


COMMENTARY

Open Access



# Integration of SARS-CoV-2 RNA in infected human cells by retrotransposons: an unlikely hypothesis and old viral relationships

Nicole Grandi<sup>1\*</sup> , Enzo Tramontano<sup>1,2</sup> and Ben Berkhout<sup>3\*</sup>

## Abstract

Zhang et al. (Proc Natl Acad Sci 118:e2105968118, 2021) recently reported that SARS-CoV-2 RNA can be retrotranscribed and integrated into the DNA of human cells by the L1 retrotransposon machinery. This phenomenon could cause persistence of viral sequences in patients and may explain the prolonged PCR-positivity of SARS-CoV-2 infected patients, even long after the phase of active virus replication has ended. This commentary does critically review the available data on this topic and discusses them in the context of findings made for other exogenous viruses and ancestral endogenous retroviral elements.

**Keywords:** SARS-CoV-2, COVID19, LINE, L1, Retrotransposition, HERV, HERV-W, Chimeric RNA

## Text body

The COVID pandemic that started at the end of 2019 led to a remarkable mobilization of scientific efforts as evidenced by >175.000 publications to date. Among these, the work by Zhang et al. triggered an animated debate in the scientific community [1]. Based on studies performed in cultured cells transfected with DNA encoding the retrotransposon L1 (long interspersed nuclear elements 1), authors proposed that SARS-CoV-2 RNA, in particular the subgenomic RNA encoding the nucleocapsid (NC), can be converted into dsDNA and integrated into the cellular genome by the L1 retrotransposition machinery [1]. These SARS-CoV-2 sequences can be expressed in patients as chimeric cellular-viral transcripts, which could explain the long-term PCR positivity for viral RNA in patients who recovered from COVID. A similar

hypothesis was proposed by Yin and co-workers who observed that infection by SARS-CoV-2 (as well as other human coronaviruses) causes upregulation of retrotransposon expression, leading to the formation of chimeric virus-retrotransposon transcripts [2].

These original reports opened a heated debate on the correctness of the findings and their relevance for recovered COVID patients and subsequent work was initiated to test alternative explanations. It was proposed early on that the observed chimeric RNAs could be artifacts generated during cDNA library preparation. Two findings hint at this possibility. First, the directionality of the observed chimeric transcripts, in which a large fraction of SARS-CoV-2 RNA derives from the (–) strand, in contrast to the predominance of (+) strand RNAs in SARS-CoV-2 natural infection. Second, the absence of the 3' end and polyA tail of the viral genome, which are commonly present in integrated sequences processed by L1 elements.

The origin of the chimeric human-SARS-CoV-2 reads in RNA-seq libraries was subsequently investigated in a dedicated study [3], which showed that such hybrid sequences arose also between SARS-CoV-2 RNA and

\*Correspondence: nicole.grandi@unica.it; b.berkhout@amsterdamumc.nl

<sup>1</sup> Laboratory of Molecular Virology, Department of Life and Environmental Sciences, University of Cagliari – Cittadella Universitaria di Monserrato, SS554 Monserrato (Cagliari), Italy

<sup>3</sup> Laboratory of Experimental Virology, Department of Medical Microbiology, Amsterdam Medical Centers, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

transcripts encoded by mitochondrial DNA or episomal adenoviral DNA in transfected cells, thus being unlikely the result of genuine SARS-CoV-2 integration. Other studies focused on detecting SARS-CoV-2 retrotransposition events in deep sequencing data, confirming the absence of genuine L1-mediated integration events and suggesting that the observed chimeric transcripts had emerged during RNA-seq library construction [4–7]. Importantly, such chimeric reads were also identified when RNA from infected human cells was mixed before library preparation with RNA from uninfected or unrelated vertebrate cells [4, 7]. In addition, the lack of reproducibility of the observed host-virus chimeric transcripts across SARS-CoV-2 patient samples corroborated the idea that these sequences arose from stochastic, artifactual events at the RNA-seq level (e.g. random ligations, template switching and/or sequence alignment errors). Consistent with this notion, the chimeric reads are mostly composed of abundantly expressed cellular and viral transcripts. Together, the collective evidence for genuine SARS-CoV-2 DNA formation and integration remains sparse [7].

To put these recent reports in a broader context, some consideration should be given to the molecular biology of L1 elements and their interplay with viruses. L1 elements represent the most abundant subfamily of non-LTR retrotransposons, accounting for ~17% of the human genome. L1 elements are autonomous for self-mobilization by encoding two proteins (ORF1 and ORF2) that together mediate reverse transcription of their own RNA and subsequent integration of the resulting dsDNA in the cellular genome [8]. This process shows some cross-activity on non-autonomous retrotransposons. Despite the accumulation of inactivating mutations, a subset of 80–100 L1 elements remains active in the human genome. Accordingly, L1 retrotransposition has been observed at early stages of embryonic development, and >100 *de novo* L1 insertions have been linked to heritable genetic disorders [9]. Beyond the germ line and pluripotent stem cells, L1 activity has been reported at the somatic level in neuronal progenitors and various human tumors, possibly being responsible for mutagenic events [9]. For these reasons, L1 elements are intensively being studied in diverse diseases and they were reported to be upregulated in different pathologies and especially cancer. However, there is no direct evidence for retrotransposition as a cause of disease. This also holds true for the multi-step process of tumorigenesis, where putative LINE contributions could be due to indirect effects, e.g. by non-specific epigenetic changes in cancer cells.

