Mechanism of Strand-Biased and Codon-Contexted Somatic Hypermutation in Immunity and Cancer: Implications of the finding that ADAR-deaminases can directly mediate A-to-I Editing on both the DNA and RNA moieties of RNA:DNA hybrids.

Edward J Steele[1] and Robyn A Lindley[2,3]

[1,a] CYO'Connor ERADE Village Foundation Inc. Piara Waters, WA, AUSTRALIA
[2] GMDxCo Pty Ltd, Hawthorn Vic, AUSTRALIA; and [3] Department of Pathology
Faculty of Medicine, Dentistry & Health Sciences, University of Melbourne Vic, AUSTRALIA

Running head: *Mechanism A-to-I at Transcription Bubbles*

Correspondence:
Associate Professor Edward J Steele PhD
Honorary Research Associate
CYO'Connor ERADE Village Foundation Inc.
24 Genomics Rise, Piara Waters, WA 6112, AUSTRALIA

Tel (+61) (0) 420 863 551
Email: ejsteele@cyo.edu.au

Abstract

This paper simply links the recent findings of Zheng, Lorenzo and Beal on ADAR-mediated DNA and RNA deamination at RNA:DNA hybrids, to our previous work on strand-biased and codon-context mutation signatures in B lymphocytes (Ig SHM) and codon-contexted exome-wide point mutation patterns in cancer genomes. We conclude that *in vivo* the A-to-I DNA editing component at RNA: DNA hybrids occurring in Transcription Bubbles, while important, is of far lower A-to-I editing efficiency than in dsRNA substrates (as shown in Zheng et al). Indeed the RNA moiety of RNA:DNA hybrids is also edited at similar lower frequency (relative to dsRNA substrates). Further if the A-to-I DNA editing at RNA:DNA hybrids were the sole cause of A-to-I (read as A-to-G) mutation events *in vivo* then the exact opposite strand biases at A:T base pairs (T>>>A) of what is actually observed (A>>>T) would be predicted. Thus we conclude that the extreme strand-biased somatic mutation patterns documented by us *in vivo* should be logically interpreted by the predicted sequential steps of the RNA/RT-based mechanism.

Abbreviations used in this paper:
Aag, alkyladenine DNA glycosylase; **ADAR,** Adenosine Deaminase that acts on RNA; **AID**, activation induced cytidine deaminase, a APOBEC family member, initiating via C-to-U lesions in ssDNA of class switch recombination (CSR) and somatic hypermutation (SHM) processes at somatically rearranged Ig V(D)J gene loci, and known to activate cytidine mutagenic deamination during transcription in other somatic tissues, particularly in cancer; **APOBEC family**, generic abbreviation for the deoxyribonucleic acid, or dC-to-dU, deaminase family (APOBEC3 A, B, C, D, F, G, H) similar in DNA sequence to the "apolipoprotein B RNA editor" APOBEC1, and known to activate mutagenic cytidine deamination during transcription in somatic tissues, particularly in cancer; **AP**, an Abasic, or apurinic/apyrimidinic, site; **APE**, AP endonuclease; **A-to-I**, adenosine-to-inosine RNA editing; **BER**, base excision repair; **dA**, deoxyadenosine; **dA-to-dI**, deoxyadenosine to deoxyinosine DNA editing; **dC**, deoxycytosine; **Deaminase**, catalytic domain in ADAR and AID/APOBEC enzymes; **Ig-SHM-like** response, strand-biased somatic mutation patterns similar to that observed in Ig SHM; **MMR**, mismatch repair; **Motif**, 4 to 6 nucleotide (N) sequence defining specificity of deaminase targeting; **MSH2-MSH6**, MutSα heterodimer recognising mispaired bases in DNA duplex; **NTS**, the non-transcribed, or "Top", 5' to 3' strand; **Pol-η** or DNA polymerase-η (eta); **RNA Pol II**, RNA Polymerase II; **rA**, adenosine in RNA; **RT**, reverse transcriptase; **RT-Pol-η**, reverse transcriptase activity displayed by Pol-η; **SHM**, somatic hypermutation; **T**, Thymine; **TS**, the transcribed, or "Bottom", 3' to 5' strand, in context of a Transcription Bubble; **TSM**, targeted somatic mutations : the process of targeting C and A nucleotides for deamination in actively transcribed genes that results in a dominant type of mutation caused by a deaminase binding domain (DBD) or an Inferred-DBD at a particular codon position; **TSRT**, target site reverse transcription; **U**, uracil; **UNG**, uracyl DNA glycosylase involved in BER at dU sites in DNA resulting in either an Abasic site (AP) or APE-mediated ssDNA nicks (above); **V[D]J**, generic symbol for a rearranged immunoglobulin (or T cell receptor, TCR) variable region gene in the Adaptive Immune System; **W**, weak base pair involving A or U/T; **Y,** pyrimidines T/U or C.

