RNA modifications and gene expression

Addition of non-templated nucleotides:

- A-tail (PAP and TRAMP)
- U-tail

Changes in the nucleotide sequence:

- RNA editing

Chemical modification of nucletides:

- 2'-O-Methylation, N^5-methylcytidine, N^6-Methyladenosine, N^1-Methyladenosine, N^7-methylguanine….
- Pseudourydilation

RNA editing

RNA editing

RNA editing is a post-transcriptional process that changes the nucleotide sequence of an RNA transcript from the DNA sequence encoded by the corresponding gene, excluding changes due to splicing, capping and polyadenilation

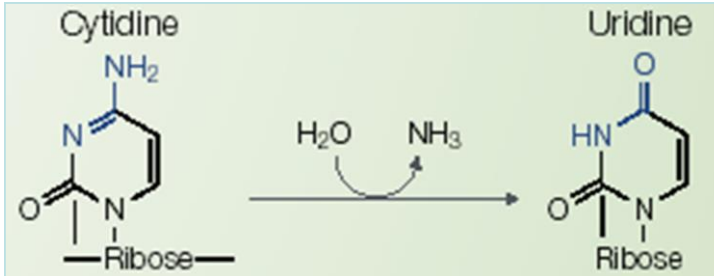
RNA editing can create, delete, alter codons, splicing sites, RNA structures changing the sequence defined by the DNA

RNA editing can affect coding (mRNAs) and non-coding RNAs (tRNAs, rRNAs, lncRNAs and microRNAs)

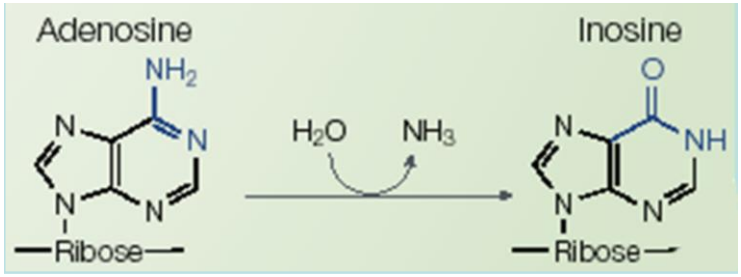
RNA editing

The best-characterized types of RNA editing that are found in mammals are the conversion of cytosine (C) to uracil (U) and the conversion of adenosine (A) to inosine (I), which are performed by **APOBEC1** and **ADARs** proteins, respectively. The A-> I editing is the most prevalent in higher eukaryotes.

***APOB* mRNA-editing enzyme catalytic polypeptide 1(APOBEC1)**



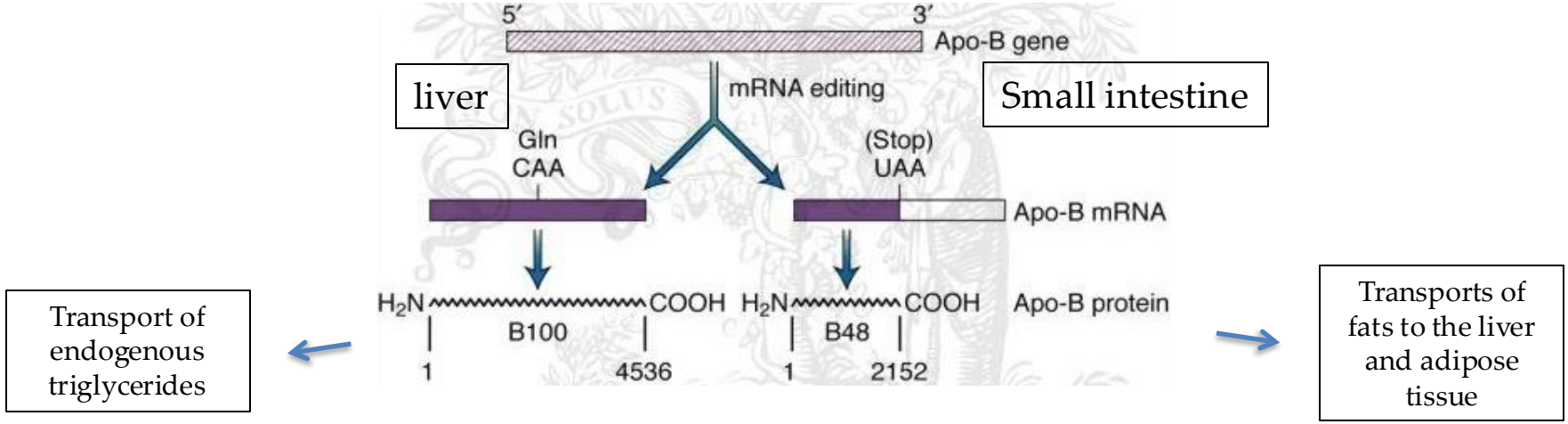
Adenosine Deaminase Acting on RNA (ADAR) proteins



Mammalian C->U RNA editing

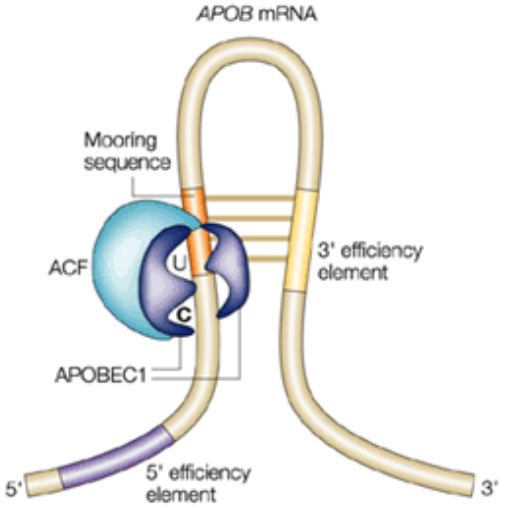
Apolipoprotein B (ApoB):

APOB is a component of the plasma lipoproteins and is crucial for the transport of cholesterol and of triglycerides in the plasma. There are two forms of APOB: **APOB100** and the shorter **APOB48** isoform, which results from the deamination of C->U at nucleotide position 6666 (C^{6666}) in the *APOB* mRNA, which causes the change of a glutamine to a translational stop codon. Because, in humans, this editing event occurs in the small intestine but not in the liver, the APOB100 isoform is synthesized only in the liver and is used to assemble the **very-low-density lipoprotein (VLDL)** that is necessary for the transport of endogenously synthesized triglycerides. Conversely, APOB48 is generated in the small intestine and is necessary for the synthesis and secretion of **CHYLOMICRON,** a large lipoprotein complex formed in the intestine that transports fats from the intestine to the liver and to adipose tissue.



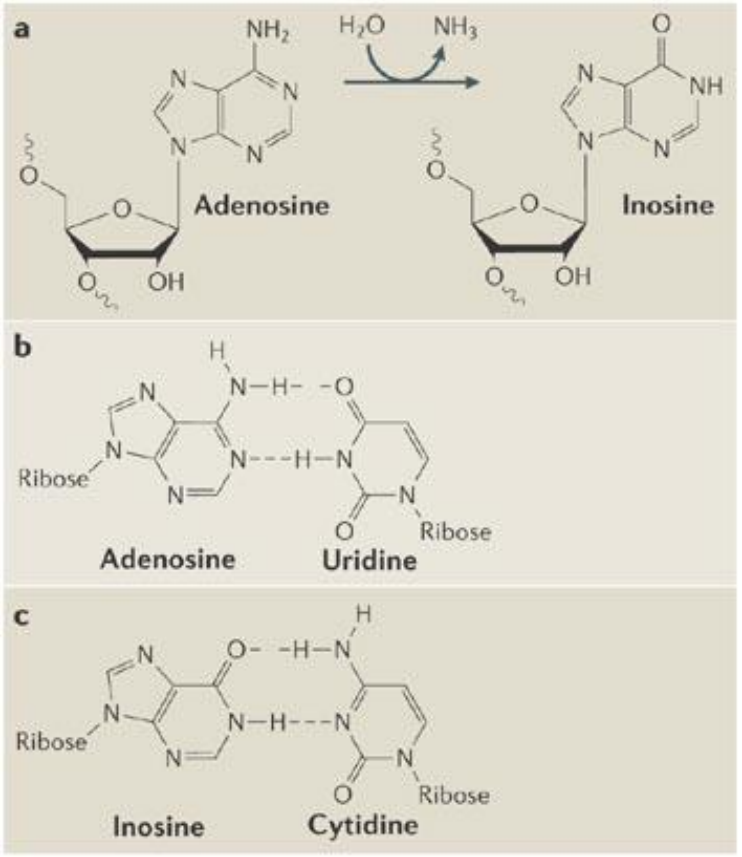
Mammalian C->U RNA editing

The human *APOB100* locus spans 43 kb, has 29 exons and encodes one of the largest known proteins (4,536 amino acids). Although the mRNA is 14 kb, editing occurs with exact precision at C^{6666} and requires both ***trans*-acting factors** and ***cis*-acting sequence elements** that surround the cytosine that is edited. The **mooring sequence** and the **3' efficiency element** form a double-stranded (ds) stem that is predicted to position the edited cytosine in a favourable configuration for deamination. The cytidine deaminase, **APOBEC1** (*APOB* mRNA-editing enzyme catalytic polypeptide 1) which catalyses the deamination reaction, and an auxiliary factor **ACF** (APOBEC1 complementation factor) form a complex that is the minimum requirement for editing of *APOB in vitro*. **RBM47** can substitute for ACF *in vitro* and is indispensable for *Apob* editing *in vivo*. <u>APOBEC1 is specifically expressed in the small intestine in humans.</u>



