No evidence of human genome integration of SARS-CoV-2 found by long-read DNA sequencing

Authors
Nathan Smits, Jay Rasmussen, Gabriela O. Bodea, ..., Daniel Watterson, Adam D. Ewing, Geoffrey J. Faulkner

Correspondence
faulknergj@gmail.com

In brief
In response to a recent claim that SARS-CoV-2 can be incorporated into human DNA by LINE-1 retrotransposon proteins, Smits et al. apply long-read DNA sequencing to cultured HEK293T cells infected with SARS-CoV-2. They find no evidence for genomic integration of the virus, despite demonstrated availability of the LINE-1 protein machinery.

Highlights
- Nanopore sequencing of SARS-CoV-2-infected HEK293T cells and control samples
- No SARS-CoV-2 genomic integrants were found in any sample
- LINE-1 insertions carrying retrotransposition hallmarks were detected in HEK293T cells
- Hepatitis B virus integration was observed by sequencing liver cancer samples
No evidence of human genome integration of SARS-CoV-2 found by long-read DNA sequencing

Nathan Smits,1,10 Jay Rasmussen,2,10 Gabriela O. Bodea,1,2,10 Alberto A. Amarilla,3,10 Patricia Gerdes,1 Francisco J. Sanchez-Luque,4,5 Prabha Ajjikuttira,2 Naphak Modhiran,3 Benjamin Liang,3 Jamila Faivre,6 Ira W. Deveson,7,8 Alexander A. Khromykh,3,9 Daniel Watterson,3,5 Adam D. Ewing,1 and Geoffrey J. Faulkner1,2,11,*

1Mater Research Institute, University of Queensland, TRI Building, Woolloongabba, QLD 4102, Australia
2Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia
3School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, QLD 4072, Australia
4GENYO, Pfizer-University of Granada-Andalusian Government Centre for Genomics and Oncological Research, PTS Granada 18016, Spain
5MRC Human Genetics Unit, Institute of Genetics and Cancer (IGC), University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK
6INSERM, U1193, Paul-Brousse University Hospital, Hepatobiliary Centre, Villejuif 94800, France
7Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia
8St Vincent’s Clinical School, Faculty of Medicine, University of New South Wales, Sydney, NSW 2052, Australia
9Australian Infectious Diseases Research Centre, Global Virus Network Centre of Excellence, Brisbane, QLD 4072, Australia
10These authors contributed equally
11Lead contact
*Correspondence: faulknergj@gmail.com

SUMMARY

A recent study proposed that severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) hijacks the LINE-1 (L1) retrotransposition machinery to integrate into the DNA of infected cells. If confirmed, this finding could have significant clinical implications. Here, we apply deep (>50×) long-read Oxford Nanopore Technologies (ONT) sequencing to HEK293T cells infected with SARS-CoV-2 and do not find the virus integrated into the genome. By examining ONT data from separate HEK293T cultivars, we completely resolve 78 L1 insertions arising in vitro in the absence of L1 overexpression systems. ONT sequencing applied to hepatitis B virus (HBV)-positive liver cancer tissues located a single HBV insertion. These experiments demonstrate reliable resolution of retrotransposon and exogenous virus insertions by ONT sequencing. That we find no evidence of SARS-CoV-2 integration suggests that such events are, at most, extremely rare in vivo and therefore are unlikely to drive oncogenesis or explain post-recovery detection of the virus.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a positive-sense single-stranded ~30-kbp polyadenylated RNA betacoronavirus (V’kovski et al., 2021; Wu et al., 2020). SARS-CoV-2 does not encode a reverse transcriptase (RT) and therefore is not expected to integrate into genomic DNA as part of its life cycle. This assumption is of fundamental importance to the accurate diagnosis and potential long-term clinical consequences of SARS-CoV-2 infection, as demonstrated by other viruses known to incorporate into genomic DNA, such as human immunodeficiency virus 1 (HIV-1) and hepatitis B virus (HBV) (Bill and Summers, 2004; Fujimoto et al., 2012; Jiang et al., 2012; Nagaya et al., 1987). Notably, a recent work by Zhang et al. reported potential evidence of SARS-CoV-2 integration into the genome of infected human cells (Zhang et al., 2021). Prior analyses of mammalian genome sequences, as well as in vivo and in vitro experimental data, indicate single-stranded RNA viruses can act as templates for endogenous RTs (Belyi et al., 2010; Feschotte and Gilbert, 2012; Geuking et al., 2009; Horie et al., 2010; Kawasaki et al., 2021; Klenerman et al., 1997). These studies provide a conceptual basis to further investigate the genomic integration of SARS-CoV-2, as pursued by Zhang et al.

