy of Infancy, or Dravet Syndrome), is present in S1 within fifteen amino acids of the edited residues (red on white helix in Figure 5.1). This observation raises the intriguing possibility of steric interaction between the edited residues and the residues involved in epilepsy.

Further, the GEFS+ mutation was introduced into the endogenous locus via HR and is downstream of the loxP. This feature allows easy Cre-mediated *trans* recombination (**Appendix A**) with any *cis* element or editing mutation upstream of the loxP. Combining these mutations into a single *para* allele would reveal synergistic or antagonistic effects of changes conferred by editing on the epilepsy phenotype. Although behavioral assays were not sensitive enough to detect any subtle phenotype conferred by mutations in *para* editing, these changes might be revealed in the background of the temperature-sensitive phenotype.

Appendix A

A new hammer in the homologous recombination toolbox: Cre-mediated *trans* recombination

Homologous recombination (HR) is widely used in mouse and yeast genetics, but rarely in *Drosophila* engineering. Our lab has recently streamlined HR [Staber et al., 2011], and the use of this important technique is rapidly expanding. Often, as in this manuscript, we seek to introduce multiple mutations into a single genetic locus. These mutations are designed to fall on homology arms (please see Figure 1.17) bisected by a LoxP site. After HR the LoxP site remains in the genome as the single remnant of the HR process. Previously, the desire to combine mutations, which may reveal synergistic or antagonistic molecular interactions, necessitated creation of a completely new construct containing both mutations, followed by introduction of the construct into *Drosophila* embryos and the subsequent HR process. However, we reasoned that mutations on opposite targeting arms (within the same gene) might be recombined *in vivo* using Cre Recombinase (Figure A.1 **a**). "Cre-mediated *trans* recombination" might easily allow for combination of mutations using simple *Drosophila* crossing, and would likely be useful in other model organisms in which HR is used.

The final step of HR produces male flies expressing Cre Recombinase and containing the targeted mutation. These flies lack the *mini-white* marker due to the action of Cre in the zygote or early embryo [Staber et al., 2011]. These flies, therefore, may be immediately used in Cre-mediated *trans* recombination, if the addition of another mutation is desired. To test this concept we used "DCS delete" (arm 1) male flies produced during HR and crossed them to females with a GEFS+ epilepsy mutation (Figure A.1 **b**). GEFS+ (Generalized Epilepsy with Febrile Seizures +) is a human epilepsy disorder, often caused by a mutation in the *SCN1A* sodium channel gene. One known human GEFS+ mutation has been introduced into *Drosophila paralytic* (K1270T) to create a fly epilepsy disease model [Sun et al., 2012]. We chose this mutation because it is located in arm 2 and introduces a SexAI restriction site. The DCS delete mutation similarly introduces a PacI site, and therefore the presence of both mutations may be scored by double digest (Figure A.1 **c**).

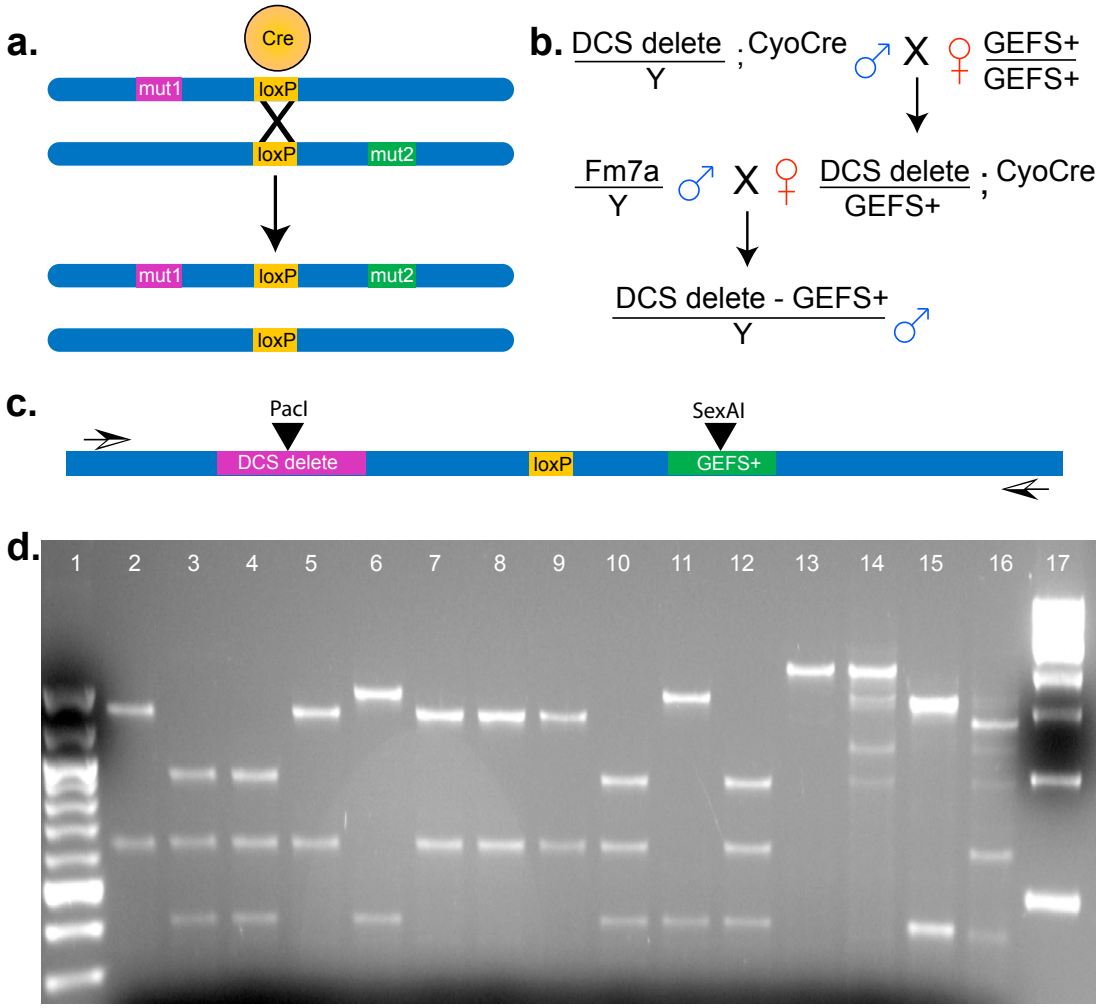


Figure A.1: Cre-mediated *trans* recombination. **a.** In the presence of Cre Recombinase (yellow), genomic material (blue) may be exchanged between chromosomes via the LoxP sites (yellow). This facilitates combining mutations (purple and green) introduced via HR within a single locus, provided they map to opposite arms. **b.** To test the theory of Cre-mediated *trans* recombination, we used the DCS delete mutation on *para* arm 1, which introduces a PacI restriction site, and the GEFS+ epilepsy mutation on *para* arm 2, which introduces a SexAI restriction site. Through crossing, we created transheterozygous females expressing Cre and assayed the genotypes of their sons via PCR and restriction digest (**c**). **d.** Proof of concept genotyping agarose gel for Cre-mediated *trans* recombination. Lane 1: 100 bp ladder. Lanes 2-13: progeny males from transheterozygous mothers. Lanes 14-16: control flies (14: wild type, 15: DCS delete single mutant, 16: GEFS+ single mutant). Lane 17: 1 Kb ladder

Results from this experiment are shown in Figure A.1 **d**, in which lanes 2-13 represent progeny males of unknown genotypes. Lanes 14-16 represent control animals (14: wild type X chromosome from a CyoCre individual, 15: DCS delete, 16: GEFS+). Lanes 3,4,10, and 12 represent animals in which the two mutations have been combined through Cre-mediated *trans* recombination. Lanes 2,5,7,8 and 9 represent males with the GEFS+ mutation, while lanes 6 and 11 show digest of the DCS delete mutation. Lane 13 represents a wild type chromosome with neither the GEFS+ or DCS delete mutation, which is also produced by Cre-mediated *trans* recombination (Figure A.1 **a**).

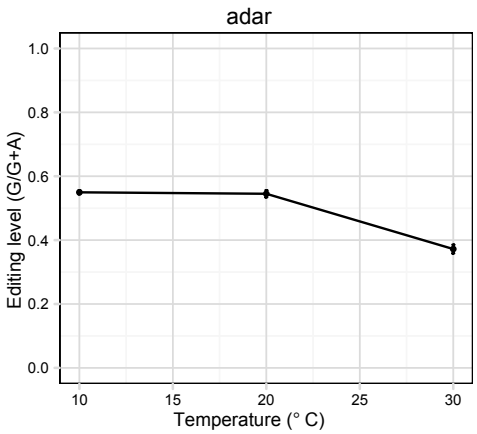
To my knowledge this is the first example of Cre inducing interchromosomal recombination that preserves overall chromosome sequence and order, although a similar technique has been used in mouse embryonic stem cells to exchange nonhomologous chromosome sequences [Collins et al., 2000, Smith et al., 1995, Van Deursen et al., 1995]. In fission yeast, the Cre/LoxP system has been used to demonstrate genomic organization, as interactions and LoxP-mediated recombinations between *nonhomologous* loci are dependent on proximity [Molnar and Kleckner, 2008, Burgess and Kleckner, 1999].

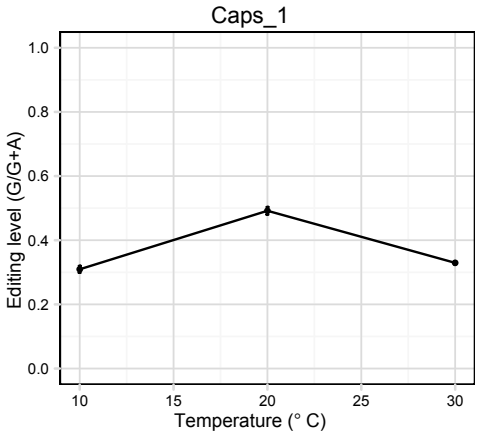
Cre-mediated *trans* recombination will be invaluable in *Drosophila* HR, as it may avoid many months of work to obtain interesting and informative double mutant alleles, which may reveal synergistic, antagonistic, or rescue interactions. Now that this technique has been demonstrated, HR targeting constructs may be specifically designed so that mutations thought to interact are contained on opposite targeting arms.

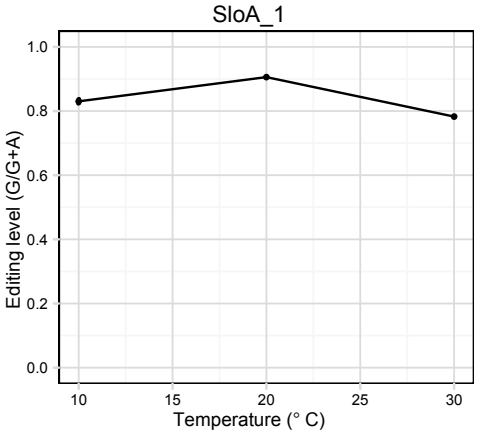
Appendix B

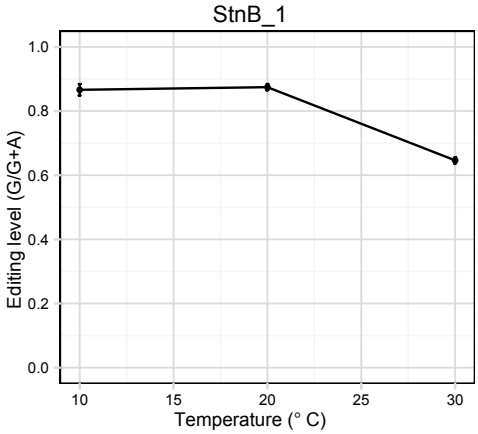
Temperature-responsiveness of individual editing sites

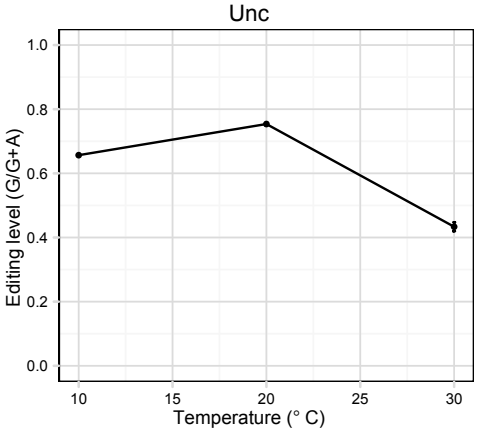
We investigated fifty-five editing sites in sixteen *Drosophila* transcripts that had been previously characterized [Savva et al., 2012]. The results of each site are presented below.

