A perplexing mystery since the early days of the COVID-19pandemic has been the recurrence of positive SARS-CoV-2 PCR tests in patients who have long recovered from their initial infection and have not been re-exposed to the virus.

Two eminent biology professors at the Whitehead Institute and the Massachusetts Institute of Technology (MIT)—Rudolf Jaenisch, PhD, and Richard Young, PhD, both members of the National Academy of Sciences—were struck by this curious phenomenon.

Rudolf Jaenisch, PhD, professor of biology, Whitehead Institute, MIT

"Since both of us have a background in retroviruses, which integrate into the genome as part of their normal life cycle, we hypothesized that the SARS-CoV-2, which does not integrate into the genome as part of its normal life cycle, could be hijacked by retroviral-like transposable elements and get integrated. This could give rise to long-term expression of viral sequences, detectable by PCR, in the absence of infectious virus," Jaenisch told *GEN* in an exclusive interview.

The resulting study, recently published in the *Proceedings of the National Academy of Sciences*—"Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues"—builds on key questions that the team probed as part of an earlier preprint posted on *bioRxiv*, in their attempts to find a solution to this mystery.

Using three independent sequencing approaches, Jaenisch and his team showed the presence of human–viral chimeric transcripts in infected human HEK293T cells in culture and

patient-derived tissues to demonstrate that DNA copies of fragments of SARS-CoV-2's genomic RNA sequences can integrate into the human genome and can be transcribed into RNA.

At least one of the mechanisms by which this integration occurs is through LINE1 elements—autonomous retrotransposons that can insert themselves (and other sequences) into genomic sites through the back-coding of DNA from an RNA template (reverse transcription). The Whitehead authors showed the presence of consensus or variant LINE1 recognition sites in a majority of human DNA sequences that flank the integrated viral sequences.

These novel findings, no doubt compounded by the pressures of the pandemic, have generated a heated debate in the scientific community.

Gaetan Burgio, MD, PhD, scientist at the Australian National University, has co-led a consortium that has unsuccessfully tried to replicate earlier work from Jaenisch's team.

Burgio told *GEN*: "The virus segments inserted into the human genome are rather atypical… some sort of random virus fragments that have attached to a human genome, leading me to think that these sequences are artifactual from the library preparation, contamination during PCR process and sequencing. For example, the polyA tail of the viral genome is missing in these sequences and none of the 3' end of the viral genome was found in these chimeric sequences. The orientation of the integration of the viral genome doesn't make sense to me."

A recently released *bioRxiv* preprint also suggested that the human SARS-CoV-2 chimeric transcripts detected through RNA sequencing may arise as a sample prep artifact rather than true reverse transcription, integration, and expression.

Ellen Foxman, MD, PhD, assistant professor in the department of laboratory medicine at Yale School of Medicine, reiterated the same concerns. "My overall take on this [*PNAS*] paper is that it is exploratory research, and that while the data suggests that fragments of the SARS-CoV-2 genome can become integrated in human cells, other explanations for the data are also possible, as the authors discussed." Foxman's lab focuses on investigating virus infections in airway passages. She commented here as an expert unrelated to the study.

Foxman added, "Exploratory basic research is very important over the long term, it helps us develop a better understanding of how our bodies and viruses work."

Jaenisch admitted that the conclusions made in the original bioRxiv preprint were "ill-fated" and the findings, although correct, were not supported by strong evidence.

"Chimeric RNA could be artifactually generated when you prepare your cDNA library for sequencing because the reverse transcriptase jumps between different templates. [In the preprint] we didn't have direct evidence to show that [viral] DNA integrated into the genome. In the new [*PNAS*] paper, we now have unambiguous evidence that these viral sequences are integrated into the genome. The most common mechanism [for this] is what's called LINE1-

mediated retroposition, coming from the footprints of the viral sequence in the genome. It's irrefutable. They can integrate."

The key question that Jaenisch and his team probe is whether viral sequences integrate and express in patients. This question can only be answered indirectly, as it is technically challenging to obtain direct evidence of rare integration events in the genomic DNA. Analyzing the "sense" or directionality of the sequences of transcribed viral RNA, which are present in greater copy numbers in cells, provides important clues.

The genomic material in SARS-CoV-2 is a positive-sense RNA strand (5'-3') that has the same polarity as viral mRNA and can be directly translated into viral protein. When SARS-CoV-2 infects and replicates in a host cell, therefore, the plus strand predominates.

"We asked how much minus strand is in [infected] cells. It's very little. Then we asked, how could minus strands arise? If viral sequences integrate into the genome they integrate in both orientations. DNA analysis shows it is 50–50," said Jaenisch.

In infected cells where the virus is actively replicating, minus-strand viral RNA counts are 1,000 or less but in patient-derived tissues where there is no clinical evidence for viral replication, up to 50% of the viral transcripts are minus strands.

