Commentary for *Journal Carcinogenesis & Mutagenesis*
Submitted 18 November 2020

Presumptive evidence for ADAR1 A-to-I deamination at WA-sites as the mutagenic genomic driver in hepatocellular and related ADAR1-Hi cancers

Robyn A. Lindley [1,2]
Edward J. Steele [3,4*]

[1] GMDx Group Ltd, Melbourne Vic, AUSTRALIA
[2] Department of Clinical Pathology
Faculty of Medicine, Dentistry & Health Sciences, University of Melbourne
Vic, AUSTRALIA
[3] Melville Analytics Pty Ltd, Melbourne, Vic, 3004, AUSTRALIA
[4] C. Y. O'Connor Foundation, 24 Genomics Rise, Piara Waters, WA 6112 AUSTRALIA

***Correspondence**:
Edward J Steele, Corresponding Author
[2] Melville Analytics Pty Ltd. Level 2, 517 Flinders Lane, Melbourne, VIC 3000 Australia
Tel: +61 (0) 420 863 551
emails: e.j.steele@bigpond.com and ejsteele@cyo.edu.au

Running Head: *ADAR1 RNA editing signature in cancer genomes*

Abstract

In hepatocellular cancer (HCC) there is an over expression of the RNA editing enzyme ADAR1. Further, the prominent genomic somatic mutation signature in HCC is almost exclusively focused on mutations at A:T base pairs where A-to-G mutations far exceed T-to-C mutations (when read on the non-transcribed strand). A clear mechanism for this extreme transcriptional strand biased mutation signature, putatively associated with over expression of ADAR1 deaminase, is yet to be explicitly demonstrated. The standard description of this strand bias has been nominally called "Transcription Coupled Damage" (TCD) to distinguish it from more conventional "Transcription Coupled Repair" (TCR). We show that the TCD description does not satisfy all features of the molecular evidence. The conventional view is that ADAR1 is thought to target adenosines at WA-sites for editing to inosine (I) in double stranded RNA stem-loop structures in transcripts. Here we show that the totality of the molecular and cellular data on these mutation signatures provides strong presumptive evidence for a clear role for ADAR1-mediated A-to-I deamination at WA-sites as the mutagenic driver in hepatocellular and possibly other related ADAR1-Hi cancers displaying biased mutation features at A:T base pairs.

Key words: ADAR1 deaminase; A-to-G versus T-to-C strand biased signature; Hepatocellular cancer; DNA polymerase-η; reverse transcription; somatic hypermutation

For a number of years we have been marshalling evidence in support of the reverse transcriptase mechanism of somatic hypermutation (SHM) in rearranged immunoglobulin (Ig) genes [1-4] and in non-Ig genes undergoing "dysregulated" SHM Ig-like responses in non-lymphoid cancer genomes [3, 4-9]. The most compelling data from cancer genomes are from those mutagenic processes targeting the DNA binding region of the TP53 gene in a variety of common cancers including those of breast, lung, bladder, ovary curated in the IARC TP53 database [6,7].

The main components of the reverse transcriptase mechanism following initiation by AID of cytosine to uracil (C-to-U) DNA deamination lesions in genomic DNA [10,11] are:
1). DNA polymerase-η is the sole error-prone DNA polymerase targeting A:T base pairs at W<u>A</u>-sites during SHM *in vivo* [12,13] and also an efficient cellular reverse transcriptase (RT) together with potential back-up RT activity from Y family relative DNA polymerase-κ [1,4,14,15].
ii). The signature for ADAR deaminase A-to-I(G) by the RNA editor ADAR1, also targeting W<u>A</u>-sites, where the edits appear in pre-mRNA co-transcriptionally [16,17]. Indeed, the analyses show that ADAR1 introduces A-to-I(G) modifications in the nascent pre-mRNA in post-transcriptional DNA [2,6]. Also recently, such editing events have been identified in the RNA and DNA moieties of RNA: DNA hybrids *in vitro* [18], implying that adenines in both RNA and DNA moieties of RNA:DNA hybrids are A-to-I (G) edited at stalled transcription bubbles *in vivo* [3].

