

Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination

Riccardo Pecori¹, Salvatore Di Giorgio¹, J. Paulo Lorenzo and F. Nina Papavasiliou¹

Abstract | The AID/APOBEC polynucleotide cytidine deaminases have historically been classified as either DNA mutators or RNA editors based on their first identified nucleic acid substrate preference. DNA mutators can generate functional diversity at antibody genes but also cause genomic instability in cancer. RNA editors can generate informational diversity in the transcriptome of innate immune cells, and of cancer cells. Members of both classes can act as antiviral restriction factors. Recent structural work has illuminated differences and similarities between AID/APOBEC enzymes that can catalyse DNA mutation, RNA editing or both, suggesting that the strict functional classification of members of this family should be reconsidered. As many of these enzymes have been employed for targeted genome (or transcriptome) editing, a more holistic understanding will help improve the design of therapeutically relevant programmable base editors.

Genetic information is encoded by DNA, transcribed into RNA and translated into protein. When originally proposed¹, this foundational tenet assumed faithful transmission of information such that mRNA accurately reflects what is encoded at the DNA level. However, it is now clear that RNA molecules can undergo several processing events that diversify the genomic information, resulting in different transcripts that, in some cases, encode different protein isoforms. Examples of such processes are alternative splicing², alternative polyadenylation³ and base modifications.

Most RNA base modifications are not easily detectable via synthesis-based RNA sequencing⁴, making it exceedingly difficult to distinguish between modified and unmodified RNA molecules⁵. One exception is RNA base deamination (also known as RNA editing), a widespread set of modifications that lead to a change in the RNA sequence itself. RNA editing can be detected simply by comparing the sequence of the transcript with that of its cognate gene.

In mammals, RNA editing refers specifically to the deamination of adenosine to inosine (A-to-I) or cytosine to uracil

(C-to-U); for the purposes of this Perspective we will exclude the phenomenon of uracil insertion or deletion that was described as RNA editing in mitochondria of *Trypanosoma brucei*⁶. A-to-I editing is catalysed by the adenosine deaminase acting on RNA (ADAR) protein family^{7–9}. C-to-U editing is performed by numerous cytosine deaminases, the best known of which belong to a family of mammalian enzymes known as the ‘activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like’ (AID/APOBEC) protein family¹⁰ (BOX 1).

The first member of the AID/APOBEC family to be characterized was the bona fide RNA editing enzyme APOBEC1 (FIG. 1a). Since then, additional RNA editing deaminases belonging to this family have been described, including APOBEC3A (A3A) and A3G. These RNA editors have the peculiar ability to also deaminate DNA, leading to single-nucleotide variant mutations that often occur processively in genomic DNA or reverse-transcribed viral cDNA (FIG. 1b). By contrast, other family members seem to have lost their ability to deaminate RNA: some, instead, catalyse mutation of viral

DNA (or cDNA); others have very specific genomic DNA substrates — for example, AID edits the expressed immunoglobulin gene (FIG. 1c). Finally, APOBEC2 cannot edit RNA or DNA but has the ability to bind DNA with affinities much higher than those reported for any other family member¹¹.

The interplay between RNA editing and DNA mutation and the types of molecular restrictions that determine substrate range and selectivity is the focus of this Perspective. We first summarize the main determinants of AID/APOBEC substrate selectivity in members that are able to deaminate only DNA (we call these ‘specialists’) and those that can deaminate both RNA and DNA (we term these ‘generalists’) (TABLE 1); note that family members where activity has not yet been tested on both substrates remain unassigned in this scheme. We then provide examples of how these different functionalities have allowed specific members of the AID/APOBEC family to drive evolution in different contexts. Finally, we discuss how AID/APOBEC enzymes have been co-opted into synthetic biology — specifically into the genome and transcriptome engineering technologies broadly known as programmable base editing, which have enormous therapeutic potential¹². A broad understanding of the molecular features that drive AID/APOBEC selectivity will be key to the development of such precision therapeutics.

Determinants of substrate selectivity

AID/APOBEC enzymes share three major functional elements: they all contain the catalytic domain (comprising the enzymatic pocket that, in part, overlaps with the substrate binding surface), whereas some also contain a cofactor interaction region (that can also multimerize) and sequence elements that define the subcellular localization of each protein (FIG. 2a). Sequence and/or structural variations in any of these features can change nucleic acid preference, for example through minor alterations in the substrate binding groove, or through restricted subcellular localization, such as exclusion from the nucleus through interaction with cofactors or through intramolecular oligomerization (reviewed elsewhere^{13,14}).

Box 1 | AID/APOBECs and their cellular functions

Reported functions for key members of the activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (AID/APOBEC) family of enzymes are discussed below.

APOBEC1

Human APOBEC1 expression is confined to the small intestine¹⁶⁰ and its only confirmed physiologic target is the apolipoprotein B (APOB) mRNA⁵⁷ (FIG. 1a). However, mouse APOBEC1 (mAPOBEC1) is more broadly expressed, mostly in immune cells but also in the small intestine and liver⁶³. The availability of genetic models has enabled the activity of mAPOBEC1 to be studied more thoroughly than human APOBEC1, and hundreds of edited transcripts have been detected in numerous mouse tissues^{161,162}. Interestingly, different cofactors drive mAPOBEC1 activity to different mRNAs (FIG. 1a). Although a specific function for APOBEC1 (other than its role in *ApoB* editing) has been hard to discern, loss of mAPOBEC1 activity confirms its importance for the function of innate immune cells of the monocytic lineage (including macrophages, microglia and dendritic cells among others) and microglia-mediated nervous system homeostasis^{63,163}. APOBEC1 can also deaminate DNA, a process that has been linked to cancer^{142,164,165} (FIG. 1a).

AID

AID is a single-stranded DNA (ssDNA) deaminase with a strong target preference for unmodified dC in the sequence context 5'-WRC-3' (W = dA or dT, R = dA or dG)^{166–168}. AID catalyses the deamination of immunoglobulin (*Ig*) genes. When targeted to the rearranged V gene, it initiates somatic hypermutation (SHM); when it is targeted to the switch regions upstream of the constant gene, it initiates a deletional recombination programme that results in antibody isotype switching (FIG. 1c). Both of these mechanisms increase antibody diversification in the host¹⁶⁹. Although AID can bind both RNA and ssDNA, it only deaminates ssDNA; no catalytic activity towards RNA has been reported to date^{170,171}. Additionally, it targets cytosines within the expressed antibody gene at a higher frequency and specificity than all other genomic loci^{172,173}; although the underlying mechanism is unknown, it likely relies both on the locus architecture and on region-restricted binding cofactors. Finally, unregulated AID activity can catalyse off-target mutations and chromosomal translocations^{107,108} (FIG. 1), although at rates substantially lower than those reported within the *Ig* gene.

APOBEC2

APOBEC2 is expressed in cardiac and skeletal muscle¹⁷⁴. All APOBEC2 knockout animal models tested (from zebrafish to mouse) display a myopathy^{175,176}. The strong conservation of APOBEC2 suggests an undiscovered essential molecular function^{158,177}, which does not seem to require deamination⁴⁸. Indeed, we have reported in a preprint that APOBEC2 does not deaminate RNA or DNA but has retained the ability to bind DNA at specific promoter regions — and, through this functionality, to act as a transcriptional repressor¹¹.

APOBEC3

APOBEC3 is encoded by a single gene in mice, which has expanded and diverged in humans and primates into a seven-gene subfamily encoding seven proteins (designated APOBEC3A (A3A), A3B, A3C, A3D, A3F, A3G and A3H)¹⁷⁸. The major role of these proteins seems to be the restriction of viruses and genomic mobile elements^{81,179}, through the deamination of their obligatory single-stranded cDNA intermediates¹⁸⁰ (FIG. 1b). Such damaged cDNA can be cleaved and degraded¹⁸¹. However, if the viral genome is not degraded, deamination by APOBEC3 results in mutations that support viral evolution (FIG. 4). Certain APOBEC3 family members have also been reported to target the host genome leading to point mutations, DNA breaks and chromosomal instability^{120,121} (FIG. 1b). These effects may drive cancer evolution (FIG. 4) through an increase in tumour heterogeneity¹²⁴.

APOBEC4

APOBEC4 is expressed in the mammalian testis¹⁸². Two recent works suggest a role for APOBEC4 in promoter modulation within mammalian cells¹⁸³ and the antiviral response in birds¹⁸⁴. However, as very little is known about this protein, it will not be discussed further in this Perspective.

Key enzymatic features

Here, we describe common structural features of the generalists (APOBEC1, A3A and A3G) and how they relate to their ability to bind to and deaminate RNA and DNA. We compare and contrast these features with those of mammalian AID and APOBEC2, two specialists that have lost this substrate flexibility and have assumed unique functions. We specifically focus on the substrate binding groove of these

proteins, which is largely defined by four loops surrounding the active site (loops 1, 3 and 7, with minor contributions from loop 5) (reviewed elsewhere¹³) (FIG. 2b). These loops have been demonstrated to be responsible for the interaction of AID/APOBECs with their substrates^{15–21}, but they also help delineate the contours of the catalytic pocket. Therefore, together they define most enzymatic functionality — from substrate binding to dinucleotide preference to catalysis.

Loop 7 residues and nucleic acid interaction within generalists.

Recent co-crystal structures of A3A bound to a six-nucleotide single-stranded DNA (ssDNA) substrate (PDB:5SWW) (TABLE 2) revealed that the deoxycytidine to be deaminated (C_0) is located at the bottom of a substrate binding groove shaped by loops 1, 3 and 7, where it forms a π -stacking interaction with Y130 in loop 7 (FIG. 3a). A3A protein variants in which Y130 is replaced by an alanine (Y130A) lacked deaminase activity in vitro, proving the catalytic importance of this residue²⁰. Y130 of A3A corresponds to residues Y315 of A3G and F120 of APOBEC1 (FIG. 3b,c). The co-crystal structure of A3G bound to a nine-nucleotide ssDNA substrate containing a 5'-TCCCA-3' target sequence (PDB:6BUX) (TABLE 2) confirmed that Y315 forms a π -stacking interaction with C_0 (REF.²²) (FIG. 3b). Moreover, Y315A variants of A3G have significantly reduced binding to both ssDNA and RNA substrates in in vitro binding assays²³. Co-crystal structures of APOBEC1 bound to nucleic acid substrates are not yet available, but evidence suggests it too interacts with C_0 through a π -stacking interaction: F120A variants have little or no deaminase activity towards RNA or DNA substrates in in vitro assays²¹; and alignment of A3A and A3G co-crystal structures with the structure of APOBEC1 (PDB:6X91) (TABLE 2) shows an almost perfect overlap of APOBEC1 F120 with A3A Y130 and A3G Y315 (FIG. 3b). Molecular dynamics simulations suggest that this catalytic pocket 'breathes', often occluding or restricting substrate entrance^{20,24}, suggesting it could enforce local sequence preference and might even be selectively druggable.