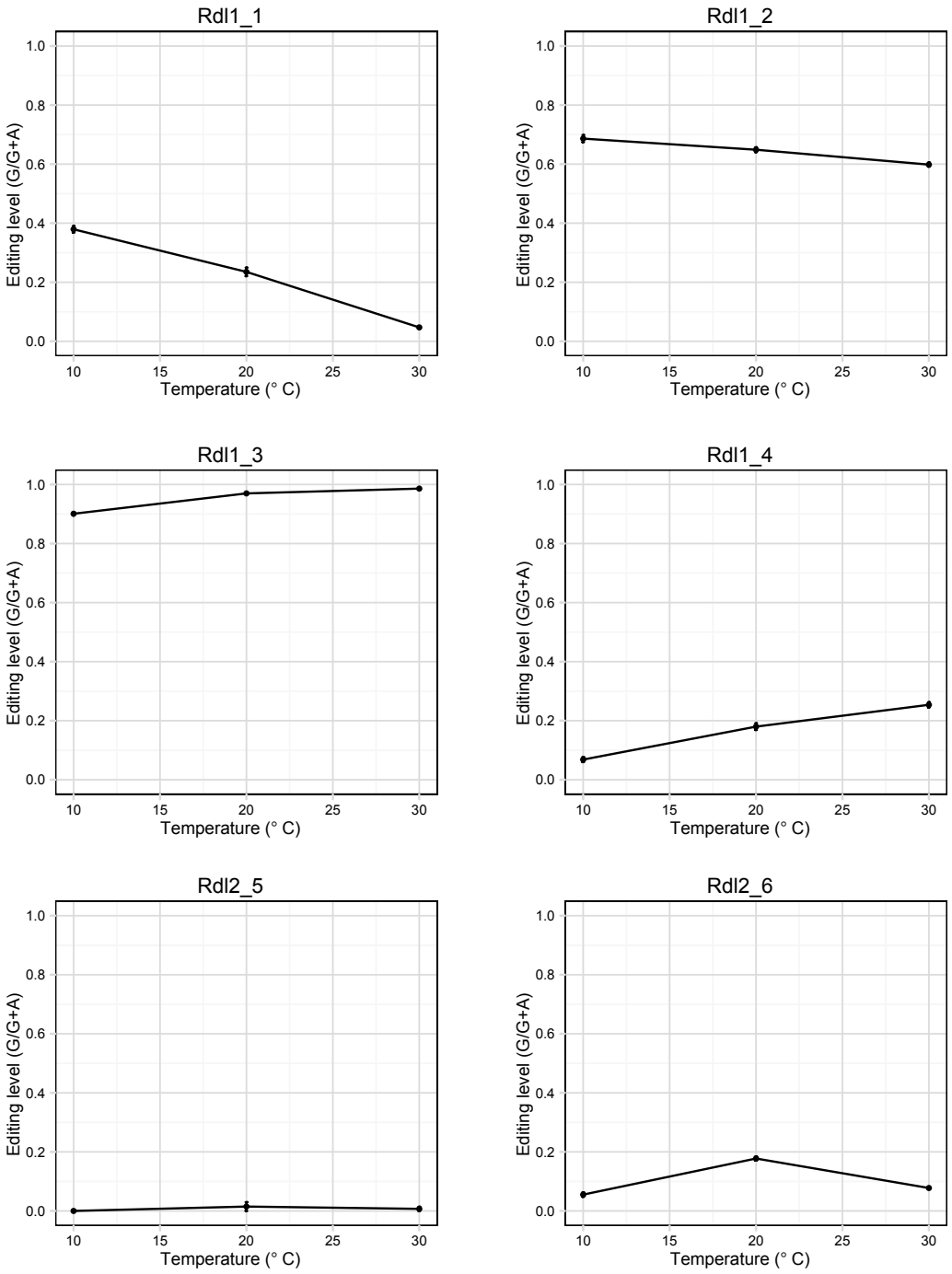


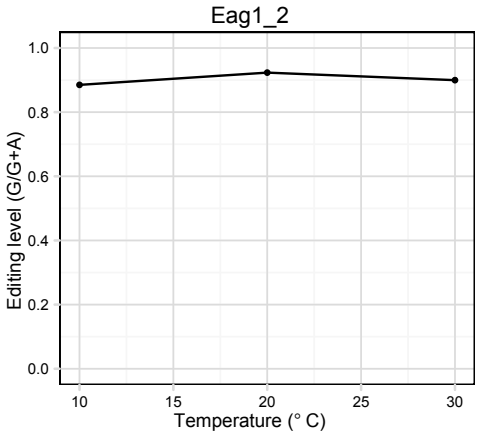


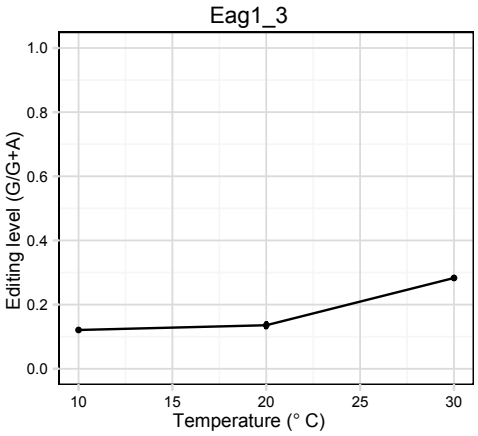


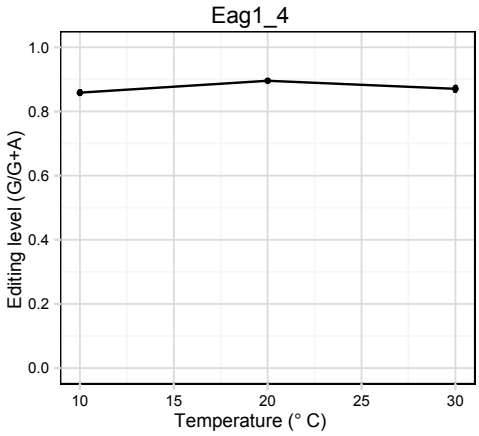


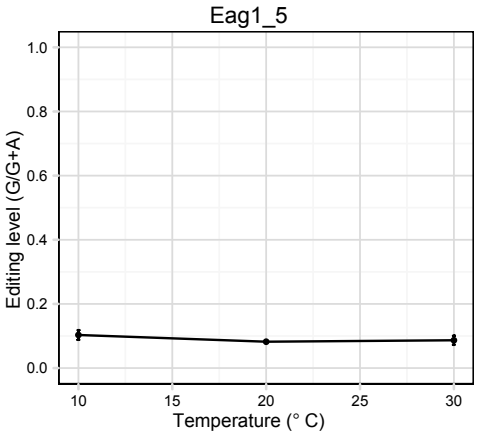


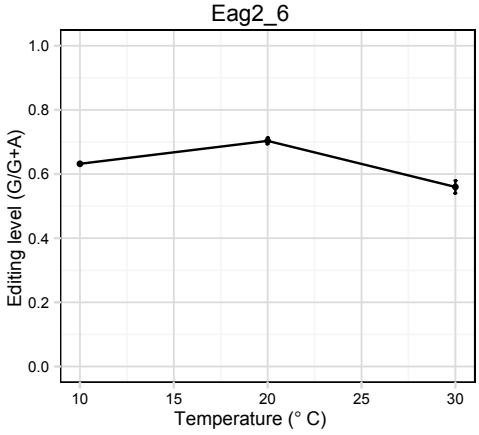

Rdl1_1
Rdl1_2
Rdl1_3
Rdl1_4
Rdl2_5
Rdl2_6
Editing level (G/G+A)
Temperature (° C)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30