"This was very convincing evidence for two conclusions: 1)

there can indeed be these viral sequences integrating in patients and 2) they can be expressed," said Jaenisch. The authors detected chimeric sequences that contain minus strand RNA in both infected human HEK293T cells and patient-derived tissues.

One of the concerns regarding evidence of viral integration obtained in human cell culture is whether the same holds true in living organisms. The scarcity of viral integration in host cells poses a technical challenge in answering this question. Jaenisch said, "We would argue human cellular RNA fused to minus strand viral RNA cannot have another explanation. There is integration. RNA is easier to detect since it produces many copies but to detect DNA, you have to see it at the single cell level and that we can't do."

Some critics contend that the physiological levels of reverse transcription machinery in human cells are very low and insufficient for cellular integration of SARs-CoV-2. In their cell line experiments, Jaenisch's team transfected HEK293T cells with LINE1 prior to infection with SARS-CoV-2 to increase the likelihood of detecting rare integration events. This has prompted critics to question the biological relevance of the detected chimeric transcripts.

Burgio said, "These integration events were found in a context of strong overexpression of LINE1 transposable elements, which is not seen in a real-life setting."

"LINE1 expression is induced in cells under stress," countered Jaenisch. "Stress can be induced by infection with a virus, by exposure cytokines, by aging, in cancer. So, you

could argue when a patient is infected with [SARs-CoV-2] virus it induces LINE1 and that promotes integration… it is a very clear possibility." The team presented evidence supporting the induction of LINE1 by virus infection. Analogous results have also been observed by other groups.

Young added: "What would be a reasonable hypothesis here is that the stress of the virus infection has elevated the level of reverse transcriptase." The Whitehead team is pursuing experiments to confirm this hypothesis.

Richard Young, PhD, professor of biology, Whitehead Institute, MIT

Whether there is any specificity to the site of integration of these viral fragments is currently unresolved, largely due to technical challenges. "These are rare integrations. You cannot clone these cells because the cells die. You have a snapshot of a population. You can't do the experiments here that you can do with retroviruses," said Jaenisch.

Young added, "Preliminary data suggests that there are many sites of integration."

Expressing concerns

Much debate centers around the possibility of these integrated viral sequences generating infectious virus and altering host DNA with deleterious consequences.

"There is no evidence presented that the extremely rare events proposed in this paper would be harmful to human health or could result in live SARS-CoV-2 viruses being produced," said Foxman.

Jaenisch agreed and emphasized, "The biggest piece of [viral] DNA we find is 5% of the viral genome, 1,600 bps. There is absolutely no way in which infectious virus can be made from these integrated sequences."

The interpretation of diagnostic PCR tests for SARS-CoV-2, however, acquire a layer of complexity in the light of these findings. "The clear conclusion is that if you are PCR positive it does not mean that you are shedding virus and that you are infectious. You really have to detect infectious virus to make that statement," said Jaenisch.

A major reason for the charged emotions regarding this study stems from the wider debate about whether mRNA vaccines could similarly integrate into human DNA with potentially deleterious consequences.

Foxman said, "A controversial result such as this one can be important in motivating new areas of research that ultimately lead to big discoveries. However, it would be a mistake to over-interpret this paper as having significance for patient care or vaccines in the current pandemic."

“There is absolutely no reason to believe that any of the vaccine mRNA is doing the same thing. The viral spike protein mRNA is a tiny piece. Vaccines are not inducing LINE element RTs,” said Young. “Vaccines are protecting against the possibility of long-term seriously debilitating diseases or death.”

The turmoil surrounding the Jaenisch/Young preprint also raises questions regarding the utility of preprints in the scientific process. “In hindsight, preprints are a good thing but to have a review discussion with the public of a paper (which I would have liked to have with the reviewers) is not something I wish to do in the future. If a paper comes out that is not peer reviewed, I don’t think I would like to have a public discussion about it. And that was a very traumatic discussion,” said Jaenisch.

Young added, “Scientific criticism is a key and valuable part of scientists coming to a view about the truth, whether that happens in the form of a preprint or later on… In an emotionally laden environment where some scientific observations are being used for political or other purposes, it is a bit of a perversion of the otherwise quite natural and useful process.”

The Whitehead team’s findings leave numerous questions unanswered. “The key question,” said Jaenisch, “that we do not have an answer to is: are these integrated sequences translated? If they are translated, they could be presented on the cell surface and provoke an immune or autoimmune reaction.” Studies have shown elevated levels of new autoantibodies in some COVID-19 patients.

The Jaenisch/Young team aims to determine which tissues are prone to express viral sequences in patients and are collaborating with Brigham and Women's Hospital to analyze human tissue samples.

Young said, "Ideas that do not conform to the current way we think about a process are very valuable in exploring a new understanding. It was not thought that these coronaviruses have the means to integrate in the genome. We are proposing that that can happen by a mechanism that is LINE1/RT mediated. It's naturally going to provoke some interesting discussion. We welcome that discussion."

"We welcome scientific discussions, not politically motivated distortions," said Jaenisch.

TAGS COVID-19 viral integration