Contrary to previous data and expectations that there are no known deoxyadenosine (dA) deaminases which act directly on adenine in polynucleotide DNA strands (whether as duplex DNA or ssDNA during Ig SHM [1-3] and reviewed in Lindley [4]), the laboratory of Peter Beal has recently published incontrovertible evidence showing that A-to-I DNA editing of deoxyadenines in the DNA moiety in RNA:DNA hybrid substrates does occur [5]. Here we link these findings to our previous work on strand-biased and codon-context mutation signatures in immunoglobulin-targeted somatic hypermutation in B lymphocytes (Ig SHM) and codon-contexted exome-wide somatic point mutation patterns in cancer genomes [4,6]. We have rationalized the logical implications of these findings, derived from *in vitro* biochemical assays, and compared the expected outcome(s) of such dA-to-dI DNA deaminations with the known set of strand-biased mutation patterns observed at A:T and G:C base pairs for *in vivo* Ig SHM in B lymphocytes [6, 7] as well as seen at numerous A-sites and C-sites across the exomic regions in cancer genomes [8,9]. We have reasoned that *in vivo* the A-to-I DNA editing component most likely to occur at RNA:DNA hybrids, occurs in the context of an open Transcription Bubble on the bottom transcribed strand (TS) (Figure 1, Figure 2). While this activity is likely to be important in some mutagenic-caused diseases, the A-to-I editing efficiency would be far lower than in dsRNA substrates, and confirmed by Zheng et al [5]. Based on our mutation data-driven analyses we predict that *in vivo,* this targeting occurs routinely in Ig SHM at dsRNA WA sites in stem-loops in nascent pre-mRNA molecules emerging from Transcription Bubbles [10]. The conversion of these WA-to-WI RNA deamination events as A-to-G mutations in DNA following reverse transcription and replication (Figure 1) causes the extreme strand biased mutation patterns documented by us *in vivo* [6- 11]. If ADAR-mediated DNA deaminations *in vivo* were only due to the dA-to-dI events documented by Zheng et al 2017 [5] it would logically result in strand-biased mutations at A-sites, but these strand-biases would be predicted to be the *opposite* of what is actually observed *in vivo* i.e. we would expect to see mutations off T exceeding mutations off A (T>>>A), when read by convention in the sequence of the Top or non-transcribed strand (NTS). The observed strand-biases found in both Ig SHM and most (if not all) cancer genomes that we have analyzed unequivocally show A>>>T (Figure 1). Further Zheng et al [5] also show that the RNA moiety in RNA:DNA hybrids (in the correct *in vivo* Transcription Bubble orientation) is also A-to-I edited at a similar lower efficiency (about 10-20 fold lower) than dsRNA substrates. Thus, we conclude that the implications of the quantitative biochemical A-to-I deamination assays of Zheng et al 2017 [5] should be logically rationalized by the predicted sequential steps of a RNA/RT-based mechanism (Figure 1). This is important because, it is still not widely acknowledged that SHM processes involve any form of RNA intermediary – and hence a now crucial reverse transcription (RT) step.

In Figure 2, a summary of the molecular events at "stalled" Transcription Bubbles (in codon-register) that we believe play a key role *in vivo* during Ig SHM and tissue-wide oncogenesis is shown [4]. The assumption that the AID-deaminase mediated C-to-U and ADAR-mediated A-to-I occurs at stalled Transcription Bubbles in codon-register, also provides a rational explanation for the codon-context
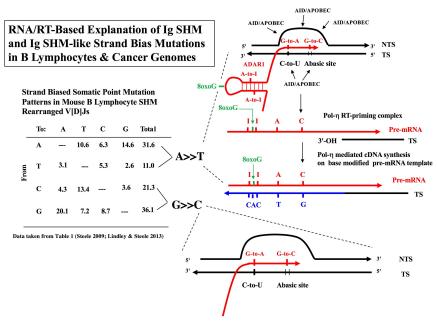