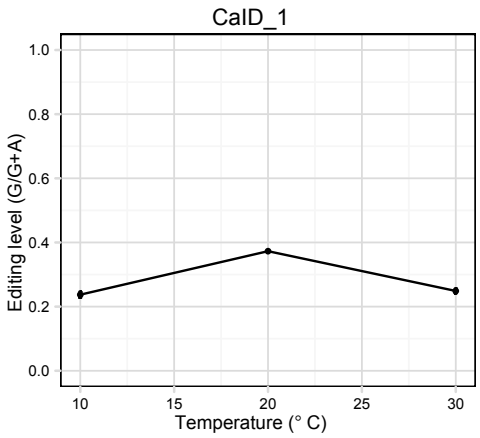

Eag1_2
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

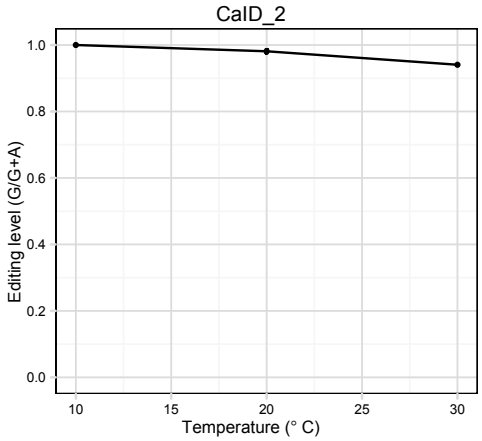

Eag1_3
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

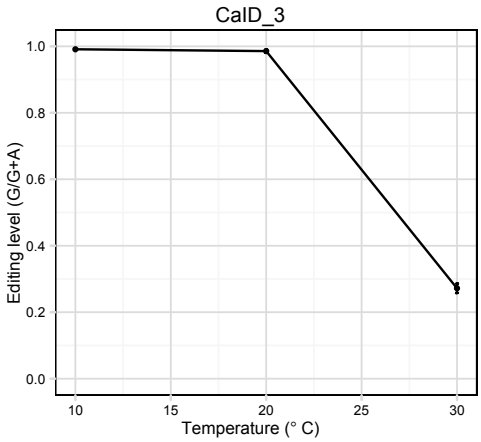

Eag1_4
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

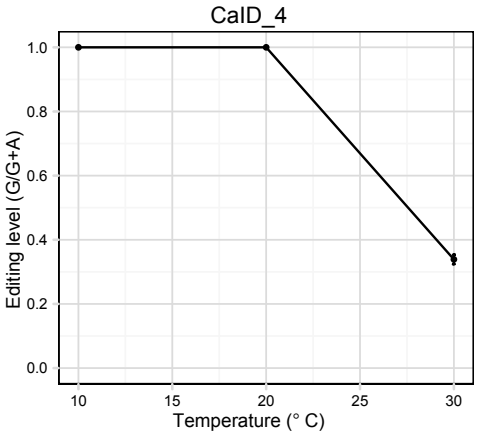

Eag1_5
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

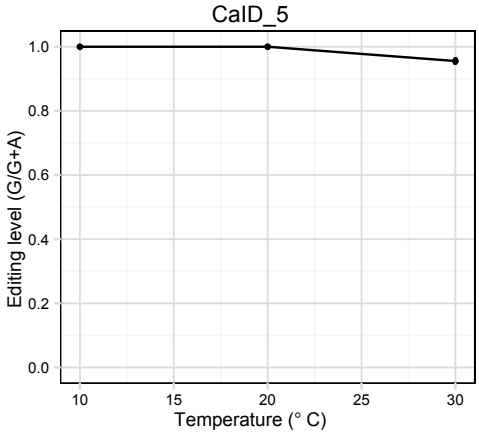

Eag2_6
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

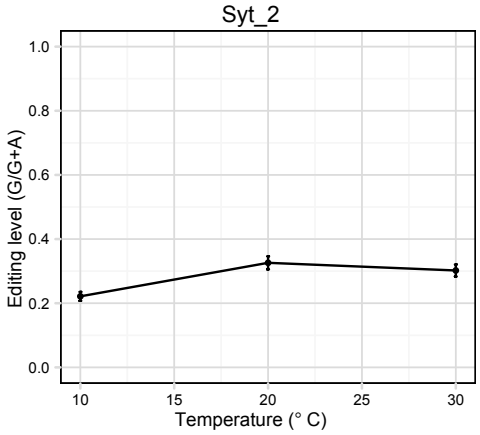

CaID_1
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

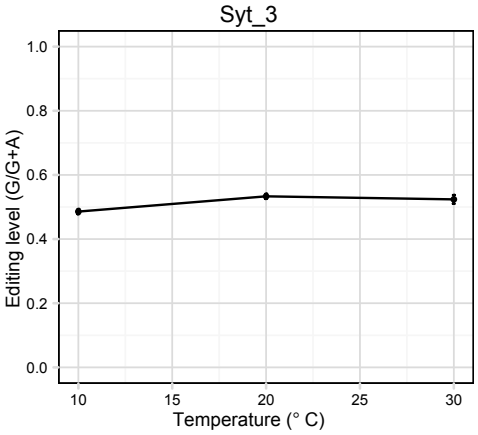

CaID_2
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

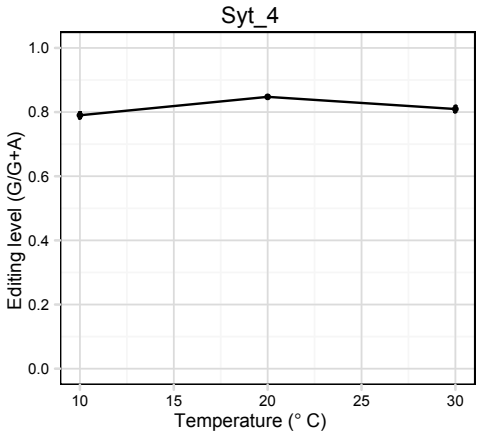

CaID_3
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

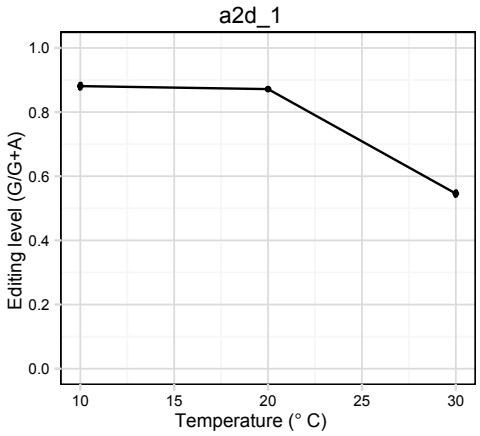

CaID_4
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

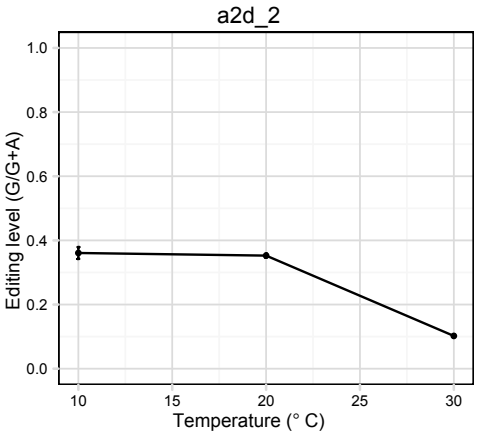

CaID_5
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

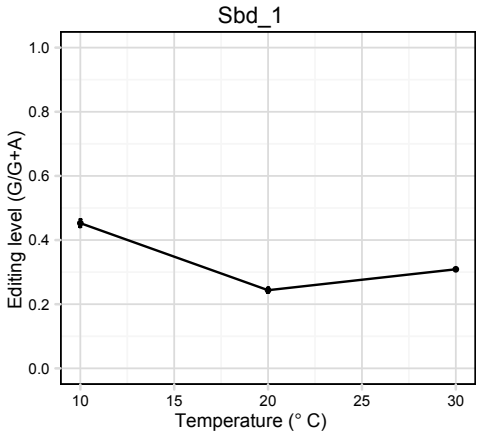

Syt_2
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

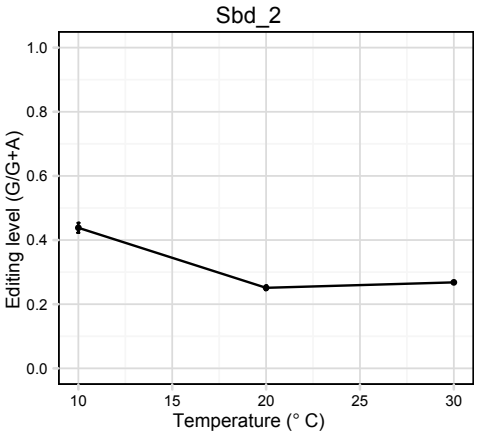

Syt_3
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

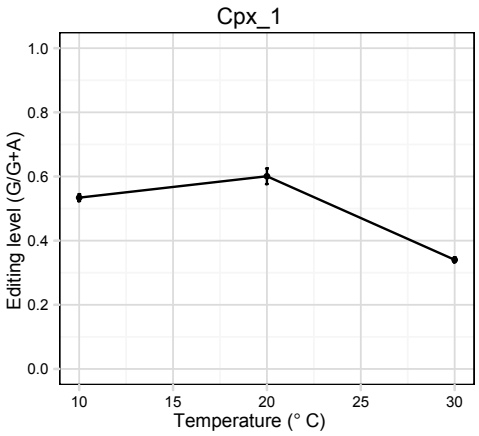

Syt_4
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

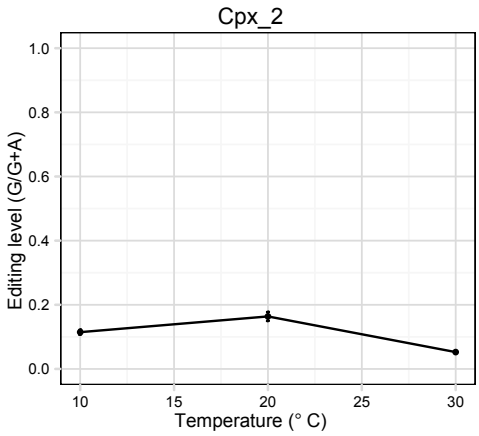

a2d_1
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

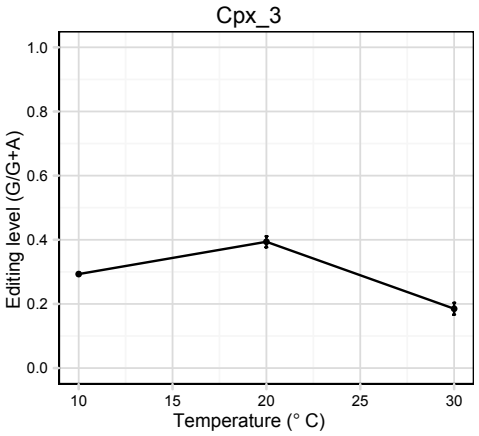

a2d_2
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

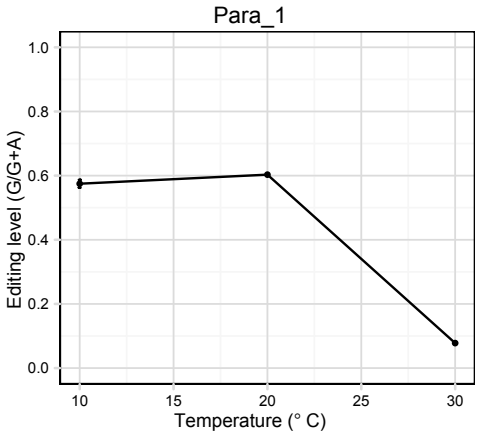

Sbd_1
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

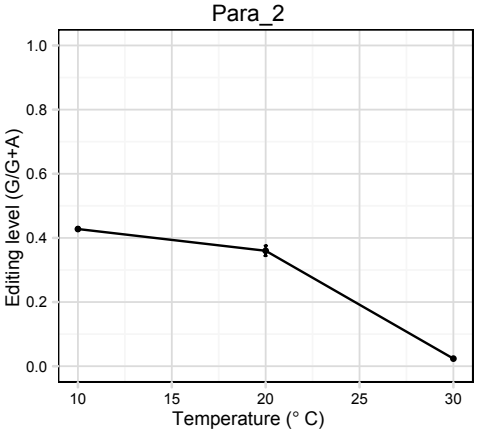

Sbd_2
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

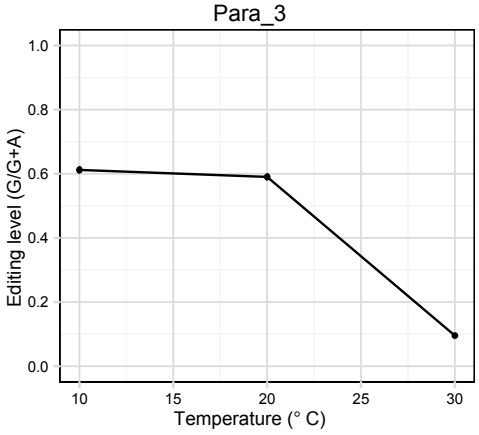

Cpx_1
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

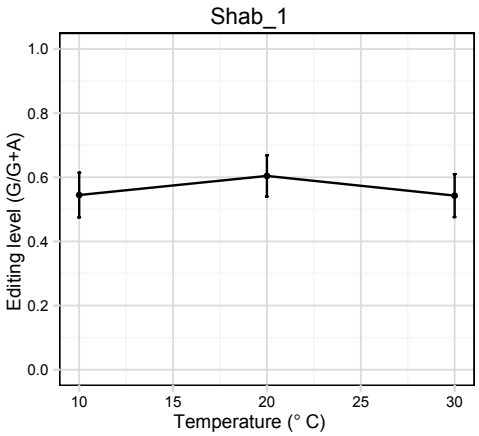

Cpx_2
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

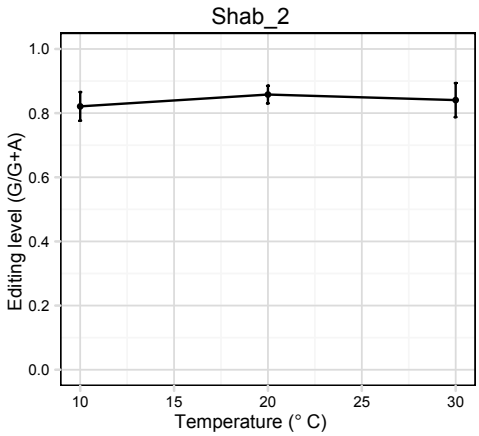

Cpx_3
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

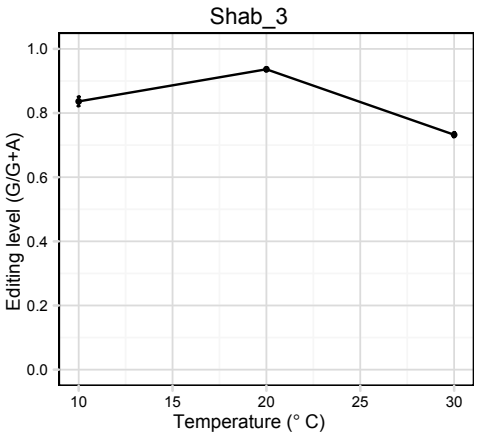

Para_1
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

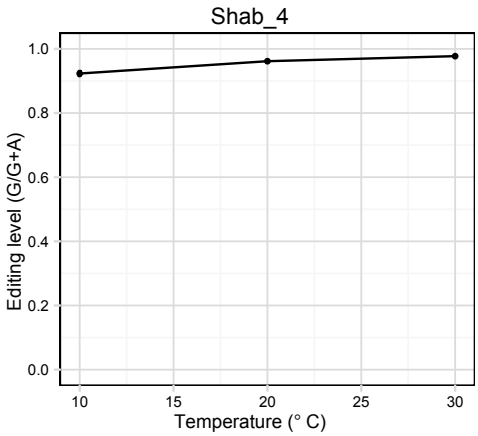

Para_2
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

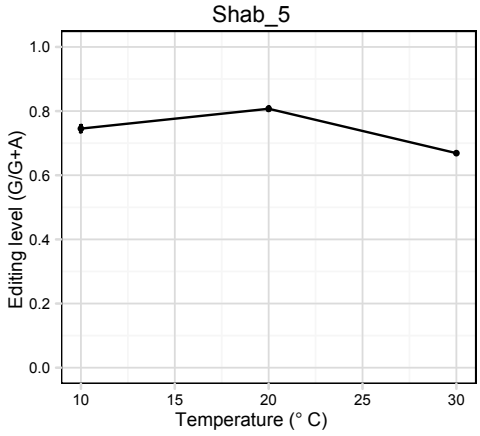

Para_3
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

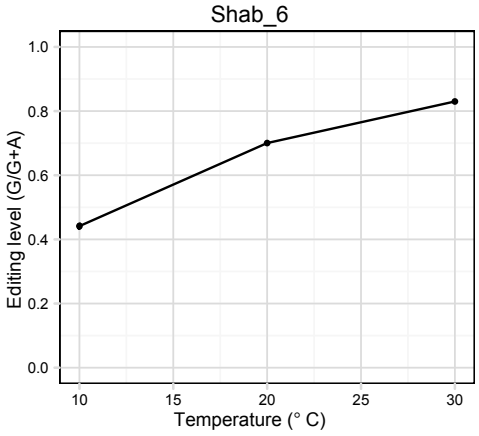

Shab_1
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

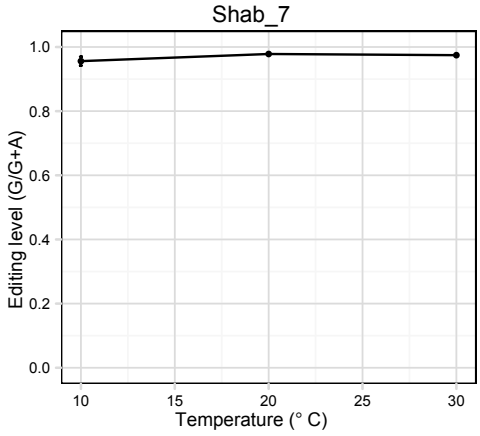

Shab_2
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

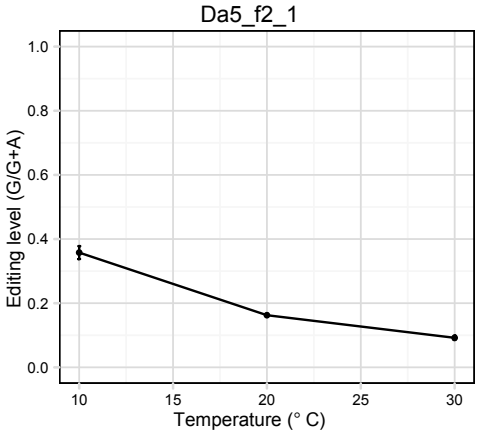

Shab_3
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

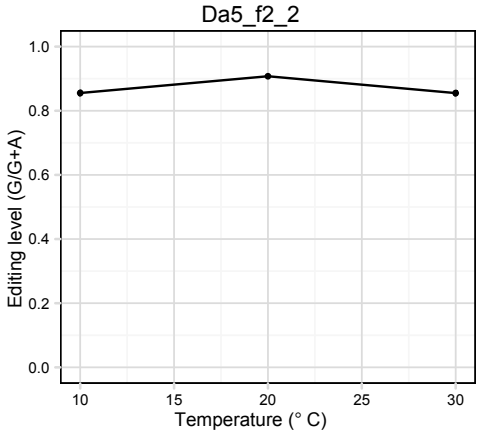

Shab_4
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