These findings together infer a transcription-linked model and evoking a RNA to DNA copying process back into genomic DNA. That is, involving an implied mRNA intermediary by a reverse transcription (RT) step at the genomic target site.

Consistent this conclusion it has been found that amongst several strand-biased *in vivo* mutation signatures observed in Ig SHM [19] and in Ig SHM-like patterns in cancer genomes [5-7] the somatic mutations are found at W<u>A</u> motifs. These are *always strand biased* with A-to-G mutations exceeding T-to-C mutations by 2 to 3 fold (mutations read by convention 5' to 3' on the non-transcribed strand). A-to-C and A-to-T transversion mutations also exceed their T-to-G and T-to-A Watson-Crick pyrimidine complements. More recently, genome-wide analyses have also confirmed wide-spread W<u>A</u>-site/DNA Polymerase-η associated mutation signatures in pan-cancer genomes [20,21].

We ask: Does evidence already exist for a strong signal of ADAR1 A-to-I(G) RNA editing in the DNA of ADAR1 over-expressing hepatocellular carcinomas? In these (and some related) cancers it has been

established that ADAR1 over-expression is the cause of significantly elevated A-to-I RNA editing and ADAR1 acts as an oncogene promoting cancer progression causing increased risk of liver cirrhosis and postoperative recurrence with poor prognoses [22]. The same appears to be happening in gastric cancer [23]. The cause of this over expression could be ADAR1 gene amplification at 1q21 as demonstrated in other cancers [24].

A striking feature of genome-wide sequencing of cancer genomes [25] are the single base substitution (SBS) hepatocellular mutations annotated as signatures 12 and 16 which exhibit an extremely strong transcriptional strand bias with A-to-G mutations greatly exceeding T-to-C mutations mainly at WA motifs, about 5 fold higher in the case of signature 16 (see the updates in [26] and at the COSMIC Signatures website: https://cancer.sanger.ac.uk/cosmic/signatures.) Note that SBS signatures 12 and 16 have also been identified in some other cancers (Kidney-RCC, Biliary-AdenoCA, Prost-AdenoCA, Head -SCC). A similar A:T focused yet now reversed orientation strand-biased signature in SBS 21 is evident in stomach cancers [25], suggesting to us the interpretation [3] that A-to-I deamination has occurred directly on the transcribed DNA strand at the annealed RNA:DNA hybrid in these cancers. However with respect to the striking adenine-targeted strand bias on the non-transcribed strand in HCC Alexandrov et al [25] discuss the striking strand biases as follows:

> " .. On the assumption that the transcriptional strand biases in signatures 12 and 16 are introduced by transcription-coupled NER, these currently unexplained signatures may be the result of bulky DNA helix-distorting adducts on adenine. However, there is no previous basis for invoking transcription-coupled NER in the genesis of these signatures and other causes of transcriptional strand bias may exist".

This is a key qualification. The formal explanations at the COSMIC website are (author addition and emphasis in italics):

> • For signature 12 (SBS12): " Transcriptional strand-bias … with more mutations of A than T on the untranscribed strands of genes consistent with damage to adenine and repair by transcription-coupled nucleotide excision repair" *[27]* and ,
>
> • For signature 16 (SBS16): " Extremely strong transcriptional strand bias with more mutations of A than T (*by five fold and mainly A-to-G exceeding T-to-C*) on the untranscribed strands of genes consistent with damage to adenine and *differential repair of the transcribed strand* by transcription coupled nucleotide excision repair *[27]* . There is also evidence for SBS16 of transcription coupled DNA damage, with more damage to A on untranscribed than on transcribed strands (in addition to transcription coupled nucleotide excision repair)." The latter conclusion is to a type of mechanism we infer in the later work of Haradhvala et al [28].

It needs to be noted here that there is a preference in both signatures for A-to-G mutations to occur at W<u>A</u>-sites where the 5' nucleotide is A or T i.e. consistent with the well known ADAR signature motif W<u>A</u>, where W = T or A).