The residues in loop 7 are also essential for the interaction with the nucleotide preceding C_0 (the -1 position), thereby defining the local dinucleotide sequence preference (FIG. 3c). This was demonstrated by cell-based assays in which loop 7 of A3G was replaced with the corresponding region from A3A, which changed its dinucleotide preference from 5'-CC to 5'-TC (the preference of A3A)¹⁸. Single amino acid exchanges demonstrated the critical role of D317 of A3G as a main determinant of dinucleotide substrate preference. Interestingly, D317W led to stronger preferences towards 5'-TC than did D317Y, suggesting that amino acids with larger aromatic side chains may be even more favourable to 5'-TC deamination¹⁸. The currently available co-crystal structures of A3A and A3G bound to nucleic acids (TABLE 2) provide an explanation of these findings. D131 and Y132 of A3A have a

primary role in defining the A3A preference for a 5'-TC sequence motif: although these residues have the potential to interact

extensively with either T₋₁ or C₋₁, the size of the -1 pocket precludes access of the larger purine^{16,20}. Taken together, these

findings highlight the importance of the residues in loop 7 in determining the local dinucleotide preference of a generalist,

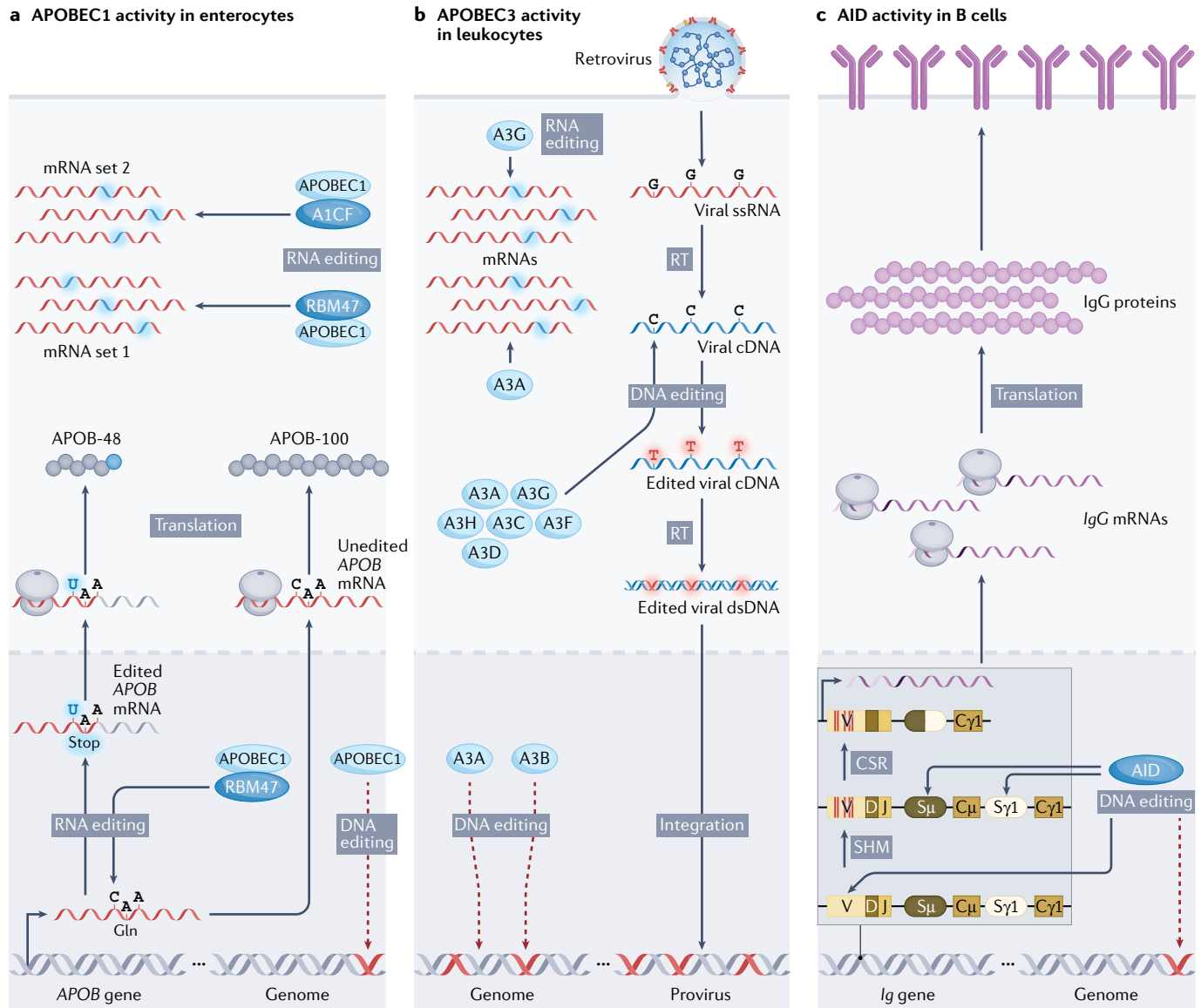


Fig. 1 | Physiological and aberrant functions of the AID/APOBEC deaminases. a | APOBEC1 in humans and mice acts in the nucleus of enterocytes, together with its cofactor RNA-binding motif protein 47 (RBM47), to edit apolipoprotein B (APOB) mRNA¹⁵⁵. Editing leads to C-to-U base change that converts Gln (CAA) to a stop codon (UAA). Edited and unedited APOB mRNAs are then translated in the cytoplasm, generating two distinct isoforms: short (APOB-48) and long (APOB-100) isoform, respectively. APOB-100 is the major component of plasma low-density lipoproteins whereas APOB-48 is essential for secretion of chylomicrons. In mice, APOBEC1, together with RBM47, catalyses RNA editing of a large set of additional transcript targets (mRNA set 1). A change of cofactor from RBM47 to APOBEC1 complementation factor (A1CF) leads to RNA editing of a different set of transcript targets (mRNA set 2), suggesting that target specificity resides with the cofactor. Finally, in mice and humans, APOBEC1 is also able to induce DNA editing within the genome of the cells, leading to undesired mutations (dashed red arrow)^{109,141,142}. **b** | In humans, APOBEC3 family members play an essential role during retroviral infections (for example, in leukocytes). Specifically, once a retrovirus infects a cell, it releases its viral genome as single-stranded RNA (ssRNA), which is retro-transcribed (RT)

to cDNA. APOBEC3 proteins can deaminate this single-stranded DNA (ssDNA) leading to C-to-T base changes and mutations within the viral genome. This edited viral genome can be degraded (if heavily edited) or integrated into the genome as a provirus. APOBEC3A (A3A) and A3G are also able to perform RNA editing on RNA viruses (such as SARS-CoV-2) as well as host mRNAs. Aberrant activity of A3A and A3B can also induce DNA mutations within the genome of cells (dashed red arrows). **c** | AID plays an essential role in B cell antibody diversification, where it catalyses deamination either within transcribed (black arrow) antibody variable region V(D)J gene segments of the immunoglobulin (Ig) gene leading to somatic hypermutation (SHM) (mutations represented as red bars) or within repetitive 'switch' regions upstream of the constant region gene segments, leading to class switch recombination (CSR) (switch regions S μ or S γ 1 shown). Resulting mRNA encodes an IgG1 protein that contains a hypermutated variable region and a γ 1 heavy chain. Unregulated AID activity can also result in mutations and translocations elsewhere in the genome (dashed red arrow). AID/APOBEC, activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like; dsDNA, double-stranded DNA.

Table 1 | AID/APOBEC members are classified as generalists or specialists based on their substrate flexibility and functional restriction

Protein	Substrate				Subcellular localization	Tissue expression ^a	Target sequence	Known biological functions
	Host		Viral					
	RNA	DNA	RNA	DNA				
Generalists								
A3A	ss	ss	ss	ss	N/C	Monocytes and macrophages, bone marrow and lymphoid tissue, urinary tract and bladder	5'-TC-3'	Protection from viruses
A3G	ss	No	ss	ss	C	Peripheral blood cells, IFN α -activated cells, bone marrow and lymphoid tissue, breast tissue, reproductive system, urinary tract and bladder, gastrointestinal tract, liver and gallbladder, respiratory tract	5'-CC-3'	Protection from viruses and retroelements
APOBEC1	ss	ss	ss	No	N/C	Gastrointestinal tract (mouse and human); immune cells (mouse only)	5'-AC-3'	Regulation of cholesterol metabolism (mouse and human); modulation of monocyte transcriptome (mouse only); potentially antiviral ⁹⁵
Specialists								
AID	No	ssDNA and RNA/DNA hybrid	No	No	N/C	Germinal centre B cells, lymphoid tissue	5'-WRC-3'	Secondary antibody diversification
APOBEC2	No	ss/ds binding only	No	No	N/C	Muscle (skeletal and cardiac); some B cell/T cell subsets	ND	ND
Unassigned								
A3B	ND	ss	No	ss	N	IFN α -activated liver cells, bone marrow, gastrointestinal tract, kidney, urinary tract and bladder	5'-TC-3'	Protection from viruses
A3C	No	ND	No	ss	N/C	Peripheral blood cells, bone marrow and lymphoid tissue, skin, muscle, breast tissue, reproductive system, kidney, urinary tract and bladder, gastrointestinal tract, liver and gallbladder, proximal digestive tract, respiratory tract, endocrine tissues	5'-TC-3'	Protection from viruses
A3D	No	No	No	ss	C	Peripheral blood cells, bone marrow and lymphoid tissue, gastrointestinal tract, female reproductive system	5'-TC-3'	Protection from viruses
A3F	No	No	No	ss	C	Peripheral blood cells, IFN α -activated cells, bone marrow and lymphoid tissue, reproductive system, endocrine tissues, muscle	5'-TC-3'	Protection from viruses
A3H	ND	ND	No	ss	N/C	Peripheral blood cells, bone marrow and lymphoid tissue.	5'-TC-3'	Protection from viruses

APOBECs that require further studies on substrate preference and activity are classified as 'unassigned'. The underline indicates the C that is deaminated by each enzyme. A3A, APOBEC3A; AID/APOBEC, activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like; C, cytoplasmic; ds, double-stranded; N, nuclear; N/C, nuclear and cytoplasmic; ND not determined; ss, single-stranded. ^aTissue expression courtesy of the [Human Protein Atlas](#)¹³⁴.

most likely by affecting the size of the substrate binding pocket.

Although these experiments indicate that the catalytic pocket confers some degree of substrate sequence specificity, small-molecule inhibitors that can discriminate between AID/APOBEC family members have yet to be identified — implying that the catalytic pockets of these enzymes share strong common features (recently reviewed elsewhere²⁵) and that the main determinant of substrate selectivity may, in fact, be the substrate binding groove.

The substrate binding groove of generalists is U-shaped. In vitro assays indicate that generalist AID/APOBECs prefer structured substrates; disrupting stem-loops in ssDNA and RNA substrates directly alters the frequency with which they are deaminated by A3A and A3G (REFS^{26–29}). Moreover, the co-crystal structures of A3A and A3G bound to nucleic acid demonstrate that the U-shaped substrate binding groove formed by loops 1, 3, 5 and 7 (with the catalytic pocket located at the bottom of the U shape where the π -stacking

interaction occurs)^{16,20} (FIG. 3d) optimally accommodates a stem-loop structure. Although crystal structures of APOBEC1 bound to ssDNA or single-stranded RNA (ssRNA) are not currently available, its best-studied target, apolipoprotein B (APOB) mRNA, is predicted to form a stem-loop secondary structure^{30–32}. Thus, generalists bind substrates with similar conformations to tRNA, the substrate of the distantly related tRNA adenosine deaminases (such as Tada³³; FIG. 2a), suggesting that the shape of the binding

groove in generalists has co-evolved with the structure of their nucleic acid substrate. Moreover, the shape of this groove could be predictive of generalists. We note here that a similarly shaped groove (termed ‘patch 1’) is evident in co-structures of A3H with ssDNA and RNA³⁴. For the purposes of this Perspective, A3H is considered ‘unassigned’ because its ability to catalyse RNA editing has not yet been tested. However, given the shape of its substrate binding groove and its demonstrated ability to bind RNA (see below), we would predict

that it too might function as a bona fide RNA editor.

Groove residues help generalists discriminate between RNA and DNA substrates. Residues in the loops forming the substrate binding groove of AID/APOBECs have key roles in substrate discrimination (that is, binding of RNA versus DNA). For example, a W121A substitution in loop 7 of APOBEC1 almost completely abolishes deamination of RNA while retaining activity on DNA, indicating an essential role of this

amino acid in substrate differentiation²¹. Notably, alignment with other APOBECs reveals that W121 in APOBEC1 corresponds to Y113 in A3H (FIG. 3c), a residue that directly interacts with a ribose 2'-hydroxyl of bound RNA^{15,19} (PDB:6B0B, PDB:5W3V) (TABLE 2). The same residue also corresponds to D131 in A3A and D316 in A3G. As discussed above, these residues have been shown to be important for deamination activity on ssDNA and for local dinucleotide sequence preference^{16,18,20,35}, but no evidence yet exists for their function on