LINE-1 (L1) retrotransposons reside in all mammalian genomes (Kazazian and Moran, 2017). In humans, L1 transcribes a bicistronic mRNA encoding two proteins, namely, ORF1p and ORF2p, which are essential to L1 mobility (Moran et al., 1996). ORF2p possesses endonuclease (EN) and RT activities and exhibits a strong cis preference for reverse transcription of L1 mRNA (Doucet et al., 2015; Garcia-Perez et al., 2007; Kulpa and Moran, 2006; Monot et al., 2013; Moran et al., 1996; Wei et al., 2001). Nonetheless, the L1 protein machinery can transmobilize polyadenylated cellular RNAs, especially those produced by Alu and SINE-VNTR-Alu (SVA) retrotransposons but also including protein-coding gene mRNAs (Dewannieux et al., 2003; Ennault et al., 2000; Garcia-Perez et al., 2007; Hancks et al., 2011; Raiz et al., 2012). Somatic L1 mobilization in cis is observed in embryonic cells, the neuronal lineage, and various cancers (Evrony et al., 2015; Feusier et al., 2019;
Rodriguez-Martín et al., 2020; Sanchez-Luque et al., 2019; Schauer et al., 2018; Scott et al., 2016). In contrast, somatic L1-mediated trans mobilization is apparently rare in vivo (Evory et al., 2015; Rodriguez-Martín et al., 2020; Sanchez-Luque et al., 2019) and is likely repressed by various mechanisms (Ahl et al., 2015; Deniz et al., 2019; Doucet et al., 2015; Ewing et al., 2020; Sanchez-Luque et al., 2019). Although Alu and, to a lesser extent, SVA are readily mobilized in cultured cell assays by L1 proteins, the same machinery produces less than one non-retrotransposon cellular RNA insertion for every 2,000 L1 insertions (Dewannieux et al., 2003; Hancks et al., 2011; Wei et al., 2001). Both cis and trans L1-mediated insertions incorporate target site duplications (TSDs) and a 3' polyA tract and integrate at the degenerate L1 EN motif 5'-TTTT/AA (Dewannieux et al., 2003; Esnault et al., 2000; Garcia-Perez et al., 2007; Gilbert et al., 2005; Hancks et al., 2009; Jurka, 1997; Moran et al., 1996; Raiz et al., 2012). These sequence hallmarks can together discriminate artifacts from genuine insertions (Faulkner and Billon, 2018).

In their work, Zhang et al. overexpressed L1 in HEK293T cells, infected the cells with SARS-CoV-2, and identified DNA fragments of the virus through PCR amplification. These results, alongside other less direct (Kazachenka and Kassiotis, 2021; Yan et al., 2021) analyses, were interpreted as evidence of SARS-CoV-2 genomic integration. Crucially, Zhang et al. then detected 63 putative SARS-CoV-2 integrants by Oxford Nanopore Technologies (ONT) long-read sequencing. Of these integrants, only one on chromosome X was spanned by an ONT read aligned to one locus and was flanked by potential TSDs (Figure 1A). However, this SARS-CoV-2 integrant did not incorporate a 3' polyA tract, as is expected for an L1-mediated insertion, and involved an unusual 28-kb internal deletion of the SARS-CoV-2 sequence. Overall, the SARS-CoV-2 integrants reported by Zhang et al. were 26-fold enriched in exons, despite the L1 EN showing no preference for these regions (Flasch et al., 2019; Sultana et al., 2019). Zhang et al. also used Illumina short-read sequencing to map putative SARS-CoV-2 integration junctions in HEK293T cells without L1 overexpression. A lack of spanning reads and the tendency of Illumina library preparation to produce artifacts (Treiber and Waddell, 2017) leave this analysis open to interpretation.