Deamination of adenosine to inosine by ADAR

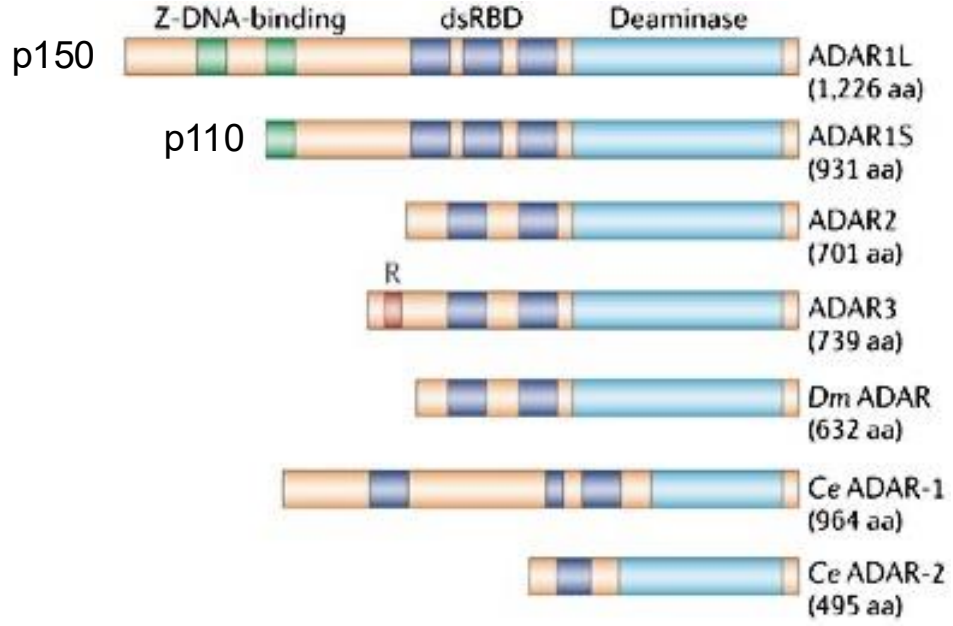
Adenosine Deaminase Acting on RNA (ADAR) proteins convert adenosine in inosine**.** Adenosine base-pairs with uridine. By contrast, inosine base-pairs, as guanosine, with cytidine.



Copyright © 2006 Nature Publishing Group
Nature Reviews | Molecular Cell Biology

ADAR proteins

Three human **ADAR**-family members (ADAR1–3) have been identified. Both **ADAR1** and **ADAR2** are present in many tissues, whereas **ADAR3** is expressed only in the brain and lacks enzymatic activity *in vitro*. Two isoforms of ADAR1, a full-length ADAR1L (inducible) and a shorter, N-terminal-truncated ADAR1S (constitutive), are known. ADAR1L is detected mainly in the cytoplasm, whereas ADAR1S localizes in the nucleoplasm and nucleolus. ADAR2 localizes predominantly in the nucleolus. Both ADAR1 and ADAR2 function as dimers and are subject to dominant negative effects by ADAR3.



Cyt (inducible)

Nucl and No (constitutive)

No

The significance of the nucleolar localization of ADAR1S and ADAR2 is not currently clear.

No ADAR genes have been identified in the genomes of plants, fungi or yeast

RNA-editing deficiencies

The inactivation of ADAR-gene-family members has significant physiological consequences, reported as phenotypic alterations of ADAR-gene mutants created in various species.

- Flies with a homozygous deletion in the *Adar* gene exhibit brain-related changes such as a lack of coordinated locomotion and neurodegeneration.

- Strains of *C. elegans* that contain homozygous deletions of both *adar-1* and *adar-2* display defective chemotaxis.

- Mice with a homozygous *ADAR2*-null mutation die several weeks after birth. These mice experience repeated episodes of epileptic seizures that originate from excess influx of Ca^{2+} and consequent neuronal death. The inactivation of ADAR1 leads to an embryonic lethal phenotype that is caused by defective erythropoiesis and widespread apoptosis.

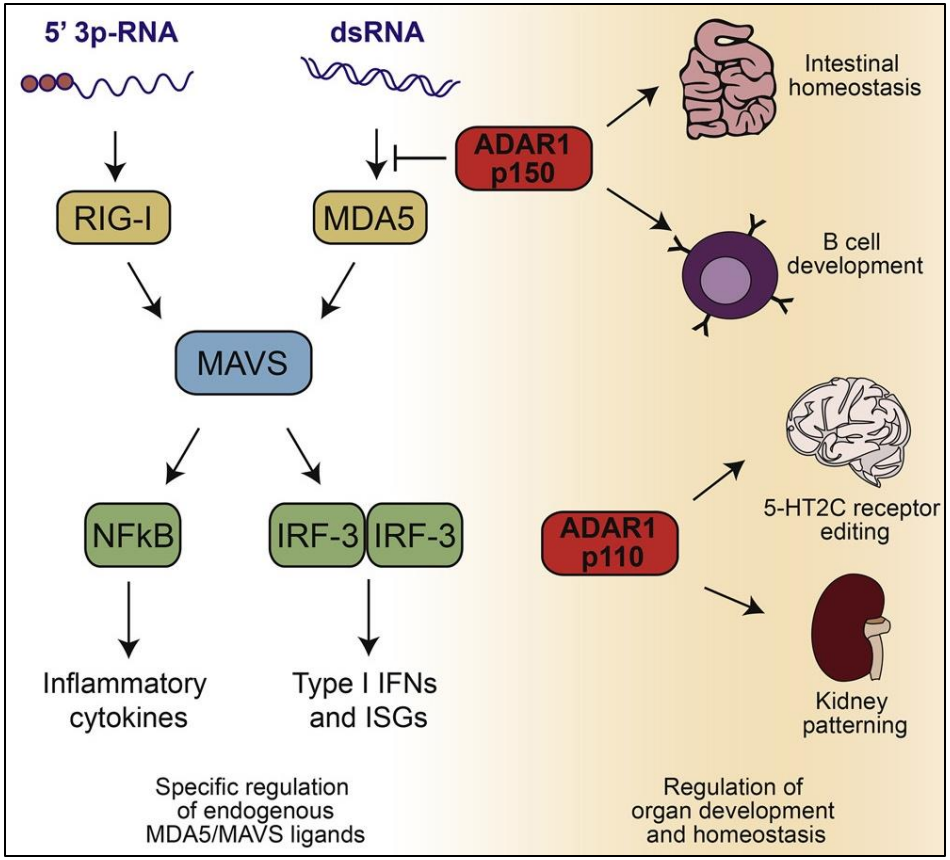
- Human diseases or pathophysiologies can also be caused by dysfunction of the RNA editing mechanism.

RNA editing and human diseases

- epilepsy and neurodegeneration (Palladino et al., Cell 2000;Higuchi et al., Nature, 2000).
- ataxia (Ma et al., J Clinical Investigation 2001).
- retinal defects (Palladino et al., Cell 2000; Blackshaw et al., Cell 2001).
- acute myeloid leukaemia (PTPN6) (Beghini et al., Hum. Mol. Gen. 2000).
- brain malignant gliomas (Maas et al., PNAS 22001).
- Hirschsprung disease (HSCR) (iper-editing of the endothelin B receptor) (Tanoue A et al., J Biol Chem 2002).
- schizophrenia (hypoediting of serotonin receptor) (Sodhi et al., Mol. Psychiat. 2001).
- depression (drugs and serotonin receptor) (Niswender et al., Nueropsychopharm. 2001).
- dyscromatosis symmetrica hereditaria (Miyamara et al., Am J hum genet 2003)
- sporadic amyotrophic lateral sclerosis (ALS) (Kawahara et al., Nature 2004)
- ischaemia (Peng et al., Neuron 2006)
- glucose modulation pancreatic islets and beta-cells (Levanon et al., NAR, 2006)
- inflammation (Rabinovici et al., Immunology 2003)

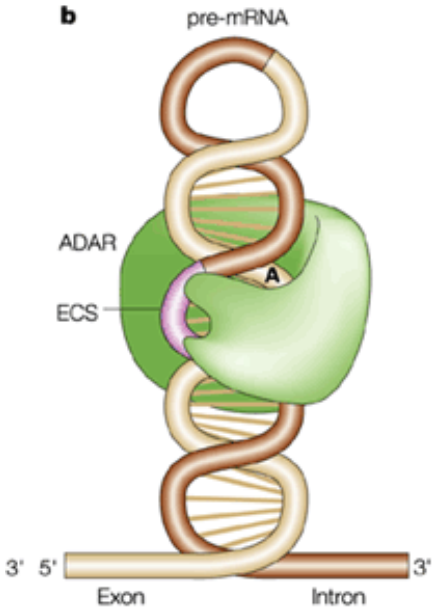
ADAR1 is required to prevent activation of innate immunity by endogenous dsRNA

Mutations in the ADAR1 enzyme cause Aicardi-Goutières syndrome (AGS), a severe autoimmune disease associated with an aberrant type I interferon response. The p150 isoform of ADAR1 is a specific and essential negative regulator of the MDA5-MAVS RNA sensing pathway.