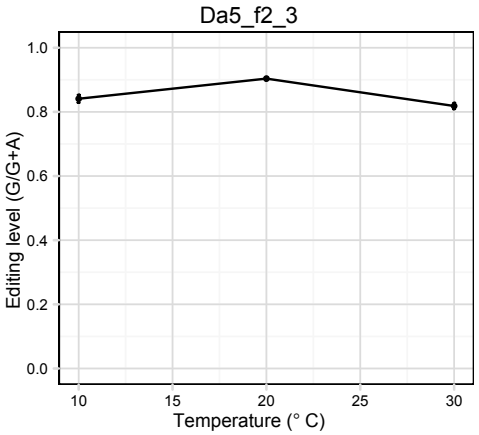

Shab_5
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

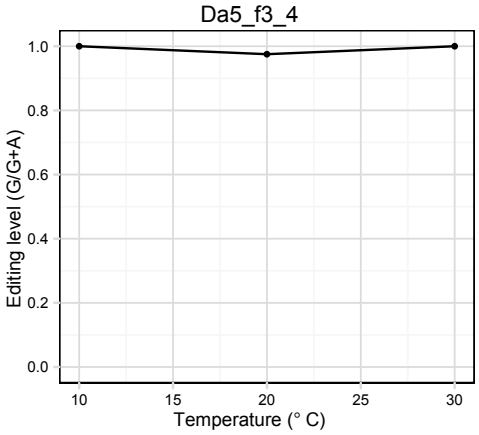

Shab_6
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

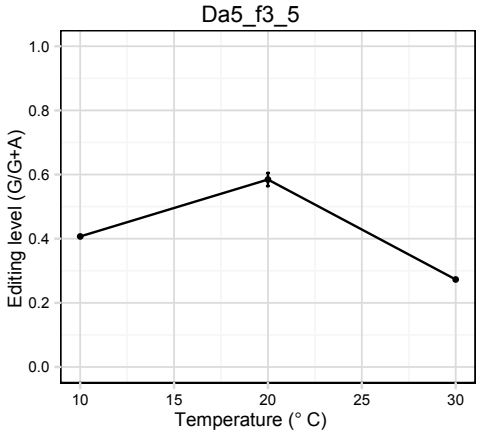

Shab_7
Editing level (G/G+A)
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
Temperature (° C)

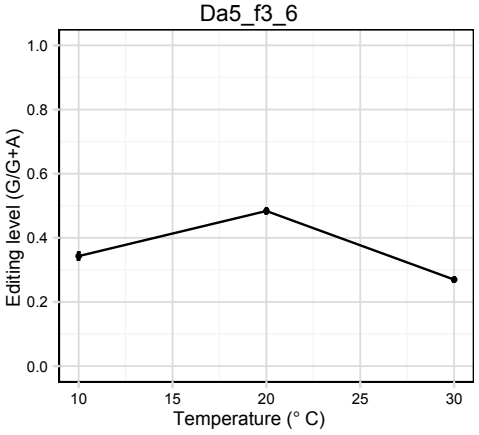

Da5_f2_1
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

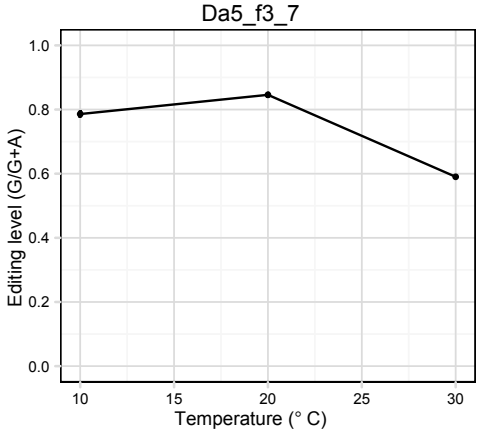

Da5_f2_2
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

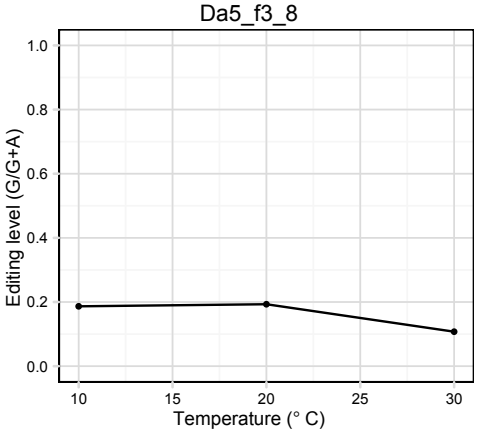

Da5_f2_3
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

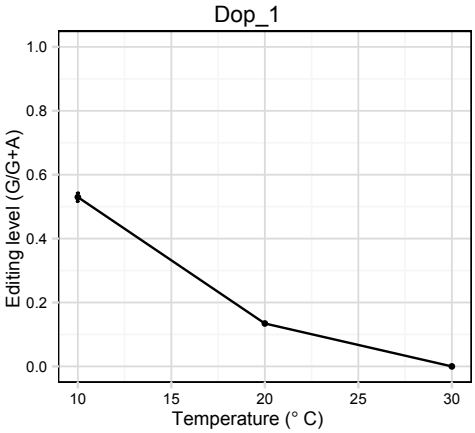

Da5_f3_4
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

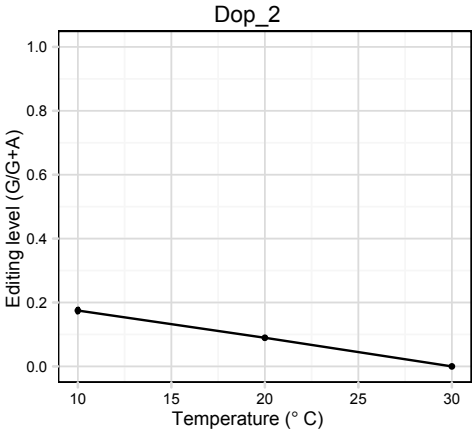

Da5_f3_5
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

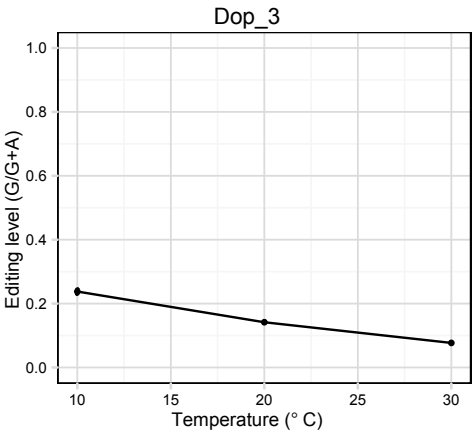

Da5_f3_6
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

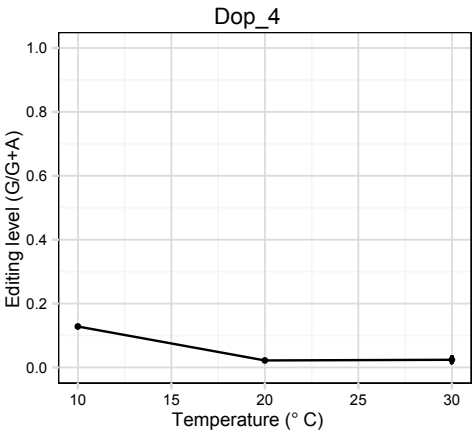

Da5_f3_7
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

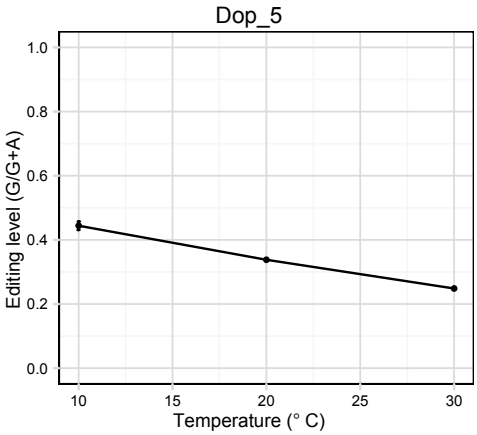

Da5_f3_8
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

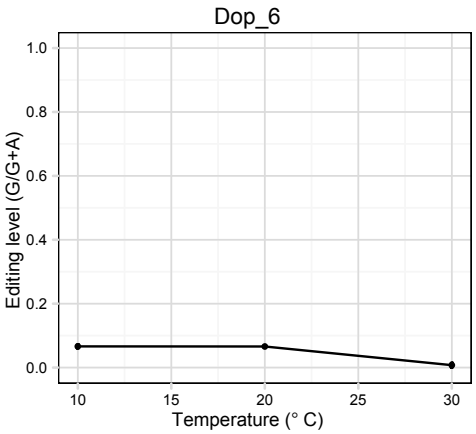

Dop_1
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)


Dop_2
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)


Dop_3
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)


Dop_4
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)


Dop_5
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)


Dop_6
Editing level (G/G+A)
1.0
0.8
0.6
0.4
0.2
0.0
10
15
20
25
30
Temperature (° C)

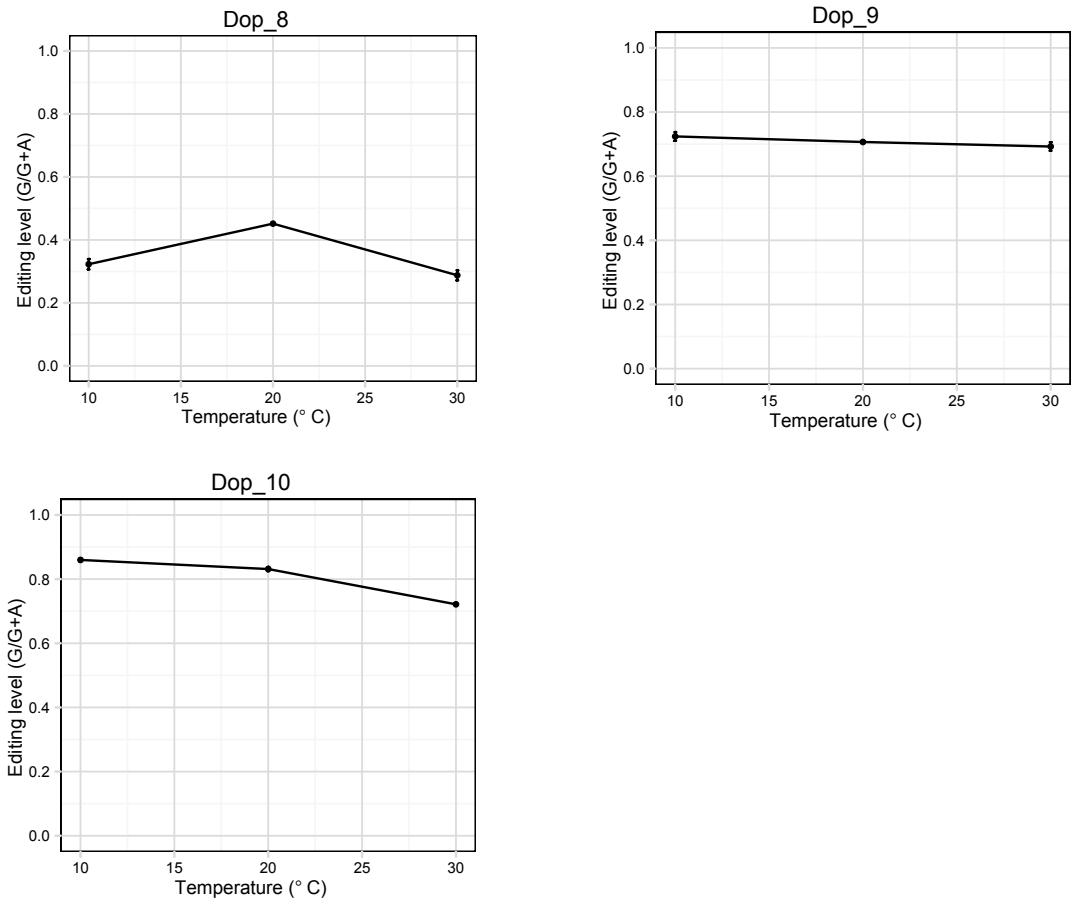


Figure B.1: Editing response to temperature of 55 individual *Drosophila* editing sites. Adr, (*dAdar*, the RNA editing enzyme). Caps (*calcium-activated protein for secretion*, a regulator of exocytosis). SloA (*slowpoke A*, a calcium-activated potassium channel). StnB (*stoned-B*, a regulator of endocytosis). Unc (*uncoordinated-13*, a regulator of exocytosis). Rdl (*resistant to dieldrin*, a GABA receptor α subunit). Eag (*ether a go-go*, a voltage-gated potassium channel α subunit). CaID (*Ca-alpha1D*, a voltage-gated calcium channel α subunit). Syt (*synaptogagmin-1*, a regulator of exocytosis. Ard (*ard*, a nicotinic acetylcholine receptor α subunit). Sbd (*Sbd*, a nicotinic acetylcholine receptor β subunit. Cpx (*complexin*, a regulator of exocytosis). Para (*paralytic*, a voltage-gated sodium channel α subunit). Shab (*shab*, a voltage-gated potassium channel). Da5 (*Dα5*, a nicotinic acetylcholine receptor α subunit). Da6 (*Dα6*, a nicotinic acetylcholine receptor α subunit). Dop (*DopEcR*, a Dopamine/Ecdysteroid receptor). Editing sites as in Savva et al. 2011.

