Protein family review

The AID/APOBEC family of nucleic acid mutators
Silvestro G Conticello

Address: Core Research Laboratory - Istituto Toscana Tumori, Florence, Via Cosimo il Vecchio 2, 50139 Firenze, Italy.
Email: silvo.conticello@ittumori.it

Published: 17 June 2008

Genome Biology 2008, 9:229 (doi:10.1186/gb-2008-9-6-229)
The electronic version of this article is the complete one and can be
found online at http://genomebiology.com/2008/9/6/229
© 2008 BioMed Central Ltd

Summary

The AID/APOBECs, a group of cytidine deaminases, represent a somewhat unusual protein family that can insert mutations in DNA and RNA as a result of their ability to deaminate cytidine to uridine. The ancestral AID/APOBECs originated from a branch of the zinc-dependent deaminase superfamily at the beginning of the vertebrate radiation. Other members of the family have arisen in mammals and present a history of complex gene duplications and positive selection. All AID/APOBECs have a characteristic zinc-coordination motif, which forms the core of the catalytic site. The crystal structure of human APOBEC2 shows remarkable similarities to that of the bacterial tRNA-editing enzyme TadA, which suggests a conserved mechanism by which polynucleotides are recognized and deaminated. The AID/APOBECs seem to have diverse roles. AID and the APOBEC3s are DNA mutators, acting in antigen-driven antibody diversification processes and in an innate defense system against retroviruses, respectively. APOBEC1 edits the mRNA for apolipoprotein B, a protein involved in lipid transport. A detailed understanding of the biological roles of the family is still some way off, however, and the functions of some members of the family are completely unknown. Given their ability to mutate DNA, a role for the AID/APOBECs in the onset of cancer has been proposed.

Gene organization and evolutionary history

The AID/APOBEC proteins are found in vertebrates and share the ability to insert mutations in DNA and RNA by deaminating cytidine to uridine. The first family member to be identified and characterized was the apolipoprotein B editing complex 1 (APOBEC1), a protein involved in the editing of the apolipoprotein B (ApoB) pre-mRNA [1,2]. Further members were identified as DNA mutators. Activation-induced deaminase (AID) was revealed to be essential for the antigen-driven diversification of already rearranged immunoglobulin genes in the vertebrate adaptive immune system [3], and the APOBEC3s were shown to be involved in the restriction of retrovirus propagation in primates [4,5]. The other members of the family, APOBEC2 and APOBEC4, have not yet been characterized. Table 1 lists the human AID/APOBEC paralogs; family members from other species are listed in Additional data files 1 and 2.

All the AID/APOBECs share the structural and catalytic backbone of the zinc-dependent deaminases, a large gene superfamily encoding enzymes involved in the metabolism of purines and pyrimidines (Figure 1). Of these deaminases, the tRNA adenosine deaminases (Tad/ADAT2) edit adenosine to inosine at the anticodon of various tRNAs in both eukaryotes and prokaryotes [6] and are thought to be the group from which the AID/APOBEC family originated (Figure 1). Indeed, as well as having functional and structural similarities to the AID/APOBECs [7,8], ADAT2 from trypanosomes seems to be able to deaminate cytidine in DNA [9].

The rise of the AID/APOBEC gene family appears to have been concurrent with the appearance of the vertebrate lineage and the evolution of adaptive immunity and AID is thought to be one of the ancestral family members (Figure 2). AID homologs able to trigger somatic hypermutation and class-switch recombination in B cells have been described in bony fish [10,11], while bona fide AID homologs have been identified both in cartilaginous fish [10], which have immunoglobulin genes, and in the sea
lamprey, a jawless vertebrate [12], which does not. The presence of AID in the lamprey is remarkable, as its system of ‘adaptive immunity’ is based not on immunoglobulins but on variable lymphocyte receptors (VLRs), a large family of proteins containing leucine-rich repeats, which undergo at least one round of diversification [12,13]. It will be interesting to know whether the sea lamprey AID homolog is involved in this process.

The gene structure for most AID/APOBEC genes includes five exons and is reminiscent of that of the DCDT (dCMP deaminases)/ADAT2 genes, in which the catalytic site is encoded in the third exon. In contrast, the other ancestral AID/APOBEC genes, namely APOBEC4 [14] and APOBEC2 [15,16] (found in all jawed vertebrates, see Figure 2), have two and three exons respectively, with the coding sequence being mostly confined to the second exon. The few amino acids encoded by the first exon of APOBEC2 have no similarity to any known sequence. These observations provide clues to the overall evolution of the gene family: the absence of introns in the deaminase-like region of APOBEC4 and APOBEC2 suggests that these genes might be the result of early retrotranspositional events. Given the position of the sea lamprey deaminase genes (AID-CDA1 and CDA2) in the phylogenetic tree (see Figure 2), the APOBEC4 clade seems to have evolved independently from that of AID, while the clustering of APOBEC2 raises the possibility that AID provided its evolutionary scaffold. The phylogenetic relationships and gene structure of the later-evolved members of the family (APOBEC1 and APOBEC3) indicate that they have originated from sequential duplications of the AID locus.