Substrate and editing-site selectivity

The ADARs recognize the adenosine to be edited not by a surrounding consensus sequence but by the structure of the duplex that is formed between the **editing site** and an **editing site complementary sequence** (ECS) that is usually located in a downstream intron. The enzymes bind to the double-stranded (ds)RNA through their dsRBDs. Both intermolecular and intramolecular dsRNAs of >20 bp (two turns of the dsRNA helix) can serve as substrates for ADAR. Many adenosine residues of a long, completely base-paired dsRNA are edited non-selectively (perfect dsRNAs). By contrast, dsRNA with mismatched bases, bulges and loops (imperfect dsRNAs) are edited selectively; only a few adenosines are specifically chosen, indicating that the secondary structure in ADAR substrates dictates editing-site selectivity. Furthermore, some editing sites are preferentially edited only by ADAR1 or ADAR2.



Consequences of RNA editing

1. **A to I conversion can change sequence and/or codons**

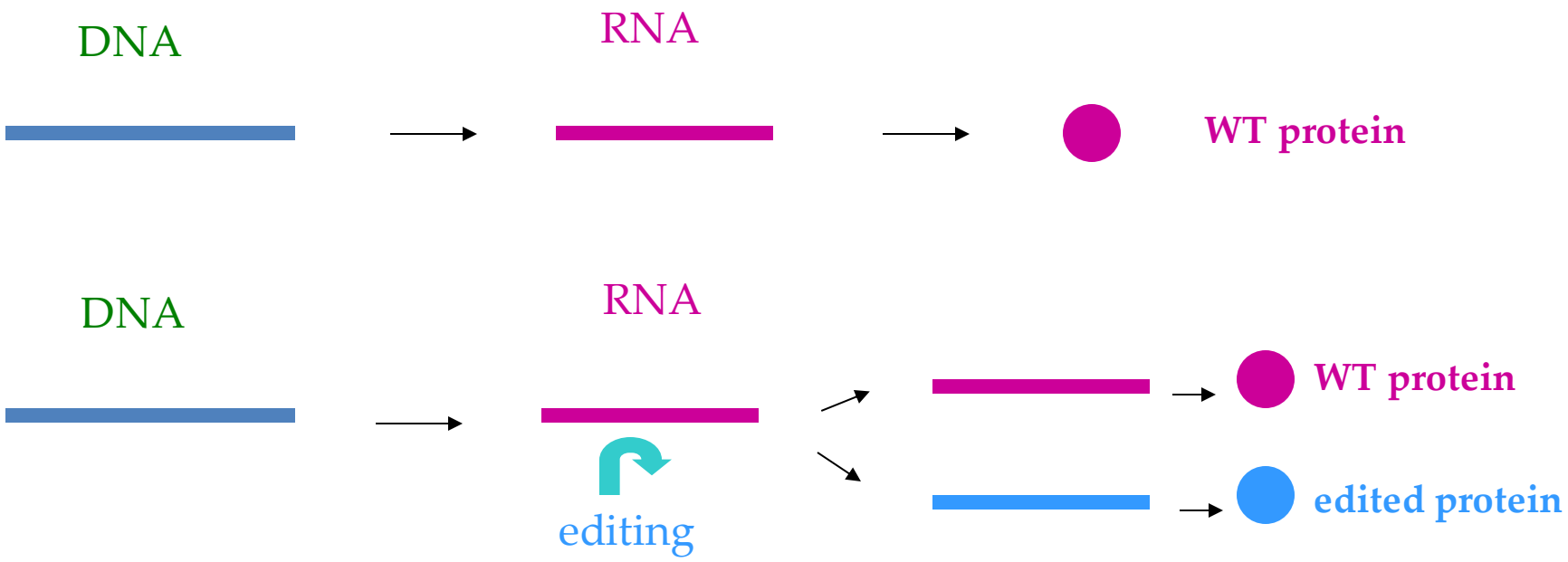
2. **A to I conversion can alter splice site/binding sites**

- 5′/GU-AG/3′ to 5′/GU-IG/3′
- 5′/AU-AG/3′ to 5′/IU-AG/3′

3. **A to I conversion can change RNA structure**

Editing sites found in protein-coding regions

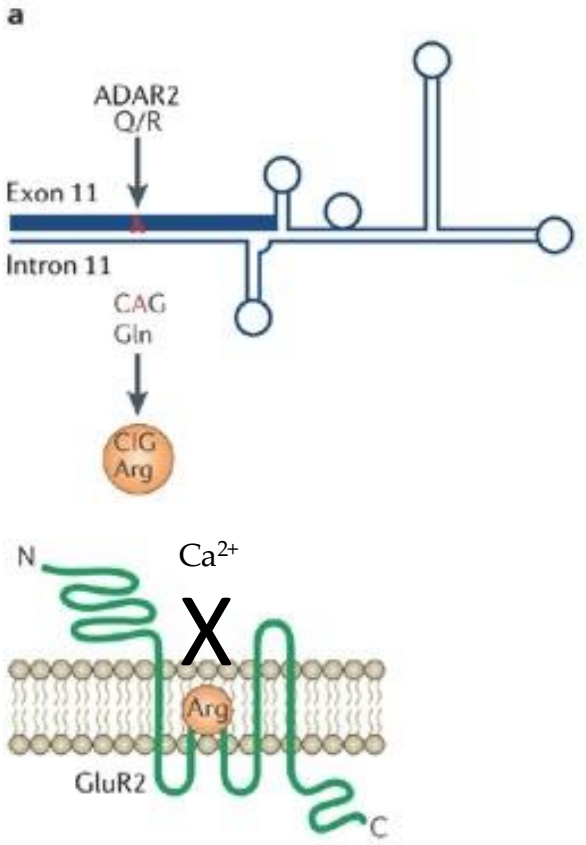
RNA editing is a mechanism to create protein variants by "recoding" the DNA information



In most cases RNA editing does not occur with 100% efficiency leading to a mix of both the original as well as the edited versions.

Functional changes by A→I RNA editing of coding sequences

L-glutamate is the predominant excitatory neurotransmitter in vertebrate nervous systems, and the **glutamate receptor (GluR)** has been implicated in neuronal plasticity and higher functions such as memory and learning. A->I RNA editing of the Gln/Arg (Q/R) site leads to the replacement of a Gln by an Arg residue. Ion-channel receptors that contain the edited GluR2 subunit are impermeable to Ca^{2+}, whereas channels that lack the edited subunit permit influx of Ca^{2+}. Q/R-site editing also regulates the tetramerization and intracellular trafficking of the receptor protein



Glutamate receptor is nearly 100% **edited**

NATURE | VOL 406 | 6 JULY 2000

Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2

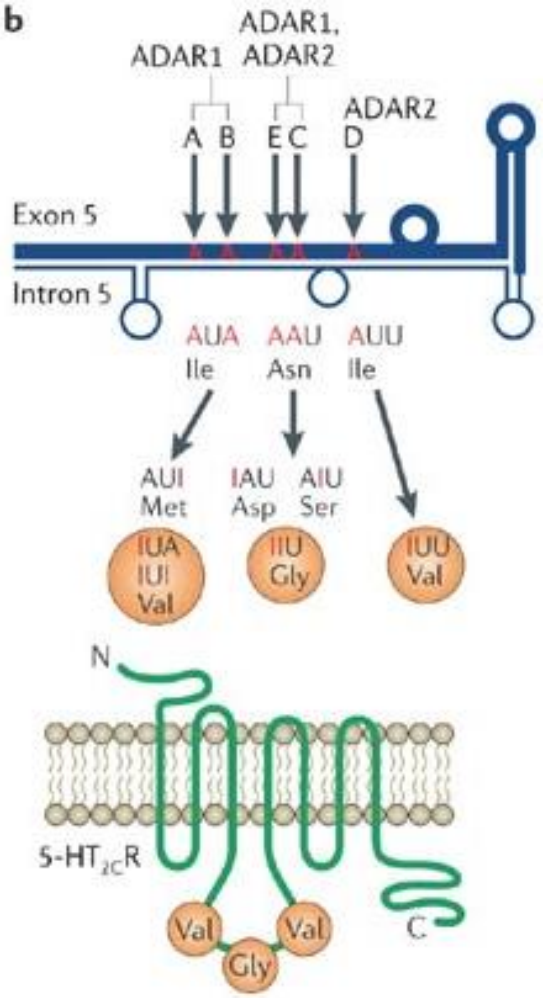
Miyoko Higuchi*, Stefan Maas*†‡, Frank N. Single*†, Jochen Hartner*, Andrei Rozov§, Nail Burnashev§, Dirk Feldmeyer§, Rolf Sprengel* & Peter H. Seeburg*

The impaired phenotype appeared to result entirely from a single underedited position, as it reverted to normal when both alleles for the underedited transcript were substituted with alleles encoding the edited version of GluR2.