There are few reports on L1-mediated mobilization of viral transcripts. In hepatocellular carcinoma (HCC) induced by hepatitis B virus (HBV), recurrent integration

of HBV subgenomic RNAs was reported to yield a chimeric long non-coding RNA between the HBV mRNA for the X antigen (HBx) and L1 RNA in >23% of patient samples [10]. Of note, this HBx-L1 chimeric RNA is reported to promote malignant transformation and hepatic injury [11]. Unlike for retroviruses, integration is not a mandatory step in the HBV replication cycle and the mechanism of HBV integration in HCC cells remains poorly characterized. The observation that ~90% of HBV-induced HCC cells contain at least one integrated HBV-DNA fragment, combined with their preferential localization in or near repetitive elements, could cautiously suggest a possible role of L1 elements in the mobilization of short HBV transcripts [12]. This scenario is consistent with the fact that HBV replication occurs in the nucleus and is corroborated by the presence of HBV-integrations in most HCC samples, whose abundance seems to negatively correlate with patient survival [12]. Of note, ~40% of viral breakpoints observed upon HBV integration are restricted to an 1800-bp genome portion including the viral enhancer, X gene and core gene, which may contain features that are recognized by the L1 machinery. Perhaps a coincidence, but the size of the above mobilized HBV genome portion is comparable to that of the mobilized SARS-CoV-2 RNA fragment (1,662 bp) reported by Zhang et al. [1]. Specific breaks in the viral genome also occur during SV40-BK virus oncogenesis, leading to upregulated expression of the viral oncogene. It is important to stress that integration as detected in tumor cells does NOT occur during normal virus replication.

The most remarkable case of L1-virus interplay does however not involve “modern” human viruses, but rather a group of human endogenous retroviruses (HERVs) that were acquired by the primate genome some 20–43 million years ago through infection of the germ line by now-extinct retroviruses [13]. The hallmark of all retroviruses is reverse transcription of their RNA genome into dsDNA that integrates in the genome of the infected cells. Hence, germline integration of these ancestral retroviruses allowed their inheritance as Mendelian genes and vertical transmission to the offspring. HERV retrotransposons currently constitute ~8% of our genome and have occasionally been used to develop novel and important physiological processes like placenta formation [14, 15]. The HERV-W group is unique for its colonization dynamics: among the 213 members, 135 (63%) are not direct retroviral integrations, but rather processed pseudogenes that were generated through mobilization of HERV-W transcripts by the L1 machinery [13, 16]. Only this HERV group shows such L1-dependency, although the determinants for the specific interaction with L1 remain unclear. Sequence analyses indicated that mobilization is 2.5-fold

more efficient for subgroup 1 HERV-W members, suggesting the presence of preferential sequence signatures for L1 recognition [13]. Besides retroviruses, which have reverse transcription and integration as a stable biological feature, an example of human endogenous viral elements (EVEs) that have likely involved L1 in their formation are the bornavirus-like elements, i.e. the only non-retroviral RNA virus-derived EVEs [17]. This scenario is supported by the fact that most of such elements originate from reverse-transcription and integration of the mRNA coding for ancient bornavirus nucleoprotein, with genomic localization and flanking sequences being consistent with L1 action [18].

## Conclusions

Overall, it seems unlikely that the L1 machinery is responsible for “genuine” SARS-CoV-2 genomic integration in infected cells. In fact, this process of *trans*-mobilization has been infrequent even for exogenous retroviruses that use integration as a key feature of their replication cycle. Accordingly, despite the evidence that the primate DNA genome has been invaded repeatedly by exogenous retroviral infections during evolution, only a single HERV group was copied-and-pasted by the L1 machinery, and the molecular signatures that facilitate viral RNA retrotransposition by the L1 apparatus are still poorly defined. It is known that SINE non-autonomous retrotransposons exploit the L1 machinery for retrotransposition, which is based on 3' end sequence similarity between LINEs and SINEs [8]. Given the nucleotide sequence diversity of L1 3' ends and the candidate viral sequences that were mobilized by L1 (e.g. HERV-W pseudogenes and HBx mRNA), it seems likely that retrotransposition involved the recognition of RNA secondary structures and other spatial features instead of a specific sequence. The cell type in which integration occurs may also have a considerable impact. Whereas HERV-W pseudogenes are formed in germ line cells, which are known to have high L1 physiological activity, HBV and possibly SARS-CoV-2 integrations were described in somatic cells. This still-unveiled selectivity of L1 mobilization towards certain viral transcripts, in addition to the concrete possibility that chimeric SARS-CoV-2/cellular RNAs can artefactually arise during the amplification/sequencing procedures, remain major confounding factors in the characterization of putative *de novo* retrotransposition events of SARS-CoV-2 mRNA.