The APOBEC1 locus derives from an inverted duplication of the AID locus on the same chromosome, located 40 kb away in most mammals. APOBEC1 homologs with the same genomic orientation are found in marsupials. In primates, owing to an inversion, the APOBEC1 locus is located approximately 1 Mb away from the AID locus [10]. The main difference between APOBEC1 and the other AID/APOBEC genes is an extended coding sequence at its 3’ end, whose significance has yet to be understood.

The APOBEC3 locus [4] originated after the divergence of the marsupial and placental lineages and is located in the region syntenic with human chromosome 22q13. A duplication event in the original ‘placental’ locus formed the two ancestral APOBEC3 genes, from which all other APOBEC3s have evolved through a complex history of gene

| Name      | Genomic location | Deaminase domains | Expression                  | Cellular localization  | Editing activity | Target                           | References |
|-----------|------------------|-------------------|-----------------------------|------------------------|-----------------|----------------------------------|------------|
| AID       | 12p13            | 5                 | 1                           | Activated B cells, testis | Mainly cytoplasmic, acts in the nucleus | DNA               | Immunoglobulin gene               | [3,31]     |
| APOBEC1   | 12p13.1          | 5                 | 1                           | Small intestine         | Cytoplasmic/nuclear, acts in the nucleus | RNA, DNA           | Apolipoprotein B mRNA             | [1,2]      |
| APOBEC2   | 6p21             | 3                 | 1                           | Skeletal muscle, heart  | Cytoplasmic/nuclear                           | Unknown           | Unknown                          | [15,16]    |
| APOBEC3A  | 22q13.1          | 5                 | 1                           | Keratinocytes, blood    | Predominantly nuclear                          | DNA               | Adeno-associated virus, retrotransposons | [4,104]    |
| APOBEC3B  | 22q13.1          | 8                 | 2                           | Intestine, uterus, mammary gland, keratinocytes, other | Cytoplasmic/nuclear | DNA               | Retroviruses, retrotransposons, HBV | [4,104]    |
| APOBEC3C  | 22q13.1          | 4                 | 1                           | Many tissues            | Cytoplasmic/nuclear                           | DNA               | Retroviruses, retrotransposons, HBV | [4]        |
| APOBEC3DE | 22q13.1          | 7                 | 2                           | Thyroid, spleen, blood  | Unknown                                       | DNA               | Retroviruses                      | [4,10,107] |
| APOBEC3F  | 22q13.1          | 8                 | 2                           | Many tissues            | Cytoplasmic                                    | DNA               | Retroviruses, retrotransposons, HBV | [4]        |
| APOBEC3G  | 22q13.1          | 8                 | 2                           | Many tissues, T cells   | Cytoplasmic                                    | DNA               | Retroviruses, retrotransposons, HBV | [4,5]      |
| APOBEC3H  | 22q13.1          | 5                 | 1                           | Blood, thymus, thyroid, placenta | Unknown                                       | DNA               | Retroviruses                      | [10,107]   |
| LOC196469 | 12q23            | 1                 | 2                           | Pseudogene              | -                                            | -                  | -                                | [10,107]   |
| APOBEC4   | 1q25.3           | 2                 | 1                           | Testis                   | Unknown                                       | Unknown            | Unknown                          | [14]       |

*This pseudogene originated from a recent retrotranspositional event. HBV, hepatitis B virus.
Characteristic structural features

Until very recently a crystal structure for a functionally characterized AID/APOBEC was not available and many of the structural features of this protein family have been ascertained by comparing their primary and secondary structure with the crystal structures of the functionally uncharacterized APOBEC2 [20] and of other zinc-dependent deaminases, especially TadA from the Tad/ADAT2 family [21]. The three-dimensional structure of the carboxy-terminal domain of APOBEC3G has now been published [22]. This model shows the closeness between the APOBEC3G structure and those already known.