GLUR2 is a subunit of AMPA receptor

Functional changes by A→I RNA editing of coding sequences

Serotonin receptors have important roles in physiological and behavioural processes such as circadian rhythms, emotional control and feeding behaviour. G-protein-coupling functions of **serotonin (5-HT) receptor-2C (5-HT_{2C}R)** are dramatically reduced by A->I RNA editing that occurs at five sites (A, B, C, D and E sites). For example, the potency of the agonist-stimulated G-protein-coupling activity of the fully edited receptor isoform (Val-Gly-Val) is reduced by 20-fold compared with the unedited receptor isoform (Ile-Asn-Ile).



Copyright © 2006 Nature Publishing Group
Nature Reviews | Molecular Cell Biology

A-to-I editing in selected mammalian protein-coding sequences and its functional consequences

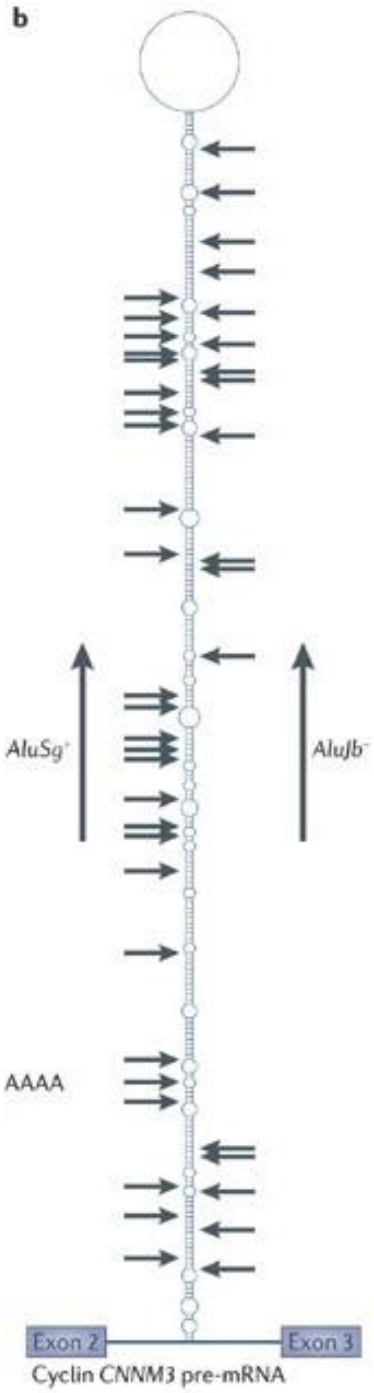
Gene	Protein	Recoding	ADAR responsible	Function
GRIA2	GluR2 subunit of AMPA glutamate receptor	Q→R	ADAR2	Change in Ca^{2+} permeability
		R→G	ADAR1, ADAR2	Change in receptor desensitization
GRIA3	GluR3 subunit of AMPA glutamate receptor	R→G	ADAR1, ADAR2	Change in receptor desensitization
GRIA4	GluR4 subunit of AMPA glutamate receptor	R→G	ADAR1, ADAR2	Change in receptor desensitization
GRIK1	GluR5 subunit of kainate glutamate receptor	Q→R	ADAR1, ADAR2	Change in Ca^{2+} permeability
GRIK2	GluR6 subunit of kainate glutamate receptor	Q→R	ADAR1, ADAR2	Change in Ca^{2+} permeability
		I→V	ADAR1, ADAR2	
		Y→C	ADAR2	
HTR2C	Serotonin receptor 2C	I→V, M	ADAR1	Change in G protein-coupling functions
		N→S, G, D	ADAR1, ADAR2	
		I→V	ADAR2	
KCNA1	Voltage-gated K^+ channel (Kv1.1)	I→V	ADAR2	Change in channel inactivation
GABRA3	$GABA_A$ receptor, subunit α3	I→M	ADAR1, ADAR2	Kinetics of activation and inactivation, receptor trafficking
BLCAP	Bladder cancer-associated protein	Y→C	ADAR1, ADAR2	Not determined
		Q→R	ADAR1, ADAR2	
		K→R	ADAR1, ADAR2	
CYFIP2	Cytoplasmic FMR1-interacting protein 2	K→E	ADAR2	Not determined
FLNA	Filamin-α	Q→R	ADAR1, ADAR2	Not determined
FLNB	Filamin-β	M→V	ADAR1, ADAR2	Liver cancer progression
COPA	Coatomer protein complex subunit-α	I→V	ADAR2	Suppression of liver cancer
IGFBP7	Insulin-like growth factor-binding protein 7	K→R R→G	–	Proteolytic cleavage sensitivity
AR	Androgen receptor	T→A	ADAR1, ADAR2	Prostate cancer progression Inhibition of interaction with androgen ligands
AZIN1	Antizyme inhibitor 1	S→G	ADAR1	Liver cancer progression Change in affinity for antizyme
NEIL1	DNA repair enzyme NEI-like protein 1	K→R	ADAR1	Change in efficiency or specificity of damaged base removal
GLI1	Glioma-associated oncogene 1	R→G	ADAR1, ADAR2	Increased transcription enhancement
RHOQ	RAS homology family member Q	N→S	–	Colorectal cancer metastasis Disruption of interaction with Rap–RapGAP

Screening for A->I RNA editing sites

Computational methods for the genome-wide identification of new A->I RNA editing sites have been developed. Reverse transcriptase recognizes inosine as if it were guanosine. Therefore, an A->I RNA editing site can be identified when a cDNA sequence and the corresponding genome sequence are aligned, given that guanosine residues reverse-transcribed from inosines are detected in place of gene-encoded adenosines. The screening strategy consists of an algorithm to align a cluster of AG mismatches in cDNAs to the genome sequence and to assemble them into clusters that contain complete or partial genes in the dsRNA regions. This is followed by the elimination of single nucleotide polymorphisms (SNPs).

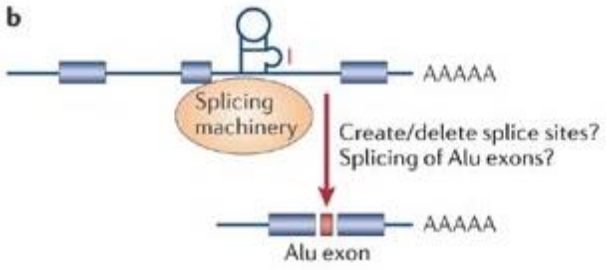
Global editing of non-coding RNAs

The majority of A->I editing sites that were identified in the human transcriptome (15,000 sites, mapped in 2,000 different genes) reside in non-coding regions that consist of inversely oriented repetitive elements, mostly **Alu repeats** (90%) and some **LINE repeats** (10%), representing 13% and 21% of the human genome, respectively. On the basis of this analysis, it is predicted that >85% of pre-mRNAs, are possibly edited, with the vast majority being targeted in introns (90%) and the rest in UTRs. The protein re-coding as a result of A-> I RNA editing is rare.



Implications of repetitive RNA editing

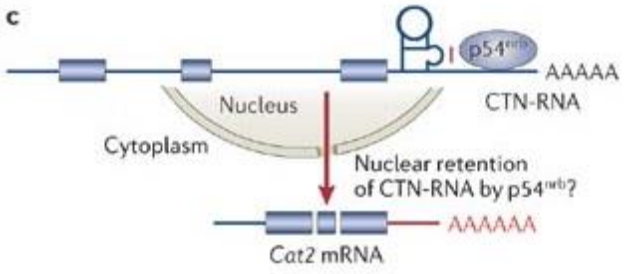
An inosine is interpreted by the splicing machinery as a guanosine. Therefore, splice sites might be created or deleted due to A->I editing of intronic Alu fold-back double-stranded (ds)RNAs, leading to the inclusion or exclusion of Alu exons.



Modulating splicing sites

A->I editing of a SINE fold-back dsRNA present in the 3' UTR of CTN-RNA and its binding to $p54^{nrb}$ might be involved in the regulatory mechanism that retains this RNA in nuclear speckles.

