Mechanism of Somatic Hypermutation in Immunity and Cancer:

Critical Analysis of "DNA Editing in DNA/RNA hybrids by adenosine deaminases that act on RNA" by Zheng Y, Lorenzo C and Beal PA (2017) Nucl Acids Research 2017 In Press doi: 10.1093/nar/gkx050

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*

Keywords:
Somatic Hypermutation at Transcription Bubbles
Strand-Biased Mutations
DNA Polymerase -η
A-to-I RNA editing
Targeted Somatic Mutation (TSM) and Codon-Context Mutations
AID/APOBEC-Deaminase Oncogenesis

[a] ***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 findings of Zheng, Lorenzo and Beal (2017) 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 in Transcription Bubbles, while important is of far lower A-to-I editing efficiency than in dsRNA substrates (as shown in Zheng et al 2017), and that the extreme strand biased mutation patterns documented by us *in vivo* should be understood and logically rationalized 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; **ADAT**, Adenosine Deaminase that acts on tRNA; **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; **Deaminase**, catalytic domain in ADAR and AID/APOBEC enzymes; **DSB**, double strand DNA breaks; **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; **N**, any nucleotide; **NTS**, the non-transcribed, or "Top", strand; **NGS,** Next Generation Sequencing; **Pol-η** or DNA polymerase-η (eta); **R**, Adenosine (A) or Guanine (G) , purines; **RNA Pol II**, RNA Polymerase II; **RT**, reverse transcriptase; **RT-Pol-η**, reverse transcriptase activity displayed by Pol-**η; S**, strong base pair involving Cytosine (C) or Guanine (G); **SHM**, somatic hypermutation; **T**, Thymine; **TS**, the transcribed, or "Bottom", 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 DBD or Inf-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); **UTR**, untranslated regions in the upstream (5') and downstream (3') regulatory regions of protein coding genes; **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; **X**, C or A ; **Y,** pyrimidines T/U or C.

The arresting and very important paper just sent to us as a PDF (by Professor Liam Keegan of CEITEC – Central European Institute of Technology - Masaryk Universityin Brno) on January 31st 2017 published by Zheng, Lorenzo and Beal (2017) has just appeared In Press in *Nucleic Acid Research*. These beautiful and striking biochemical data are *very relevant* to previous data and analyses published by us on the *in vivo* deamination mechanisms (C-to-U, A-to-I) which generate the characteristic somatic point mutation patterns in rearranged immunoglobulin (Ig) variable genes (B lymphocyte, V[D]Js, Steele et al 2006; Steele 2009; Steele 2016) and in the exomic regions of the human cancer genome (Steele & Lindley 2010; Lindley & Steele 2013; Lindley 2013; Lindley et al 2016; Lindley 2017 In Preparation, about to be submitted to *Molec. Genet. Genomic Med*.).

Our molecular explanation of the extreme strand biased mutation patterns in both Ig SHM in B lymphocytes and in the Ig-SHM-Like responses in human cancer genomes is summarised in Figure 1 (and Legend). Whilst in Figure 2 is the summary of the molecular events at "stalled" Transcription Bubbles we believe play a key role *in vivo* during Ig SHM and tissue-wide Oncogenesis. The data summarised in these Figures are covered at some detail and at great length in our previous publications (reviewed in Lindley and Steele 2013; Steele 2016; Lindley 2017). Our purpose here is to show the connection between our work on these mechanisms and the paper just published by the group of Peter Beal at Department of Chemistry, University of California, at Davis (Zhen et al 2017).