Like all zinc-dependent deaminases, the main structural feature of the AID/APOBECs is the domain responsible for their catalytic activity. In the amino-acid sequence, the signature for this domain is a H[AV]E-x[24-36]-PCxxC motif (where x is any amino acid) (Figure 3). The histidine (H) and the two cysteines (C) coordinate a zinc atom and form the catalytic core of the deaminase (Figure 4). The cytidine is bound in this pocket and is deaminated through nucleophilic attack on the ammonium group on its carbon 4 by an activated water molecule (coordinated by the zinc atom) and the nearby glutamate, which acts as a proton donor.

The overall structure of the AID/APOBECs resembles that of other zinc-dependent deaminases. A series of five β strands forms the backbone of the molecule and α helices 2 and 3 hold the histidine and the cysteines in place and thus shape the catalytic pocket (Figure 4). Structural similarities with the Tad/ADAT2s in particular provide clues to the ability of the AID/APOBECs to deaminate cytidine. A comparison with the crystal structure of the bacterial TadA protein bound to its substrate [21] reveals the presence of a conserved loop (labeled in orange in Figure 4) that may play a role in substrate recognition [7,8]. A serine-tryptophan-serine (SWS) motif (corresponding to SSS in APOBEC2) located before the PCxxC motif is necessary for catalytic activity [23] (labeled in pink in Figure 4). This structural arrangement forms a trough where the DNA strand could be positioned and recognized. Recognition of the substrate through these loops might explain the observation that different AID/APOBECs display sequence-context preferences in regard to the nucleotides immediately upstream of the cytidine to be deaminated (see for example [24-26]).

Dimerization/oligomerization of the AID/APOBECs has been reported, often occurring in an RNA-dependent manner ([20, 27–29] and references therein), but, in the case of AID and APOBEC3G, the quaternary structure does not seem to be necessary for the enzymatic activity (see for example [30]).

duplications and fusions [10,17]. In some species, such as rodents, pigs and cattle, the two original genes have merged to form a single gene with a double zinc-coordinating domain, whereas in other species - primates, horses, bats, and felines - one of the two genes has been repeatedly duplicated to form an array of APOBEC3 genes. In primates in particular, the locus has rapidly expanded to seven genes. This rapid evolution of the APOBEC3 locus is thought to be the result of selective pressure on the APOBEC3s from their targets (retroviruses and retrotransposons) [18,19].

Figure 1
Schematic representation of the evolutionary relationships between the AID/APOBECs and the rest of the zinc-dependent deaminases. The only other zinc-dependent deaminase families widely expressed in metazoans and from which the AID/APOBECs (shaded in red) could have originated are the cytidine deaminases (CDA), the dCMP deaminases (DCDT) or the tRNA adenosine deaminases (Tad/ADAT2) (all shown in orange). CDAs and DCDTs act on free pyrimidines in the salvage pathway, the Tad/ADAT2s edit adenosine 34 at the anticodon of various tRNAs to inosine and are essential in bacteria, yeast and metazoans [6]. AID/APOBECs are unlikely to have originated from CDAs because of the differences in gene organization and catalytic domain [7,10]; DCDTs, despite the similar secondary structure, differ substantially from the other zinc-dependent deaminase families widely expressed in metazoans [108]. Phylogenetic data [10], species representation, and structural/functional features favor the tRNA-editing enzymes as the origin of the AID/APOBECs [7,8], a model supported by the observation that ADAT2 from trypanosomes can deaminate DNA [9]. The tRNA\textsuperscript{34} adenosine 37 deaminases type 1 (ADAT1) and the mRNA adenosine deaminases 1, 2, and 3 (ADARs) (shaded in green) are thought to have originated from the Tad/ADAT2 family independently of the AID/APOBECs. CoDA, cytosine deaminases; RibD, riboflavin deaminases; GuanineD, guanine deaminases.
Figure 2
Phylogenetic relationships within the AID/APOBEC gene family. The neighbor-joining tree shown here is generated from a protein alignment of the exon encoding the zinc-coordinating motif (the alignment is provided as Additional data file 1). The position of the agnathan (sea lamprey) AID (indicated by the arrow), separated from the clade comprising all the other AID/APOBECs, could suggest that all family members have originated from the ancestral AID. The different clusters in the AID/APOBEC family are identified, with the APOBEC3 cluster further divided into Z1a, Z1b, and Z2 clades (for the nomenclature of the APOBEC3 subgroups see [10]). Each domain of the double-domained APOBEC3s is included individually, with the amino-terminal and carboxy-terminal domains labeled [N] and [C], respectively. While APOBEC1 has been described only in mammals, the APOBEC2 group is found in all jawed vertebrates, including the primitive ghost shark. The duplication of the APOBEC2 locus after an ancient genome duplication in bony fish has been maintained, resulting in two coevolving APOBEC2 genes. The organisms in which each group is found are indicated below the clade label. Clades are collapsed for clarity, and only nodes with a bootstrap value greater than 50 are shown. The sequences used are either described in [10,12] or obtained from the Ensembl Genome Browser [109]. The sequences for the ghost shark were obtained using the AID/APOBECs as queries in BLAST searches on the Callorhinchus milii genome shotgun contigs (GenBank accession numbers: AID, AAUX01329030; APOBEC2, AAUX01039499; APOBEC4, AAUX01642881).
The known functions of the AID/APOBECs revolve around their ability to get, more or less specifically, to their substrate and deaminate it. This means that, given the diverse roles that the AID/APOBECs perform, their cellular localization varies. Nonetheless, most AID/APOBECs are initially localized to the cytoplasm, a safe place considering their ability to mutate DNA.