Nuclear retention

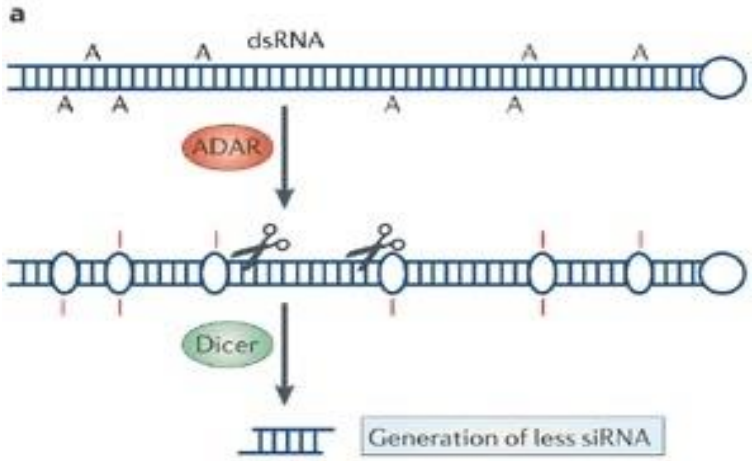


When cells are placed under stress, CTN-RNA is cleaved and de novo polyadenylated at an alternative site to release the protein-coding Cat2 mRNA, which is then translated into cationic amino-acid transporter-2 protein

Crosstalk between RNA editing and RNAi

In general, dsRNA-binding proteins lack sequence specificity. Therefore, it has been speculated that the A->I RNA editing mechanism might interact with the RNAi pathway by competing for shared dsRNA substrates and reducing RNAi efficacy. The dsRNA that is extensively edited *in vitro* by ADAR indeed becomes resistant to Dicer, resulting in the generation of less siRNA and reduced RNAi.

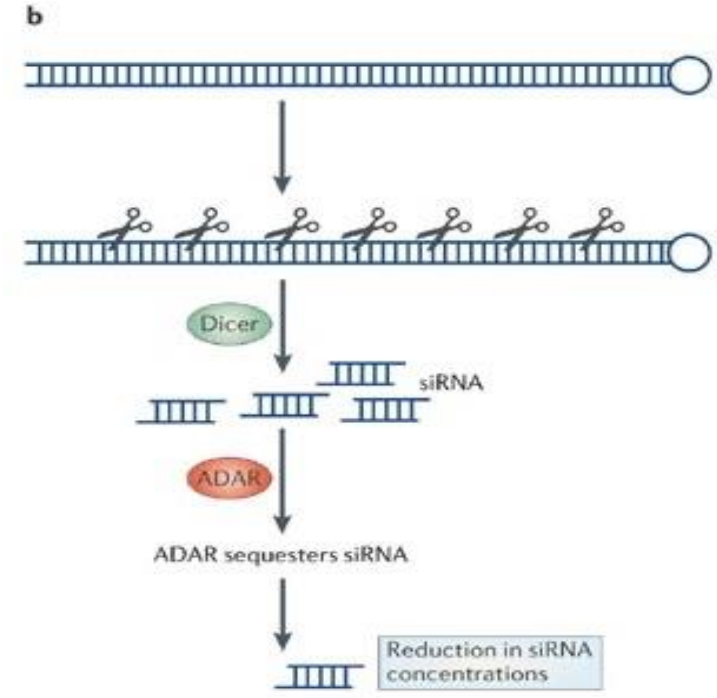
In *C. elegans* strains the ADAR-null worm phenotypes can be reverted by RNAi deficiency, indicating that these phenotypes are RNAi dependent.



Copyright © 2006 Nature Publishing Group
Nature Reviews | Molecular Cell Biology

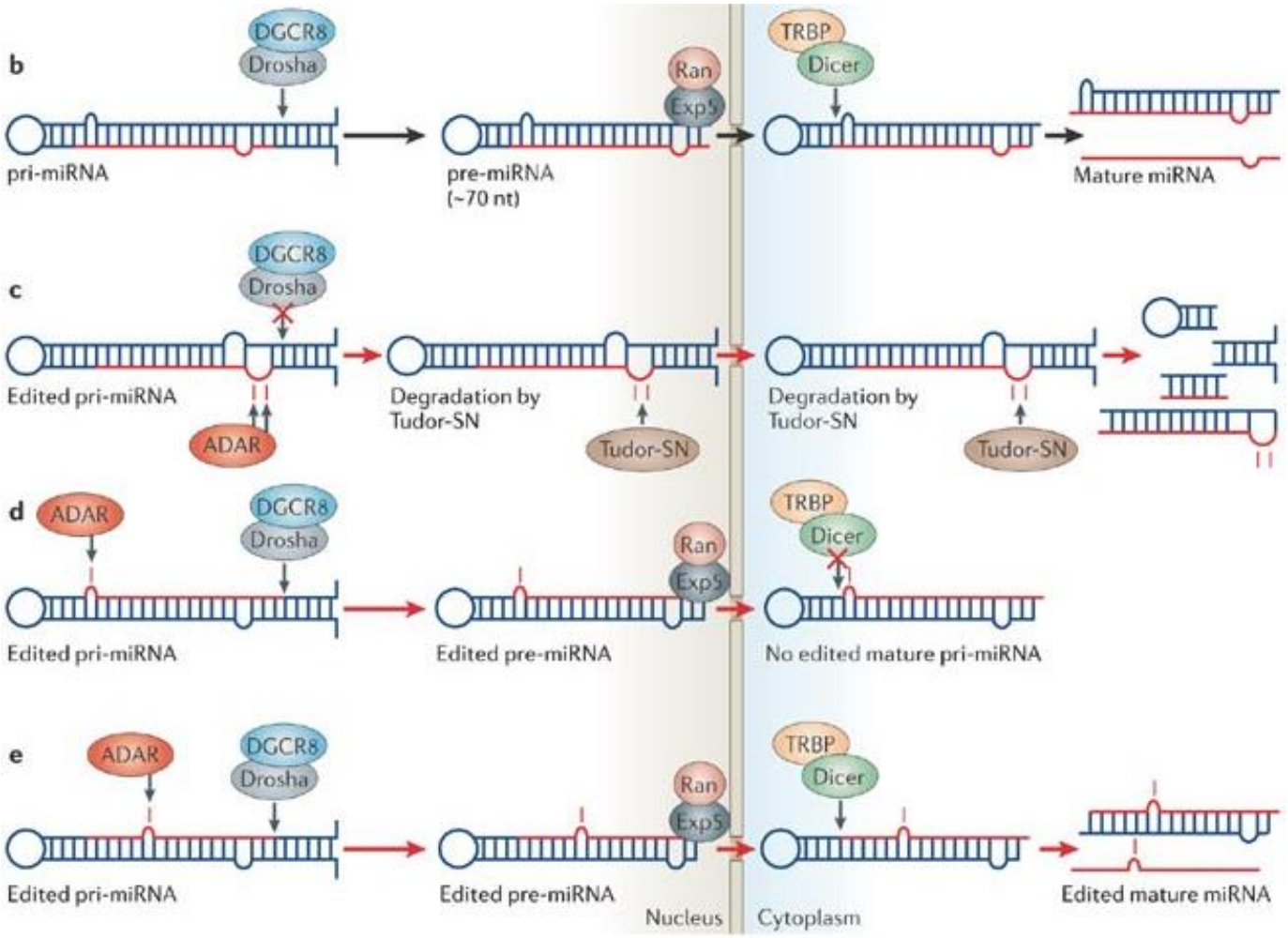
Crosstalk between RNA editing and RNAi

Cytoplasmic ADAR1L has also been reported to bind ds-siRNAs tightly. Gene silencing by siRNA is significantly more effective in mouse fibroblasts that are homozygous for an *Adar1*-null mutation than in wild-type cells. These findings implicate ADAR1L as a cellular factor that limits siRNA potency in mammalian cells by decreasing the effective siRNA concentration and its incorporation into RISC.



Copyright © 2006 Nature Publishing Group
Nature Reviews | Molecular Cell Biology

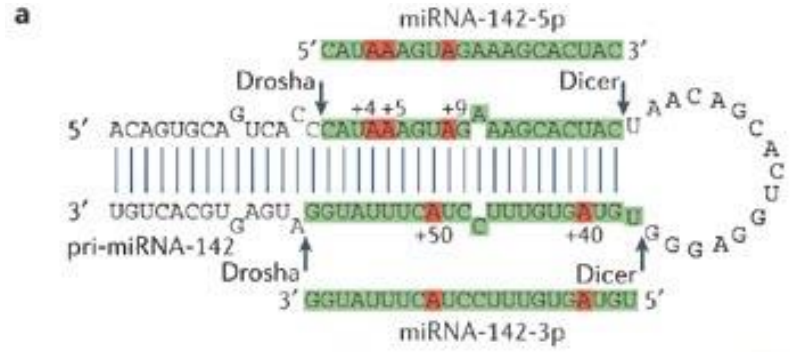
Crosstalk between RNA editing and microRNAs