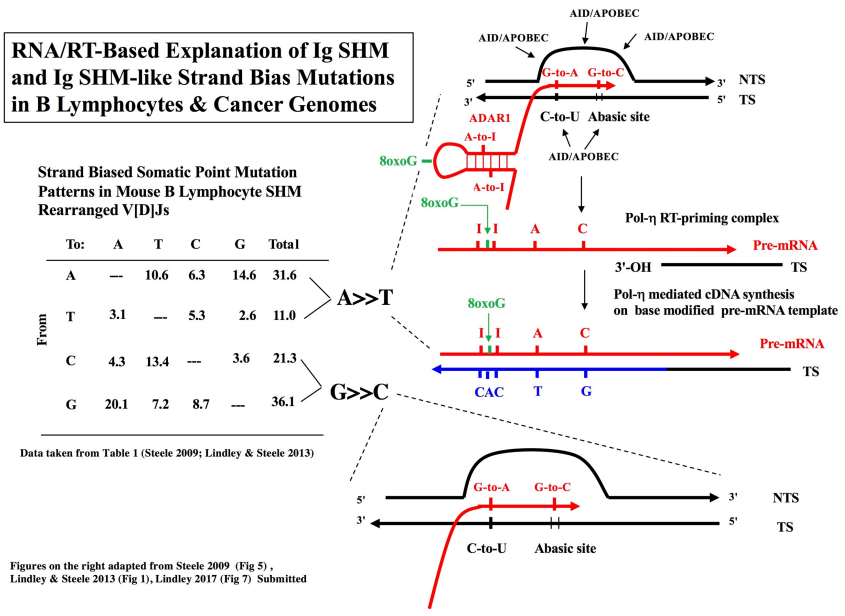

Figure 1. Pattern of Somatic Point Mutations in Ig Somatic Hypermutation in B Lymphocytes. Likely molecular explanation for the extreme strand biased somatic mutations in Ig SHM. But very similar data and explanation apply to exome-wide point mutations in cancer genomes (based on the hypothesised dysregulated 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 (2013) and Figure 7 in Lindley (2017), and reviewed again in Steele (2016). This is a variant of the target site reverse transcription (TSRT) process originally hypothesized by Luan et al (1993) and first applied to the Ig SHM process in Steele et al (1997). 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. Steps on the right show various mutated DNA and RNA intermediates and substrate complexes for both

deamination reactions, 8oxoG modifications in RNA (Wu and Li 2008), and cDNA synthesis (it is not known if 8oxoG sites generated by reactive oxygen species are preferred in unpaired loops or dsRNA regions). In over 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, (2003) 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 Trancription Bubble-proximal dsRNA stem loops may be copied back into DNA by reverse transcription via Pol-η (Franklin et al, (2004; Steele et al 2006). Also shown in green are 8oxoG modifications in mRNA which on 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 [Steele et al.,2006] or 8oxoG modifications via reactive oxygen species. Next formation of RT-priming substrates (for Y Family translesion DNA polymerase-η, Goodman 2002 now acting in it reverse transcriptase mode, Franklin et al 2004) by annealing of nicked TS strand with an exposed 3'-OH end. This could arise due to excision at a previous AID-mediated abasic site or an excision introduced by endonuclease activity associated with the MSH2-MSH6 heterodimer engaging a U:G mispaired lesion. This allows extension of 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 an unknown and indeterminant number of steps involving strand invasion(?), heteroduplex formation and/or resolution of heteroduplex (?), full length copying of newly synthesized transcribed strand (?) cDNA is locked into the genomic DNA at the the V[D}J site as envisaged by Luan et al (1993).

The primary purpose then of the RNA/RT-based mechanism of SHM (Steele and Pollard 1987; Steele and Blanden 2001; Lindley & Steele 2013; Steele 2016) is to explain the extreme strand biased mutations at A:T and G:C base pairs. This model does just that. It does not just focus on trying to explain just 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).

These two sets of concerns led us to the explanatory molecular model outlined in Figure 1. This model actually does explain the minutiae of *all the known* major mutation strand biases (Steele 2009). For this reason we adjudge all new data in the field (e.g. Zheng et al 2017) as whether their molecular implications flow through to real *in vivo* mutation phenomena - strand biased and codon-contexted mutation signatures (Lindley & Steele 2013; Lindley 2013; Lindley et al 2016; Steele 2016).

It is clear from the quantitative A-to-I editing data in Zheng et al (2017) that the editing efficiency of ADAR1/2 on the DNA component (in the RNA:DNA hybrid) is far less (an order of magnitude lower on a molar/time kinetic basis) than normal A-to-I editing of dsRNA substrates. Further,*in vivo,* the relevant configurations in a Transcription Bubble would be as shown in (Figure 2, discussed at great length in Lindley 2017). Since the evidence shows that A-to-I editing does not occur in ssDNA e.g displaced NTS in Figure 2 (Lindahl 1993; Longerich 2007; Alseth et al 2014; and analysed in detail in Lindley 2017) how