**AID**

AID was first discovered in 1999 in a subtractive hybridization screen comparing switch-induced and uninduced murine B lymphoma cells [3] and it is selectively expressed in activated B cells in germinal centers. Subsequent genetic experiments have revealed that AID is central to antigen-driven antibody diversification by class-switch recombination, somatic hypermutation, or gene conversion [31-33]. Genetic AID deficiency leads to Type 2 Hyper-IgM Syndrome [34], an immunodeficiency in which the inability to carry out class-switch recombination leads to the absence of antibodies other than those of the IgM class.

AID was initially thought to be an RNA-editing enzyme, but the discovery that it could mutate *Escherichia coli* DNA provided insight into its mechanism of action [35]. The ability of AID to deaminate C to U in DNA is in keeping with the observation that there are two mutational phases during the somatic hypermutation process. Further confirmation of its role as a DNA mutator came from evidence that uracil DNA glycosylase (UNG), the enzyme responsible for the removal of uracil in DNA, acts downstream of AID [36,37]. In humans, mutations in the UNG gene cause Type 5 Hyper-IgM Syndrome [37].
Further studies revealed that AID targets single-stranded DNA (see for example [38]) with a preference for cytidines within a sequence motif WRC (W is A or T; R is A or G) [24,39]. These observations are consistent with the presence of mutational hotspots at the immunoglobulin locus and the need for transcription (making single-stranded DNA available) for the antibody diversification processes to occur ([40] and references therein). Following deamination to uracil, recruitment of the general DNA repair machinery results in both somatic mutation and the initiation of class-switch recombination ([41] and references therein).

Whereas the function of AID is exerted in the nucleus, AID is predominantly cytoplasmic owing to the presence of a nuclear export signal (NES) at the extreme carboxyl terminus [42-44]. Accumulation of AID in the nucleus of murine B cells after ablation of the NES does not increase somatic hypermutation at the immunoglobulin locus, but causes an increase in non-physiologic hypermutation elsewhere in the genome [43]. Intriguingly, the same region is deemed necessary for the successful initiation of class-switch recombination, but it is not clear whether there is a causal relation between nuclear export and this process [45]. The presence of a weak nuclear localization signal (NLS) in the AID amino-terminal region has also been reported [42,44,45].

**APOBEC1**

APOBEC1 is expressed in the human small intestine and in the liver in rodents. It is responsible for ApoB pre-mRNA editing [1,2]: deamination of cytidine 6666 changes a glutamine codon into a stop codon, thus generating a shorter form of ApoB (ApoB48). ApoB48 is the main component in the hydrophilic shell of the chylomicrons, the very low-density lipoproteins that transport triglycerides from the intestine to the tissues.

Like AID, APOBEC1 acts in the nucleus [46] and shuttles between cytoplasm and nucleus by virtue of an amino-terminal NLS and a carboxy-terminal NES [47,48]. APOBEC complementation factor (ACF) is known to target APOBEC1 and leads to suppression of the edited ApoB mRNA nonsense-mediated decay (see for example [49]). Intriguingly, while the only phenotype in APOBEC1-deficient mice is the lack of ApoB mRNA editing (see for example [50]), ACF deficiency is lethal [51]. This, together with the conservation of the ACF gene throughout metazoa, could mean that AID/APOBECS were co-opted for ApoB mRNA editing only at a later stage in evolution, after the AID gene had been duplicated and the newly formed APOBEC1 was free to evolve.

