SARS-CoV-2 RNA reverse-transcribed and integrated into the human genome

Liguo Zhang[1], Alexsia Richards[1], Andrew Khalil[2], Emile Wogram[1], Haiting Ma[1], Richard A[3], Young[1], Rudolf Jaenisch[4]

[1]Department of Whitehead Institute for Biomedical Research, Cambridge, MA, USA.
[2]Department of Engineering John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge,MA, USA.
[3]Department of Engineering wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA,USA.
[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.

ABSTRACT

Prolonged SARS-CoV-2 RNA shedding and recurrence of PCR-positive tests have been widely reported in patients after recovery, yet these patients most commonly are non-infectious1- Here we investigated the possibility that SARS-CoV-2 RNAs can be reverse-transcribed and integrated into the human genome and that transcription of the integrated sequences might account for PCR-positive tests. In support of this hypothesis, we found chimeric transcripts consisting of viral fused to cellular sequences in published data sets of SARS-CoV-2 infected cultured cells and primary cells of patients, consistent with the transcription of viral sequences integrated into the genome. To experimentally corroborate the possibility of viral retro- integration, we describe evidence that SARS-CoV-2 RNAs can be reverse transcribed in human ells by reverse transcriptase (RT) from LINE-1 elements or by HIV-1 RT, and that these DNA sequences can be integrated into the cell genome and subsequently be transcribed. Human endogenous LINE-1 expression was induced upon SARS-CoV-2 infection or by cytokine exposure in cultured cells, suggesting a molecular mechanism for SARS-CoV-2 retro-integration in patients. This novel feature of SARS-CoV-2 infection may explain why patients can continue to produce viral RNA after recovery and suggests a new aspect of RNA virus replication.

Introduction

Continuous or recurrent positive SARS-CoV-2 PCR tests have been reported in patients weeks or months after recovery from an initial infection1-14. Although bona fide re-infection of SARS-CoV-2 after recovery has been reported lately15, cohort-based studies with strict quarantine on subjects recovered from COVID-19 suggested "re-positive" cases were not caused by re-infection16,17. Furthermore, no replication-competent virus was isolated or spread from

these PCR-positive patients1-3,5,6,12. The cause for such prolonged and recurrent viral RNA production is unknown. As positive-stranded RNA viruses, SARS-CoV-2 and other beta-coronaviruses such as SARS-CoV-1 and MERS employ an RNA-dependent RNA polymerase to replicate their genomic RNA and transcribe their sub-genomic RNAs18-20. One possibility is SARS-CoV-2 RNAs could be reverse-transcribed and integrated into the human genome, and transcription of the integrated DNA copies could be responsible for positive PCR tests. Endogenous reverse transcriptase (RT) activity has been observed in human cells, and the products of reverse transcription have been shown to become integrated into the genome. For example, APP transcripts have been shown to be reverse-transcribed by

resultant APP fragments integrated into the genome of neurons and transcribed22. Human LINE- 1 elements (~17% of the human genome), a type of autonomous retrotransposons, are a potential source of endogenous RT, able to retro-transpose themselves and other non-autonomous elements such as Alu.

correlates with retro-integration

Human LINE-1 elements are autonomous retro-transposons with their encoded reverse transcriptase (ORF2p) and supporting protein (ORF1p) also aiding non-autonomous elements to retro-transpose, such as Alu and other cellular RNAs21. We found that expression of LINE-1 elements was significantly up-regulated in published RNA-Seq data of cells upon infection with SARS-CoV-2 and correlated with chimeric read abundance (Fig. 3a-b, S4a-d, compare Calu3.

Cells that are efficiently infected versus NHBE cells that are resistant to infection). Although the upregulation in Calu3 was not higher than that in NHBE, multiple LINE-1 elements were upregulated as compared to just one in NHBE (Fig. 3a, S4b, d). Expression analysis using LINE- 1 specific primers 33,34 showed a ~3-4-fold up-regulation of LINE-1 in Calu3 cells when infected by SARS-CoV-2 (Fig. 3c). Moreover, PCR analysis on Calu3 cellular DNA showed retro-integration of SARS-CoV-2 N sequences after infection (Fig. 3d-e), possibly by the activate LINE-1 reverse transcriptase. Patients infected with SARS-CoV-2 and other corona viruses show evidence of cytokine.

Induction associated with the immune response, and in severe cases experience a cytokine storm, prompting us to investigate

whether cytokines alone can induce LINE-1activation. We treated cells with cytokine-containing conditioned media from Myeloid, Microglia, or CAR- T cell cultures and found a ~2-3-fold upregulation of endogenous LINE-1 expression by PCR analysis (Fig. 3f, S5b). Expressed LINE-1 protein (ORF1p) was also confirmed by immunofluorescence staining (Fig. 3g-h, S5a). In summary, our results show induced LINE-1 expression in cells stressed by viral infection or exposed to cytokines, suggesting a molecular mechanism for SARS- CoV-2 retro-integration in human cells.