RNA/RT-Based Explanation of Ig SHM and Ig SHM-like Strand Bias Mutations in B Lymphocytes & Cancer Genomes

Strand Biased Somatic Point Mutation Patterns in Mouse B Lymphocyte SHM Rearranged V[D]Js

From \ To:	A	T	C	G	Total
A	---	10.6	6.3	14.6	31.6
T	3.1	---	5.3	2.6	11.0
C	4.3	13.4	---	3.6	21.3
G	20.1	7.2	8.7	---	36.1

Data taken from Table 1 (Steele 2009; Lindley & Steele 2013)

A>>T

G>>C

Figure 1. Pattern of Somatic Point Mutations in Ig Somatic Hypermutation in B Lymphocytes. Likely molecular explanation for the extreme strand biased somatic mutations in Ig SHM. But very similar data and explanation apply to exome-wide point mutations in cancer genomes (based on the hypothesised dysregulated Ig-like SHM process operating across the cancer genome involving DNA and RNA deaminations coupled to reverse transcription). Adapted from Figure 1 in Lindley and Steele [9] and Figure 7 in Lindley [4] , and reviewed again in Steele [6]. This is a variant of the target site reverse transcription (TSRT) process originally hypothesized by Luan et al [14] and first applied to the Ig SHM process in Steele et al [15]). Shown for the generation of the main A-site and G-site strand biased mutation components is a Transcription Bubble and sequelae showing some hypothesised DNA and RNA intermediates highlighted for the generation of the main strand- biased mutation signatures involving A-to-G, G-to-A, G-to-T and G-to-C. Black lines are DNA strands, red lines are mRNA, blue lines are cDNA strands copied off mRNA by a cellular reverse transcriptase such as DNA polymerase η [16]. Steps on the right show various mutated DNA and RNA intermediates and substrate complexes for both deamination reactions, 8oxoG modifications in RNA [17], and cDNA synthesis. Note that it is still not known if 8oxoG sites generated by reactive oxygen species are preferred in unpaired loops or dsRNA regions. In our view, mutations are first introduced at the DNA level by AID/APOBEC family-mediated C-to-U deaminations, and then uracil DNA glycosylase (UNG)-generated abasic sites in the TS (which can further mature into single strand nicks via the action of AP endonuclease (APE). These template sites are transcribed into mRNA by RNA Pol II generating G-to-A and G-to-C modifications respectively in the pre-mRNA Kuraoka et al [18] which on TSRT-mediated reverse transcription, integration and DNA replication result in G-to-A and G-to-C mutations in the NTS. Separately, adenosine-to-inosine (A-to-I) RNA editing events at WA targets in the nascent and Transcription Bubble-proximal dsRNA stem loops may be copied back into DNA by reverse transcription via Pol-η [16 ,10]. Also shown in green are 8oxoG modifications in mRNA, which during reverse transcription, integration and DNA replication would result in strand-biased G-to-T transversions on the NTS. The strand invasion and integration of newly synthesised cDNA TS are hypothesized necessary steps (not shown here). In more detail: RNA Pol II introduces mutations in mRNA as it copies the AID/APOBEC lesions in TS DNA, concurrently A-to-I RNA edited sites appear in RNA stem(-loops) forming in nascent mRNA near the transcription bubble [10] or 8oxoG RNA modifications via reactive oxygen species [9]. Next, the RT-priming substrates are formed by annealing the nicked TS strand with an exposed 3'-OH end (for Y Family translesion DNA polymerase-η [19], now acting in it reverse transcriptase mode [16]). This could arise due to excision at a previous AID-mediated abasic site, or due to an excision introduced by endonuclease activity associated with the MSH2-MSH6 heterodimer engaging a U:G mispaired lesion [20] . This allows extension of a new TS by cDNA synthesis from the 3'-OH end copying the already base modified mRNA template (with I base pairing preferentially, like G, with C; and 8oxoG mispairing with A). Then it is envisaged that an unknown and indeterminant number of steps previously predicted by Luan et al [14] occur. It is possible that these steps involve forms of strand invasion, heteroduplex formation and/or resolution of heteroduplex, full length copying of newly synthesized transcribed strand cDNA being locked into the genomic DNA at the V[D}J re-arranged genomic site.
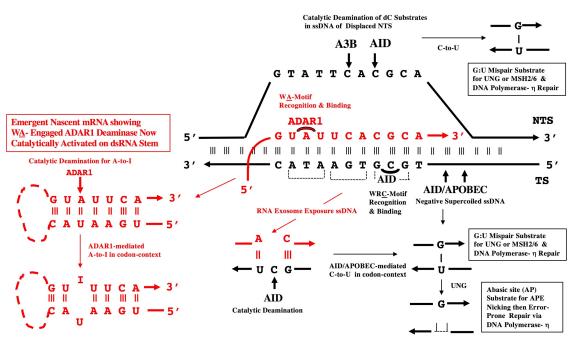
Targeted Somatic Mutation (TSM) signatures [12,13] observed in Ig SHM [6] and cancer exomes [4,12,13]. Thus both C-to-U and A-to-I events are anticipated to occur, in one model [4] in two main steps at RNA:DNA hybrids stalled in codon-register. First, motif recognition (binding); and second, the sequelae of steps leading to catalytic deamination of dC and rA nucleotide targets (Figure 2).