While AID and the APOBEC3s have a loose sequence context preference for cytidine deamination, APOBEC1 is part of a complex that strictly recognizes a sequence 3’ to the cytidine to be deaminated (the mooring sequence). After binding of the editing complex to an AU-rich motif, overlapping with the mooring sequence, APOBEC1 edits the C6666. The efficiency of editing is also mediated by a number of other cis-acting elements ([52] and references therein). While there is no doubt on the physiological role for APOBEC1, its overexpression causes deamination of various RNAs in a promiscuous manner (see for example [53,54]). Intriguingly, APOBEC1 can also deaminate cytidine in DNA [23,55], which might suggest additional functions for it, maybe more related to those of the other AID/APOBECS.

**The APOBEC3s**

The APOBEC3s were first identified as paralogs of APOBEC1 by Jarmuz *et al* [4], but attained the limelight when human APOBEC3G was identified as the factor involved in HIV restriction [5]. HIV mutants lacking the viral infectivity factor (Vif) are non-infective in certain cell lines (so-called nonpermissive cell lines) but will propagate in others (permissive cell lines). APOBEC3G mRNA was isolated through a cDNA subtraction screen between CEM (nonpermissive) and CEM-SS (permissive) cells; its overexpression in CEM-SS cells reverses the permissive phenotype to nonpermissive [5].

APOBEC3G is packaged into the HIV virion and exerts its action on the nascent first DNA strand produced by reverse transcription in the target cell [56-60]. As a consequence, the viral genome is prevented from integrating into the cell’s genome and those rare retrotranscripts that do succeed in integrating are heavily mutated and nonfunctional. APOBEC3G produces characteristic G to A mutations on the viral plus-strand cDNA, and in experimental conditions the mutation load on the viral genome can be as high as 3%.

In the presence of Vif, however, APOBEC3G is not able to prevent HIV propagation as it is ubiquitinated and targeted for proteasomal degradation - via a Cul5-SCF complex - when it interacts with Vif through its amino-terminal domain (see for example [61,62]). It is interesting to note that the interaction with Vif has shaped the evolution of APOBEC3G: a single amino-acid change among primate APOBEC3Gs confers resistance to other primate lentiviral Vif proteins (see for example [63]).

Like APOBEC3G, all the primate APOBEC3 paralogs are able to restrict retroviruses with varying efficiency (Table 1). APOBEC3F, which has similar activity and expression pattern to APOBEC3G, preferentially deaminates cytidines, but in a different sequence context (see for example [25,26,64]). Interestingly, an analysis of HIV sequences hypermutated in vivo reveals a mutational bias toward the sequence preferences of APOBEC3G and F [24,65].

Most cellular APOBEC3G is kept inactive in high molecular weight ribonucleoprotein complexes [28,29,66]. Its packaging into virions is mediated by both viral and cellular RNAs [67-72], although the HIV Gag protein increases packaging efficiency [70,72,73]. For enzymatic activity to
be displayed, the balance between high molecular weight and low molecular weight APOBEC3G complexes must be reversed [28] and, even after being packaged into virions, APOBEC3G must be freed by the action of RNase H during retrotranscription [66]. Indeed, the avidity of APOBEC3G for RNA and its localization in mRNA-processing bodies in the cytoplasm (see for example [74]) could serve as both regulatory and targeting mechanisms, and these properties might shed light on novel functions for APOBEC3s, such as involvement in microRNA regulation [75].

Given the similarities in the replication mechanisms, the APOBEC3s are able to restrict both retrotransposons and viruses with a reverse transcription step during their replication cycle ([7,76] and references therein).

**Uncharacterized APOBECs**

APOBEC2 [15,16], the only AID/APOBEC member until very recently for which a crystal structure was available [20], is expressed specifically in skeletal muscle and heart. It has proved the most elusive AID/APOBEC to characterize functionally, mainly because it has none of the enzymatic activities typical of its paralogs [15,16,23,77]. APOBEC2 does not seem necessary for mouse development [77], but it is noteworthy that its appearance during metazoan evolution coincides with the evolution of slow/fast striated muscle and cardiac muscle [78,79]. Moreover, as with AID, the purifying selection driving the evolution of APOBEC2 at both the inter-species and intra-species level (bony fish have two copies of the gene) [7,18] suggests an evolutionary history constrained by function.

Very little is known about APOBEC4, the most recently identified AID/APOBEC [14]. Its low sequence similarity to the other AID/APOBECs casts doubt on its ability to deaminate cytidine [7], but its ancestry might reveal novel links to the tRNA-editing enzymes and provide clues to the origin of the AID/APOBECs.