Thus, Haradhvala et al. [28] have asserted that adenines are somehow specifically targeted for damage during transcription by a process that they refer to as "Transcriptional Coupled Damage" (TCD). They elucidate further by adding " we propose that this is due to two separate processes operating on different strands-TCD and TCR (*transcription-coupled repair*) ..This explains the extreme transcriptional asymmetry of liver A/G compared to other signatures … these results suggest that the A/G signature is caused by a mutational process distinct from typical bulky-adduct damage. Finally, we noticed that one colorectal patient (''CRC-8'') from an earlier study of nine colorectal whole genomes … showed the same signature of TCD. Thus, this phenomenon may be enriched in liver but not exclusive to it."

Thus, their explanation is that the extreme bias of A-to-G over T-to-C mutations in hepatocellular cancer is caused by some unspecified damage to adenine on the non-transcribed strand (NTS) coupled to transcription itself. This targeting is over and above any differential Transcription Coupled Repair that preferentially clears bulky adducts [27] of the template or transcribed strand (TS) leaving an excess of unrepaired A-to-G mutations on the displaced coding or NTS. They call this Transcription Coupled Damage (restricted to adenines) but suggest no plausible molecular mechanism to explain the data as it is assumed to be a part of the TCD signature.

We find that our own data analyses are consistent with the Alexandrov et al. [25] and Haradhvala et al.[28] empirical observations. However, we now offer a molecular explanation for the strong strand biases reported. From our perspective, the strand biased signatures at W<u>A</u>-sites, where mutations of A-to-G far exceed mutations of T-to-C by some factors of 4-fold or more, are expected and readily explained by the ADAR1/Pol-η coupled target-site reverse transcriptase (TSRT) mechanism premised on a DNA-to-RNA-to-DNA transcription linked process [1-4]. Given that several other cancers also display ADAR1 over-expression e.g. gastric cancers [23] and lung adenocarcinomas [24, 29], this new interpretation of the causative molecular mechanisms in ADAR1-over expressing cancers is worth investigating from the point of view of therapeutic targeting of the presumptive causative deaminase, ADAR1. Our explanation is also consistent in part with the recent work by Tasakis et al. [30] for ADAR1 driving mutagenic progression in multiple myeloma at larger R-loops across the cancer genome. In this case the A-to-I(G) editing on the transcript appears also to generate A-to-I (G) mutations in the transcribed (-) DNA strand, identifying as T-to-C mutations on the reference (+) strand. In conclusion we believe that our interpretation is also consistent with, and is an extension of, the DNA

damage and repair mechanisms proposed by Haradhvala et al. [28] to explain the observed mutational strand asymmetries of APOBEC signatures in cancer genomes.

References

1. Franklin A, Milburn PJ, Blanden RV, Steele EJ. (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.

2. Steele EJ, Lindley RA, Wen J, Weiller GF. (2006) Computational analyses show A-to-G mutations correlate with nascent mRNA hairpins at somatic hypermutation hot- spots, DNA Repair. 5,1346–1363.

3. Steele EJ, Lindley RA. (2017) ADAR deaminase A-to-I editing of DNA and RNA moieties of RNA:DNA hybrids has implications for the mechanism of Ig somatic hypermutation. DNA Repair 55, 1-6.

4. Franklin A, Steele EJ, Lindley RA. (2020) A proposed reverse transcription mechanism for (CAG)n and similar expandable repeats that cause neurological and other diseases. Heliyon 6 , e03258.

5. Steele EJ, Lindley RA. (2010) Somatic mutation patterns in non-lymphoid cancers resemble the strand biased somatic hypermutation spectra of antibody genes DNA Repair. 9, 600–603.

6. Lindley, RA. (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.

7. Lindley RA, Steele EJ. (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 2013; Article ID 921418, 18 pages. https://www.hindawi.com/journals/isrn/2013/921418/

8. Lindley RA, Humbert P, Larmer C, Akmeemana EH, Pendlebury CRR.(2016) Association between targeted somatic mutation (TSM) signatures and HGS-OvCa progression. Cancer Med. 5, 2629–2640.