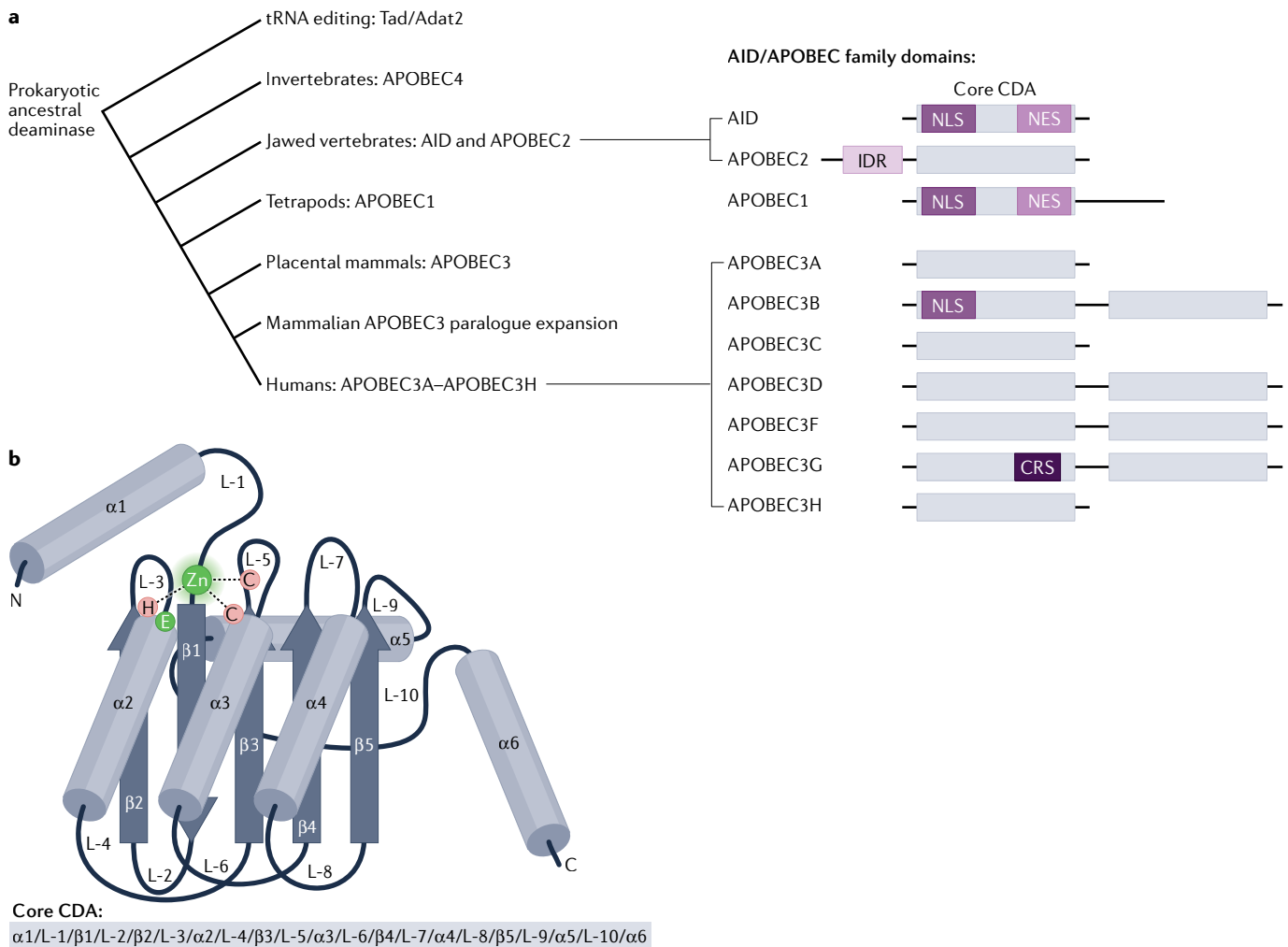


Fig. 2 | The emergence of the AID/APOBEC family and the conserved core cytidine deaminase domain. **a** | Left: simplified phylogenetic tree of the AID/APOBEC family, which is believed to have emerged by co-opting prokaryotic tRNA editing enzymes (Tad/ADAT2) to deaminate DNA¹⁵⁶. In the vertebrate-specific branch, AID and APOBEC2 are the most ancient members (present in cartilaginous and bony fish)¹⁵⁷, APOBEC1 emerged later in the tetrapod–lungfish divergence; and APOBEC3 appeared even later, in placental mammals. Both are believed to have evolved from AID gene duplications^{82,158}. Paralogue expansion within placental mammals led to emergence of several APOBEC3 subfamily members, with the seven members of the human subfamily being among the most diverse^{82,159}. More recently, orthologues of APOBEC4 have been found in invertebrates, suggesting it predates rest of family members and forms a separate invertebrate branch¹⁵⁸. Right: domain delineation of members of the vertebrate-specific AID/APOBEC family.

Each member of the family contains the core zinc-dependent cytidine deaminase domain (core CDA). Specific members contain accessory motifs within core CDA that determine subcellular localization, including nuclear localization signal (NLS), nuclear export signal (NES) and cytoplasmic retention signal (CRS). Some members contain additional accessory regions that provide specific molecular properties: for example, APOBEC2 contains an amino-terminal intrinsically disordered region (IDR) whereas the carboxy terminus of APOBEC1 is hydrophobic. **b** | Core CDA composed of a five-stranded β -sheet ($\beta 1$ – $\beta 5$) surrounded by six α -helices ($\alpha 1$ – $\alpha 6$). Several loops found within the deaminase fold (L-1 to L-10) with loops 1, 3, 5 and 7 forming the substrate binding groove. Catalytic pocket coordinates a zinc ion (Zn; green sphere) with the His-Glu (H and E) and Cys-Cys (C) motifs found on $\alpha 2$ and L-5/ $\alpha 3$, respectively. AID/APOBEC, activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like.

Table 2 | Available crystal structures of AID/APOBECs

Protein	Species	Ligand sequence	PDB	Ref.
APOBEC2	Human	NA	2NYT	45
A3A	Human	5'-dTdTdCdTdT-3'	5KEG	16
A3A	Human	5'-dAdTdTdCdGdGdG-3'	5SWW	20
A3H	Human	Endogenous RNA (8-mer mixed sequence)	6B0B	19
A3H	Pig-tailed macaque	Endogenous RNA (9-mer mixed sequence)	5W3V	15
AID	Human	dCMP	5W3V	37
A3G	Human	5'-dAdAdTdTdCdCdAdAdA	6BUX	22
APOBEC1	Human	NA	6X91	21

A3A, APOBEC3A; AID/APOBEC, activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like; NA, not available.

RNA. However, two A3A protein variants were recently described that exclusively deaminate RNA. Both variants have a Y132G substitution combined with additional substitutions in loop 1 or helix 6, implicating these amino acids in substrate discrimination³⁶. Nonetheless, much remains to be learned about how generalist APOBECs discriminate between RNA or DNA substrates; additional structures of the proteins bound to DNA or RNA, in combination with genetic studies targeting specific amino acids, will be necessary to pinpoint the residues that define substrate selectivity.

Structural differences in grooves of specialists reflect functional differences.

The structure of AID (with a dCMP bound within the catalytic pocket) (PDB:5W3V) (TABLE 2) revealed a bifurcated, rather than U-shaped, substrate binding surface. Residues of loops 1, 3 and 7 are essential in shaping the substrate channel where the dCMP coordinates³⁷ (FIG. 3e), which is connected to a second groove, termed the 'assistant patch'³⁷ (FIG. 3e). Positively charged basic residues in these channels form a binding surface, which is separated near their point of convergence by negatively charged residues in loop 7 (the 'separation wedge') (FIG. 3e). Groove residues are highly conserved in AID proteins from different species, but not among other APOBECs, highlighting that this structure is specific to AID. Interestingly, similar separation wedge structures have been observed for proteins that recognize branched nucleic acids, such as T4 RNase H³⁸ or Cas9 (REF.³⁹), suggesting that AID recognizes structured substrates. Although AID targeting mechanisms are still not fully clarified, the conformation of the substrate binding region agrees with recent experiments that reveal a possible role for G-quadruplex structures in guiding and targeting AID, at least in the context of

immunoglobulin class switch recombination (CSR)^{37,40}. These data also highlight the importance of the substrate binding groove structure in allowing different AID/APOBECs to discriminate substrates based on their secondary structure. It must also be noted that AID, similar to other specialists, can bind RNA⁴¹, especially within RNA–DNA hybrids⁴², but cannot deaminate it⁴³, suggesting again that binding is required but not sufficient for catalysis.

The structure of APOBEC2 was the first among the AID/APOBEC family to be published^{44,45} (PDB:2NYT) (TABLE 2), but little is known about its molecular substrate and so co-crystal structures are currently unavailable. As such, it is not possible to assess the conformation of the substrate binding groove, but the APOBEC2 structure does provide some insight into its lack of deaminase activity. E60 in APOBEC2 forms a point of coordination with the zinc ion that is absent from catalytically active AID, A3A, A3G or APOBEC1, and this may affect catalytic activity by disrupting coordination of an essential water molecule or by modulating substrate affinity⁴⁶ (FIG. 3f). Deamination could also be prevented by obstruction of the nucleic acid binding pocket by loop 1 (FIG. 3f). However, given the flexibility of this loop seen in its solution structures⁴⁴, intermolecular interactions affecting its conformation may allow transient access to the deaminase active site and transient interactions with nucleic acid. Recent work from our laboratory, which has been made available as a preprint, strongly suggests that APOBEC2 has retained the ability to interact with ssDNA containing GC-rich motifs; moreover, this interaction seems to affect gene expression¹¹. It is tempting to speculate that, similar to AID, APOBEC2 may interact with G-quadruplex structures found within these GC-rich promoter sequences. Alternatively, APOBEC2 may interact with transient

ssDNA structures resulting from RNA polymerase promoter melting, in a manner similar to other APOBECs⁴⁷. We currently speculate that transcriptional repression through chromatin interaction may be an evolutionarily conserved function of APOBEC2 (REF.⁴⁸), especially in the context of cellular reprogramming.

Taken together, the available AID/APOBEC structures illustrate the flexibility of their core structure and how it maintains the active site requirements of the family while enabling substrate restriction and functional specialization in some members or broader substrate preference and functional plasticity in others.

Subcellular localization

Regardless of the innate capacity of an AID/APOBEC protein to bind and deaminate DNA, RNA or both substrates, its ability to do so in cells will depend on its subcellular localization and its access to the specific substrate. Whereas mRNA, viral RNA and viral DNA can all be deaminated in either the nucleus or the cytoplasm, the host genome can only be deaminated by nuclear-localized family members. For example, despite having DNA binding and deamination capabilities, the generalist A3G cannot mutate genomic DNA because it is confined to the cytoplasm.

The subcellular localization of each member of the AID/APOBEC family may depend on active or passive cellular mechanisms. Transit of AID and APOBEC1 between the nucleus and the cytoplasm relies on both an amino-terminal bipartite basic nuclear localization signal (NLS) sequence and a strong carboxy-terminal leucine-rich nuclear export signal (NES) sequence^{49,50} (FIG. 2a). APOBEC1 also contains a C-terminal hydrophobic domain, which is involved in intramolecular interactions that can play a part in further defining subcellular localization²¹. An extensive study of AID and APOBEC2 protein chimaeras showed that nuclear import of AID involves residues in addition to the N-terminal NLS, whereas APOBEC2 lacks NLS or NES motifs and, instead, passively diffuses between the cytoplasmic and nuclear compartments⁵¹. Unlike the rest of the AID/APOBEC family, APOBEC2 contains an N-terminal glutamate-rich acidic intrinsically disordered region (IDR), which could further restrict its subcellular localization through intermolecular interactions with shuttling proteins or cofactors (FIG. 2a).

Single-domain human APOBEC3 paralogues, A3A, A3C and A3H, are small

enough (~25 kDa) to passively enter and exit the nucleus, and are generally found throughout the cell during interphase⁵² (FIG. 2a; TABLE 1). For example, A3H lacks an NLS but enters the nucleus through passive diffusion and is retained within the nucleolar subcompartment⁵³. By contrast, the larger (>50 kDa) double-domain APOBEC3 paralogues cannot passively enter the nucleus; A3B is constitutively nuclear owing to its N-terminal NLS^{52–54}, whereas A3D, A3F and A3G lack an NLS and are mostly found within the cytoplasm (FIG. 2a). Interestingly, A3G seems to contain a novel cytoplasmic retention signal (CRS)⁵⁵. All human APOBEC3 paralogues are excluded from chromatin during mitosis when the nuclear envelope breaks down, which presumably inhibits genome mutagenesis⁵² (REF.¹⁴ offers an in-depth review on trafficking kinetics of the AID/APOBEC family of proteins).

Cofactors

AID/APOBEC enzymes interact with numerous protein cofactors that enable them to carry out their functions in the cell. Here, we focus on cofactors that affect substrate targeting or modulate catalytic activity.