**Frontiers**

Despite the rapid progress in research on the AID/APOBECs, many questions remain. Apart from AID, the ancestral AID/APOBECs have not been functionally characterized. Moreover, while the enzymatic mechanisms of the characterized AID/APOBECs are now well known, the upstream and downstream events that mediate their action, and their involvement in other biological pathways are not yet known.

**The physiological targets of APOBEC3s and retroviral inactivation**

There have been a number of reports suggesting that the antiretroviral activity of the APOBEC3s could be dissociated from their ability to deaminate DNA ([80] and references therein), but with a finer calibration of the experimental system, the only significant antiviral activity is likely to be due to the deaminase activity [81-83]. This highlights the difficulty in assaying the relevance of potential targets of APOBEC3s: the typical experimental system is based on transient overexpression of the enzyme together with the relevant retrovirus, followed by assessment of the infectivity of the viral particles in target cells. While this system can be used to test for novel APOBEC3 targets, it cannot be easily tuned to simulate the endogenous levels of the APOBEC3s. Thus the only known physiological target for endogenous APOBEC3s is HIV, and the G to A mutational bias observed in mobile elements [84,85] is the only indication of an involvement of the APOBEC3s in inhibiting their transposition in vivo. New tools to study the targets of the APOBEC3s in a more physiological manner are needed.

As discussed above, retroviral inactivation by APOBEC3s is due to the resulting inability of the retroviruses (or other mobile elements) to be integrated into the target-cell genome [26,57]. While a role for DNA glycosylases in trash the APOBEC3-modified viral genome was initially hypothesized, this is not the case [82,83,86,87], and other hypotheses need to be tested (for example, inefficient retrotranscription [88]). But it will be difficult to prove this without being able to assay APOBEC3s at their endogenous levels.

**Targeting AID to the immunoglobulin locus**

Little is known of the mechanisms that lead AID to act specifically on the rearranged variable regions of immunoglobulin genes in antigen-activated B cells. Although cis-acting elements that might help determine specificity have been identified (for a review see [40]), a trans-acting machinery is likely to play a major role in this targeting. Few proteins that interact with AID have been identified so far: MDM2, a regulatory protein shuttling between cytoplasm and nucleus [89] and replication protein A (RPA), a ubiquitous protein that binds single-stranded regions of DNA in DNA replication and repair [90]. While this property of RPA makes its association with AID intriguing, its lack of specificity cannot explain the physiological targeting to the immunoglobulin locus. Moreover, murine AID needs to be phosphorylated in order to trigger antibody diversification and AID is associated with protein kinase A (PKA) [91-93]. Yet, AID phosphorylation is not specific to B cells [94], it is not required for the fish homolog to act [95], and phosphorylation-defective AID mutants show delayed activity in somatic hypermutation and its substantial decrease [92]. These findings suggest that phosphorylation might be more related to AID modulation than to its targeting.

**AID/APOBECs and cancer**

While the AID/APOBECs are powerful tools for improving the immune response, it is clear that their unique activity - inserting mutations in nucleic acids - represents a double-edged sword in cellular metabolism. Transgenic mice overexpressing APOBEC1 and AID develop tumors [96,97],
and the mutational context of C to T changes in genes commonly mutated in cancer is consistent with the action of these deaminases [24]. In addition, the AID/APOBECs are widely expressed in cancer tissues and cell lines [4,23,98].

It was known even before the identification of AID that a number of genes were mutated as a byproduct of antibody diversification processes and that mutations and aberrations in some of these genes were specific to cancers of the B-cell lineage. AID has subsequently been proved to trigger c-myc/Igh translocations (a common trait in Burkitt’s lymphoma) in Balb/c mice [99-101]. Furthermore, expression of AID is needed in order to develop germinal-center-derived lymphomas in cancer-prone mice [102], and its aberrant expression might also have a role in the development of cancer (see for example [103]).

While there is no experimental evidence for the involvement of the APOBECs in cancer, they were first identified in keratinocytes treated with PMA, a phorbol ester known to be a skin tumor promoter [104]. Moreover, the induction of these mutators by viral infection [105] or antiviral pathways [106] could be the key to their role in cancer.

In the end, given that an association between the AID/APOBECs and the onset of cancer has been established, it needs to be ascertained whether this is due to stochastic events - unavoidable side effects of a mutational machinery - or if there are specific conditions that might induce aberrant function. This will only be achieved by an in-depth knowledge of the physiological roles of the AID/APOBECs.

Additional data files
Additional data is available online with this article. Additional data file 1 contains the alignment of the protein sequences used to calculate the phylogenetic tree shown in Figure 2. Additional data file 2 is a detailed version of the phylogenetic tree shown in Figure 2.