The application of ONT sequencing to HEK293T cells nonetheless held conceptual merit. ONT reads can span germline and somatic retrotransposition events end-to-end and resolve the sequence hallmarks of L1-mediated integration (Ewing et al., 2020; Sludeja et al., 2021). Through this approach, we previously found two somatic L1 insertions in the liver tumor sample of an individual positive for hepatitis C virus (HCV), a ~10-kbp positive-sense single-stranded non-polyadenylated RNA virus (Lauer and Walker, 2001), including one PCR-validated L1 insertion spanned by a single ONT read (Ewing et al., 2020; Shukla et al., 2013). Moreover, HEK293T cells are arguably a context favorable to L1 activity. They express L1 ORF1p (Philippe et al., 2016), accommodate engineered L1-mediated retrotransposition in cis and in trans (Hancks et al., 2011; Kubo et al., 2006; Niewiadomska et al., 2007; Sanchez-Luque et al., 2019), and support SARS-CoV-2 viral replication (Figure S1). Endogenous L1-mediated insertions can be detected in cell culture by genomic analysis of separate cultivars derived from a common population (Klawitter et al., 2016; Nguyen et al., 2018). Based on this experimental rationale, we sought to replicate the central findings of Zhang et al. and, after deeply ONT sequencing SARS-CoV-2-infected HEK293T cells, we did not detect SARS-CoV-2 genomic integration.

**RESULTS**

We ONT sequenced (~54x genome-wide depth, read length N50 of ~39 kb) genomic DNA harvested from HEK293T cells infected with SARS-CoV-2 at a multiplicity of infection (MOI) of 1.0, as well as mock-infected cells (~28x depth, N50 of ~47 kb) (Figures 2A and S1; Table S1). As positive controls, we ONT sequenced the tumor and non-tumor liver tissue of a HBV-positive hepatocellular carcinoma patient (Shukla et al., 2013). HBV is a DNA virus known to be integrated into sites of genomic damage by DNA double-strand break repair (Bill and Summers,
Figure 2. Detection of endogenous L1-mediated retrotransposition in human cells

(A) Experimental design. HEK293T cells were divided into two populations (cultivars), which were then either SARS-CoV-2 infected or mock infected. DNA was extracted from each cultivar, as well as from hepatocellular carcinoma patient samples, and subjected to ONT sequencing. ONT reads were used to call non-reference L1 and virus insertions with TLDR, which also resolves TSDs and other retrotransposition hallmarks. TSDs, red triangles; polyA tract, green rectangle; ONT read, blue rectangle. Some illustrations are adapted from a previous study (Ewing et al., 2020).

(B) TSD size distribution for non-reference L1 insertions, as annotated by TLDR, as well as cultivar-specific L1 insertions found only in either our HEK293T cells infected with SARS-CoV-2 or our mock infected cells. Pie charts indicate the percentages of exonic, intronic, and intergenic insertions, annotated by RefSeq coordinates.

(C) As for (B), except showing data for Alu insertions.