Bibliography

[Agrawal and Stormo, 2005] Agrawal, R. and Stormo, G. D. (2005). Editing efficiency of a Drosophila gene correlates with a distant splice site selection. RNA *11*, 563–6.

[Anderson et al., 1998] Anderson, S., Schlegel, B., Nakajima, T., Wolpin, E. and Parvin, J. (1998). BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. Nature Genetics *19*, 254–256.

[Ansai et al., 2013] Ansai, S., Sakuma, T., Yamamoto, T., Ariga, H., Uemura, N., Takahashi, R. and Kinoshita, M. (2013). Efficient targeted mutagenesis in medaka using custom-designed transcription activator-like effector nucleases. Genetics *193*, 739–749.

[Aruscavage and Bass, 2000] Aruscavage, P. J. and Bass, B. L. (2000). A phylogenetic analysis reveals an unusual sequence conservation within introns involved in RNA editing. RNA *6*, 257–69.

[Athanasiadis et al., 2004] 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 et al., 2011] Bahn, J. H., Lee, J. H., Li, G., Greer, C., Peng, G. and Xiao, X. (2011). Accurate Identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res *22*.

[Baloh, 2012] Baloh, R. H. (2012). How do the RNA-binding proteins TDP-43 and FUS relate to amyotrophic lateral sclerosis and frontotemporal degeneration, and to each other? Curr Opin Neurol *25*, 701–7.

[Bartel, 2004] Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281–97.

[Basillo et al., 1962] Basillo, C., Wahba, A., Lengyel, P., Speyer, J. and Ochoa, S. (1962). Synthetic polynucleotides and the amino acid code. V. Proc Natl Acad Sci U S A *48*, 613–616.

[Bass et al., 2012] Bass, B., Hundley, H., Li, J. B., Peng, Z., Pickrell, J., Xiao, X. G. and Yang, L. (2012). The difficult calls in RNA editing. Nat Biotechnol *30*, 1207–9.

[Bass, 1997] Bass, B. L. (1997). RNA editing and hypermutation by adenosine deamination. Trends Biochem Sci *22*, 157–62.

[Bass, 2002] Bass, B. L. (2002). RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem *71*, 817–46.

[Bass and Weintraub, 1987] Bass, B. L. and Weintraub, H. (1987). A Developmentally Regulated Activity That Unwinds Rna Duplexes. Cell *48*, 607–613.

[Bedell et al., 2012] Bedell, V., Wang, Y., Campbell, J., Poshusta, T., Starker, C., Krug, R., Tan, W., Penheiter, S., Ma, A., Leung, A., Fahrenkrug, S., Carlson, D., Voytas, D., Clark, K., Essner, J. and Ekker, S. (2012). In vivo genome editing using a high-efficiency TALEN system. Nature *491*, 114–118.

[Bernstein et al., 2001] Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–6.

[Bhalla et al., 2004] Bhalla, T., Rosenthal, J. J., Holmgren, M. and Reenan, R. (2004). Control of human potassium channel inactivation by editing of a small mRNA hairpin. Nat Struct Mol Biol *11*, 950–6.

[Blackburn and Collins, 2011] Blackburn, E. H. and Collins, K. (2011). Telomerase: an RNP enzyme synthesizes DNA. Cold Spring Harb Perspect Biol *3*.

[Blow et al., 2004] Blow, M., Futreal, P. A., Wooster, R. and Stratton, M. R. (2004). A survey of RNA editing in human brain. Genome Res *14*, 2379–87.

[Blow et al., 2006] Blow, M. J., Grocock, R. J., van Dongen, S., Enright, A. J., Dicks, E., Futreal, P. A., Wooster, R. and Stratton, M. R. (2006). RNA editing of human microRNAs. Genome Biol *7*, R27.

[Boch et al., 2009] Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509–1512.

[Bonetti and Carninci, 2012] Bonetti, A. and Carninci, P. (2012). Structure by sequencing: discovery of hidden messages in the noncoding RNA fraction. Mol Cell *48*, 153–5.

[Bratt and Ohman, 2003] Bratt, E. and Ohman, M. (2003). Coordination of editing and splicing of glutamate receptor pre-mRNA. RNA *9*, 309–18.

[Brierley et al., 2007] Brierley, I., Pennell, S. and Gilbert, R. J. (2007). Viral RNA pseudoknots: versatile motifs in gene expression and replication. Nat Rev Microbiol *5*, 598–610.

[Brink et al., 1993] Brink, M. F., Verbeet, M. P. and de Boer, H. A. (1993). Formation of the central pseudoknot in 16S rRNA is essential for initiation of translation. EMBO J *12*, 3987–96.

[Buratti and Baralle, 2004] Buratti, E. and Baralle, F. E. (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol *24*, 10505–14.

[Burgess and Kleckner, 1999] Burgess, S. and Kleckner, N. (1999). Collisions between yeast chromosomal loci in vivo are governed by three layers of organization. Genes and Development *13*, 1871–1883.

[Burnett et al., 2012] Burnett, J. C., Zaia, J. A. and Rossi, J. J. (2012). Creating genetic resistance to HIV. Curr Opin Immunol *24*, 625–32.

[Burns et al., 1997] 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*, 303–8.

[Bycroft et al., 1995] Bycroft, M., Grunert, S., Murzin, A. G., Proctor, M. and St Johnston, D. (1995). NMR solution structure of a dsRNA binding domain from Drosophila staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. Embo J *14*, 3563–71.

[Carlson et al., 2012] Carlson, D., Tan, W., Lillico, S., Stverakova, D., Proudfoot, C., Christian, M., Voytas, D., Long, C., Whitelaw, B. and Fahrenkrug, S. (2012). Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci U S A *109*, 17382–17387.

[Casey, 2011] Casey, J. L. (2011). Control of ADAR1 Editing of Hepatitis Delta Virus RNAs. Curr Top Microbiol Immunol *353*.

[Cattaneo, 1994] Cattaneo, R. (1994). Biased (A–¿I) hypermutation of animal RNA virus genomes. Curr Opin Genet Dev *4*, 895–900.

[Cermak et al., 2011] Cermak, T., Doyle, E., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J., Somia, N., Bogdanove, A. and Voytas, D. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Research *39*, e82.

[Chen and Garcia, 2012] Chen, A. A. and Garcia, A. E. (2012). Mechanism of enhanced mechanical stability of a minimal RNA kissing complex elucidated by nonequilibrium molecular dynamics simulations. Proc Natl Acad Sci U S A *109*, E1530–9.

[Chen et al., 2000] Chen, C. X., Cho, D. S., 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*, 755–67.

[Chen et al., 2008] Chen, L. L., DeCerbo, J. N. and Carmichael, G. G. (2008). Alu element-mediated gene silencing. EMBO J *27*, 1694–705.

[Chen et al., 2013] Chen, S., Oikonomou, G., Chin, C., Niles, B., Liu, J., Lee, D., Antosheckin, I. and Prober, D. (2013). A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. Nucleic Acids Res *41*, 2769–2778.

[Cho et al., 2003] 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*, 17093–102.

[Chung et al., 2008] Chung, W. J., Okamura, K., Martin, R. and Lai, E. C. (2008). Endogenous RNA interference provides a somatic defense against Drosophila transposons. Curr Biol *18*, 795–802.

[Cisse et al., 2012] Cisse, I., Kim, H. and Ha, T. (2012). A rule of seven in Watson-Crick base-pairing of mismatched sequences. Nat Struct Mol Biol *19*, 623–7.

[Collins et al., 2000] Collins, E. C., Pannell, R., Simpson, E. M., Forster, A. and Rabbitts, T. H. (2000). Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. EMBO Rep *1*, 127–32.

[Czech et al., 2008] Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J. A., Sachidanandam, R., Hannon, G. J. and Brennecke, J. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798–802.

[Daniel et al., 2012] 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*, 9876–86.

[de la Mata et al., 2003] de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D. and Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell *12*, 525–32.

[Decher et al., 2010] Decher, N., Streit, A. K., Rapedius, M., Netter, M. F., Marzian, S., Ehling, P., Schlichthorl, G., Craan, T., Renigunta, V., Kohler, A., Dodel, R. C., Navarro-Polanco, R. A., Preisig-Muller, R., Klebe, G., Budde, T., Baukrowitz, T. and Daut, J. (2010). RNA editing modulates the binding of drugs and highly unsaturated fatty acids to the open pore of Kv potassium channels. EMBO J *29*, 2101–13.

[Desterro et al., 2003] Desterro, J. M., Keegan, L. P., Lafarga, M., Berciano, M. T., O'Connell, M. and Carmo-Fonseca, M. (2003). Dynamic association of RNA-editing enzymes with the nucleolus. J Cell Sci *116*, 1805–18.

[Dillon et al., 2009] Dillon, M., Wang, G., Garrity, P. and Huey, R. (2009). Thermal preference in Drosophila. Journal of Thermal Biology *34*, 109–119.

[Ding et al., 2013] Ding, Q., Lee, Y., Schaefer, E., Peters, D., Veres, A., Kim, K., Kuperwasser, N., Motola, D., Meissner, T., Hendriks, W., Trevisan, M., Gupta, R., Moisan, A., Banks, E., Friesen, M., Schinzel, R., Xia, F., Tang, A., Xia, Y., Figueroa, E., Wann, A., Ahfeldt, T., Daheron, L., Zhang, F., Rubin, L., Peng, L., Chung, R., Musunuru, K. and Cowan, C. (2013). A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell *12*, 238–251.

[Ding et al., 2004] Ding, Y., Chan, C. Y. and Lawrence, C. E. (2004). Sfold web server for statistical folding and rational design of nucleic acids. Nucleic Acids Res *32*, W135–41.

[Dingledine et al., 1999] Dingledine, R., Borges, K., Bowie, D. and Traynelis, S. F. (1999). The glutamate receptor ion channels. Pharmacol Rev *51*, 7–61.

[Doria et al., 2009] Doria, M., Neri, F., Gallo, A., Farace, M. G. and Michienzi, A. (2009). Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. Nucleic Acids Res *37*, 5848–58.

[Dracheva et al., 2008] Dracheva, S., Patel, N., Woo, D. A., Marcus, S. M., Siever, L. J. and Haroutunian, V. (2008). Increased serotonin 2C receptor mRNA editing: a possible risk factor for suicide. Mol Psychiatry *13*, 1001–10.

[Draper, 1995] Draper, D. E. (1995). Protein-RNA recognition. Annu Rev Biochem *64*, 593–620.

[Eckmann et al., 2001] Eckmann, C. R., Neunteufl, A., Pfaffstetter, L. and Jantsch, M. F. (2001). The human but not the Xenopus RNA-editing enzyme ADAR1 has an atypical nuclear localization signal and displays the characteristics of a shuttling protein. Mol Biol Cell *12*, 1911–24.

[Economos and Lints, 1986] Economos, A. C. and Lints, F. A. (1986). Developmental temperature and life span in Drosophila melanogaster. II. Oscillating temperature. Gerontology *32*, 28–36.

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

[Eisenberg, 2011] Eisenberg, E. (2011). Bioinformatic Approaches for Identification of A-to-I Editing Sites. Curr Top Microbiol Immunol *353*.

[Eisenberg et al., 2005] Eisenberg, E., Nemzer, S., Kinar, Y., Sorek, R., Rechavi, G. and Levanon, E. Y. (2005). Is abundant A-to-I RNA editing primate-specific? Trends in Genetics *21*, 77–81.

[Enstero et al., 2009] Enstero, M., Daniel, C., Wahlstedt, H., Major, F. and Ohman, M. (2009). Recognition and coupling of A-to-I edited sites are determined by the tertiary structure of the RNA. Nucleic Acids Res *37*, 6916–26.

[Fagegaltier et al., 2009] Fagegaltier, D., Bouge, A. L., Berry, B., Poisot, E., Sismeiro, O., Coppee, J. Y., Theodore, L., Voinnet, O. and Antoniewski, C. (2009). The endogenous siRNA pathway is involved in heterochromatin formation in Drosophila. Proc Natl Acad Sci U S A *106*, 21258–63.

[Feng et al., 1995] Feng, G., Deak, P., Chopra, M. and Hall, L. M. (1995). Cloning and functional analysis of TipE, a novel membrane protein that enhances Drosophila para sodium channel function. Cell *82*, 1001–11.

[Feng et al., 2006] Feng, Y., Sansam, C. L., Singh, M. and Emeson, R. B. (2006). Altered RNA editing in mice lacking ADAR2 autoregulation. Mol Cell Biol *26*, 480–8.