However Zheng et al [5] show that the RNA moiety of the DNA:DNA hybrid is also A-to-I edited, just like the DNA moiety in a RNA:DNA hybrid - and at similar lower efficiency (compared to dsRNA substrates, the conventional target of ADAR deaminases [33]). Therefore the extrusion of the nascent pre-mRNA from the stalled Transcription Bubble as shown in Figure 2, to allow catalytic A-to-I deamination on dsRNA stem-loops may not be necessary if the ADAR can gain direct access to the RNA in the annealed RNA:DNA hybrid in the open Transcription Bubble. Thus the observed extreme strand bias of A>>>T mutations observed *in vivo* could be a direct result of A-to-I RNA editing within the Transcription Bubble on the nascent RNA annealed to the transcribed DNA strand. However the quantitative significant correlations in the VκOx1Jκ5 mutation dataset with the likely frequency of nascent stem-loop formation related to short unpaired loops (i.e. nascent RNA stem-loops forming near the Transcription Bubble) are the substrates A-to-I edited at highest frequency[10], supporting a two-step process as shown in Figure 2. The data summarised in Figures 1 and 2 are covered in detail in our previous publications (reviewed in Lindley and Steele [9], Steele [6], and Lindley [4]).

Our purpose here then has been to show the consistency between our work on these A-to-I mechanisms, and the just published findings of the Beal lab: the primary function of a RNA/RT-based mechanism of SHM [21, 22,16, 11,10, 9, 6] is to explain the extreme strand biased mutations at A:T and G:C base pairs. This model does just that. It does not focus on trying to explain only the dominance of A-to-G over T-to-C, but the totality of A-site strand dominance over T-site mutations (A>>>T) *and* the same type of extreme excess of G-site mutations over those at C-sites (G>>>C, as shown in Figure 1).

We conclude that *in vivo,* the A-to-I DNA editing component at RNA: DNA hybrids in the context of a Transcription Bubble, while important, will be predicted to exhibit far lower A-to-I editing efficiency than in dsRNA substrates and as shown *in vitro* in Zheng et al [5]. In summary, the extreme strand-biased mutation patterns documented by us *in vivo,* can only be understood in the context of the RNA/RT-based model proposed by us, because, *in vivo* the edited dA would be on the TS strand in the

Transcription Bubble (Figure 2) and such DNA A-to-I editing *cannot generate* the A>>>T strand biases evident in all extant *in vivo* datasets as described in Figure 1.

Figure 2. Elongation Stalls in Codon Register Allowing Access to Potential ssDNA and dsRNA Deamination Substrates. See text for details. Adapted from Lindley [4]. For discussion on contributions of negative supercoiled ssDNA regions to SHM patterns on both NTS and TS see Shen and Storb [23]; Wright et al [24,25,26]; Franklin and Blanden [27, 28]. For background discussions on stalling of transcription elongation see Mooney et al [29] and Moore and Proudfoot [30]; for the normal preference for the displaced NTS strand for AID/APOBEC strand -biased C-to-U deamination, see Sohail et al [31]; for background on Y family DNA translesion polymerases see Goodman [19]; for Y family polymerases, particularly DNA Polymerase-η, as reverse transcriptases, see Franklin et al [16]; for background on A-to-I RNA editing see Bass [32] , and for the action of the RNA exosome, see Basu et al [33].