Acknowledgements
Helpful discussions with C Rada, MA Langlois, M Wang and JM Di Noia have influenced the writing of this review. This work was supported by an institutional grant from the Istituto Toscano Tumor. References
1. Navaratnam N, Morrison JR, Bhattacharya S, Patel D, Funahashi T, Giannini F, Teng BB, Davidson NO, Scott J: The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem 1993, 268:20709-20712.
2. Teng B, Burant CF, Davidson NO: Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science 1993, 260:1816-1819.
3. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T: Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem 1999, 274:18470-18476.
4. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N: An anthrapoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22, Genomics 2002, 79:285-296.
5. Sheehy AM, Hadgis NC, Choi JD, Malim MH: Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 2002, 418:646-650.
6. Gerber AP, Keller W: An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science 1999, 286:1146-1149.
7. Conticello SG, Langlois MA, Yang Z, Neuberger MS: DNA deamination in immunity: AID in the context of its APOBEC relatives. Adv Immunol 2007, 94:37-73.
8. Conticello SG, Langlois MA, Neuberger MS: Insights into DNA deaminasins. Nat Struct Mol Biol 2007, 14:7-9.
9. Rubin MA, Pastar I, Gaston KW, Ragone FL, Janzen CJ, Cross GA, Papavassiliou FN, Alfonso JD: An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. Proc Natl Acad Sci USA 2007, 104:7821-7826.
10. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS: Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol 2005, 22:367-377.
11. Saunders HL, Magor BG: Cloning and expression of the AID gene in the channel catfish. Dev Comp Immunol 2004, 28:657-663.
12. Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov VL, Aravin LD, Pancer Z: Evolution and diversification of lamprey antigen receptors: evidence for an ancient origin of the APOBEC family cytidine deaminase. Nat Immunol 2007, 8:674-675.
13. Nagawa F, Kishishita N, Shimizu K, Hirose S, Miyoshi M, Nezu J, Nishimura T, Nishizumi H, Takahashi Y, Hashimoto S, Takeuchi M, Miyajima A, Takemori T, Otsuka AJ, Sakano H: Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. Nat Immunol 2007, 8:206-213.
14. Rogozin IB, Basu MK, Jordan IK, Pavlov VL, Koonin EV: APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle 2005, 4:1281-1285.
15. Liao W, Hong SH, Chan BH, Rudolph FB, Clark SC, Chan L: APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochem Biophys Res Commun 1999, 260:398-404.
16. Anant S, Henderson JO, Mukhopadhyay D, Navaratnam N, Kennedy S, Min J, Davidson NO: Novel role for RNA-binding protein CUGBP2 in mammalian RNA editing, CUGBP2 modulates C to U editing of apolipoprotein B mRNA by interacting with splicing-1 and AC8, the apobec-1 complementation factor. J Biol Chem 2001, 276:47338-47351.
17. Muenk C, Beck T, Zielonka J, Hotz-Wagenblatt A, Chareza S, Battenberg M, Thiebelein J, Cichutek K, Bravo IG, O’Brien S, Lorechelt M, Yukihi N: Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. Genome Biol 2008, 9:R48.
18. Sawyer SL, Emerman M, Malik HS: Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol 2004, 2:26-37.
19. Zhang J, Webb DM: Rapid evolution of primate antiviral enzyme APOBEC3G. Hum Mol Genet 2004, 13:1785-1791.
20. Prochnow C, Brandsteitter R, Klein MG, Goodman MF, Chen XS: The APOBEC-2 crystal structure and functional implications for the deaminase AID, Nature 2007, 445:447-451.
21. Losey HC, Rustenburg AJ, Verdine GL: Crystal structure of Staphylococcus aureus RNA adenosine deaminase TadA in complex with RNA, Nat Struct Mol Biol 2006, 13:153-159.
22. Chen KM, Harjes E, Gross PJ, Fahmy A, Lu Y, Shindo K, Harris RS, Matsuo H: Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature 2008, 452:116-119.
23. Harris RS, Petersen-Mahrt SK, Neuberger MS: RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol Cell 2002, 10:1247-1253.
24. Beale RC, Petersen-Mahrt SK, Wutz IN, Harris RS, Rada C, Neuberger MS: Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J Mol Biol 2004, 337:585-596.
25. Liddament MT, Brown WL, Schumacher AJ, Clark SC, Chan L: Mutational comparison of the single-domain APOBEC3G and double-domain APOBEC3F/G anti-retroviral cytidine deaminases provides.
insight into their DNA target site specificities. Nucleic Acids Res 2003, 31:1913-1923.