Copyright © 2006 Nature Publishing Group
Nature Reviews | Molecular Cell Biology

Crosstalk between RNA editing and microRNAs

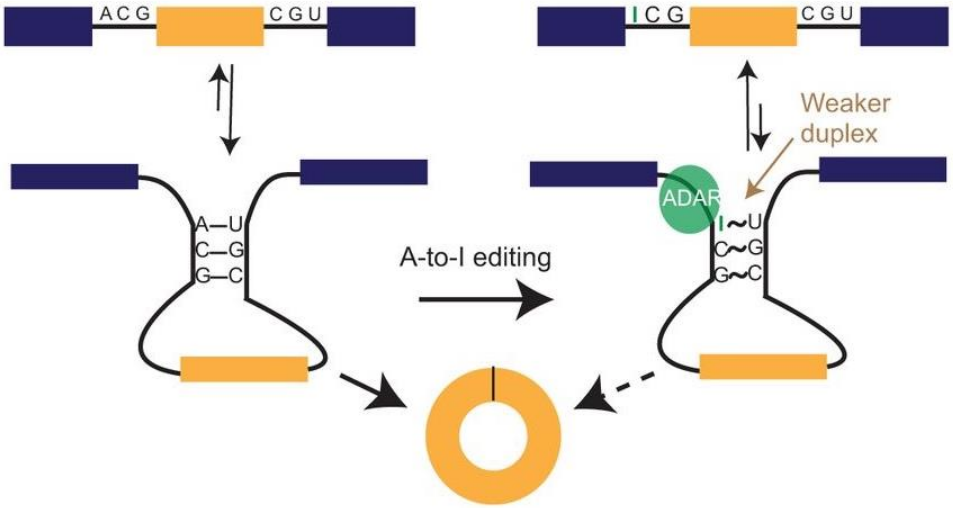
Recent studies showed that certain miRNA precursors are edited by ADAR. A systematic survey of human pri-miRNA sequences identified A->I RNA editing sites in 6% of all pri-miRNAs examined. The editing of miRNA precursors could have important implications for their processing, as well as the expression and the functions of mature miRNAs. A->I RNA editing alters the fold-back dsRNA structure of miRNA precursors; this might affect their subsequent processing and export steps, and change targets specificity.



circRNAs are upregulated upon knockdown of Adar1

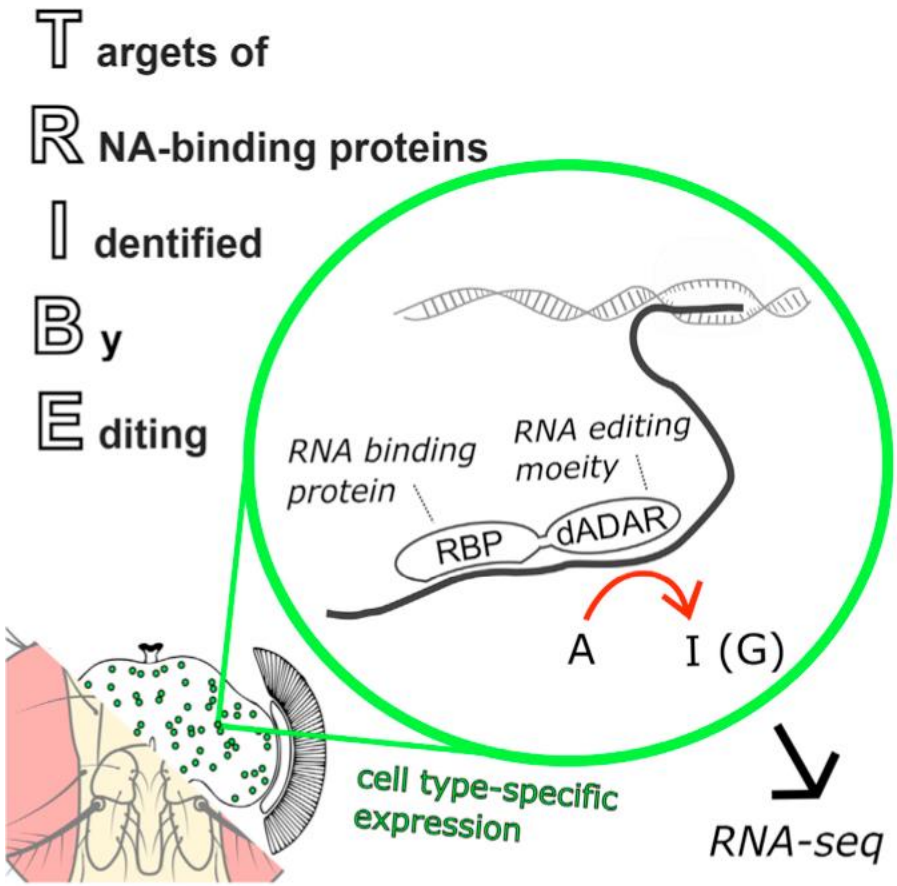
These results link A-to-I editing to circRNA biogenesis and predict that A-to-I editing events ‘‘melt’’ the stems that are formed across introns that bracket a circRNA.

Since it has been shown that circular splicing and linear splicing can compete with each other, it is possible that regulation of circRNA biogenesis by ADAR1 serves as a mechanism to regulate expression of the linear isoforms.



TRIBE method for mapping RNA-protein interatcions

The RBP of interest is fused to the enzyme ADAR, which deaminates nearby adenosines, after which deaminated bases are subsequently identified by sequencing.

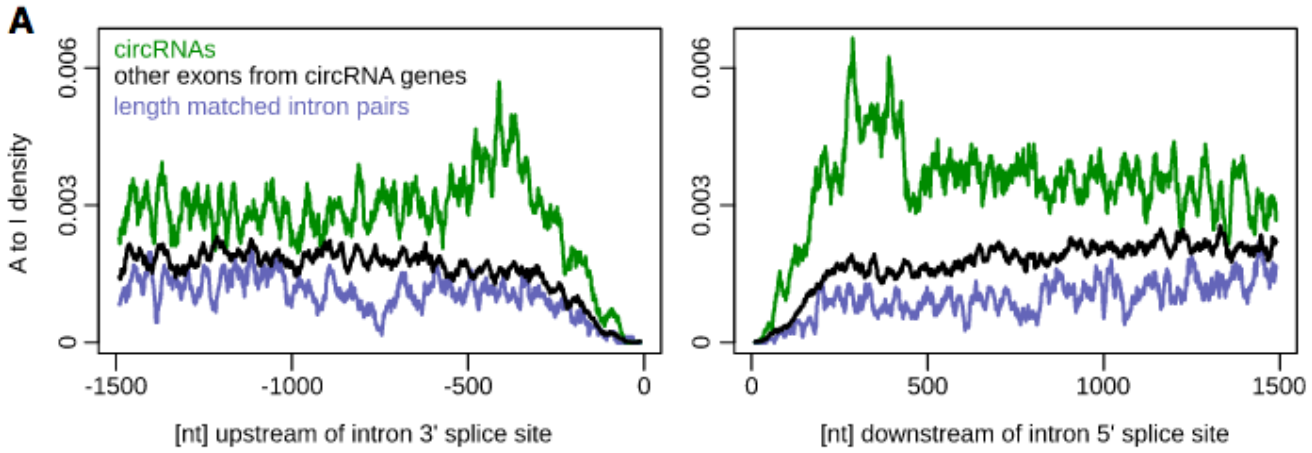


- No need to immunoprecipitate the protein of interest
- no dependence on UV cross-linking

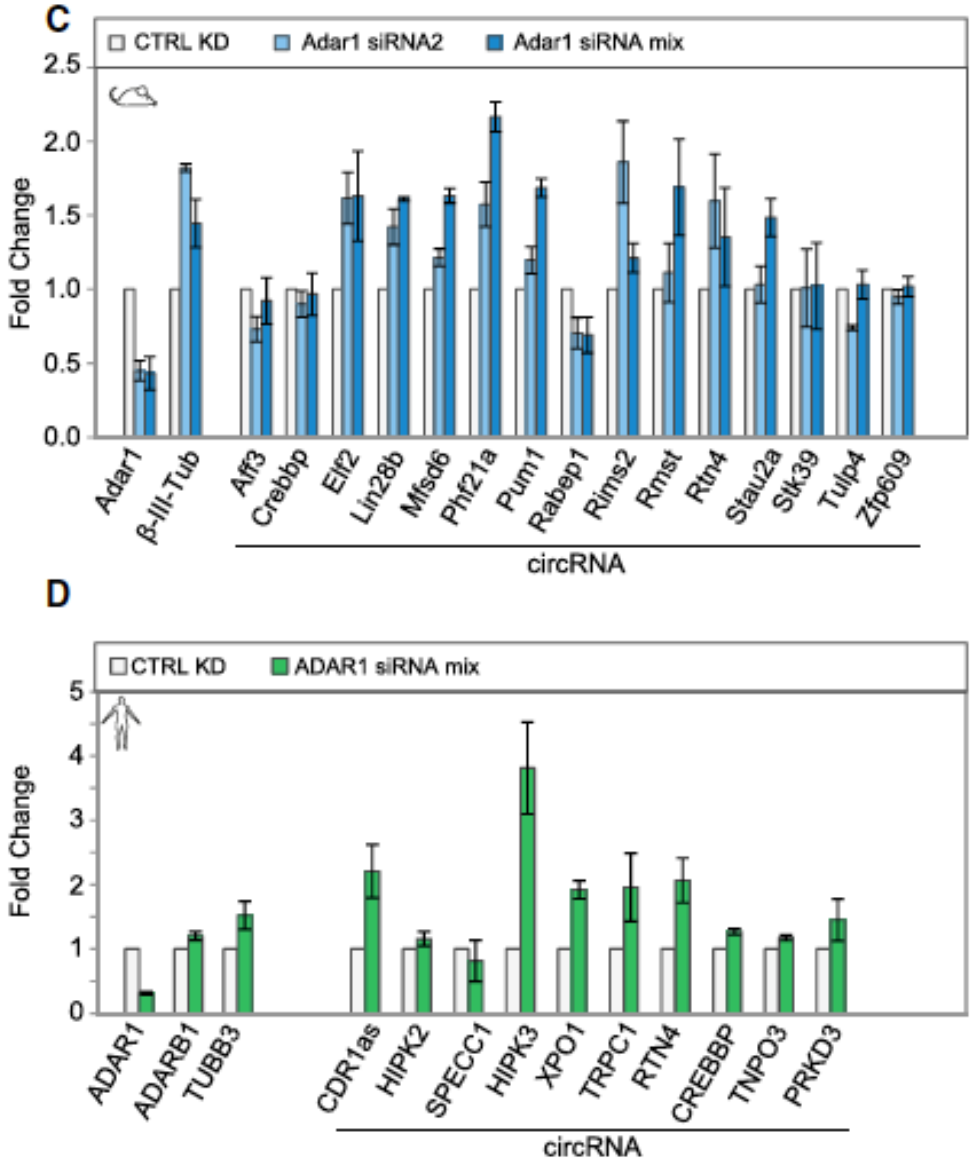
Crosstalk between RNA editing and circRNAs