To date, APOBEC1 is the only AID/APOBEC protein for which specific cofactors have been demonstrated to modulate its catalytic activity. In mice, APOBEC1 is expressed in the small intestine and the liver, where it edits a specific cytosine within the *APOB* pre-mRNA. The C-to-U RNA editing event recodes a CAA codon to a stop codon, resulting in a truncated form of the *APOB* protein, called APOB-48 (REFS^{56,57}) (BOX 1; FIG. 1a). Two cofactors of mouse APOBEC1 (mAPOBEC1) — APOBEC1 complementation factor (A1CF)^{58,59} and RNA-binding motif protein 47 (RBM47)⁶⁰ — have so far been identified, but given that

doubly mutant mice lacking both of these cofactors still retain some C-to-U editing activity, other cofactors are likely to exist^{61,62}. A1CF and RBM47 bind RNA, interact directly with APOBEC1 protein^{58–60} and have an essential role in defining which RNAs are targeted for editing as well as determining the level of editing per target^{61,62}. Elegant genetic dissection in a mouse system suggests that cofactors ‘recruit’ different (sometimes partially overlapping) sets of transcripts to the editing complex (FIG. 1a) and that cofactor dominance is associated with editing frequency^{61,62}. Together with the fact that APOBEC1 exerts its biological function by deaminating target cytosines within cohorts of transcripts that define common pathways⁶³, these experiments support the idea that distinct tissues drive APOBEC1 to specific sets of transcripts through the provision of different sets of cofactors⁶⁴.

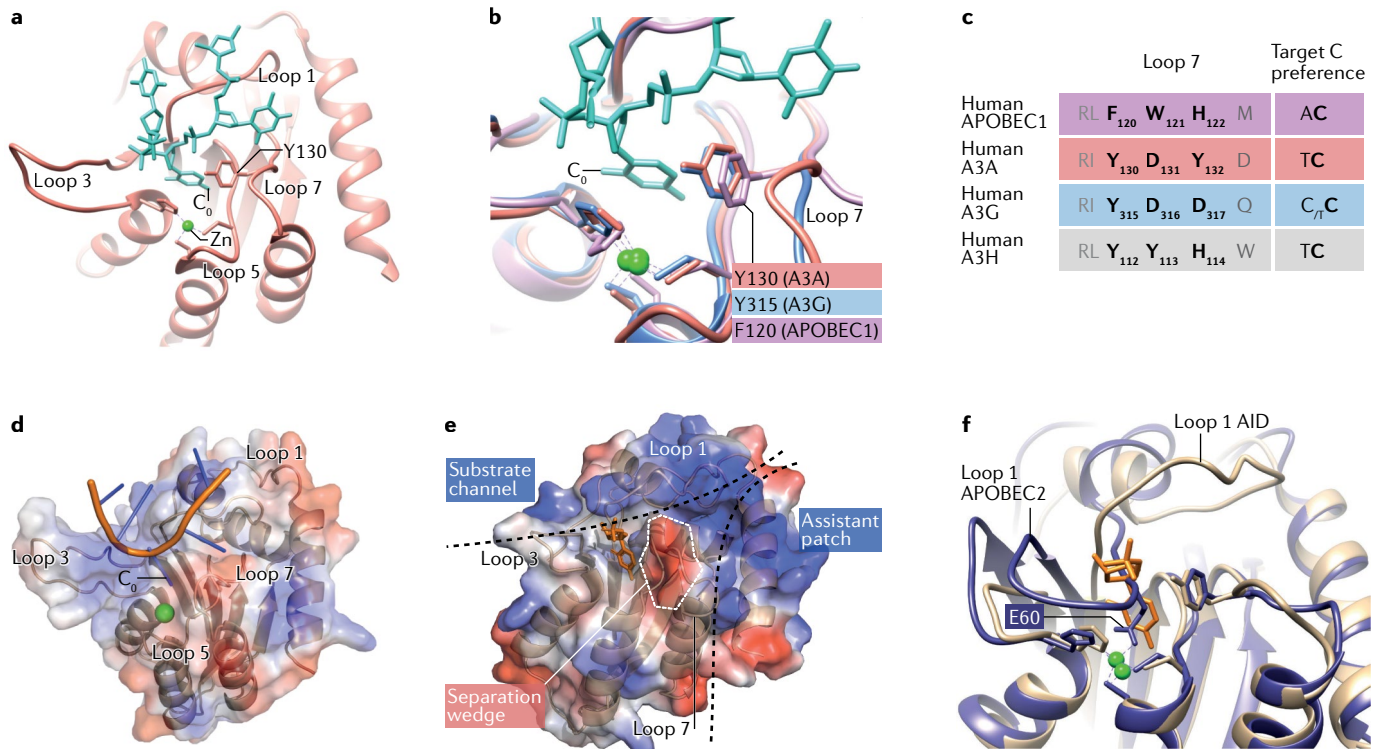
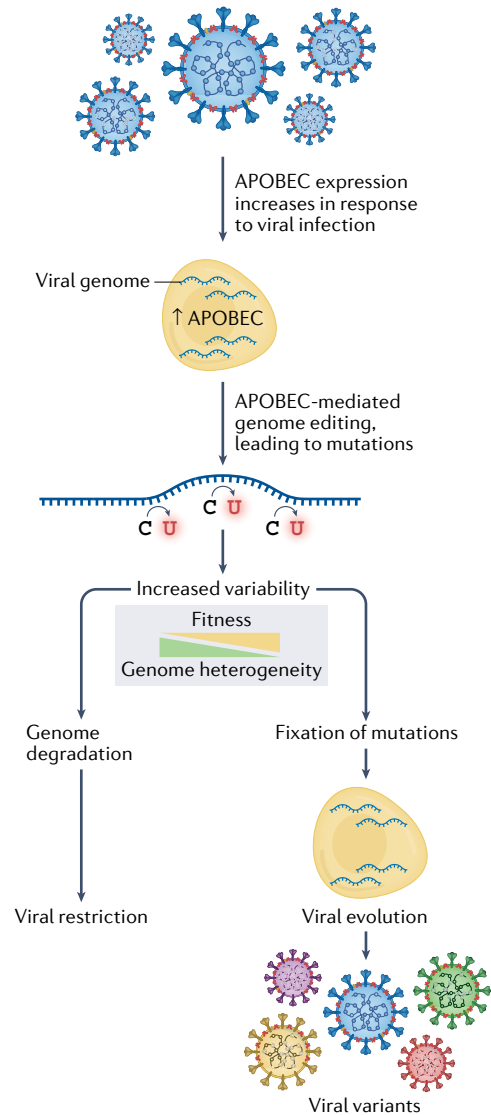


Fig. 3 | Structural insights from generalists and specialists. a | Co-crystal structure of APOBEC3A (A3A) bound to six-nucleotide single-stranded DNA (ssDNA; turquoise) (PDB:5SWW) (TABLE 2). Target deoxycytidine (C₀) located at bottom of the substrate binding groove formed by loops 1, 3 and 7 and forms a π -stacking interaction with Y130 (PDB:5SWW) (TABLE 2). **b** | Overlapping crystal structures of APOBEC1 (purple) (PDB:6X91) (TABLE 2), A3A bound to ssDNA (pink) (PDB:5KEG) (TABLE 2) and A3G bound to ssDNA (blue) (PDB:6BUX) (TABLE 2). Residues F120 (APOBEC1), Y130 (A3A) and Y315 (A3G) form critical aromatic π -stacking interactions with target C (C₀, from co-crystal structure of A3A bound to ssDNA; turquoise) (PDB:5SWW) (TABLE 2). **c** | Alignment of amino acids present in loop 7 for different APOBECs and target C preference motif for each. **d** | A3A binds U-shaped substrates, such as ssDNA (orange) (nucleobases represented as blue sticks) (PDB:5SWW) (TABLE 2). **e** | AID in co-crystal

structure with dCMP ligand (PDB:5W0U) (TABLE 2) shown as a molecular surface. AID loops 1, 3 and 7 form positively charged (blue) bifurcated substrate binding surface, comprising ‘substrate channel’ (which hosts dCMP) and a second groove, termed the ‘assistant patch’. The two grooves are separated near the point of convergence by negatively charged residues in loop 7 (red) known as the ‘separation wedge’. **f** | Co-crystal structure AID (light brown) with a dCMP ligand (orange) (PDB:5W0U) (TABLE 2) overlaid with crystal structure of APOBEC2 (blue) (PDB:2NYT) (TABLE 2). Loop 1 of APOBEC2 obstructs substrate (orange) at the active site. Residue E60 in APOBEC2 forms a fourth point of coordination with zinc ion (Zn; green sphere). AID/APOBEC, activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like. Part **e** is adapted from REF.¹³, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>).

a APOBECs in viral restriction and evolution



b AID/APOBECs in tumour restriction and evolution

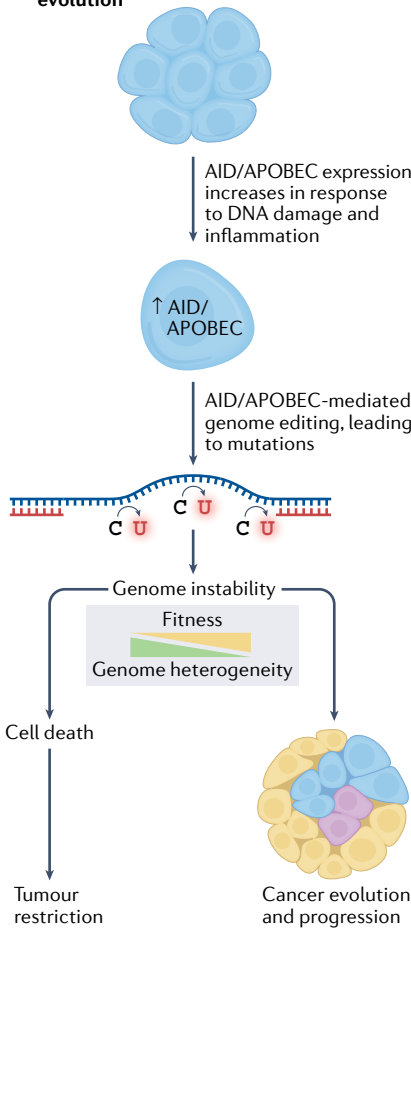


Fig. 4 | The consequences of deamination for adaptive evolution. **a** | APOBEC levels are upregulated in response to viral infection, and both single-stranded DNA (ssDNA) and single-stranded RNA (ssRNA) viral genomes can undergo APOBEC-mediated deamination resulting in mutations. When mutational load in the viral genome is so high that the genome cannot perform its function, the genome is degraded and viral particles are not produced; this process is known as viral restriction. However, mutations resulting from lower levels of deamination can become fixed in the viral genome after replication, increasing the probability of producing viral variants with altered characteristics (compared with the original strain). **b** | If AID/APOBEC-catalysed C-to-U deamination events are not repaired by the base excision repair pathway during replication, resulting C-to-T transitions can induce massive DNA damage and genome instability. Cell death can be triggered if accumulation of transitions leads to an excessive mutational load, resulting in tumour restriction. However, lower levels of mutation can fuel genome variability and tumour cellular heterogeneity that, upon selection, can result in adaptive evolution and cancer progression. AID/APOBEC, activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like.

Several potential cofactors have been identified for AID^{65–70}, but none has been proven to be the key determinant in targeting AID to the immunoglobulin locus, its physiological target. Finally, a secondary Zn²⁺ ion has been shown to allosterically modulate catalysis of A3A and A3G (REF⁷¹). Although not a cofactor in the traditional sense, this

functionality points to possible surfaces that could be occupied by more traditional cofactors to regulate enzymatic function.

AID/APOBECs drive adaptive evolution RNA editing and DNA mutations have very different features; editing is transient and tunable, whereas mutations are irreversible

and heritable. Despite these differences, both mechanisms create genetic variability that has an essential role in adaptive evolution^{72–74}. In this section, we discuss how AID/APOBEC proteins can drive adaptive evolution in viral and cancer genomes owing to their ability to deaminate both RNA and DNA.

APOBEC3 proteins in viral genome evolution

Early experiments predicted that T cells express a factor that blocks the replication of viral infectivity factor (Vif)-deficient human immunodeficiency virus type 1 (HIV-1)^{75,76}. A3G was later identified as one of the factors responsible for this HIV-1 restriction through active deamination of nascent retroviral cDNA^{77–79}, with subsequent studies highlighting the involvement of A3D, A3F and A3H (REFS^{80,81}). Although many of these experiments were performed in APOBEC3-overexpressing cells infected with pseudotyped HIV and may not fully reflect in vivo conditions, the general consensus is that several APOBEC3 proteins individually and synergistically restrict viral infectivity of HIV and many other viruses during natural infections, a view that is supported by the substantial expansion of the APOBEC3 family in organisms that support large infection loads, such as bats^{82,83}. In a process known as hypermutation, APOBEC3 proteins can deaminate a substantial proportion of the total cytosines in the HIV cDNA in a single round of viral replication, with reports of up to 10% in in vitro or cell culture experiments and up to 98% in HIV sequences isolated from peripheral blood mononuclear cells. The resulting uracils are recognized and excised by the host uracil DNA N-glycosylase (UNG) protein, which initiates the base excision repair pathway and, ultimately, leads to heavily damaged genomes containing multiple abasic sites. These genomes can be further cleaved and degraded, thereby decreasing viral infectivity^{77,83}. However, genomes with less extensive damage (and fewer abasic sites) can simply be repaired, often resulting in mutations that can support viral evolution⁸⁴ and the acquisition of drug resistance⁸⁵, altered transmission and immune escape^{85,86} (FIG. 4a).