Further studies are necessary to assess the actual impact of active retrotransposons in the mobilization of viral and host transcripts and to characterize the molecular mechanisms underlying their integration in the host genome and their subsequent expression.

## Abbreviations

NC: Nucleocapsid; L1: Long interspersed nuclear elements 1; HCC: Hepatocellular carcinoma; HBV: Hepatitis B virus; HBx: Hepatitis B virus X antigen; HERVs: Human endogenous retroviruses; EVEs: Endogenous viral elements.

## Acknowledgements

Nothing to declare.

## Authors' contributions

BB conceived the study. NG, ET and BB designed the study. NG collected the data and was a major contributor in writing the manuscript. BB and ET participated in the writing. All authors revised the manuscript and approved the final version.

## Funding

Nothing to declare.

## Availability of data and materials

Not applicable.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Laboratory of Molecular Virology, Department of Life and Environmental Sciences, University of Cagliari – Cittadella Universitaria di Monserrato, SS554 Monserrato (Cagliari), Italy. <sup>2</sup>Istituto di Ricerca Genetica e Biomedica (IRGB), CNR – SS554, Monserrato (Cagliari), Italy. <sup>3</sup>Laboratory of Experimental Virology, Department of Medical Microbiology, Amsterdam Medical Centers, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands.

Received: 30 September 2021 Accepted: 19 October 2021

Published online: 29 October 2021

## References

- Zhang L, Richards A, Barrasa MI, Hughes SH, Young RA, Jaenisch R. Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proc Natl Acad Sci*. 2021;118:e2105968118.
- Yin Y, Liu XZ, He X, Zhou LQ. Exogenous coronavirus interacts with endogenous retrotransposon in human cells. *Front Cell Infect Microbiol*. 2021;11:1–11.
- Kazachenka A, Kassiotis G. SARS-CoV-2-host chimeric RNA-sequencing reads do not necessarily arise from virus integration into the host DNA. *Front Microbiol*. 2021. <https://doi.org/10.3389/fmicb.2021.676693>.
- Chen Y-S, Lu S, Zhang B, Du T, Li W-J, Lei M, et al. Comprehensive analysis of RNA-seq and whole genome sequencing data reveals no evidence for SARS-CoV-2 integrating into host genome. *Protein Cell*. 2021. <https://doi.org/10.1007/s13238-021-00861-8>.
- Parry R, Gifford RJ, Lytras S, Ray SC, Coin LJM. No evidence of SARS-CoV-2 reverse transcription and integration as the origin of chimeric transcripts in patient tissues. *Proc Natl Acad Sci*. 2021;118:e2109066118.
- Smits N, Rasmussen J, Bodea GO, Amarilla AA, Gerdes P, Sanchez-Luque FJ, et al. No evidence of human genome integration of SARS-CoV-2 found by long-read DNA sequencing. *Cell Rep*. 2021;36:109530.
- Yan B, Chakravorty S, Mirabelli C, Wang L, Trujillo-Ochoa JL, Chauss D, et al. Host-virus chimeric events in SARS-CoV-2-infected cells are infrequent and artifactual. *J Virol*. 2021. <https://doi.org/10.1128/JVI.00294-21>.
- Ferrari R, Grandi N, Tramontano E, Dieci G. Retrotransposons as drivers of mammalian brain evolution. *Life*. 2021;11:1–29.

9. Suarez NA, Macia A, Muotri AR. LINE-1 retrotransposons in healthy and diseased human brain. *Dev Neurobiol.* 2018;78:434–55.
10. Lau CC, Sun T, Ching AKK, He M, Li JW, Wong AM, et al. Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell.* 2014;25:335–49.
11. Liang HW, Wang N, Wang Y, Wang F, Fu Z, Yan X, et al. Hepatitis B virus-human chimeric transcript HBx-LINE1 promotes hepatic injury via sequestering cellular microRNA-122. *J Hepatol.* 2016;64:278–91.
12. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet.* 2012;44:765–9.
13. Grandi N, Cadeddu M, Blomberg J, Tramontano E. Contribution of type W human endogenous retroviruses to the human genome: characterization of HERV-W proviral insertions and processed pseudogenes. *Retrovirology.* 2016;13:67.
14. Grandi N, Tramontano E. HERV envelope proteins: Physiological role and pathogenic potential in cancer and autoimmunity. *Front Microbiol.* 2018;9:462.
15. Grandi N, Tramontano E. Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. *Front Immunol.* 2018;9:1–16.
16. Pavlíček A, Paces J, Elleder D. Processed pseudogenes of human endogenous retroviruses generated by LINEs: their integration, stability, and distribution. *Genome Res.* 2002;12:391–9.
17. Honda T. Potential links between hepadnavirus and bornavirus sequences in the host genome and cancer. *Front Microbiol.* 2017;8:2537.
18. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, et al. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature.* 2010;463:84–7.

### Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