As inverted *Alu* repeats are known to be a target for RNA editing, the formation of circRNAs can be regulated by RNA editing.

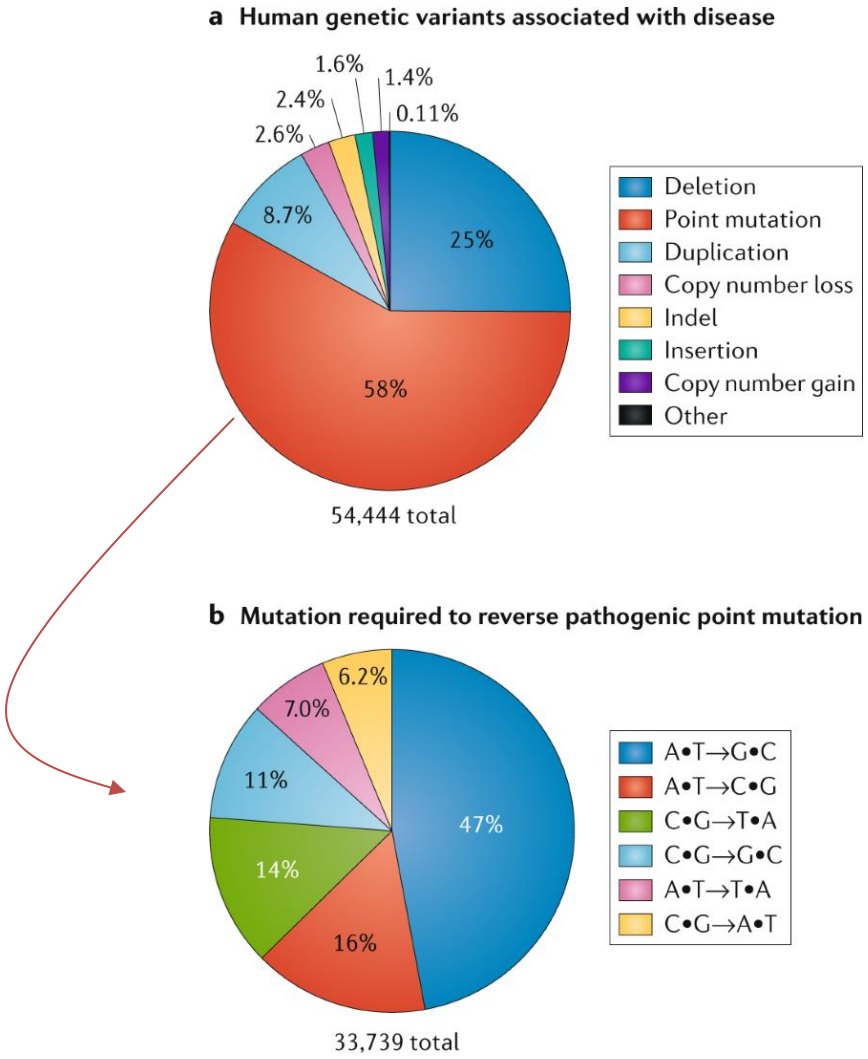
Introns surrounding circRNAs are highly edited



circRNAs are upregulated upon knockdown of Adar1



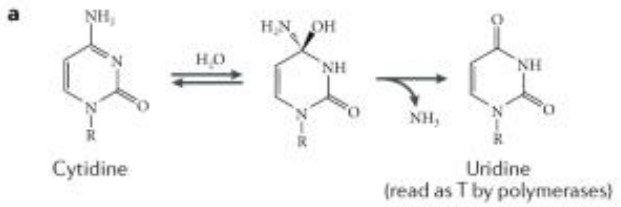
Distribution of human pathogenic genetic variants, including point mutations



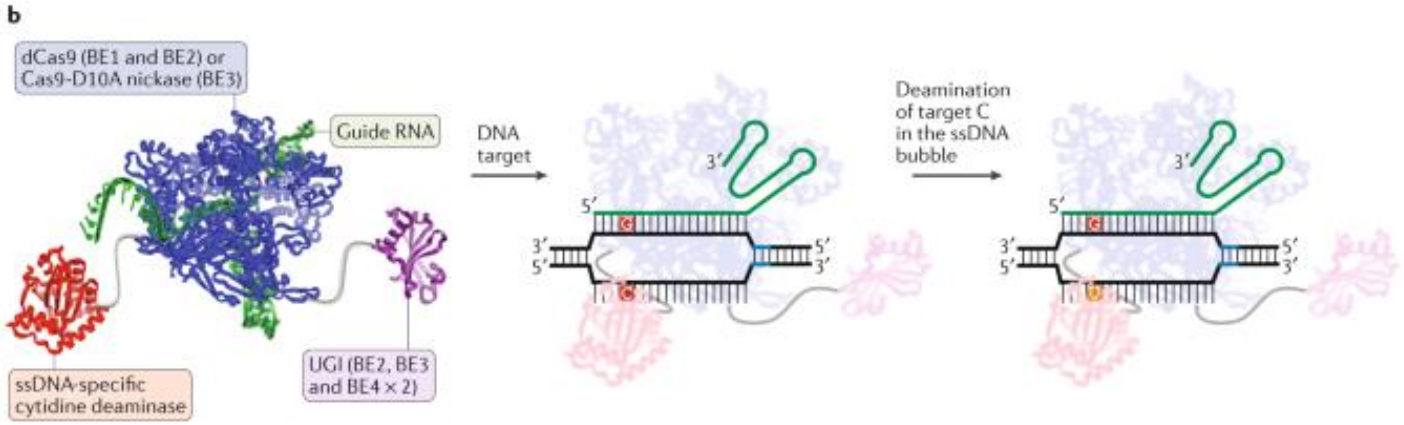
Rees and Liu , 2018. Nature Rev. Genetics

Development of cytosine base editors

Cytosine base-editing strategy by base editor 1 (BE1), BE2, BE3 or BE4. R loop formation exposes a region of single-stranded DNA (ssDNA) to the cytidine deaminase domain. Target cytosines in this region are deaminated to uracil



Although BE1 mediates efficient, RNA-programmed deamination of target cytosines in vitro, it is npt able to counteract cellular BER



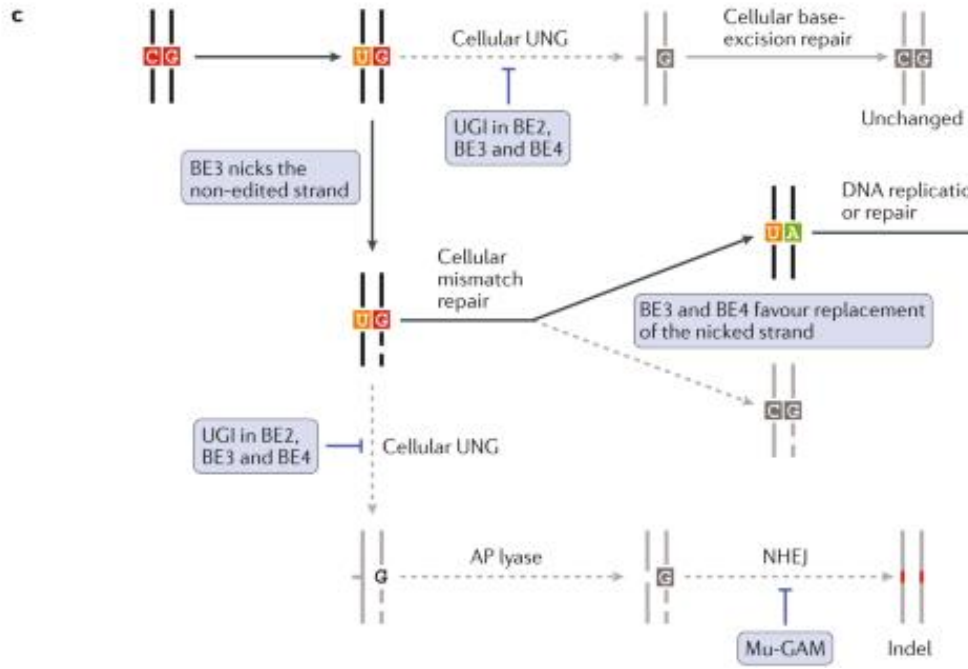
UGI = uracil DNA glycosylase inhibitor

Rees and Liu , 2018. Nature Rev. Genetics

Development of cytosine base editors

The cellular response to cytosine base editing is depicted. Uracil *N*-glycosylase (UNG)-mediated excision of the uracil generated in genomic DNA is inhibited by BE2, BE3 and BE4.