Analysis of HIV genomes that have undergone hypermutation or are associated with immune escape reveals an enrichment of APOBEC3-defined mutational signatures, which, in conjunction with biochemically derived triplet preferences, strongly support a physiologic role for specific APOBEC3

family enzymes in both viral restriction and viral evolution (reviewed elsewhere⁸⁷). Although the majority of knowledge surrounding APOBECs and viral restriction comes from the study of retroviruses, DNA viruses such as hepatitis B virus (HBV) and human papilloma virus (HPV) are also restricted by APOBEC3 enzymes^{88–90}. Additionally, some APOBEC3 proteins can also deaminate viral genomes composed solely of RNA, such as the positive-sense RNA genome of the betacoronavirus SARS-CoV-2. Soon after the beginning of the COVID-19 pandemic, RNA sequencing data from bronchoalveolar lavage fluid of patients with COVID-19 was used to monitor the mutational signatures shaping the viral genome before fitness selection^{91,92}. The most common mutations detected in these sequencing data were A-to-G and T-to-C changes (possibly the outcome of ADAR1 activity on the positive-sense and negative-sense strands, respectively, during viral replication) followed by C-to-T and G-to-A changes, likely mediated by APOBEC3 proteins, the only AID/APOBEC family members known to bind and deaminate viral RNA^{91–93}. The involvement of APOBEC3 proteins is further supported by the frequent occurrence of edited Cs within the motif 5'-U/ACU/A-3' (REFS^{91,94}) (although a recent preprint indicates this could also be explained by APOBEC1-mediated deamination⁹⁵) and in terminal loop rather than stem sequences⁹⁶, and the upregulation of APOBEC3 proteins in samples from patients with COVID^{97–100}. Analysis of SARS-CoV-2 genomic sequences largely acquired through the process of viral genome surveillance of variants of interest over the course of the pandemic has revealed that, after fitness selection, about 40% of all mutations involve C-to-T changes (reviewed elsewhere^{100,101}), which are at least partially confined to a group of mutational hotspots¹⁰², a pattern consistent with APOBEC3 activity. Numerous other ssRNA viruses (including human T cell leukaemia virus type 1 (HTLV-1) and rubella) have been shown to be targeted by APOBEC3 proteins (reviewed elsewhere⁸¹). Overall, deep sequencing data strongly support a functional role of APOBEC3 family members in the restriction of ssRNA viruses in natural settings.

Taken together, these studies clearly show the effects of APOBEC3 mutagenesis on viral genomes and its relevance to virus evolution^{84,103,104}. As generalists with a preference for viruses with ssRNA and DNA genomes¹⁰⁵, A3A and A3G contribute to restriction of a range of viruses but

can also drive evolution of retroviruses (such as HIV-1 (REF.⁸⁵)), DNA viruses (such as herpesviruses⁷⁴) and also ssRNA viruses that lack ssDNA intermediates (including SARS-CoV-2 and rubella among others⁹⁶).

AID/APOBECs and cancer evolution

The first solid piece of genetic evidence linking any AID/APOBEC family member to cancer was the finding that APOBEC1 overexpression in the liver of transgenic animals induces hepatocellular carcinoma¹⁰⁶, although whether this was the result of RNA editing or DNA mutation remained unclear. Ectopic expression of AID was later shown to catalyse off-target DNA mutations and chromosomal translocations^{107,108}, albeit at rates substantially lower than those reported for its true target, the immunoglobulin genes. Subsequently, some APOBEC3 family members (chiefly those with access to the nucleus) were reported to be a cause of DNA damage and mutagenesis^{109,110}. Indeed, based on mutational signatures found in cancer genomes, AID/APOBEC-derived mutations are present in more than 50% of human cancer types, and account for 5–90% of all substitution mutations^{111,112}. In addition, AID/APOBEC mutations can occur in clusters over kilobase-sized regions^{113,114}. These hypermutated clusters are termed kataegis mutations^{113,115} and have been reported in more than 60% of cancers¹¹⁶. They are especially prominent in cancer types where APOBEC3 mutagenesis is active¹¹⁷. Expression of some AID/APOBEC enzymes in tumours (such as AID in chronic myeloid leukaemia¹¹⁸ or A3B in tamoxifen-resistant breast cancer¹¹⁹) has been correlated with increased tumour evasion and drug resistance, suggesting that they drive tumour evolution. Independently of kataegis mutations, APOBEC3-catalysed mutagenesis can also lead to chromosomal instability^{120,121} and, thus, to either cell-autonomous lethality^{122,123} or to cancer evolution through increased tumour heterogeneity¹²⁴. Given that these outcomes mirror those of APOBEC3-mediated viral restriction, we hypothesize that expression of APOBEC3 proteins is induced by the inflammatory cancer microenvironment in an attempt to kill malignant cells via localized hypermutation. However, when APOBEC3-mediated mutation is not successful in achieving tumour restriction^{125,126}, the tumour cells that have evaded cell death (that is, those with non-lethal levels of mutation) can drive cancer evolution, thus leaving behind a mutational signature in the genome at sites that are likely directly related to the original drive to restrict²⁶ (FIG. 4b).

The RNA editing capacity of some AID/APOBEC deaminases has also been directly linked to the generation of heterogeneity essential to tumour evolution^{127–131} (for a comprehensive recent review on the AID/APOBEC but also ADAR contribution to tumour evolution, see REF.¹²⁸). For example, loss of editing (through ablation of *ApoBec1*) in the small intestine of a mouse model of intestinal cancer (the APC^{min} mouse) leads to substantial tumour reduction¹³². Additionally, deletion of *ApoBec1* from the germline of a mouse model of testicular cancer (in which around 8% of male mice succumb to testicular teratocarcinomas by 4 weeks of age) ablates susceptibility¹³³. Finally, it has recently been demonstrated that the location of A3A-catalysed DNA mutations in cancer genomes can be predicted in clinical samples by monitoring the frequency of A3A RNA editing at the same loci²⁸. This finding supports the notion that editing precedes mutation and that RNA editors induced under inflammatory conditions can also inflict DNA damage, such as kataegis mutation. More generally, these data imply that the RNA editing state of a cell determines the fate of that cell, even in the absence of a heritable genomic mutation. Indeed, both A3A expression and RNA editing were detected in cancers such as acute myeloid leukaemia and myeloproliferative neoplasm²⁸, yet APOBEC-associated genomic signatures are only a minor component of the mutational signatures present in these tumours¹¹¹, further implying that A3A activity on RNA could precede DNA mutagenesis in cancer.

AID/APOBECs as base-editing tools

In this section we will discuss how AID/APOBEC enzymes have been used in genome and transcriptome engineering technologies, broadly known as programmable base editing (FIG. 5), to revert T-to-C or A-to-G transitions in DNA or mRNA, and how a fuller understanding of their substrate specificities can inform the design and optimization of these tools. As this Perspective is focused on AID/APOBECs, we will not discuss mRNA base-editing technologies that are based on adenosine deaminase enzymes (reviewed extensively elsewhere^{134–139}).

DNA-directed base-editing tools

The first members of the AID/APOBEC family to be used as the basis of a cytosine base editor (CBE) were AID, rat APOBEC1 (rA1) and A3G. A seminal paper from

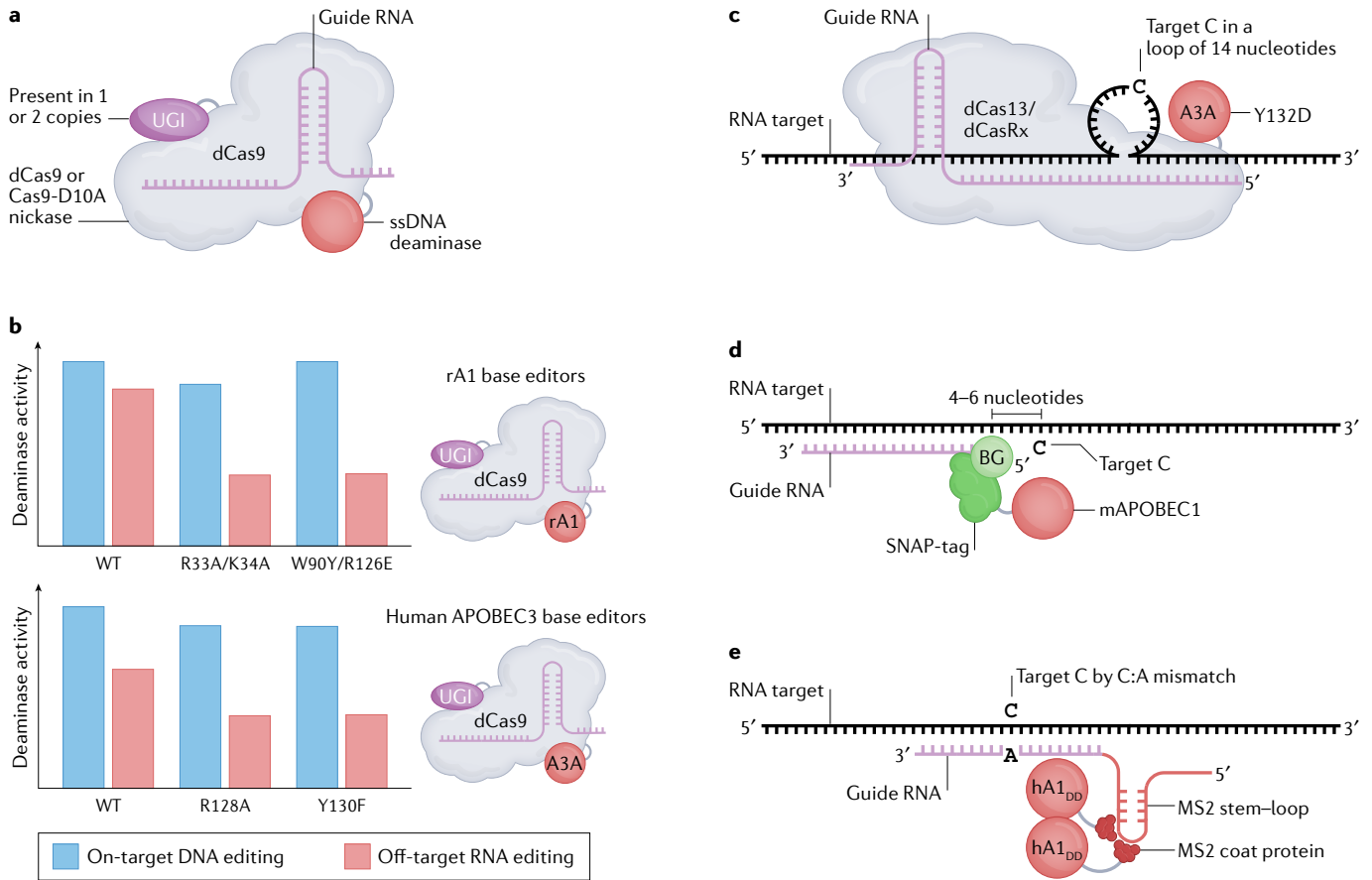


Fig. 5 | APOBEC-derived DNA and RNA base-editing tools. a | DNA-directed cytosine base editor (CBE) comprises catalytically dead Cas9 (dCas9), guide RNA (gRNA), single-stranded DNA (ssDNA) deaminase and uracil DNA glycosylase inhibitor (UGI). **b** | Quantification of on-target editing (on DNA) and off-target editing (on RNA), allowing quick visual understanding of features of each genome base editor. Several CBE variants illustrated, which differ with respect to the deaminase used (rat APOBEC1 (rA1) or human APOBEC3A (A3A)) and specific mutations within the deaminase. Note that amino acids identified in the human APOBEC1 structure²¹ as likely to be functionally important can inform base-editing work with rA1. **c–e** | RNA-directed CBE tools based on APOBEC proteins. APOBEC variants are directed to a specific nucleotide in a transcript of interest via an antisense gRNA, design of which varies according to the editing system: RNA base editing by CURE (cytidine-specific C-to-U RNA Editor) — here,

targeting mediated by gRNA that recruits a chimeric protein comprising either dCas13 or dCasRx and a Y132D variant of A3A to the RNA target — and gRNA creates a 14-nucleotide loop containing C to be edited (part **c**); RNA base editing with a SNAP-tagging system — mouse APOBEC1 (mAPOBEC1)-SNAP chimaera recruited to target RNA via covalent linkage to a benzylguanine (BG)-modified gRNA — and unlike other systems, C to be deaminated is positioned four to six nucleotides downstream of region bound by gRNA (part **d**); and RNA base editing with an MS2-tagging system — a human APOBEC1 deamination domain (hA1_{DD})-MS2 chimaera is recruited to a specific location on target RNA by binding MS2 coat proteins to MS2 stem-loop on gRNA — and in this system, C to be edited is specified by a C:A mismatch between target RNA and gRNA (part **e**). AID/APOBEC, activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like; WT, wild type.

the Liu laboratory used catalytically dead CRISPR-associated endonuclease (dCas) fused to these AID/APOBEC family members, together with appropriate Cas9 guide RNAs (gRNAs), to target deaminase activity to specific loci and induce single base changes in the absence of a DNA break¹⁴⁰. Given the substantial activity of rA1 as a DNA mutator^{109,141,142}, its fusion with dCas9 was the most efficient at generating specific C-to-T (or G-to-A) substitutions within DNA, constituting the first CBE¹⁴⁰. Several variations of this system were soon developed to increase base-editing efficiency (by fusion with a uracil DNA glycosylase inhibitor (UGI)), to reduce indel generation

(for example, by using Cas9-D10A, a nickase mutant of Cas9) and to reduce off-target editing (by using A3A or AID instead of APOBEC1) (reviewed elsewhere¹³⁸).