[Filbin et al., 2013] Filbin, M., Vollmar, B., Shi, D., Gonen, T. and Kleft, J. (2013). HCV IRES manipulates the ribosome to promote the switch from translation initiation to elongation. Nat Struct Mol Biol *20*, 150–158.

[Fischer et al., 2008] Fischer, S. E., Butler, M. D., Pan, Q. and Ruvkun, G. (2008). Trans-splicing in C. elegans generates the negative RNAi regulator ERI-6/7. Nature *455*, 491–6.

[Fuller-Pace, 2000] Fuller-Pace, F. (2000). DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic Acids Research *34*, 4206–4215.

[Galeano et al., 2011] Galeano, F., Tomaselli, S., Locatelli, F. and Gallo, A. (2011). A-to-I RNA editing: The "ADAR" side of human cancer. Semin Cell Dev Biol *23*.

[Gallo et al., 2003] Gallo, A., Keegan, L. P., Ring, G. M. and O'Connell, M. A. (2003). An ADAR that edits transcripts encoding ion channel subunits functions as a dimer. EMBO J *22*, 3421–30.

[Ganetzky, 1984] Ganetzky, B. (1984). Genetic studies of membrane excitability in Drosophila: lethal interaction between two temperature-sensitive paralytic mutations. Genetics *108*, 897–911.

[Garber et al., 2012] Garber, G., Smith, L. A., Reenan, R. A. and Rogina, B. (2012). Effect of sodium channel abundance on Drosophila development, reproductive capacity and aging. Fly (Austin) *6*, 57–67.

[Garrett and Rosenthal, 2012] Garrett, S. and Rosenthal, J. J. (2012). RNA editing underlies temperature adaptation in K+ channels from polar octopuses. Science *335*, 848–51.

[Gelinas et al., 2011] Gelinas, J. F., Clerzius, G., Shaw, E. and Gatignol, A. (2011). Enhancement of replication of RNA viruses by ADAR1 via RNA editing and inhibition of RNA-activated protein kinase. J Virol *85*, 8460–6.

[Gerber and Keller, 2001] Gerber, A. P. and Keller, W. (2001). RNA editing by base deamination: more enzymes, more targets, new mysteries. Trends Biochem Sci *26*, 376–84.

[Ghildiyal et al., 2008] Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z. and Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077–81.

[Gilbert, 1986] Gilbert, D. M. (1986). Temporal order of replication of Xenopus laevis 5S ribosomal RNA genes in somatic cells. Proc Natl Acad Sci U S A *83*, 2924–8.

[Glanzer et al., 2005] Glanzer, J., Miyashiro, K. Y., Sul, J. Y., Barrett, L., Belt, B., Haydon, P. and Eberwine, J. (2005). RNA splicing capability of live neuronal dendrites. Proc Natl Acad Sci U S A *102*, 16859–64.

[Goguel et al., 1993] Goguel, V., Wang, Y. and Rosbash, M. (1993). Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. Mol Cell Biol *13*, 6841–8.

[Goldin, 2002] Goldin, A. L. (2002). Evolution of voltage-gated Na(+) channels. J Exp Biol *205*, 575–84.

[Goldstrohm et al., 2001] Goldstrohm, A. C., Greenleaf, A. L. and Garcia-Blanco, M. A. (2001). Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. Gene *277*, 31–47.

[Gommans et al., 2009] Gommans, W. M., Mullen, S. P. and Maas, S. (2009). RNA editing: a driving force for adaptive evolution? BioEssays *31*, 1137–45.

[Gong and Golic, 2003] Gong, W. J. and Golic, K. G. (2003). Ends-out, or replacement, gene targeting in Drosophila. Proc Natl Acad Sci U S A *100*, 2556–61.

[Graveley, 2001] Graveley, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. Trends Genet *17*, 100–7.

[Graveley et al., 2011] Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., Artieri, C. G., van Baren, M. J., Boley, N., Booth, B. W., Brown, J. B., Cherbas, L., Davis, C. A., Dobin, A., Li, R., Lin, W., Malone, J. H., Mattiuzzo, N. R., Miller, D., Sturgill, D., Tuch, B. B., Zaleski, C., Zhang, D., Blanchette, M., Dudoit, S., Eads, B., Green, R. E., Hammonds, A., Jiang, L., Kapranov, P., Langton, L., Perrimon, N., Sandler, J. E., Wan, K. H., Willingham, A., Zhang, Y., Zou, Y., Andrews, J., Bickel, P. J., Brenner, S. E., Brent, M. R., Cherbas, P., Gingeras, T. R., Hoskins, R. A., Kaufman, T. C., Oliver, B. and Celniker, S. E. (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473–9.

[Gu et al., 2012] Gu, T., Buaas, F. W., Simons, A. K., Ackert-Bicknell, C. L., Braun, R. E. and Hibbs, M. A. (2012). Canonical A-to-I and C-to-U RNA editing is enriched at 3'UTRs and microRNA target sites in multiple mouse tissues. PLoS One *7*, e33720.

[Hammond and Wood, 2011] Hammond, S. M. and Wood, M. J. (2011). Genetic therapies for RNA mis-splicing diseases. Trends Genet *27*, 196–205.

[Hannon, 2002] Hannon, G. J. (2002). RNA interference. Nature *418*, 244–51.

[Hanrahan et al., 2000] Hanrahan, C. J., Palladino, M. J., Ganetzky, B. and Reenan, R. A. (2000). RNA editing of the Drosophila para Na(+) channel transcript. Evolutionary conservation and developmental regulation. Genetics *155*, 1149–60.

[Harris-Kerr et al., 1993] Harris-Kerr, C. L., Zhang, M. and Peebles, C. L. (1993). The phylogenetically predicted base-pairing interaction between alpha and alpha' is required for group II splicing in vitro. Proc Natl Acad Sci U S A *90*, 10658–62.

[Hideyama et al., 2012] Hideyama, T., Yamashita, T., Aizawa, H., Tsuji, S., Kakita, A., Takahashi, H. and Kwak, S. (2012). Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. Neurobiol Dis *45*, 1121–8.

[Higuchi et al., 2000] 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 et al., 1993] 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.

[Hodges et al., 2002] Hodges, D., Lee, D., Preston, C., Boswell, K., Hall, L. and O'Dowd, D. (2002). tipE regulates Na^+-dependent repetitive firing in Drosophila neurons. Molecular and Cellular Neuroscience *19*, 402–416.

[Hoffman, 2010] Hoffman, A. (2010). Physiological climatic limits in Drosophila: patterns and implications. The Journal of Experimental Biology *213*, 870–880.

[Hoopengardner et al., 2003] Hoopengardner, B., Bhalla, T., Staber, C. and Reenan, R. (2003). Nervous system targets of RNA editing identified by comparative genomics. Science *301*, 832–6.

[Huxley, 1959] Huxley, A. F. (1959). Ion movements during nerve activity. Ann N Y Acad Sci *81*, 221–46.

[Huynen et al., 1997] Huynen, M., Gutell, R. and Konings, D. (1997). Assessing the reliability of RNA folding using statistical mechanics. J Mol Biol *267*, 1104–12.

[Ingleby et al., 2009] Ingleby, L., Maloney, R., Jepson, J., Horn, R. and Reenan, R. (2009). Regulated RNA editing and functional epistasis in Shaker potassium channels. J Gen Physiol *133*, 17–27.

[Jacobs et al., 2009] Jacobs, M. M., Fogg, R. L., Emeson, R. B. and Stanwood, G. D. (2009). ADAR1 and ADAR2 expression and editing activity during forebrain development. Dev Neurosci *31*, 223–37.

[Jepson et al., 2011] Jepson, J. E., Savva, Y. A., Yokose, C., Sugden, A. U., Sahin, A. and Reenan, R. A. (2011). Engineered alterations in RNA editing modulate complex behavior in Drosophila: regulatory diversity of adenosine deaminase acting on RNA (ADAR) targets. J Biol Chem *286*, 8325–37.

[Jin et al., 2007] Jin, Y., Tian, N., Cao, J., Liang, J., Yang, Z. and Lv, J. (2007). RNA editing and alternative splicing of the insect nAChR subunit alpha6 transcript: evolutionary conservation, divergence and regulation. BMC Evol Biol *7*, 98.

[Jin et al., 2009] Jin, Y., Zhang, W. and Li, Q. (2009). Origins and evolution of ADAR-mediated RNA editing. IUBMB Life *61*, 572–8.

[Joung and Sander, 2013] Joung, J. and Sander, J. (2013). TALENs: a widely applicable technology for targeted genome editing. Nature Reviews *14*, 49–55.

[Kawahara et al., 2004] Kawahara, Y., Ito, K., Sun, H., Aizawa, H., Kanazawa, I. and Kwak, S. (2004). Glutamate receptors: RNA editing and death of motor neurons. Nature *427*, 801.

[Kawahara et al., 2007] 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*, 1137–40.

[Kawamura et al., 2008] Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T. N., Siomi, M. C. and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature *453*, 793–7.

[Keegan et al., 2005] Keegan, L. P., Brindle, J., Gallo, A., Leroy, A., Reenan, R. A. and O'Connell, M. A. (2005). Tuning of RNA editing by ADAR is required in Drosophila. EMBO J *24*, 2183–93.

[Kim et al., 2004] Kim, D. D., Kim, T. T., 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*, 1719–25.

[Kim et al., 2013] Kim, Y., Kweon, J., Kim, A., Chon, J., Yoo, J., Kim, H., Kim, S., Lee, C., Jeong, E., Chung, E., Kim, D., Lee, M., Go, E., Song, H., Kim, H., Cho, N., Bang, D., Kim, S. and Kim, J. (2013). A library of TAL effector nucleases spanning the human genome. Nat Biotechnol *31*, 251–258.

[Kuroda et al., 1991] Kuroda, M., Kernan, M., Kreber, R., Ganetzky, B. and Baker, B. (1991). The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila. Cell *66*, 935–947.

[Kuttan and Bass, 2012] Kuttan, A. and Bass, B. L. (2012). Mechanistic insights into editing-site specificity of ADARs. Proc Natl Acad Sci U S A *109*, E3295–304.

[Lachaise et al., 1988] Lachaise, D., Cariou, M., David, J., Lemeunier, F., Tsacas, L. and Ashburner, M. (1988). Evolutionary Biology vol. 22, chapter Historical Biogeography of the Drosophila-melanogaster species subgroup, pp. 159–225. New York: Plenum Press.

[Lalonde et al., 2011] Lalonde, E., Ha, K. C., Wang, Z., Bemmo, A., Kleinman, C. L., Kwan, T., Pastinen, T. and Majewski, J. (2011). RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. Genome Res , *21*, 545–54.

[Larschan et al., 2011] Larschan, E., Bishop, E. P., Kharchenko, P. V., Core, L. J., Lis, J. T., Park, P. J. and Kuroda, M. I. (2011). X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature , *471*, 115–8.

[Laurencikiene et al., 2006] Laurencikiene, J., Kallman, A. M., Fong, N., Bentley, D. L. and Ohman, M. (2006). RNA editing and alternative splicing: the importance of co-transcriptional coordination. EMBO Rep , *7*, 303–7.

[Lee et al., 1997] Lee, C., Chang, K., Kuroda, M. and Hurwitz, J. (1997). The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation. EMBO , *16*, 2671–2681.

[Lehmann and Bass, 2000] Lehmann, K. A. and Bass, B. L. (2000). Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry , *39*, 12875–84.

[Lev-Maor et al., 2007] 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 et al., 2004] Levanon, E. Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z. Y., Shoshan, A., Pollock, S. R., Sztybel, D., Olshansky, M., Rechavi, G. and Jantsch, M. F. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nature Biotechnology , *22*, 1001–1005.

[Li et al., 2009] Li, J. B., Levanon, E. Y., Yoon, J. K., Aach, J., Xie, B., Leproust, E., Zhang, K., Gao, Y. and Church, G. M. (2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science , *324*, 1210–3.

[Li et al., 2011] Li, M., Wang, I. X., Li, Y., Bruzel, A., Richards, A. L., Toung, J. M. and Cheung, V. G. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. Science , *333*, 53–8.

[Lin et al., 2009] Lin, W. H., Wright, D. E., Muraro, N. I. and Baines, R. A. (2009). Alternative splicing in the voltage-gated sodium channel DmNav regulates activation, inactivation, and persistent current. J Neurophysiol , *102*, 1994–2006.

[Littleton and Ganetzky, 2000] Littleton, J. T. and Ganetzky, B. (2000). Ion channels and synaptic organization: analysis of the Drosophila genome. Neuron , *26*, 35–43.

[Liu et al., 2012] Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., Zhu, N., Shen, Y., Chen, Y., Zhang, B., Deng, W. and Jiao, R. (2012). Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. Journal of Genetics and Genomics , *39*, 209–215.

[Liu et al., 1997] Liu, Y., George, C. X., Patterson, J. B. and Samuel, C. E. (1997). Functionally distinct double-stranded RNA-binding domains associated with alternative splice site variants of the interferon-inducible double-stranded RNA-specific adenosine deaminase. J Biol Chem , *272*, 4419–28.

[Lomeli et al., 1994] Lomeli, H., Mosbacher, J., Melcher, T., Hoger, T., Geiger, J. R., Kuner, T., Monyer, H., Higuchi, M., Bach, A. and Seeburg, P. H. (1994). Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. Science , *266*, 1709–13.