BE3 and BE4 are designed to nick the non-edited strand (containing the G of the original C•G target base pair), stimulating cellular DNA repair of that strand to replace the G with an A, completing the conversion of the original C•G base pair to a U•A base pair or, following DNA replication or repair, to a T•A base pair



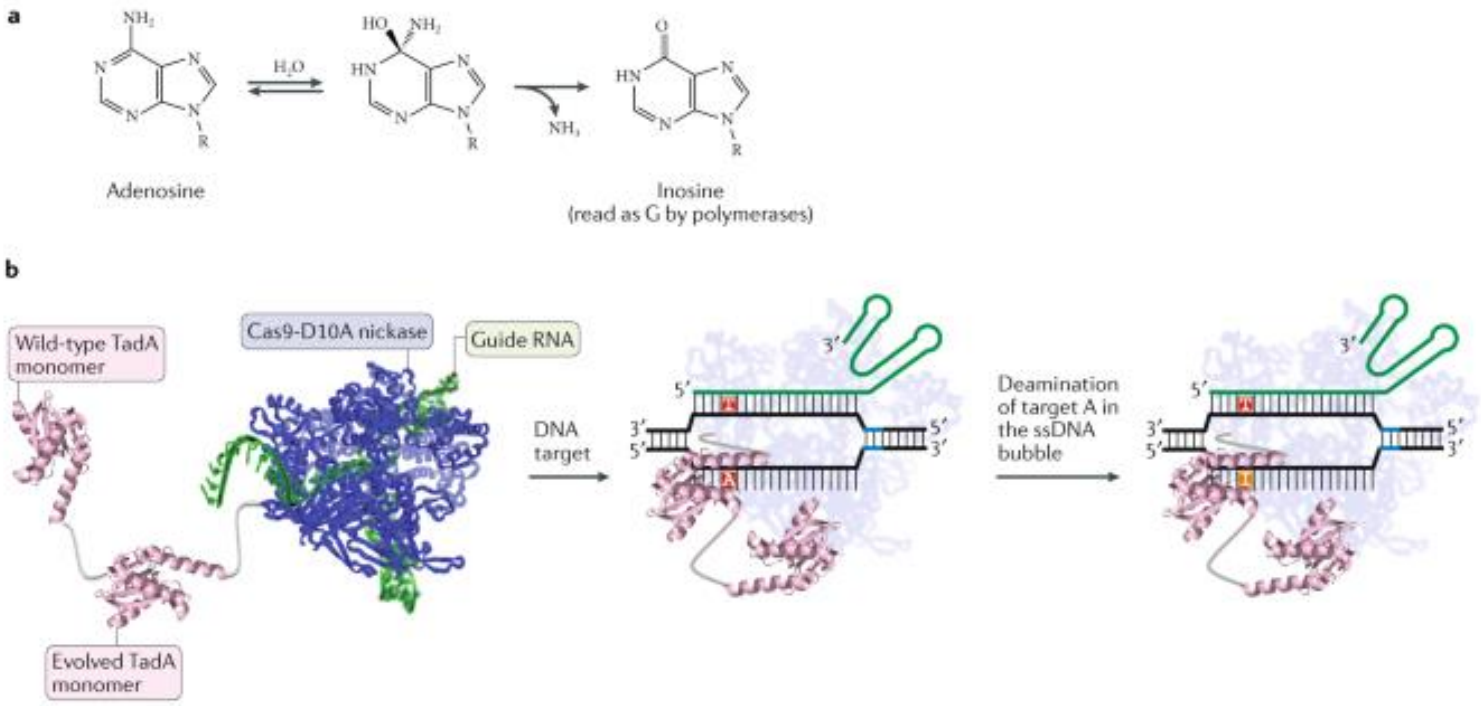
To inhibit UNG, uracil DNA glycosylase inhibitor (UGI), a small protein from bacteriophage PBS, has been fused to BE1, generating BE2.

third- generation base editors (BE3) that specifically nick the non- edited DNA strand. Nicking the non- edited DNA strand biases cellular repair of the U•G mismatch to favour a U•A outcome

Rees and Liu , 2018. Nature Rev. Genetics

Adenine base editing in DNA

Adenine base editor (ABE)-mediated DNA base editing to convert an A•T base pair to a G•C base pair is shown. Current ABEs contain one wild-type TadA (*E.coli* tRNA adenosine deaminase enzyme) structural monomer and one evolved TadA catalytic monomer. R loop formation exposes a small region of single-stranded DNA (ssDNA), within which A is deaminated to I by the heterodimeric wild-type TadA–evolved TadA heterodimer



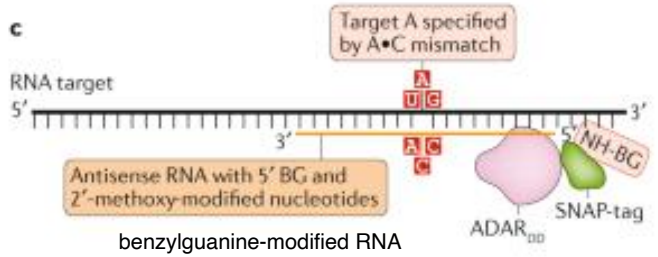
TadA: tRNA adenosine deaminase enzyme,

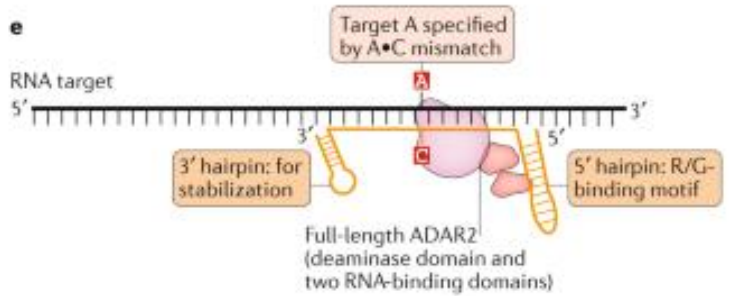
Rees and Liu , 2018. Nature Rev. Genetics

Adenine base editing in RNA

ADAR deaminase variants or ADAR deaminase domain ($ADAR_{DD}$) are localized to the RNA transcript of interest through an antisense RNA with variable lengths of homology to the target transcript. The target A is specified with an A•C mismatch in the mRNA–antisense RNA duplex.

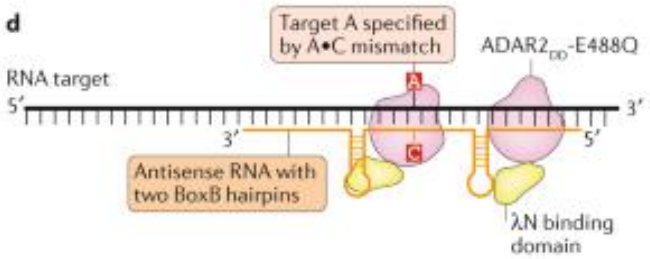
ADAR deaminase domain ($ADAR_{DD}$)–SNAP tag fusion to a benzylguanine (BG)-modified antisense RNA



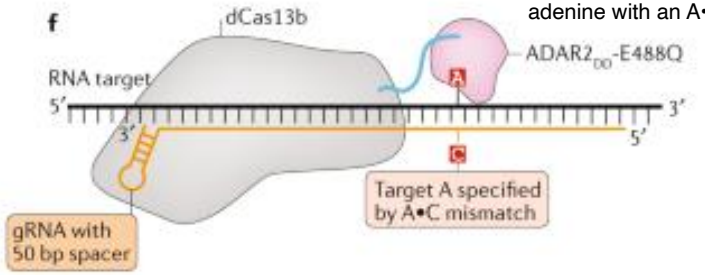


The antisense RNA comprises a 5′ R/G-binding motif hairpin and the native binding sequence for full-length ADAR2

The ADAR2$_{DD}$ is fused to λN protein; λN binds to one of the two BoxB hairpins integrated into the antisense RNA



hyperactive ADAR2DD-E488Q mutant and specifying the target adenine with an A•C mismatch



A complex of nuclease dead Cas13b (dCas13b) and guide RNA (gRNA) is guided to the target RNA by a 50-nucleotide spacer. The target A is specified by an A•C mismatch centrally located within the 50-nucleotide spacer

Rees and Liu , 2018. Nature Rev. Genetics