Given that rA1 and A3A are generalists, it was unavoidable that DNA editing systems based on these deaminases would also lead to several thousand unwanted RNA editing events^{143,144}. However, this off-target activity was almost entirely eliminated by introducing specific amino acid changes into rA1 and A3A. Two different two-amino acid changes to rA1 (R33A/K34A and W90Y/R126E) each resulted in reduced off-target activity on RNA while retaining efficient base editing

on DNA^{143,144} (FIG. 5b). Similarly, off-target RNA editing by A3A was reduced by introducing either an R128A or a Y130F amino acid change¹⁴⁴ (FIG. 5b). R128A and Y130F of A3A and R126E of rA1 occur in loop 7, emphasizing the importance of residues in this loop for deamination of RNA. Moreover, R33A/K34A changes were shown to affect the capability of APOBEC1 to bind RNA⁴⁹. These mutations illustrate how a better understanding of the features that determine whether an AID/APOBEC protein acts as a generalist or a specialist might enable specificity issues to be avoided by facilitating more informed CBE design and optimization at the outset.

RNA-directed base-editing tools

The development of AID/APOBECs as RNA-directed CBEs has proven to be more difficult than DNA-directed CBEs, leading one group to, instead, evolve ADAR proteins to induce C-to-U editing¹⁴⁵. One possible explanation for these difficulties is that RNA deamination by APOBEC1, A3A and A3G requires the target RNA to adopt specific secondary structures^{26,29–31,146}. This theory is supported by studies using the recently developed CURE (cytidine-specific C-to-U RNA Editor) system, which uses gRNAs to target a Y132D mutant version of A3A fused either to dPspCas13b or dCasRx to specific locations in a target transcript. Interestingly, A3A was only able to elicit RNA editing at the desired location when these gRNAs induced the target transcripts to form a loop¹⁴⁷ (FIG. 5c). Importantly, no off-target DNA editing was detected using CURE, although a few hundred off-target RNA edits were found¹⁴⁷. Two other recently reported RNA-directed CBE approaches used mAPOBEC1 or human APOBEC1 in combination with either SNAP-tagged or MS2-tagged gRNAs to target specific target mRNAs^{148,149} (FIG. 5d,e). Neither of these two methods was checked for off-target DNA editing, but the mAPOBEC1-SNAP system demonstrated that integration of an inducible editing enzyme reduces global off-target RNA editing, as had previously been shown for ADAR RNA base-editing technologies^{149,150}. This method was not benchmarked against CURE, making a direct comparison difficult, but it is important to note that the reported RNA off-target activity of CURE (measured as a simple sum of sites and noting that CURE enzymes are overexpressed) is much lower than that of mAPOBEC1-SNAP (REFS^{147,149}). Despite these recent developments, APOBEC1-based RNA-directed CBE systems still suffer from moderate levels of global off-target RNA editing and, owing to the inherent dinucleotide preference of A3A, CURE can only edit Cs present in a 5'-UC-3' motif (FIG. 3c). A better understanding of how APOBEC1, A3A and A3G interact specifically with RNA will help improve the current systems and facilitate the development of new ones.

Expanding the potential of base editing

An important limitation of the RNA-directed CBE systems described here is that editing is restricted to locations that match the sequence context preferences of the enzymes used. In particular, no currently known APOBECs naturally edit Cs within a 5'-GC-3' context (FIG. 3c). Therefore,

it will be necessary to develop additional context-specific base editors to complete the spectrum of Cs that can be edited. Considering the importance of residues in loop 7 (but also in loops 1 and 3) in defining the substrate and sequence context preference of the AID/APOBEC enzymes, it seems reasonable to hypothesize that altering residues within these loops may be a way to change the local motif preferences and alleviate target motif limitations. Finally, recruitment of endogenous AID/APOBECs for base-editing purposes (as has been done for ADAR^{151,152}) remains an unexplored field. Further developments in this area are important because endogenous AID/APOBEC enzymes are generally overexpressed in contexts (such as cancer) in which therapeutic editing could be beneficial.

Conclusion and future perspectives

Here, we have argued that, under certain conditions, several AID/APOBEC deaminases can act on both RNA and DNA

substrates whereas other family members are substrate-restricted. Through the analysis of recently published co-crystal structures we have attempted to describe the features that allow these enzymes to 'toggle' between substrates (as APOBEC1 and some APOBEC3 proteins do) and how such activity can be restricted (as in the case of AID and, perhaps, APOBEC2). With the advent of programmable base editors, it will be important to analyse all known AID/APOBEC deaminases (not only all mammalian family members but also distant relatives that seem to exist in marine organisms¹⁵³) for their properties, in order to develop CBEs that can selectively target RNA or DNA and to expand the local sequence preference of such tools.

Such analyses can also help answer biological questions arising from the close mechanistic relationship between RNA editing and DNA mutation. For example, it is well understood that DNA mutators of the APOBEC3 family are upregulated in

Glossary

Base modifications

Chemically altered nucleotides within mature RNA molecules.

dCasRx

Catalytically dead RNA-guided RNA-targeting CasRx from *Ruminococcus flavefaciens* XPD3002. CasRx is another member of the CRISPR family (class 2, VI-D).

dPspCas13b

Catalytically dead RNA-guided RNA-targeting CRISPR-Cas13b from *Prevotella* sp. P5-125. Cas13b is a member of the CRISPR family (class 2, type VI). Physiologically, Cas13b catalyses site-specific cleavage of single-stranded RNA.

G-quadruplex

A non-canonical four-stranded secondary structure of guanine-rich DNA sequences.

Guide RNAs

(gRNAs). Short RNA sequences used in base-editing technologies to target the base editor to a specific sequence in DNA or RNA. Depending on the tagging system used, the base editor can be recruited by the gRNA using specific scaffolds (for Cas proteins), sequences (MS2 coat protein) or chemical modifications (for SNAP).

Intrinsically disordered region

(IDR). An unstructured domain of proteins that are believed to have roles in intermolecular and intramolecular interactions, such as complex formation and phase separation.

MS2-tagged

Refers to a molecule labelled using a tagging system based on the natural interaction between the MS2 bacteriophage coat protein and a stem-loop structure from the phage genome. The sequence forming the stem-loop can be attached to a guide RNA (gRNA) to target an MS2-tagged base editor.

Nuclear export signal

(NES). A short peptide motif enriched for hydrophobic residues (such as Leu) recognized by exportins (such as XPO1/CRM1) that tags a protein for nuclear exit.

Nuclear localization signal

(NLS). A short peptide motif enriched for positively charged residues that tags a protein for nuclear import.

Pseudotyped HIV

Chimaeric viruses composed of the envelope glycoprotein of vesicular stomatitis virus (VSV-G) and the human immunodeficiency virus type 1 (HIV-1) core; these viruses are more infectious than non-pseudotyped HIV-1 viruses.

SNAP-tagged

Refers to a molecule labelled using a tagging system based on the SNAP-tag self-labelling protein derived from the human O⁶-alkylguanine-DNA alkyltransferase. As a SNAP-tag will form a covalent linkage with benzylguanine (BG)-modified nucleotides, a SNAP-tagged base editor can be directed to specific targets by BG-modified guide RNAs (gRNAs).

Stem-loops

Specific structures that may occur in single-stranded RNA (ssRNA) when complementary sequences base pair to form a double helix that ends in an unpaired (single-stranded) loop. Stem-loops are also known as hairpin structures or hairpin loops.

π -Stacking

Attractive non-covalent interactions between aromatic rings.

Tumour restriction

The limitation of tumour growth and/or tumour suppression or ablation by numerous distinct molecular mechanisms. Here, we specifically refer to the limitation of tumour growth owing to cell death after activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (AID/APOBEC)-mediated hypermutation.

cancer tissue — a holistic (but yet to be fully tested) view of the field would argue that these enzymes are actually upregulated in the context of a programmed RNA editing response to inflammation, and that DNA mutation is an off-target outcome of this response^{28,143}. If, as implied, RNA is the preferred substrate for these enzymes, it will be important to understand the physiologic role of RNA editing in the context of an early host response to tumour inflammation. Finally, if kataegis mutations (detected in the majority of human cancers) are simply the by-product of the host's attempt to limit tumour growth, then RNA editing could be used diagnostically as an early biomarker for ongoing tumour diversification and relapse²⁸.

Riccardo Pecori¹, Salvatore Di Giorgio¹, J. Paulo Lorenzo^{1,2} and F. Nina Papavasiliou^{1,2,3}

¹Division of Immune Diversity (D150), Program in Cancer Immunology, German Cancer Research Centre (DKFZ), Heidelberg, Germany.

²Faculty of Biosciences, Heidelberg University, Heidelberg, Germany.