(D) Detailed characterization of an L1 insertion detected in SARS-CoV-2-infected HEK293T cells by a single spanning ONT read aligned to chromosome 14. Nucleotides highlighted in red correspond to the integration site TSD. Underlined nucleotides correspond to the L1 EN motif. The cartoon indicates a full-length L1HS insertion flanked by TSDs (red triangles) and a 3′ polyA tract (green), with the underneath numeral representing the 5′-end position relative to the mobile L1HS sequence L1.3 (Dombroski et al., 1993). The relevant spanning ONT read, with identifier, is also positioned underneath the cartoon. Symbols (α, β, δ, and γ) represent the approximate position of primers used for empty/filled site and L1-genome junction PCR validation reactions. They are displayed in gel images if
As negative controls, we used the aforementioned HCV-positive hepatocellular carcinoma samples and a normal liver sample (Ewing et al., 2020; Table S1). We viewed HCV-infected samples as a suitable negative control because HCV and SARS-CoV-2 are both positive-sense single-stranded RNA viruses, and yet HCV is not polyadenylated and is therefore unlikely to attract the L1 machinery and has not been found to integrate into infected hepatocytes or liver tumor genomes (Fujimoto et al., 2012; Lauer and Walker, 2001). To these data, we added those of Zhang et al. and then used the transposons from Long DNA Reads (TLDR) (Ewing et al., 2020) software to call SARS-CoV-2, HBV, HCV, and retrotransposon insertions spanned by at least one uniquely aligned ONT read. TLDR detected no SARS-CoV-2, HBV, or HCV insertions.

In total, TLDR identified 575 non-reference human-specific L1 (L1HS) insertions, which were typically flanked by TSDs with a median length of 14 bp (Figure 2B; Table S2). No tumor-specific L1 insertions were found, apart from the two previously detected in the HCV-infected liver tumor (Ewing et al., 2020; Shukla et al., 2013). Seventy-eight L1 insertions were found only in our SARS-CoV-2-infected HEK293T cells (n = 66) or the mock infected control (n = 12) and produced TSDs with a median length of 14 bp (Figure 2B). Of the 78 events, 69 (88.5%) were detected by a single spanning read and 13 carried a 3’ transduction (Holmes et al., 1994; Moran et al., 1999; Table S2). After random down-sampling, the more deeply sequenced SARS-CoV-2-infected HEK293T cells still had more than 2-fold more putative cultivar-specific L1 insertions than the mock-infected HEK293T cells. Next, we chose at random 6/69 L1 insertions detected by one spanning read for manual curation and PCR validation. All 6 L1 insertions bore a TSD and a 3’ polyA tract and integrated at a degenerate L1 EN motif (Figures 2D, 2E, and S2A–S2D). Three insertions were 5’ inverted (Kazazian et al., 1988; Ostertag and Kazazian, 2001; Figures 2E, S2C, and S2D) and one carried a 3’ transduction (Holmes et al., 1994) traced to a mobile (Rodriguez-Martín et al., 2020) full-length non-reference L1HS (L1HS) sequence as per prior reports (Fuji-moto et al., 2012). The HBV insertion, we PCR amplified and capillary sequenced its 3’ junction (Figure 3B). The HBV sequence was linearized and rearranged (Figure 3B) as per prior reports (Fujimoto et al., 2012; Jiang et al., 2012; Nagaya et al., 1987). Direct inspection of ONT read alignments thus recovered a HBV integrant, which are found in ~1 per 10^4–10^5 infected hepatocytes (Mason et al., 2016; Tu et al., 2018), and yet did not reveal reads alignable to the SARS-CoV-2 genome in our ONT datasets.