[Loughney et al., 1989] Loughney, K., Kreber, R. and Ganetzky, B. (1989). Molecular analysis of the para locus, a sodium channel gene in Drosophila. Cell , *58*, 1143–54.

[Luciano et al., 2004] Luciano, D. J., Mirsky, H., Vendetti, N. J. and Maas, S. (2004). RNA editing of a miRNA precursor. Rna , *10*, 1174–7.

[Luo and Reed, 1999] Luo, M. J. and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. Proc Natl Acad Sci U S A , *96*, 14937–42.

[Lyddon et al., 2012] Lyddon, R., Dwork, A. J., Keddache, M., Siever, L. J. and Dracheva, S. (2012). Serotonin 2c receptor RNA editing in major depression and suicide. World J Biol Psychiatry , *Epub*.

[Maas and Gommans, 2009] Maas, S. and Gommans, W. M. (2009). Identification of a selective nuclear import signal in adenosine deaminases acting on RNA. Nucleic Acids Res , *37*, 5822–9.

[Macbeth et al., 2005] 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*, 1534–9.

[Mandal and Breaker, 2004] Mandal, M. and Breaker, R. (2004). Gene regulation by riboswitches. Molecular Cell Biology , *5*, 451–463.

[Mantegazza et al., 2010] Mantegazza, M., Curia, G., Biagini, G., Ragsdale, D. S. and Avoli, M. (2010). Voltage-gated sodium channels as therapeutic targets in epilepsy and other neurological disorders. Lancet Neurol , *9*, 413–24.

[Marcucci et al., 2011] 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*.

[Marcucci et al., 2009] Marcucci, R., Romano, M., Feiguin, F., O'Connell, M. A. and Baralle, F. E. (2009). Dissecting the splicing mechanism of the Drosophila editing enzyme; dADAR. Nucleic Acids Res , *37*, 1663–71.

[Martins et al., 2011] Martins, S. B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J. M., de Almeida, S. F. and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. Nat Struct Mol Biol , *18*.

[Masliah et al., 2012] Masliah, G., Barraud, P. and Allain, F. H. (2012). RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. Cell Mol Life Sci , *Epub*.

[May et al., 2011] May, G., Olson, S., McManus, C. and Graveley, B. (2011). Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. RNA , *18*, 222–229.

[McManus and Graveley, 2011] McManus, C. J. and Graveley, B. R. (2011). RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev , *21*, 373–9.

[Meisler and Kearney, 2005] Meisler, M. H. and Kearney, J. A. (2005). Sodium channel mutations in epilepsy and other neurological disorders. J Clin Invest , *115*, 2010–7.

[Meister et al., 2004] Meister, G., Landthaler, M., Dorsett, Y. and Tuschl, T. (2004). Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. Rna , *10*, 544–50.

[Melcher et al., 1996] Melcher, T., Maas, S., Herb, A., Sprengel, R., Higuchi, M. and Seeburg, P. H. (1996). RED2, a brain-specific member of the RNA-specific adenosine deaminase family. J Biol Chem , *271*, 31795–8.

[Meyer et al., 2012] Meyer, L., Zweig, A., Karolchik, D., Kuhn, R., Wong, M., Sloan, C., Rosenbloom, K., Roe, G., Rhead, B., Raney, B., Pohl, A., Malladi, V., Li, C., Lee, B., Learned, K., Kirkup, V., Hsu, F., Heitner, S., Harte, R., Haeussler, M., Guruvadoo, L., Goldman, M., Giardine, B., Fujita, P., Dreszer, T., Diekhans, M., Cline, M., Clawson, H., Barber, G., Haussler, D. and Kent, W. (2012). The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res , *41*, D64–D69.

[Michel et al., 1989] Michel, F., Umesono, K. and Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns–a review. Gene , *82*, 5–30.

[Mironov et al., 2002] Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A. and Nudler, E. (2002). Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. Cell , *111*, 747–56.

[Molnar and Kleckner, 2008] Molnar, M. and Kleckner, N. (2008). Examination of interchromosomal interactions in vegetatively growing diploid Schizosaccharomyces pmobe cells by Cre/loxP site-specific recombination. Genetics , *178*, 99–112.

[Morra et al., 2008] Morra, R., Smith, E., Yokoyama, R. and Lucchesi, J. (2008). The MLE subunit of the Drosophila MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. Molecular and Cellular Biology , *28*, 958–966.

[Morse et al., 2002] 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 U S A , *99*, 7906–11.

[Moscou and Bogdanove, 2009] Moscou, M. and Bogdanove, A. (2009). A simple cipher governs DNA recognition by TAL effectors. Science , *326*, 1501.

[Munoz et al., 2010] Munoz, M. J., de la Mata, M. and Kornblihtt, A. R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem Sci , *35*, 497–504.

[Nahvi et al., 2002] Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L. and Breaker, R. R. (2002). Genetic control by a metabolite binding mRNA. Chem Biol , *9*, 1043.

[Nakajima et al., 1997] Nakajima, T., Uchida, C., Anderson, S., Lee, C., Hurwitz, J., Parvin, J. and Montminy, M. (1997). RNA helicase A mediates association of CBP with RNA polymerase II. Cell , *90*, 1107–1112.

[Nanduri et al., 1998] Nanduri, S., Carpick, B. W., Yang, Y., Williams, B. R. and Qin, J. (1998). Structure of the double-stranded RNA-binding domain of the

protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. Embo J , *17*, 5458–65.

[Narberhaus et al., 2006] Narberhaus, F., Waldminghaus, T. and Chowdhury, S. (2006). RNA thermometers. FEMS Microbiol Rev , *30*, 3–16.

[Nekrutenko, 2004] Nekrutenko, A. (2004). Reconciling the numbers: ESTs versus protein-coding genes. Mol Biol Evol , *21*, 1278–82.

[Nelson and Wyman, 1990] Nelson, J. C. and Wyman, R. J. (1990). Examination of paralysis in Drosophila temperature-sensitive paralytic mutations affecting sodium channels; a proposed mechanism of paralysis. J Neurobiol , *21*, 453–69.

[Neupert et al., 2008] Neupert, J., Karcher, D. and Bock, R. (2008). Design of simple synthetic RNA thermometers for temperature-controlled gene expression in Escherichia coli. Nucleic Acids Res , *36*, e124.

[Nielsen et al., 2006] Nielsen, M. L., Savitski, M. M. and Zubarev, R. A. (2006). Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. Mol Cell Proteomics , *5*, 2384–91.

[Nishikura, 2006] Nishikura, K. (2006). Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat Rev Mol Cell Biol , *7*, 919–31.

[Nishikura, 2010] Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem , *79*, 321–49.

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

[Nishikura et al., 1991] 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*, 3523–32.

[Ohlson et al., 2007] Ohlson, J., Pedersen, J. S., Haussler, D. and Ohman, M. (2007). Editing modifies the GABA(A) receptor subunit alpha3. Rna , *13*, 698–703.

[Ohman et al., 2000] 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*, 687–97.

[Olson et al., 2008] Olson, R. O., Liu, Z., Nomura, Y., Song, W. and Dong, K. (2008). Molecular and functional characterization of voltage-gated sodium channel variants from Drosophila melanogaster. Insect Biochem Mol Biol , *38*, 604–10.

[Palavicini et al., 2012] Palavicini, J. P., Correa-Rojas, R. A. and Rosenthal, J. J. (2012). Extra double-stranded RNA binding domain (dsRBD) in a squid RNA editing enzyme confers resistance to high salt environment. J Biol Chem , *287*, 17754–64.

[Palavicini et al., 2009] Palavicini, J. P., O'Connell, M. A. and Rosenthal, J. J. (2009). An extra double-stranded RNA binding domain confers high activity to a squid RNA editing enzyme. Rna , *15*, 1208–18.

[Paldi and Gurevitz, 2010] Paldi, T. and Gurevitz, M. (2010). Coupling between residues on S4 and S1 defines the voltage-sensor resting conformation in NaChBac. Biophys J , *99*, 456–63.

[Palladino et al., 2000] 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*, 437–49.

[Patterson and Samuel, 1995] 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*, 5376–88.

[Paz-Yaacov et al., 2010] Paz-Yaacov, N., Levanon, E. Y., Nevo, E., Kinar, Y., Harmelin, A., Jacob-Hirsch, J., Amariglio, N., Eisenberg, E. and Rechavi, G. (2010). Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. Proc Natl Acad Sci U S A , *107*, 12174–9.

[Perard et al., 2013] Perard, J., Leyrat, C., Baudin, F., Drouet, E. and Jamin, M. (2013). Structure of the full-length HCV IRES in solution. Nat Commun , *19*, 1612.

[Peters et al., 2003] Peters, N. T., Rohrbach, J. A., Zalewski, B. A., Byrkett, C. M. and Vaughn, J. C. (2003). RNA editing and regulation of Drosophila 4f-rnp expression by sas-10 antisense readthrough mRNA transcripts. Rna , *9*, 698–710.

[Plass et al., 2012] Plass, M., Codony-Servat, C., Ferreira, P., Vilardell, J. and Eyras, E. (2012). RNA secondary structure mediates alternative 3'ss selection in Saccharomyces cerevisiae. RNA , *18*, 1103–1115.

[Polson and Bass, 1994] Polson, A. G. and Bass, B. L. (1994). Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. Embo J , *13*, 5701–11.

[Polson et al., 1991] 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*, 11507–14.

[Poulsen et al., 2001] 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*, 7862–71.

[Pullirsch and Jantsch, 2010] Pullirsch, D. and Jantsch, M. F. (2010). Proteome diversification by adenosine to inosine RNA editing. RNA Biol , *7*, 205–12.

[Ramaswami and Tanouye, 1989] Ramaswami, M. and Tanouye, M. A. (1989). Two sodium-channel genes in Drosophila: implications for channel diversity. Proc Natl Acad Sci U S A , *86*, 2079–82.

[Rebagliati and Melton, 1987] Rebagliati, M. R. and Melton, D. A. (1987). Antisense Rna Injections in Fertilized Frog Eggs Reveal an Rna Duplex Unwinding Activity. Cell , *48*, 599–605.

[Reenan, 2005] Reenan, R. A. (2005). Molecular determinants and guided evolution of species-specific RNA editing. Nature , *434*, 409–13.

[Reenan et al., 2000] Reenan, R. A., Hanrahan, C. J. and Ganetzky, B. (2000). The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na+ channel transcript in a region of RNA editing. Neuron , *25*, 139–49.

[Renton et al., 2011] Renton, A. E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orrell, R. W., Neal, J., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peuralinna, T., Jansson, L., Isoviita, V. M., Kaivorinne, A. L., Holtta-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chio, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J. D., Sendtner, M., Drepper, C., Eichler, E. E., Alkan, C., Abdullaev, Z., Pack, S. D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N. M., Heutink, P., Pickering-Brown, S., Morris, H. R., Tienari, P. J. and Traynor, B. J. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron , *72*, 257–68.

[Rieder et al., 2013] Rieder, L., Staber, C., Hoopengardner, B. and Reenan, R. (2013). Tertiary structural elements determine extent and specificity of messenger RNA editing. Nature Communications , *accepted*.

[Rieder and Reenan, 2011] Rieder, L. E. and Reenan, R. A. (2011). The intricate relationship between RNA structure, editing, and splicing. Semin Cell Dev Biol , *23*, 281–8.

[Rieu and Powers, 2009] Rieu, I. and Powers, S. J. (2009). Real-time quantitative RT-PCR: design, calculations, and statistics. Plant Cell , *21*, 1031–3.

[Rinkevich et al., 2012] Rinkevich, F., Schweitzer, P. and Scott, J. (2012). Antisense sequencing improves the accuracy and precision of A-to-I editing measurements using the peak height ratio method. BMC Res Notes , *5*, 63.

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

[Rong and Golic, 2000] Rong, Y. and Golic, K. (2000). Gene targeting by homologous recombination in Drosophila. Science , *288*, 2013–2018.

[Rosenberg et al., 2011] Rosenberg, B. R., Dewell, S. and Papavasiliou, F. N. (2011). Identifying mRNA editing deaminase targets by RNA-Seq. Methods Mol Biol , *718*, 103–19.

[Rosenthal and Bezanilla, 2002] Rosenthal, J. J. and Bezanilla, F. (2002). Extensive editing of mRNAs for the squid delayed rectifier K+ channel regulates subunit tetramerization. Neuron , *34*, 743–57.

[Rosenthal and Seeburg, 2012] Rosenthal, J. J. and Seeburg, P. H. (2012). A-to-I RNA editing: effects on proteins key to neural excitability. Neuron , *74*, 432–9.

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

[Ryman et al., 2007] Ryman, K., Fong, N., Bratt, E., Bentley, D. L. and Ohman, M. (2007). The C-terminal domain of RNA Pol II helps ensure that editing precedes splicing of the GluR-B transcript. RNA , *13*, 1071–8.

[Ryter and Schultz, 1998] 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*, 7505–13.

[Sakuma et al., 2013] Sakuma, T., Hosoi, S., Woltjen, K., Suzuki, K., Kashiwagi, K., Wada, H., Ochiai, H., Miyamoto, T., Kawai, N., Sasakura, Y., Matsuura, S., Okada, Y., Kawahara, A., Hayashi, S. and Yamamoto, T. (2013). Efficient TALEN construction and evaluation methods for human cell and animal populations. Genes Cells , *18*, 315–326.