³e-mail: n.papavasiliou@dkfz-heidelberg.de

<https://doi.org/10.1038/s41576-022-00459-8>

Published online 7 March 2022

1. Crick, F. H. On protein synthesis. *Symposia Soc. Exp. Biol.* **12**, 138–163 (1958).
2. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
3. Di Giammartino, D. C., Nishida, K. & Manley, J. L. Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* **43**, 853–866 (2011).
4. Ryvkin, P. et al. HAMR: high-throughput annotation of modified ribonucleotides. *RNA* **19**, 1684–1692 (2013).
5. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **18**, 275–291 (2017).
6. Benne, R. et al. Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**, 819–826 (1986).
7. Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
8. Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).
9. Savva, Y. A., Rieder, L. E. & Reenan, R. A. The ADAR protein family. *Genome Biol.* **13**, 252 (2012).
10. Wedekind, J. E., Dance, G. S. C., Sowden, M. P. & Smith, H. C. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet.* **19**, 207–216 (2003).
11. Lorenzo, J. P. et al. APOBEC2 is a transcriptional repressor required for proper myoblast differentiation. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.07.29.223594> (2021).
12. Reardon, S. Step aside CRISPR, RNA editing is taking off. *Nature* **578**, 24–27 (2020).
13. Salter, J. D. & Smith, H. C. Modeling the embrace of a mutator: APOBEC selection of nucleic acid ligands. *Trends Biochem. Sci.* **43**, 606–622 (2018).
14. Salter, J. D., Bennett, R. P. & Smith, H. C. The APOBEC protein family: united by structure, divergent in function. *Trends Biochemical Sci.* **41**, 578–594 (2016).
15. Bohn, J. A. et al. APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA. *Nat. Commun.* **8**, 1021 (2017).
16. Kouno, T. et al. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat. Commun.* **8**, 15024 (2017).
17. Matsuoka, T. et al. Structural basis of chimpanzee APOBEC3H dimerization stabilized by double-stranded RNA. *Nucleic Acids Res.* **46**, 10368–10379 (2018).
18. Rathore, A. et al. The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. *J. Mol. Biol.* **425**, 4442–4454 (2013).
19. Shaban, N. M. et al. The antiviral and cancer genomic DNA deaminase APOBEC3H is regulated by an RNA-mediated dimerization mechanism. *Mol. Cell* **69**, 75–86.e9 (2018).
20. Shi, K. et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* **24**, 131–139 (2017).
21. Wolfe, A. D., Li, S., Goedderz, C. & Chen, X. S. The structure of APOBEC1 and insights into its RNA and DNA substrate selectivity. *NAR Cancer* **2**, zcaa027 (2020).
22. Maiti, A. et al. Crystal structure of the catalytic domain of HIV-1 restriction factor APOBEC3G in complex with ssDNA. *Nat. Commun.* **9**, 2460 (2018).
23. Polevoda, B. et al. DNA mutagenic activity and capacity for HIV-1 restriction of the cytidine deaminase APOBEC3G depend on whether DNA or RNA binds to tyrosine 315. *J. Biol. Chem.* **292**, 8642–8656 (2017).
24. King, J. J. & Larjani, M. A novel regulator of activation-induced cytidine deaminase/APOBECs in immunity and cancer: Schrödinger's CAlytic Pocket. *Front. Immunol.* **8**, 351 (2017).
25. Olson, M. E., Harris, R. S. & Harki, D. A. APOBEC enzymes as targets for virus and cancer therapy. *Cell Chem. Biol.* **25**, 36–49 (2018).
26. Buisson, R. et al. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* **364**, eaaw2872 (2019).
27. Holtz, C. M., Sadler, H. A. & Mansky, L. M. APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. *Nucleic Acids Res.* **41**, 6139–6148 (2013).
28. Jallili, P. et al. Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. *Nat. Commun.* **11**, 2971 (2020).
29. Sharma, S. & Baysal, B. E. Stem-loop structure preference for site-specific RNA editing by APOBEC3A and APOBEC3G. *PeerJ* **5**, e4136 (2017).
30. Maris, C., Masse, J., Chester, A. N. N., Navaratnam, N. & Allain, F. H. NMR structure of the apoB mRNA stem-loop and its interaction with the C to U editing APOBEC1 complementary factor. *RNA* **11**, 173–186 (2005).
31. Richardson, N., Navaratnam, N. & Scott, J. Secondary structure for the apolipoprotein B mRNA editing site. *J. Biol. Chem.* **273**, 31707–31717 (1998).
32. Shah, R. R. et al. Sequence requirements for the editing of apolipoprotein B mRNA. *J. Biol. Chem.* **266**, 16301–16304 (1991).
33. Losey, H. C., Ruthenburg, A. J. & Verdine, G. L. Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.* **13**, 153–159 (2006).
34. Ito, F. et al. Understanding the structure, multimerization, subcellular localization and mC selectivity of a genomic mutator and anti-HIV factor APOBEC3H. *Sci. Rep.* **8**, 3763 (2018).
35. Mitra, M. et al. Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties. *Nucleic Acids Res.* **42**, 1095–1110 (2014).
36. Tang, G. et al. Creating RNA specific C-to-U editase from APOBEC3A by separation of its activities on DNA and RNA substrates. *ACS Synth. Biol.* **10**, 1106–1115 (2021).
37. Qiao, Q. et al. AID recognizes structured DNA for class switch recombination. *Mol. Cell* **67**, 361–373.e4 (2017).
38. Devos, J. M., Tomanicek, S. J., Jones, C. E., Nossal, N. G. & Mueser, T. C. Crystal structure of bacteriophage T4 5' nuclease in complex with a branched DNA reveals how flap endonuclease-1 family nucleases bind their substrates. *J. Biol. Chem.* **282**, 31713–31724 (2007).
39. Jiang, F. et al. Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
40. Zheng, S. et al. Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. *Cell* **161**, 762–773 (2015).
41. Dickerson, S. K., Market, E., Besmer, E. & Papavasiliou, F. N. AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* **197**, 1291–1296 (2003).
42. Abdouni, H. S. et al. DNA/RNA hybrid substrates modulate the catalytic activity of purified AID. *Mol. Immunol.* **93**, 94–106 (2018).
43. Fritz, E. L. et al. A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylation of activated B cells. *Nat. Immunol.* **14**, 749–755 (2013).
44. Krzysiak, T. C., Jung, J., Thompson, J., Baker, D. & Gronenborn, A. M. APOBEC2 is a monomer in solution: implications for APOBEC3G models. *Biochemistry* **51**, 2008–2017 (2012).
45. Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F. & Chen, X. S. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* **445**, 447–451 (2007).
46. Ataie, N. J. et al. Zinc coordination geometry and ligand binding affinity: the structural and kinetic analysis of the second-shell serine 228 residue and the methionine 180 residue of the aminopeptidase from *Vibrio proteolyticus*. *Biochemistry* **47**, 7673–7683 (2008).
47. Boyaci, H., Chen, J., Jansen, R., Darst, S. A. & Campbell, E. A. Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* **565**, 382–385 (2019).
48. Powell, C., Cornblath, E. & Goldman, D. Zinc-binding domain-dependent, deaminase-independent actions of apolipoprotein B mRNA-editing enzyme, catalytic polypeptide 2 (ApoBec2), mediate its effect on zebrafish retina regeneration. *J. Biol. Chem.* **289**, 28924–28941 (2014).
49. Chester, A. et al. The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J.* **22**, 3971–3982 (2003).
50. Ito, S. et al. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl Acad. Sci. USA* **101**, 1975–1980 (2004).
51. Patenaude, A.-M. et al. Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat. Struct. Mol. Biol.* **16**, 517–527 (2009).
52. Lackey, L., Law, E. K., Brown, W. L. & Harris, R. S. Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle* **12**, 762–772 (2013).
53. Salamango, D. J. et al. APOBEC3H subcellular localization determinants define zipcode for targeting HIV-1 for restriction. *Mol. Cell Biol.* **38**, e00356–18 (2018).
54. Bennett, R. P. et al. APOBEC-1 and AID are nucleocytoplasmic trafficking proteins but APOBEC3G cannot traffic. *Biochem. Biophys. Res. Commun.* **350**, 214–219 (2006).
55. Bennett, R. P., Presnyak, V., Wedekind, J. E. & Smith, H. C. Nuclear exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding. *J. Biol. Chem.* **283**, 7320–7327 (2008).
56. Navaratnam, N. et al. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J. Biol. Chem.* **268**, 20709–20712 (1993).
57. Teng, B. B., Burant, C. F. & Davidson, N. O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816–1819 (1993).
58. Lellek, H. et al. Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex. *J. Biol. Chem.* **275**, 19848–19856 (2000).
59. Mehta, A., Kinter, M. T., Sherman, N. E. & Driscoll, D. M. Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Mol. Cell Biol.* **20**, 1846–1854 (2000).
60. Fossat, N. et al. C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Rep* **15**, 903–910 (2014).
61. Blanc, V. et al. Apobec1 complementation factor (A1CF) and RBM47 interact in tissue-specific regulation of C to U RNA editing in mouse intestine and liver. *RNA* **25**, 70–81 (2019).
62. Soleymjanjahi, S., Blanc, V. & Davidson, N. APOBEC1 mediated C-to-U RNA editing: target sequence and trans-acting factor contribution to 177 RNA editing events in 119 murine transcripts in-vivo. *RNA* **27**, 876–890 (2021).
63. Rayon-Estrada, V. et al. Epitranscriptomic profiling across cell types reveals associations between APOBEC1-mediated RNA editing, gene expression outcomes, and cellular function. *Proc. Natl Acad. Sci. USA* **114**, 13296–13301 (2017).
64. Lerner, T., Papavasiliou, F. N. & Pecori, R. RNA editors, cofactors, and mRNA targets: an overview of the C-to-U RNA editing machinery and its implication in human disease. *Genes* **10**, 13 (2018).