27. Lau PP, Zhu HJ, Baldini A, Charnangaye C, Chan L: Dimeric structure of a human apolipoprotein B mRNA editing protein and cloning and characterization of its localization of its gene. Proc Natl Acad Sci USA 1994, 91:8522-8526.

28. Chiu YL, Soros VB, Kreisberg JF, Stopak K, Yonemoto W, Greene WC: Cellular AOPOBC3G restricts HIV-1 infection in resting CD4+ T cells. Nat Immunol 2003, 4:186-191.

29. De Souza J, Tanaka M, Aono T, Sato T, Tanaka M, Aono T: Activation-induced cell death, AICD, is cytotoxicity active as a monomer on single-stranded DNA. DNA Repair (Amst) 2008, 7:77-87.

30. Muramatsu M, Kinoshita K, Fagassan S, Yamada S, Shinkai Y, Honjo T: Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 2000, 102:533-563.

31. Arakawa H, Hauschild J, Buerstedde JM: Requirement of the activation-induced cytidine deaminase (AID) gene for immunoglobulin gene conversion. Science 2002, 295:1301-1306.

32. Harris RS, Sale JE, Petersen-Mahrt SK, Neuberger MS: AID is essential for immunoglobulin V gene conversion in a cultured B cell line. Curr Biol 2002, 12:435-438.

33. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, Yel L, Kinoshita K, Honjo T, Fischer A, Durandy A: Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). Cell 2000, 102:545-575.

34. Petersen-Mahrt SK, Harris RS, Neuberger MS: AID mutates E Coli suggesting a DNA deamination mechanism for antibody diversification. Nature 2002, 418:99-103.

35. Di Noia J, Neuberger MS: Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature 2002, 419:43-48.

36. Imm K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, Yel L, Forerville R, Mouget-Labouesse R, Genny A, Tzechan I, Ersy F, Kayserili H, Uzgof AG, Brousse N, Muramatsu M, Notarangelo LD, Kinoshita K, Honjo T, Fischer A, Durandy A: Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). Cell 2000, 102:545-575.

37. Petersen-Mahrt SK, Harris RS, Neuberger MS: AID mutates E Coli suggests DNA deamination mechanism for antibody diversification. Nature 2002, 418:99-103.

38. Plam P, Brauner R, Petruska J, Goodman MF: Processive AID-catalysed cytidine deamination on single-stranded DNA simulates somatic hypermutation. Nature 2003, 424:103-107.

39. Di Noia J, Neuberger MS: Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem 2007, 76:1-22.

40. Brar SS, Watson M, Diaz M: Activation-induced cytidine deaminase (AID) is expressed outside of the nucleus but retained in the induction of DNA breaks. J Biol Chem 2004, 279:26395-26401.

41. McBride KM, Barreto V, Ramiro AR, Stavropoulos P, Nussenzweig MC: Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced cytidine deaminase. J Exp Med 2004, 199:1235-1244.

42. Ito S, Nagaoka H, Shinkura R, Begum N, Muramatsu M, Nakata M, Honjo T: Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. Proc Natl Acad Sci USA 2004, 101:1975-1980.

43. Shinkura R, Ito S, Begum NA, Nagaoka H, Muramatsu M, Kinoshita K, Sakakibara Y, Hiyajita H, Honjo T: Separate domains of AID are required for somatic hypermutation and class-switch recombination. Nat Immunol 2004, 5:707-712.

44. Lau PP, Xiong WJ, Zhu HJ, Chen SH, Chan L: Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. J Biol Chem 1991, 266:20550-20554.

45. Yang Y, Smith HC: Multiple protein domains determine the cell-type specific nuclear distribution of the catalytic subunit required for apolipoprotein B mRNA editing. Proc Natl Acad Sci USA 1997, 94:13075-13080.

46. Chester A, Somasekaram A, Tzima M, Jarmuz A, Gisbourne J, O'Keefe R, Scott J, Navaratnam N: The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. EMBO J 2003, 22:3971-3982.

47. Matha A, Kinter MT, Sherman AJ, Ovitt DR: Molecular cloning of apolipoprotein B mRNA editing: characterization, mutation and expression. Rev Insan Med 2004, 32B:163-168.

48. Schaefer A, Boderer HP, Cullen BR: Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleosidase domain of the gag polyprotein. Nucce Biol 2004, 22:163-168.

49. Zunsza Z, Xu H, Mbisa J, Barr R, Gorelick R, Ono A: A novel bridge protein complex is required for HIV-1 virions to exit the nucleus. J Biol Chem 2005, 280:35822-35827.