We look forward to further experimentation on the mechanism of somatic hypermutation of Ig V[D]J genes in ADAR1 deficient Aicardi-Goutières Syndrome (AGS) patients [34,35] and possibly employing the VDJ somatic hypermutation read-out of V->DJ rearrangements of the D-proximal V_H6 gene segment employed by Patricia J Gearhart and her associates whom first demonstrated that DNA Polymerase -η is the cause of A:T mutations in SHM in humans [36].

Acknowledgement

We thank Liam Keegan for discussions on the role of A-to-I RNA editing in the generation of the somatic hypermutation pattern, and for bringing the In Press paper by Zheng et al 2017 to our attention.

Conflict of Interest

The authors have no conflict of scientific interest that harms the objectivity of this paper.

References

[1]   T. Lindahl. Instability and decay of the primary structure of DNA. Nature 362 (1993) 709 - 715.

[2]   I. Alseth, B. Dalhus, M. Bjøras. 2014 Inosine in DNA and RNA. Curr. Opin. Genet. Dev. 26 (2014) 116 - 123.

[3]   S. Longerich, L. Meira, D.Shah, L.D. Samson, U. Storb. Alkyladenine DNA glycosylase (Aag) in somatic hypermutation and class switch recombination. DNA Repair 6 (2007) 1764 - 1773.

[4]   R.A. Lindley. Targeted somatic mutation (TSM) signatures: A review. To be submitted to Molecular Genetics and Genome Medicine. February 2017.

[5]   Y.C. Zheng,C. Lorenzo, P.A. Beal. DNA Editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. Nucleic Acids Research 2017 In Press doi: 10.1093/nar/gkx050

[6]   E.J. Steele. Somatic hypermutation in immunity and cancer: Critical analysis of strand-biased and codon-context mutation signatures. DNA Repair 45 (2016) 1 – 24.

[7]   E.J. Steele. Mechanism of somatic hypermutation: Critical analysis of strand biased mutation signatures at A:T and G:C base pairs. Molec. Immunol 46 (2009) 305 - 320.

[8]   E.J. Steele, R.A. Lindley. Somatic mutation patterns in non-lymphoid cancers resemble the strand biased somatic hypermutation spectra of antibody genes. DNA Repair. 9 (2010) 600 - 603.

[9]   R.A. Lindley, E.J. Steele. Critical analysis of strand-biased somatic mutation signatures in TP53 versus Ig genes, in genome -wide data and the etiology of cancer ISRN Genomics. Vol 2013 Article ID 921418, 18 pages.

[10]   E. J. Steele, R.A. Lindley, J. Wen, G.F. Weiler. Computational analyses show A-to-G mutations correlate with nascent mRNA hairpins at somatic hypermutation hotspots. DNA Repair 5 (2006) 1346 - 1363.

[11]   E.J. Steele, A. Franklin, R.V. Blanden. Genesis of the strand biased signature in somatic hypermutation of rearranged immunoglobulin variable genes. Immunol. Cell Biol. 82 (2004) 208 - 218.

[12]   R. Lindley. The importance of codon context for understanding the Ig-like somatic hypermutation strand-biased patterns in TP53 mutations in breast cancer Cancer Genet. 206 (2013) 222 - 226.

[13]   R.A. Lindley, P. Humbert, C. Larmer, E.H. Akmeemana , C.R.R. Pendlebury CRR. Association between Targeted Somatic Mutation (TSM) signatures and HGS-OvCa progression. Cancer Med. 5 (2016) 2629 - 2640.

[14]   D.D. Luan, M.H. Korman, J.L. Jakubczak, T.H. Eichbush. 1993 Reverse transcription of R2B mRNA is primed by a nick at the chromosomal target site; A mechanism for non-LTR retrotransposition. Cell 72 91993) 595 - 605.

[15]   E.J. Steele, H.S. Rothenfluh, R.V. Blanden. Mechanism of antigen-driven somatic hypermutation of rearranged immunoglobulin V(D)J genes in the mouse. Immunol. Cell Biol. 75 (1997) 82 - 95.

[16]   A. Franklin, P.J. Milburn, R.V. Blanden, E.J. Steele. Human DNA polymerase-η an A-T mutator in somatic hypermutation of rearranged immunoglobulin genes, is a reverse transcriptase. Immunol. Cell Biol. 82 (2004) 219 - 225.

[17]   J. Wu, Z. Li. Human polynucleotide phosphorylase reduces oxidative RNA damage and protects HeLa cell against oxidative stress. Biochem Biophys Res Comm. 372 (2008) 288 - 292.

