Review of: "SARS-CoV-2 RNA reverse-transcribed and integrated into the human genome"

Pascal Pineau[1]

1 Pasteur Institute

Potential competing interests: The author(s) declared that no potential competing interests exist.

In this manuscript, L. Zhang and coworkers exposed the outcome of experiments consisting in analyzing the capacity of SARS-CoV-2 RNA to integrate into human chromosomes. Data presented can be divided into two categories: (i) in silico data consisting of RNA seq analysis of SARS-CoV-2 infected cells, and (ii) in vitro experiments performed on infected cell lines.

The issue of RNA virus genome retro-transcription and integration is very important in the case of SARS-CoV-2 due to the novel methods of immunization employed in Humans on an unprecedented scale. Furthermore, it came in a controversial context where apparently poorly informed/educated Medical Doctors (mostly clinicians though) were invited by the media to confirm that retro-transcription of non-Retroviridae viral RNA was “impossible” (sic). This is, of course, absolutely untrue. The most famous example concerns the Bornaviridae, a viral family with a negative RNA genome capable to persist in the cell nucleus. There are three sequences of Bornaviruses in the genome of H. sapiens.

Therefore, there is no conceptual opposition to conducting the investigation presented by L. Zhang and colleagues.

Major Question

The authors consider that LINE1 is instrumental in the process of SARS-CoV-2 RNA integration. In 2020, the Pan-Cancer Analysis of Whole Genome consortium published a very nice typology of the LINE1 active in human cancer (Nature, 2020, 578:82, figure 6d). Are the authors capable to tell the reader whether some of these active (volcanic) LINE1 is also involved in SARS-CoV-2 retrotransposition?

The Bornavirus case although it supports the hypothesis of Zhang and coworkers differs in at least two aspects: Bornavirus is present in the nucleus of infected cells and retrotransposition obviously occurs in germ cells (the reason why all members of our species and even all primates, in fact, bear it in their genome). Is there a previous observation of SARS-CoV-2 RNA in the nucleus? Are there RNA seq from SARS-CoV-2 infected testicles available to check if chimeric reads are more abundant there?

Is there a correlation between reads of LINE1 and those of chimeric SARS-CoV-2 in all data sets investigated.?

The authors focused exclusively on the N gene. Although N-encoding subgenomic RNA is the most abundant in SARS-CoV-2-infected cells, all SARS-CoV-2 RNA contains by definition the N sequence. Thus, there is no reason to think that N is the only region that could be integrated. Figure 1d suggests that there is maybe also, albeit to a lesser extent, the M and 7a genes that might be found under an integrated form. Why not look at these genes as well?

I suppose that the N gene once integrated presents some sites targetable by restriction enzymes. Did the authors try to

cut the integrated version of the N gene with restriction endonucleases and amplify it afterward? It should not work. I could be a cheap and demonstrative control that we are looking at DNA.

The authors extract large genome fragments from gels to focus on the integrated sequences. This is very nice but the SARS-CoV-2 genome is 30Kb. Although I have to admit that it is highly improbable, a fully reverse-transcribed genome should be found precisely in this high molecular weight DNA. Is PCR outcome without DNA gel extraction different from that obtained after gel extraction?

Minor issues:

Line 39: “no replication-competent virus was isolated or spread from these PCR positive patients”: I am not sure that this argument is solid enough. Viral isolation through infection of cell culture is in general logarithmically less efficient than molecular technique (the reason why the latter have been developed). And it is especially true when viral loads are low as in the case of persistent/long-duration SARS-CoV-2 shedding.

Line 141: was it the RT of HIV or a true HIV co-infection. I am confused

Line 147: “SRAS-CoV-2”, SRAS is in French.