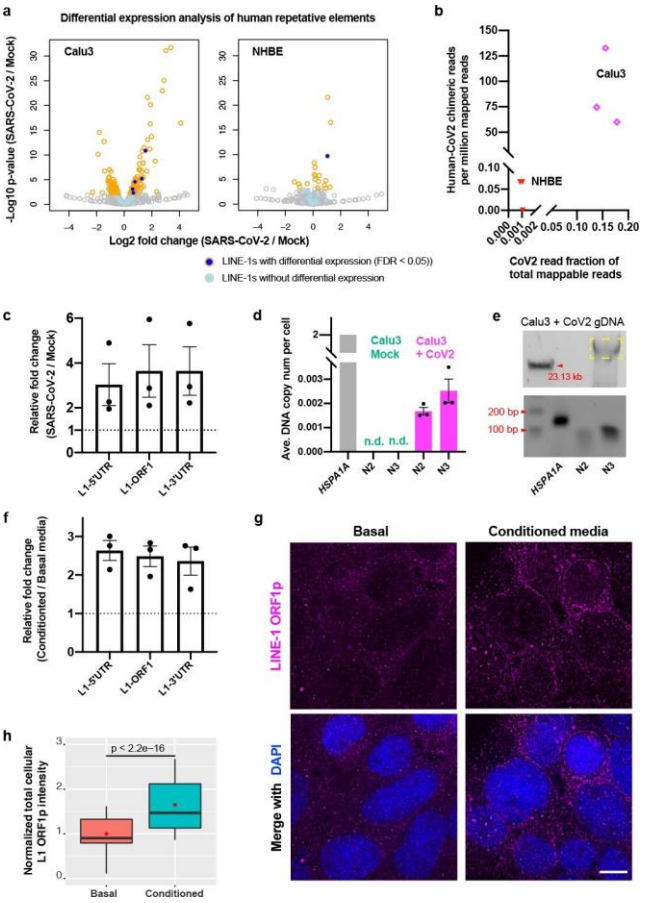

Figure 1: RNA-Seq data of cells upon infection with SARS-CoV-2

Supplementary

molecule RNA-FISH. a-b) Example images of single-molecule RNA-FISH (red/grey) targeting SARS-CoV-2 N sequence using probes shown in **Figure 2a** and merged channels with DAPI (blue) in SARS-CoV-2 infected HEK293T cells without (**a**) or with (**b**) human LINE-1 transfection. Insets in **b)**: 4x enlargement of regions in white-boxes to show nuclear signals of SARS-CoV-2 N sequence (white arrows). **c)** Comparison of nuclear N RNA-FISH signals in SARS-CoV-2 infected HEK293T cells without or with human LINE-1 transfection. Left: example images as in **a)** and **b);** Right: fraction of HEK293T cells infected by SARS-CoV-2

Figure 2: SARS-CoV-2 N RNA signals detected in cell nuclei by single.

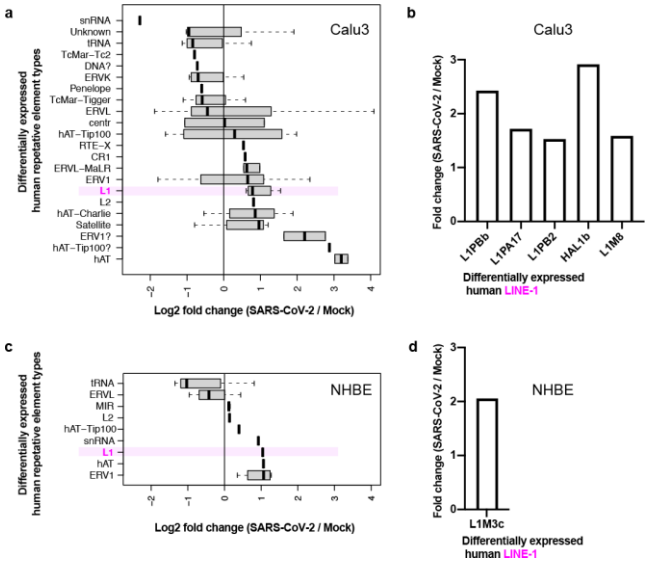

Figure 3: Primer sequences used in this study

References

1. Torres GE, Quintanilla CMR, Ruiz-EJ, Arenas R. Leishmaniasis: a review. F1000Research. 6 (F1000 Faculty Rev). 2017;750:1-15.
2. Oryan A, Akbari M. Worldwide risk factors in leishmaniasis. Asian Pac J Trop Med. 2016;9(10):925-932.
3. Salam N, Al-Shaqha WM, Azzi A. Leishmaniasis in the Middle East: incidence and epidemiology. PLOS NEGL. Trop Dis. 2014;8(10):1-8.
4. Postigo JAR. Leishmaniasis in the world health organization eastern mediterranean region. Int J Antimicrob Agents. 2010;36S: S62-S65.
5. Azizi MH, Bahadori M, Dabiri S, Meymandi SS, Azizi F. A history of Leishmaniasis in Iran from 19th century onward. Arch Iran. 2016;19(2):153-162.
6. World Health Organization (WHO). Leishmaniasis.

*Correspondence to

Liguo Zhang

Whitehead Institute for Biomedical Research Cambridge

MA

USA