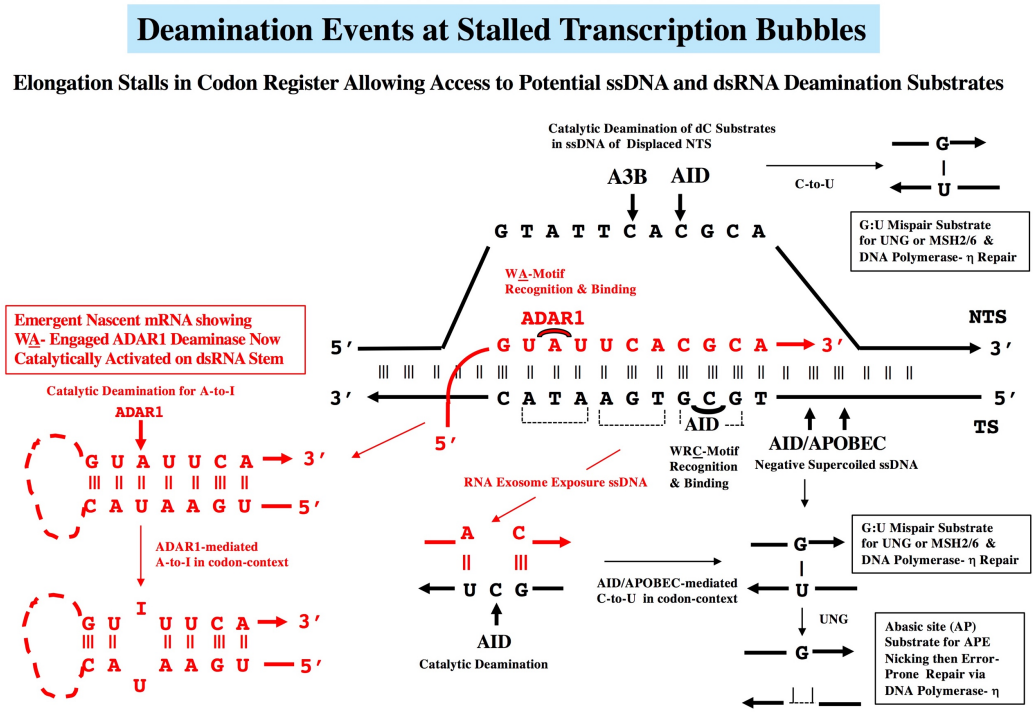

Figure 2. Elongation Stalls in Codon Register Allowing Access to Potential ssDNA and dsRNA Deamination Substrates. See text for details. Adapted from Lindley (2017). For discussion on contributions of negative supercoiled ssDNA regions to SHM patterns on both NTS and TS see Shen and Storb (2004); Wright et al (2004) and Franklin and Blanden (2004, 2005). For background discussions on stalling of transcription elongation see Mooney et al (1998) and Moore and Proudfoot (2009). See also for deep background Bass 2002; Basu et al 2011.

then do we explain, say, A>>T strand bias on basis of the Zheng et al 2017 biochemical data? It seems to us that since A-sites in the RNA:DNA hybrid (in both orientations 'top' v 'bottom') are equally poor as editing targets (compared to RNA: RNA duplexes) this means to us, that the strand biases at A-site and G-site (C-site on TS) must logically follow the rules as out lined in the RNA/RT-based model (Figure 1). We conclude that *in vivo* the A-to-I DNA editing component at RNA: DNA hybrids in Transcription Bubbles, while important is of far lower A-to-I editing efficiency than in dsRNA substrates (as shown in Zheng et al 2017), and that the extreme strand biased mutation patterns documented by us *in vivo* should be understood and logically rationalized by the predicted sequential steps of the RNA/RT-based mechanism. Because *in vivo* the edited dA would be on the TS strand in the Transcription Bubble (Figure 2) its editing, at whatever frequency, *would not generate* the A>>>T strand biases evident in all the extant *in vivo* datasets, and as explained in Figure 1. We look forward to the experimental tests that can now arise from the detailed study of somatic hypermutation of Ig V[D]J genes in ADAR1 deficient Aicardi-Goutières Syndrome (AGS) patients (O'Connell et al 2015; Rice et al 2012) possibly using the VDJ somatic hypermutation readout of the VH6 gene employed by Patricia J Gearhart and associates (Zeng et al 2001).

Acknowledgement

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

Conflict of Interest

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

References

Alseth, I., B. Dalhus, and M. Bjøras. 2014 Inosine in DNA and RNA. Curr. Opin. Genet. Dev. 26: 116 - 123

Bass, B.L. 2002 RNA editing by adenosine deaminases that act on RNA. Ann. Rev. Biochem. 71 : 817 - 846.

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

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

Franklin, A., and R.V. Blanden. 2004 On the molecular mechanism of somatic hypermutation of rearranged immunoglobulin genes. Immunol. Cell Biol. 82 : 557 - 567.

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

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

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

Lindahl, T. 1993 Instability and decay of the primary structure of DNA. Nature 362: 709 - 715.

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

Lindley, R.A. 2017 Targeted somatic mutation (TSM) signatures: A review. To be submitted to Molecular genetics and Genome Medicine. February 2017.

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

Lindley, R.A., Steele, E.J. 2013 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

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

Luan, D.D., M.H. Korman, J.L. Jakubczak, and 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 : 595 - 605

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

Moore, M. J., and N.J. Proudfoot. 2009 Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation Cell 136 : 688 - 700.

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

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

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

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

E.J. Steele E.J., and R.V. Blanden. 2001 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 : 61 - 66.

Steele, E.J., R.A. Lindley, J. Wen, and G.F. Weiller. 2006 Computational analyses show A-to-G mutations correlate with nascent mRNA hairpins at somatic hypermutation hotspots. DNA Repair 5 : 1346 - 1363.

Steele EJ, and Pollard JW (1987). Hypothesis : Somatic Hypermutation by gene conversion via the error prone DNA-to-RNA-to-DNA information loop. Molec. Immunol. 24 : 667 - 673.

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

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

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

Wright, B.E., K.H. Schmidt, and M.F. Minnick. 2004 Mechanisms by which transcription can regulate somatic hypermutation. Genes Immun. 5 : 176 - 182.

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

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

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