9. Lindley RA. (2020) Review of the mutational role of deaminases and the generation of a cognate molecular model to explain cancer mutation spectra. Med. Res. Arch. 8 (8). . https://esmed.org/MRA/mra/article/view/2177/193545632

10. Di Noia JM, Neuberger MS (2007) Molecular mechanisms of somatic hypermutation. Annu. Rev. Biochem. 76,1 - 22.

11. Teng G, Papavasiliou FN. (2007) Immunoglobulin somatic hypermutation. Annu. Rev. Genet. 41, 107–120.

12. Zeng X, Winter DB, Kasmer C, Kraemer KH, Lehmann AR, et al. (2001) DNA polymerase-η as an A-T mutator in somatic hypermutation of immunoglobulin variable genes, Nat. Immunol. 2, 537–541.

13. Delbos F, Aoufouchi S, Faili A, Weill J-C, Reynard C-A. (2007) DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. J Exp Med. 204, 17-23.

14. Su Y, Egli M, Guengerich F.P. (2017) Human DNA polymerase η accommodates RNA for strand extension. J Biol Chem. 292, 18044-18051.

15. Su Y, Ghodke PP, Egli M, Li L, Wang Y, et al. (2019) Human DNA polymerase η has reverse transcriptase activity in cellular environments. J Biol Chem. 294, 6073-6081.

16. Herbert A, Alfken J, Kim Y-G, Mian IS, Nishikura K, Rich A. (1997) A Z-binding domain present in the human editing enzyme: double-stranded RNA adenosine deaminase. Proc. Natl. Acad. Sci. U. S. A. 94, 8421–8426.

17. Herbert A, Rich A. (2001) The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. Proc. Natl. Acad. Sci. U. S. A. 98, 12132–12137.

18. Zheng YC, Lorenzo C, Beal PA. (2017) DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA, Nucleic Acids Res. 45, 3369–3377.

19. Steele EJ. (2009) Mechanism of somatic hypermutation: critical analysis of strand biased mutation signatures at A:T and G:C base pairs, Mol. Immunol. 46, 305–320.

20. Supek F, Lehner B. (2017) Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. Cell. 170, 534-547.e23.

21. Rogozin IB, Goncearenco A, Lada AG, De S, Yurchenko V, et al. (2018) DNA polymerase η mutational signatures are found in a variety of different types of cancer. Cell Cycle. 17, 348-355

22. Chan TH, Lin CH, Qi L, Fei J, Li Y, et al. (2014) A disrupted RNA editing balance mediated by ADARs (Adenosine DeAminases that act on RNA) in human hepatocellular carcinoma. Gut. 63, 832-843.

23. Chan TH, Qamra A, Tan KT, Guo J, Yang H, et al. (2016) ADAR-mediated RNA editing predicts progression and prognosis of gastric cancer. Gastroenterology. 151, 637-650.e10.

24. Amin EM, Liu Y, Deng S, tan KS, Chudgar N, et al. (2017) The RNA-editing enzyme ADAR promotes lung adenocarcinoma migration and invasion by stabilizing FAK. Sci Signal. 10(497). Pii, eaah3941.

25. Alexandrov LB, Nik-Zainal S, Wedge DC,Aparicio SAJR, Behjati S, et al. (2013) Signatures of mutational processes in human cancer Nature 500: 415-421.

26. Alexandrov LB, Kim J, Harahvala NJ, Huang MN, Ng AWT, et al. (2020) The repertoire of mutational signatures in human cancer. Nature 578, 94-101.

27. Hanawalt PC, Spivak G. (2008) Transcription-coupled DNA repair: two decades of progress and surprizes. Nat. Rev. Mol. Cell Biol. 9, 958-970.

28. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, et al. (2016) Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. Cell 164, 538–549

29. Sharpnack MF, Chane B, Aran D, Kosti I, Sharpnack DD, et al. (2018) Global Transcriptome analysis of RNA abundance regulation by ADAR in Lung Adenocarcinoma. EBioMedicine. 27, 167-175.

30. Tasakis RN, Laganà A, Stamkopoulou D, Melnekoff DT,Nedumaran P, et al., (2020) ADAR1 can drive Multiple Myeloma progression by acting both as an RNA editor of specific transcripts and as a DNA mutator of their cognate genes. https://doi.org/10.1101/2020.02.11.943845
bioRxiv preprint doi.