[Samuel, 2011] Samuel, C. E. (2011). ADARs: Viruses and Innate Immunity. Curr Top Microbiol Immunol , *353*.

[Sansam et al., 2003] Sansam, C. L., Wells, K. S. and Emeson, R. B. (2003). Modulation of RNA editing by functional nucleolar sequestration of ADAR2. Proc Natl Acad Sci U S A , *100*, 14018–23.

[Savva et al., 2012] Savva, Y. A., Jepson, J. E., Sahin, A., Sugden, A. U., Dorsky, J. S., Alpert, L., Lawrence, C. and Reenan, R. A. (2012). Auto-regulatory RNA editing fine-tunes mRNA re-coding and complex behaviour in Drosophila. Nat Commun , *3*, 790.

[Scadden, 2005] Scadden, A. D. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. Nat Struct Mol Biol , *12*, 489–96.

[Scadden and Smith, 2001] Scadden, A. D. and Smith, C. W. (2001). RNAi is antagonized by A–¿I hyper-editing. EMBO Rep , *2*, 1107–11.

[Schmucker et al., 2000] Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E. and Zipursky, S. L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell , *101*, 671–84.

[Schoft et al., 2007] Schoft, V. K., Schopoff, S. and Jantsch, M. F. (2007). Regulation of glutamate receptor B pre-mRNA splicing by RNA editing. Nucleic Acids Res , *35*, 3723–32.

[Seeburg, 2000] Seeburg, P. (2000). RNA helicase participates in the RNA editing game. Neuron , *25*, 261–263.

[Seeburg et al., 2001] Seeburg, P. H., Single, F., Kuner, T., Higuchi, M. and Sprengel, R. (2001). Genetic manipulation of key determinants of ion flow in glutamate receptor channels in the mouse. Brain Res , *907*, 233–43.

[Serganov and Nudler, 2013] Serganov, A. and Nudler, E. (2013). A decade of riboswitches. Cell , *152*, 17–24.

[Shah et al., 2009] Shah, S. P., Morin, R. D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J., Steidl, C., Holt, R. A., Jones, S., Sun, M., Leung, G., Moore, R., Severson, T., Taylor, G. A., Teschendorff, A. E., Tse, K., Turashvili, G., Varhol, R., Warren, R. L., Watson, P., Zhao, Y., Caldas, C., Huntsman, D., Hirst, M., Marra, M. A. and Aparicio, S. (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature , *461*, 809–13.

[Shepard and Hertel, 2008] Shepard, P. J. and Hertel, K. J. (2008). Conserved RNA secondary structures promote alternative splicing. RNA , *14*, 1463–9.

[Sixsmith and Reenan, 2007] Sixsmith, J. and Reenan, R. A. (2007). Comparative genomic and bioinformatic approaches for the identification of new adenosine-to-inosine substrates. Methods Enzymol , *424*, 245–64.

[Smith et al., 1995] Smith, A. J., De Sousa, M. A., Kwabi-Addo, B., Heppell-Parton, A., Impey, H. and Rabbitts, P. (1995). A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. Nat Genet , *9*, 376–85.

[Smith, 2004] Smith, L. (2004). Genetic interactions between splicing and RNA editing in Drosophila. PhD thesis,, University of Connecticut.

[Song et al., 2004] Song, W., Liu, Z., Tan, J., Nomura, Y. and Dong, K. (2004). RNA editing generates tissue-specific sodium channels with distinct gating properties. The Journal of Biological Chemistry , *279*, 32554–32561.

[Staber et al., 2011] Staber, C. J., Gell, S., Jepson, J. E. and Reenan, R. A. (2011). Perturbing A-to-I RNA editing using genetics and homologous recombination. Methods Mol Biol , *718*, 41–73.

[Stapleton et al., 2006] Stapleton, M., Carlson, J. W. and Celniker, S. E. (2006). RNA editing in Drosophila melanogaster: New targets and functional consequences. RNA , *12*, 1922–32.

[Stefl et al., 2010] Stefl, R., Oberstrass, F. C., Hood, J. L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R. B. and Allain, F. H. (2010). The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. Cell , *143*, 225–37.

[Stefl et al., 2005] Stefl, R., Skrisovska, L. and Allain, F. H. (2005). RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. EMBO Rep , *6*, 33–8.

[Stefl et al., 2006] 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*, 345–55.

[Stephens et al., 2004] Stephens, O. M., Haudenschild, B. L. and Beal, P. A. (2004). The binding selectivity of ADAR2's dsRBMs contributes to RNA-editing selectivity. Chem Biol , *11*, 1239–50.

[Stern et al., 1990] Stern, M., Kreber, R. and Ganetzky, B. (1990). Dosage effects of a Drosophila sodium channel gene on behavior and axonal excitability. Genetics , *124*, 133–43.

[Sudarsan et al., 2003] Sudarsan, N., Barrick, J. E. and Breaker, R. R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA , *9*, 644–7.

[Sun et al., 2012] Sun, L., Gilligan, J., Staber, C., Schutte, R. J., Nguyen, V., O'Dowd, D. K. and Reenan, R. (2012). A knock-in model of human epilepsy in Drosophila reveals a novel cellular mechanism associated with heat-induced seizure. J Neurosci , *32*, 14145–55.

[Sundaram et al., 2013] Sundaram, P., Kurniawan, H., Byrne, M. E. and Wower, J. (2013). Therapeutic RNA aptamers in clinical trials. Eur J Pharm Sci , *48*, 259–71.

[Sung et al., 2013] Sung, Y., Baek, I., Kim, D., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J. and Lee, H. (2013). Knockout mice created by TALEN-mediated gene targeting. Nat Biotechnol , *31*, 23–24.

[Suzuki et al., 1971] Suzuki, D. T., Grigliatti, T. and Williamson, R. (1971). Temperature-sensitive mutations in Drosophila melanogaster. VII. A mutation (para-ts) causing reversible adult paralysis. Proc Natl Acad Sci U S A , *68*, 890–3.

[Takuma et al., 1999] Takuma, H., Kwak, S., Yoshizawa, T. and Kanazawa, I. (1999). Reduction of GluR2 RNA editing, a molecular change that increases calcium influx through AMPA receptors, selective in the spinal ventral gray of patients with amyotrophic lateral sclerosis. Ann Neurol , *46*, 806–15.

[Tan et al., 2002] Tan, J., Liu, Z., Nomura, Y., Goldin, A. and Dong, K. (2002). Alternative splicing of an insect sodium channel gene generates pharmacologically distinct sodium channels. The Journal of Neuroscience , *22*, 5300–5309.

[Thackeray and Ganetzky, 1994] Thackeray, J. R. and Ganetzky, B. (1994). Developmentally regulated alternative splicing generates a complex array of Drosophila para sodium channel isoforms. J Neurosci , *14*, 2569–78.

[Thackeray and Ganetzky, 1995] Thackeray, J. R. and Ganetzky, B. (1995). Conserved alternative splicing patterns and splicing signals in the Drosophila sodium channel gene para. Genetics , *141*, 203–14.

[Tian et al., 2008] Tian, N., Wu, X., Zhang, Y. and Jin, Y. (2008). A-to-I editing sites are a genomically encoded G: implications for the evolutionary significance and identification of novel editing sites. RNA , *14*, 211–6.

[Tian et al., 2011] Tian, N., Yang, Y., Sachsenmaier, N., Muggenhumer, D., Bi, J., Waldsich, C., Jantsch, M. F. and Jin, Y. (2011). A structural determinant required for RNA editing. Nucleic Acids Res , *39*, 5669–81.

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

[Tonkin et al., 2002] Tonkin, L. A., Saccomanno, L., Morse, D. P., Brodigan, T., Krause, M. and Bass, B. L. (2002). RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans. EMBO J , *21*, 6025–35.

[Toor et al., 2008] Toor, N., Keating, K. S., Taylor, S. D. and Pyle, A. M. (2008). Crystal structure of a self-spliced group II intron. Science , *320*, 77–82.

[Tuerk and Gold, 1990] Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science , *249*, 505–10.

[Valencia et al., 2008] Valencia, P., Dias, A. P. and Reed, R. (2008). Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc Natl Acad Sci U S A , *105*, 3386–91.

[Valente and Nishikura, 2005] Valente, L. and Nishikura, K. (2005). ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. Prog Nucleic Acid Res Mol Biol , *79*, 299–338.

[Valente and Nishikura, 2007] Valente, L. and Nishikura, K. (2007). RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. J Biol Chem , *282*, 16054–61.

[Van Deursen et al., 1995] Van Deursen, J., Fornerod, M., Van Rees, B. and Grosveld, G. (1995). Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. Proc Natl Acad Sci U S A , *92*, 7376–80.

[Wahlstedt et al., 2009] Wahlstedt, H., Daniel, C., Enstero, M. and Ohman, M. (2009). Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. Genome Res , *19*, 978–86.

[Wahlstedt and Ohman, 2011] Wahlstedt, H. and Ohman, M. (2011). Site-selective versus promiscuous A-to-I editing. Wiley Interdiscip Rev RNA , *2*, 761–71.

[Wan et al., 2011] Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. and Chang, H. Y. (2011). Understanding the transcriptome through RNA structure. Nat Rev Genet , *12*, 641–55.

[Wan et al., 2012] Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., Makino, D. L., Nutter, R. C., Segal, E. and Chang, H. Y. (2012). Genome-wide measurement of RNA folding energies. Mol Cell , *48*, 169–81.

[Wang et al., 2000] Wang, Q., O'Brien, P. J., Chen, C. X., Cho, D. S., Murray, J. M. and Nishikura, K. (2000). Altered G protein-coupling functions of RNA editing isoform and splicing variant serotonin2C receptors. J Neurochem , *74*, 1290–300.

[Ward et al., 2011] Ward, S. V., George, C. X., Welch, M. J., Liou, L. Y., Hahm, B., Lewicki, H., de la Torre, J. C., Samuel, C. E. and Oldstone, M. B. (2011). RNA editing enzyme adenosine deaminase is a restriction factor for controlling measles virus replication that also is required for embryogenesis. Proc Natl Acad Sci U S A , *108*, 331–6.

[Winkler et al., 2002] Winkler, W., Nahvi, A. and Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature , *419*, 952–6.

[Wolf et al., 2002] Wolf, J., Gerber, A. P. and Keller, W. (2002). tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. Embo J , *21*, 3841–51.

[Wong et al., 2001] Wong, S. K., Sato, S. and Lazinski, D. W. (2001). Substrate recognition by ADAR1 and ADAR2. RNA , *7*, 846–58.

[Woolf et al., 1995] Woolf, T. M., Chase, J. M. and Stinchcomb, D. T. (1995). Toward the therapeutic editing of mutated RNA sequences. Proc Natl Acad Sci U S A , *92*, 8298–302.

[Wu et al., 1978] Wu, C. F., Ganetzky, B., Jan, L. Y., Jan, Y. N. and Benzer, S. (1978). A Drosophila mutant with a temperature-sensitive block in nerve conduction. Proc Natl Acad Sci U S A , *75*, 4047–51.

[Wu et al., 2011] 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*.

[Wulff et al., 2011] Wulff, B. E., Sakurai, M. and Nishikura, K. (2011). Elucidating the inosinome: global approaches to adenosine-to-inosine RNA editing. Nat Rev Genet , *12*, 81–5.

[Xia et al., 2005] Xia, S., Yang, J., Su, Y., Qian, J., Ma, E. and Haddad, G. G. (2005). Identification of new targets of Drosophila pre-mRNA adenosine deaminase. Physiol Genomics , *20*, 195–202.

[Xing and Li, 2009] Xing, D. and Li, Q. Q. (2009). Alternative polyadenylation: a mechanism maximizing transcriptome diversity in higher eukaryotes. Plant Signal Behav , *4*, 440–2.

[Yang et al., 2006] 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.

[Yang et al., 2008] Yang, Y., Lv, J., Gui, B., Yin, H., Wu, X., Zhang, Y. and Jin, Y. (2008). A-to-I RNA editing alters less-conserved residues of highly conserved coding regions: implications for dual functions in evolution. RNA , *14*, 1516–25.

[Zhang et al., 2011] Zhang, J., Kuo, C. C. and Chen, L. (2011). GC content around splice sites affects splicing through pre-mRNA secondary structures. BMC Genomics , *12*, 90.

[Zhang and Grosse, 2004] Zhang, S. and Grosse, F. (2004). Multiple functions of nuclear DNA helicase III (RNA helicase A) in nucleic acid metabolism. Acta Biochimica et Biophysica Sinica , *36*, 177–183.

[Zhang and Carmichael, 2001] Zhang, Z. and Carmichael, G. G. (2001). The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell , *106*, 465–75.

[Zu et al., 2013] Zu, Y., Tong, X., Wang, Z., Liu, D., Pan, R., Li, Z., Hu, Y., Luo, Z., Huang, P., Wu, Q., Zhu, Z., Zhang, B. and Lin, S. (2013). TALEN-mediated precise genome modification by homologous recombination in zebrafish. Nature Methods , *10*, 329–331.