Reanalyzing the ONT data generated by Zhang et al., we found 55 reads (out of ~12 million) that generated an alignment of >100 bp to the SARS-CoV-2 genome (Figure 3A). These reads (median length of 924 bp) were however 65.6% shorter than the overall dataset (2,686 bp) and were comprised of a much higher average proportion of SARS-CoV-2 sequence (52.3%) than the proportion of L1HS sequence found in reads aligned to L1HS (17.1%). Of the 555 reads, 79 generated an alignment of >100 bp to the human genome, including 1 matching the aforementioned integrant on chromosome X that lacked a 3’ polyA tract (Figure 1A). An analysis of the corresponding 79 human genome alignment breakpoints, which could be interpreted as putative SARS-CoV-2 insertion points, as per Zhang et al., indicated 36.2% were exonic (Figure 3C). By comparison, 1.1% of all non-reference L1-mediated insertions reported here were exonic (Figure 3C), as were 1.3% and 1.7% of cultivar-specific L1 and ALu insertions, respectively (Figures 2B and 2C). Finally, we investigated why TLDR called neither of the two SARS-CoV-2 insertions highlighted by Zhang et al. (Figure 1) and found ambiguity in the ONT read alignments that supported these examples, leading to their exclusion. Specifically, the putative chromosome X insertion (Figure 1A) was filtered due to its short 5’ genomic flank alignment, whereas the chromosome 22 example was filtered because the corresponding ONT read alignment was marked as supplementary to another alignment on chromosome 1 (Figure 1B). These analyses confirmed that SARS-CoV-2 alignable reads were present in the Zhang et al. ONT dataset, and yet these reads were unusually short and could include molecular artifacts interpreted by Zhang et al. as SARS-CoV-2 integrants.

**DISCUSSION**

We did not observe L1-mediated SARS-CoV-2 genomic integration in HEK293T cells, despite the availability of the L1 machinery

---

**Cell Reports**

**OPEN ACCESS**

---

4Cell Reports 36, 109530, August 17, 2021
Figure 3. ONT reads occasionally align to viral genome sequences
(A) Percentages of total ONT sequence alignable to L1HS (left), SARS-CoV-2 (middle), and HBV (right) isolate genomes. Read counts for SARS-CoV-2 and HBV are provided above histogram columns. No reads were aligned to the HCV isolate genomes. HEK293T data were generated here (SARS-CoV-2, mock) or previously by Zhang et al. (2021, Z., Z. et al.). HCV tumor/nontumor liver pairs were sequenced here (HCC32; confirmed HBV-positive) or previously (Ewing et al., 2020) (HCC33; HCV-positive). Normal liver ONT sequencing from our prior work (Ewing et al., 2020) was included as an additional control.
(B) A HBV insertion detected in non-tumor liver. In this example, an ONT read from the non-tumor liver of HCC32 spanned the 3’ junction of a HBV integrant located on chromosome 2. Of the HBV isolate genomes considered here, this read aligned best to a representative of genotype B (GenBank: AB602818). The HBV sequence was rearranged consistent with its linearization prior to integration (Fujimoto et al., 2012; Jiang et al., 2012; Nagaya et al., 1987). Numerals indicate positions relative to AB602818. Symbols (β and β) represent the approximate position of primers used to PCR validate the HBV insertion. The gel image at right shows the 3’ junction PCR results. Ladder band sizes are as indicated. The red-filled triangle indicates an on-target product confirmed by capillary sequencing.
(C) Percentages of exonic (black), intronic (gray), and intergenic (blue) genomic alignment breakpoints for ONT reads also aligned to the SARS-CoV-2 genome and also for the non-reference L1-mediated insertions reported here. Genomic features were annotated according to RefSeq coordinates. Please see Tables S1 and S2 for further information.

(Hancks et al., 2011; Kubo et al., 2006; Niewiadowska et al., 2007; Philippe et al., 2016; Sanchez-Luque et al., 2019). We did detect L1, Alu, and SVA retrotransposition events. The higher number of L1 and Alu insertions found in our SARS-CoV-2-infected HEK293T cells is of potential interest given that viral infection can repress host factors limiting L1 activity (Hrecka et al., 2011; Laguette et al., 2011; Zhao et al., 2013). This preliminary finding perhaps indicates SARS-CoV-2 infection could increase L1 or Alu retrotransposition in vitro, a possibility requiring experimental confirmation. The comparative rarity of SVA insertions and the absence of SARS-CoV-2 insertions is however congruent with the relative frequencies of L1, Alu, SVA, and non-retrotransposon cellular RNA insertions driven by L1 proteins in prior cultured cell assays (Dewannieux et al., 2003; Hancks et al., 2011; Wei et al., 2001).