65. Basu, U. et al. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* **438**, 508–511 (2005).
66. Chaudhuri, J. & Alt, F. W. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.* **4**, 541–552 (2004).
67. Conticello, S. G. et al. Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNB1. *Mol. Cell* **31**, 474–484 (2008).
68. McBride, K. M. et al. Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Biochem. Soc. Trans.* **103**, 8798–8803 (2006).
69. Pasqualucci, L., Kitaura, Y., Gu, H. & Dalla-Favera, R. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc. Natl Acad. Sci. USA* **103**, 395–400 (2006).
70. Vuong, B. Q. et al. Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nat. Immunol.* **10**, 420–426 (2009).
71. Marx, A., Galilee, M. & Alian, A. Zinc enhancement of cytidine deaminase activity highlights a potential allosteric role of loop-3 in regulating APOBEC3 enzymes. *Sci. Rep.* **5**, 18191 (2015).
72. Chen, J. & MacCarthy, T. The preferred nucleotide contexts of the AID/APOBEC cytidine deaminases have differential effects when mutating retrotransposon and virus sequences compared to host genes. *PLoS Comput. Biol.* **13**, e1005471 (2017).
73. Eisenberg, E. & Levanon, E. Y. A-to-I RNA editing — immune protector and transcriptome diversifier. *Nat. Rev. Genet.* **19**, 473–490 (2018).
74. Martinez, T., Shapiro, M., Bhaduri-McIntosh, S. & MacCarthy, T. Evolutionary effects of the AID/APOBEC family of mutagenic enzymes on human gamma-herpesviruses. *Virus Evol.* **5**, vey040 (2019).
75. Madani, N. & Kabat, D. An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein. *J. Virol.* **72**, 10251–10255 (1998).
76. Simon, J. H., Gaddis, N. C., Fouchier, R. A. & Malim, M. H. Evidence for a newly discovered cellular anti-HIV-1 phenotype. *Nat. Med.* **4**, 1397–1400 (1998).
77. Harris, R. S. et al. DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803–809 (2003).
78. Mangeat, B. et al. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99–103 (2003).
79. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
80. Dang, Y. et al. Human cytidine deaminase APOBEC3H restricts HIV-1 replication. *J. Biol. Chem.* **283**, 11606–11614 (2008).
81. Harris, R. S. & Dudley, J. P. APOBECs and virus restriction. *Virology* **479–480**, 131–145 (2015).
82. Hayward, J. A. et al. Differential evolution of antiretroviral restriction factors in pteropid bats as revealed by APOBEC3 gene complexity. *Mol. Biol. Evol.* **35**, 1626–1637 (2018).
83. Liddament, M. T., Brown, W. L., Schumacher, A. J. & Harris, R. S. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Curr. Biol.* **14**, 1385–1391 (2004).
84. Domingo, E., Sheldon, J. & Perales, C. Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev.* **76**, 159–216 (2012).
85. Kim, E.-Y. et al. Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *J. Virol.* **84**, 10402–10405 (2010).
86. Wood, N. et al. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog.* **5**, e1000414 (2009).
87. Venkatesan, S. et al. Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Ann. Oncol.* **29**, 563–572 (2018).
88. Bonvin, M. et al. Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology* **43**, 1364–1374 (2006).
89. Bulliard, Y. et al. Structure–function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J. Virol.* **85**, 1765–1776 (2011).
90. Warren, C. J. et al. APOBEC3A functions as a restriction factor of human papillomavirus. *J. Virol.* **89**, 688–702 (2014).
91. Di Giorgio, S., Martignano, F., Torcia, M. G., Mattiuz, G. & Conticello, S. G. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci. Adv.* **6**, eabb5813 (2020).
92. Graudenzi, A., Maspero, D., Angaroni, F., Piazza, R. & Ramazzotti, D. Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity. *iScience* **24**, 102116 (2021).
93. Picardi, E., Mansi, L. & Pesole, G. Detection of A-to-I RNA editing in SARS-CoV-2. *Genes* **13**, 41 (2021).
94. Simmonds, P. & Ansari, M. A. Extensive C>U transition biases in the genomes of a wide range of mammalian RNA viruses; potential associations with transcriptional mutations, damage- or host-mediated editing of viral RNA. *PLoS Pathog.* **17**, e1009596 (2021).
95. Kim, K. et al. APOBEC-mediated editing of SARS-CoV-2 genomic RNA impacts viral replication and fitness. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/2021.12.18.473309v1> (2021).
96. Klimczak, L. J., Randall, T. A., Saini, N., Li, J.-L. & Gordenin, D. A. Similarity between mutation spectra in hypermutated genomes of rubella virus and in SARS-CoV-2 genomes accumulated during the COVID-19 pandemic. *PLoS ONE* **15**, e0237689 (2020).
97. Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045.e9 (2020).
98. Cotroneo, C. E., Mangano, N., Dragani, T. A. & Colombo, F. Lung expression of genes putatively involved in SARS-CoV-2 infection is modulated in *cis* by germline variants. *Eur. J. Hum. Genet.* **29**, 1019–1026 (2021).
99. Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844 (2020).
100. Mourier, T. et al. Host-directed editing of the SARS-CoV-2 genome. *Biochem. Biophys. Res. Commun.* **538**, 35–39 (2021).
101. Ratcliff, J. & Simmonds, P. Potential APOBEC-mediated RNA editing of the genomes of SARS-CoV-2 and other coronaviruses and its impact on their longer term evolution. *Virology* **556**, 62–72 (2021).
102. van Dorp, L. et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.* **83**, 104351 (2020).
103. Matyášek, R. & Kovařík, A. Mutation patterns of human SARS-CoV-2 and Bat RaTG13 coronavirus genomes are strongly biased towards C>U transitions, indicating rapid evolution in their hosts. *Genes* **11**, E761 (2020).
104. Wang, R., Hozumi, Y., Zheng, Y.-H., Yin, C. & Wei, G.-W. Host immune response driving SARS-CoV-2 evolution. *Viruses* **12**, 1095 (2020).
105. Milewska, A. et al. APOBEC3-mediated restriction of RNA virus replication. *Sci. Rep.* **8**, 5960 (2018).
106. Yamanaka, S. et al. Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc. Natl Acad. Sci. USA* **92**, 8483–8487 (1995).
107. Franco, S. et al. H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell* **21**, 201–214 (2006).
108. Ramiro, A. R. et al. Role of genomic instability and p53 in AID-induced c-myc–lgh translocations. *Nature* **440**, 105–109 (2006).
109. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10**, 1247–1253 (2002).
110. Landry, S., Narvaiza, I., Linfesty, D. C. & Weitzman, M. D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep.* **12**, 444–450 (2011).
111. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
112. Roberts, S. A. et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
113. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
114. Roberts, S. A. et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
115. Chan, K. & Gordenin, D. A. Clusters of multiple mutations: incidence and molecular mechanisms. *Annu. Rev. Genet.* **49**, 243–267 (2015).
116. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
117. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
118. Klemm, L. et al. The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell* **16**, 232–245 (2009).
119. Law, E. K. et al. The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci. Adv.* **2**, e1601737 (2016).
120. de Bruin, E. C. et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **346**, 251–256 (2014).
121. Venkatesan, S. et al. Induction of APOBEC3 exacerbatates DNA replication stress and chromosomal instability in early breast and lung cancer evolution. *Cancer Discov.* **11**, 2456–2473 (2021).
122. Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* **9**, M57–M60 (1999).
123. Weaver, B. A. A., Silk, A. D., Montagna, C., Verdier-Pinard, P. & Cleveland, D. W. Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell* **11**, 25–36 (2007).
124. Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov.* **5**, 704–712 (2015).
125. Serebrenik, A. A. et al. The deaminase APOBEC3B triggers the death of cells lacking uracil DNA glycosylase. *Proc. Natl Acad. Sci. USA* **116**, 22158–22163 (2019).
126. Asaoka, M., Ishikawa, T., Takabe, K. & Patnaik, S. K. APOBEC3-mediated RNA editing in breast cancer is associated with heightened immune activity and improved survival. *Int. J. Mol. Sci.* **20**, 5621 (2019).
127. Ben-Aroya, S. & Levanon, E. Y. A-to-I RNA editing: an overlooked source of cancer mutations. *Cancer Cell* **33**, 789–790 (2018).
128. Christofi, T. & Zaravinos, A. RNA editing in the forefront of epitranscriptomics and human health. *J. Transl. Med.* **17**, 319 (2019).
129. Driscoll, C. B. et al. APOBEC3B-mediated corruption of the tumor cell immunopeptidome induces heteroclitic neoepitopes for cancer immunotherapy. *Nat. Commun.* **11**, 790 (2020).
130. Paz-Yaacov, N. et al. Elevated RNA editing activity is a major contributor to transcriptomic diversity in tumors. *Cell Rep.* **13**, 267–276 (2015).
131. Ramirez-Moya, J., Baker, A. R., Slack, F. J. & Santisteban, P. ADAR1-mediated RNA editing is a novel oncogenic process in thyroid cancer and regulates miR-200 activity. *Oncogene* **39**, 3738–3753 (2020).
132. Blanc, V. et al. Deletion of the AU-rich RNA binding protein apobec-1 reduces intestinal tumor burden in *Apc^{min}* mice. *Cancer Res.* **67**, 8565–8573 (2007).
133. Nelson, V. R., Heaney, J. D., Tesar, P. J., Davidson, N. O. & Nadeau, J. H. Transgenerational epigenetic effects of the Apobec1 cytidine deaminase deficiency on testicular germ cell tumor susceptibility and embryonic viability. *Proc. Natl Acad. Sci. USA* **109**, E2766–E2773 (2012).
134. Casati, B., Stamkopoulos, D., Tsakis, R. N. & Pecori, R. In *Epitranscriptomics* (eds Jurga, S. & Barciszewski, J.) 471–503 (Springer International, 2021).
135. Khosravi, H. M. & Jantsch, M. F. Site-directed RNA editing: recent advances and open challenges. *RNA Biol.* **18**, 41–50 (2021).
136. Montiel-Gonzalez, M. F., Diaz Quiroz, J. F. & Rosenthal, J. J. C. Current strategies for site-directed RNA editing using ADARs. *Methods* **156**, 16–24 (2019).
137. Park, S. & Beal, P. A. Off-target editing by CRISPR-guided DNA base editors. *Biochemistry* **58**, 3727–3734 (2019).
138. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
139. Vogel, P. & Stafforst, T. Critical review on engineering deaminases for site-directed RNA editing. *Curr. Opin. Biotechnol.* **55**, 74–80 (2018).
140. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
141. Petersen-Mahrt, S. K. & Neuberger, M. S. In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *J. Biol. Chem.* **278**, 19583–19586 (2003).

142. Saraconi, G., Severi, F., Sala, C., Mattiuz, G. & Conticello, S. G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biol.* **15**, 417 (2014).
143. Grünewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
144. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
145. Abudayyeh, O. O. et al. A cytosine deaminase for programmable single-base RNA editing. *Science* **365**, 382–386 (2019).
146. Hersberger, M., Patarroyo-white, S., Arnold, K. S. & Innerarity, T. L. Phylogenetic analysis of the apolipoprotein B mRNA-editing region. *J. Biol. Chem.* **274**, 34590–34597 (1999).
147. Huang, X. et al. Programmable C-to-U RNA editing using the human APOBEC3A deaminase. *EMBO J.* **40**, e108209 (2021).
148. Bhakta, S., Sakari, M. & Tsukahara, T. RNA editing of BFP, a point mutant of GFP, using artificial APOBEC1 deaminase to restore the genetic code. *Sci. Rep.* **10**, 17304 (2020).
149. Stroppe, A. S. et al. Harnessing self-labeling enzymes for selective and concurrent A-to-I and C-to-U RNA base editing. *Nucleic Acids Res.* **49**, e95 (2021).
150. Vogel, P. et al. Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods* **15**, 535–538 (2018).
151. Merkle, T. et al. Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. *Nat. Biotechnol.* **37**, 133–138 (2019).
152. Ou, L. et al. Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. *Nat. Biotechnol.* **37**, 1059–1069 (2019).
153. Liew, Y. J., Li, Y., Baumgarten, S., Voolstra, C. R. & Aranda, M. Condition-specific RNA editing in the coral symbiont *Symbiodinium microadriaticum*. *PLoS Genet.* **13**, e1006619 (2017).
154. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
155. Snyder, E. M. et al. APOBEC1 complementation factor (A1CF) is dispensable for C-to-U RNA editing in vivo. *RNA* **23**, 457–465 (2017).
156. Conticello, S. G., Thomas, C. J. F., Petersen-Mahrt, S. K. & Neuberger, M. S. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.* **22**, 367–377 (2005).
157. Conticello, S. G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **9**, 229 (2008).
158. Krishnan, A., Iyer, L. M., Holland, S. J., Boehm, T. & Aravind, L. Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. *Proc. Natl Acad. Sci. USA* **115**, E3201–E3210 (2018).
159. Münk, C., Willemsen, A. & Bravo, I. G. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evol. Biol.* **12**, 71 (2012).
160. Hirano, K., Min, J., Funahashi, T., Baunoch, D. A. & Davidson, N. O. Characterization of the human *apobec-1* gene: expression in gastrointestinal tissues determined by alternative splicing with production of a novel truncated peptide. *J. Lipid Res.* **38**, 847–859 (1997).
161. Blanc, V. et al. Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. *Genome Biol.* **15**, R79 (2014).
162. Rosenberg, B. R., Hamilton, C. E., Mwangi, M. M., Dewell, S. & Papavasiliou, F. N. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nat. Struct. Mol. Biol.* **18**, 230–236 (2011).
163. Cole, D. C. et al. Loss of APOBEC1 RNA-editing function in microglia exacerbates age-related CNS pathophysiology. *Proc. Natl Acad. Sci. USA* **114**, 13272–13277 (2017).
164. Niavarani, A., Shahrabi Farahani, A., Sharafkhan, M. & Rassoulzadegan, M. Pancancer analysis identifies prognostic high-APOBEC1 expression level implicated in cancer in-frame insertions and deletions. *Carcinogenesis* **39**, 327–335 (2018).
165. Rogozin, I. B. et al. Nucleotide weight matrices reveal ubiquitous mutational footprints of AID/APOBEC deaminases in human cancer genomes. *Cancers* **11**, E211 (2019).
166. Nabel, C. S. et al. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
167. Rogozin, I. B. & Diaz, M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol.* **172**, 3382–3384 (2004).
168. Rogozin, I. B. & Kolchanov, N. A. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta* **1171**, 11–18 (1992).
169. Muramatsu, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
170. Branstetter, R., Pham, P., Scharff, M. D. & Goodman, M. F. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl Acad. Sci. USA* **100**, 4102–4107 (2003).
171. Larijani, M. & Martin, A. Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Mol. Cell. Biol.* **27**, 8038–8048 (2007).
172. Betz, A. G., Rada, C., Pannell, R., Milstein, C. & Neuberger, M. S. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proc. Natl Acad. Sci. USA* **90**, 2385–2388 (1993).
173. Rajewsky, K., Forster, I. & Cumano, A. Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science* **238**, 1088–1094 (1987).
174. Liao, W. et al. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem. Biophys. Res. Commun.* **260**, 398–404 (1999).
175. Etard, C., Roostalu, U. & Strähle, U. Lack of Apobec-2-related proteins causes a dystrophic muscle phenotype in zebrafish embryos. *J. Cell Biol.* **189**, 527–539 (2010).
176. Sato, Y. et al. Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J. Biol. Chem.* **285**, 7111–7118 (2010).
177. Sawyer, S. L., Emerman, M. & Malik, H. S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* **2**, E275 (2004).
178. LaRue, R. S. et al. Guidelines for naming nonprimate APOBEC3 genes and proteins. *J. Virol.* **83**, 494–497 (2009).
179. Bogerd, H. P., Wiegand, H. L., Doehle, B. P., Lueders, K. K. & Cullen, B. R. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res.* **34**, 89–95 (2006).
180. Refsland, E. W. & Harris, R. S. The APOBEC3 family of retroelement restriction factors. *Curr. Top. Microbiol. Immunol.* **371**, 1–27 (2013).
181. Yang, B., Chen, K., Zhang, C., Huang, S. & Zhang, H. Virion-associated uracil DNA glycosylase-2 and apurinic/aprimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. *J. Biol. Chem.* **282**, 11667–11675 (2007).
182. Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I. & Koonin, E. V. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle* **4**, 1281–1285 (2005).
183. Marino, D. et al. APOBEC4 enhances the replication of HIV-1. *PLoS ONE* **11**, e0155422 (2016).
184. Shi, M. et al. Characterization and functional analysis of chicken APOBEC4. *Dev. Comp. Immunol.* **106**, 103631 (2020).

Acknowledgements

Work on RNA editing and modification in the Papavasiliou laboratory is funded by the European Research Council (ERC) (#649019) and the German Research Foundation (DFG) (TRR319-RMaP and SPP1784). The authors thank all members, past and present, of the Papavasiliou laboratory for discussions, and sincerely apologize to the many colleagues whose work could not be cited for reasons of space.

Author contributions

All authors researched data for the article, made substantial contributions to discussions of the content, and wrote the article. F.N.P. and R.P. edited the manuscript and drafted the final submission.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Genetics thanks M. Larijani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

RELATED LINKS

Human Protein Atlas: <http://www.proteinatlas.org>

© Springer Nature Limited 2022