[18]   I. Kuraoka, M. Endou, Y. Yamaguchi, Y. Wada, H. Handa, K. Tanaka. 2003 Effects of endogenous DNA base lesions on transcription elongation by mammalian RNA polymerase II. J. Biol. Chem. 278 (2003) 7294 - 7299.

[19]   M.F. Goodman, M.F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. Annu. Rev. Biochem. 71 (2002) 17 - 50.

[20]   T.M. Wilson, A. Vaisman, S.A. Martomo, P. Sullivan, L. Lan, F. Hanaoka, A. Yasui, R. Woodgate, P.J. Gearhart. MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes, J. Exp. Med. 201 (2005) 637 - 645.

[21]   E.J. Steele, J.W. Pollard JW. Hypothesis: Somatic Hypermutation by gene conversion via the error prone DNA-to-RNA-to-DNA information loop. Molec. Immunol. 24 (1987) 667 - 673.

[22]   E.J. Steele, R.V. Blanden. The reverse transcriptase model of somatic hypermutation. In *Hypermutation in antibody genes*, Eds R.D. Wood, P.J. Gearhart, M.S. Neuberger. Phil. Trans. Roy. Soc. (Series B) : Biological Sciences. 356 (2001) 61 - 66.

[23]   H.M. Shen, U. Storb. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 12997 - 13002.

[24]   B. E. Wright, K. H. Schmidt, M.F. Minnick, M. F. (2004) Mechanisms by which transcription can regulate somatic hypermutation. Genes Immun. 5 (2004)176 - 182.

[25]   B. E. Wright, K. H. Schmidt, M. F. Minnick, N. Davis I. VH gene transcription creates stabilised secondary structures for coordinated mutagenesis during somatic hypermutation. Mol. Immunol. 45 (2008) 3589 - 3599.

[26]   B. E. Wright, K. H. Schmidt, N. Davis, A.T. Hunt, M.F. Minnick II. Correlations between secondary structure stability and mutation frequency during somatic hypermutation. Mol. Immunol. 45 (2008) 3600 - 3608.

[27]   A. Franklin, R.V. Blanden. On the molecular mechanism of somatic hypermutation of rearranged immunoglobulin genes. Immunol. Cell Biol. 82 (2004) 557 - 567.

[28]   A. Franklin, R.V. Blanden. Hypothesis: Biological role for J-C intronic matrix attachment regions in the molecular mechanism of antigen-driven somatic hypermutation. Immunol. Cell Biol. 83 (2005) 383- 391.

[29]   R.A. Mooney, I. Artsinovitch, R. Landick. Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation J. Bact. 180 (1998) 3265 - 3275.

[30]   M.J. Moore, N.J. Proudfoot. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 136 (2009) 688 - 700.

[31]   A. Sohail, J. Klapacz, M. Samaranayake, A. Ullah, A. S. Bhagwat. Human activation-induced cytidine deaminase causes transcription dependent, strand-biased C to U deaminations. Nucl. Acids Res. 31 (2003) 2990 - 2994.

[32]   B.L. Bass. RNA editing by adenosine deaminases that act on RNA. Ann. Rev. Biochem. 71 (2002) 817 - 846.

[33]   U. Basu, F.L. Meng, C. Keim, V. Grinstein, E. Pefanis, J. Eccleston, et al. The RNA Exosome Targets the AID Cytidine Deaminase to Both Strands of Transcribed Duplex DNA Substrates. Cell 144 (2011) 353 - 363.

[34]   M.A. O'Connell, N.M. Mannion, L.P. Keegan. 2015 The Epitranscriptome and Innate Immunity. PLoS Genet. 2015 Dec 10;11(12):e1005687. doi: 10.1371/journal.pgen.100568

[35]   G.I. Rice, P.R. Kasher, G.M.A. Forte, N.M. Mannion, S.M. Greenwood, M. Marcin Szynkiewicz, et al. Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature Nat Genet. 44 (2012) 1243 - 1248. doi:10.1038/ng.2414.

[36]   X. Zeng, D.B. Winter, C. Kasmer, K.H. Kraemer, A.R. Lehmann, P.J. Gearhart. DNA polymerase-η as an A-T mutator in somatic hypermutation of immunoglobulin variable genes, Nat. Immunol. 2 (2001) 537 - 541.