Our approach has several notable differences and caveats when compared to that of Zhang et al. Each study used different SARS-CoV-2 isolates, and here, the MOI (1.0) was double that of Zhang et al. (MOI of 0.5). The high-molecular-weight DNA extraction method, ONT library preparation kit, and depth and quality of sequencing applied to HEK293T cells by Zhang et al. (standard isopropanol precipitation, SQK-LSK109 kit; ~21x depth, N50 of ~11 kbp) and here (Nanobind kit, SQK-LSK110 kit; ~54x depth, N50 of ~39 kbp) differed. Nevertheless, the DNA
extraction protocols of each study would limit retention of extra-chromosomal SARS-CoV-2 DNA potentially generated by ectopic L1 reverse transcription (Dhellin et al., 1997). The origins of the ONT reads aligned to the SARS-CoV-2 genome reported by Zhang et al. are therefore unclear in our view. Zhang et al. ONT sequenced only HEK293T cells transfected with an L1 expression plasmid, which human cells would not carry in vivo. We did not analyze SARS-CoV-2 patient samples, although, arguably, HEK293T cells present an environment far more conducive to L1 activity than those cells accessed in vivo by SARS-CoV-2 (Sungnak et al., 2020; Wiersinga et al., 2020). Widespread cell death post-infection also reduces the probability SARS-CoV-2 integrants would persist in the body (Karki et al., 2021; Varga et al., 2020). This view aligns with a very recent report of negligible SARS-CoV-2 DNA being detected by PCR in coronavirus disease 2019 (COVID-19) patient nasal swabs (Briggs et al., 2021).

Finally, the incredible enrichment reported by Zhang et al. for putative SARS-CoV-2 insertions in exons, which this and prior studies (Flasch et al., 2019; Sultana et al., 2019) have shown are not preferred by the L1 EN, contradicts the involvement of L1 in the events interpreted by Zhang et al. as SARS-CoV-2 genomic integrants. We conclude the L1 cis preference strongly disfavors SARS-CoV-2 retrotransposition, making the phenomenon mechanistically plausible but likely very rare, as for other polyadenylated non-retrotransposon cellular RNAs (Dewannieux et al., 2003; Doucet et al., 2015; Esnault et al., 2000; Garcia-Perez et al., 2007; Hancks et al., 2011; Kulpa and Moran, 2006; Monot et al., 2013; Moran et al., 1996; Wei et al., 2001).

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - SARS-CoV-2 infection of HEK293T cells
  - Hepatocellular carcinoma sample processing
  - ONT sequencing
  - ONT bioinformatic analyses
  - PCR validation
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109530.

ACKNOWLEDGMENTS

The authors thank the human subjects of this study who donated tissues to the Centre Hépatobiliaire, Paul-Brousse Hospital. We thank S. Richardson and R. Shukla for helpful discussions, K. Chappell and P. Young for project support, and Queensland Health for providing the SARS-CoV-2 virus isolate QLDD0. This study was funded by an Australian Government Research Training Program (RTP) Scholarships (N.S.), an NHMRC-ARC Dementia Research Development Fellowship (GNT1108258, G.O.B.), seed funding provided by the Australian Infectious Disease Research Centre to establish SARS-CoV-2 research at the University of Queensland (A.A.K), an Australian Department of Health Medical Research Future Fund (MRFF) Novel Coronavirus Vaccine Development Grant (APP1202445-2020, D.W.), an MRFF Investigator Grant (MRF1175457, A.D.E.), an NHMRC Investigator Grant (GNT1173711, G.J.F.), a CSL Centenary Fellowship (G.J.F.), and the Mater Foundation.

AUTHOR CONTRIBUTIONS

N.S., J.R., G.O.B., A.A.A., P.G., F.J.S.-L., P.A., N.M., and B.L. performed experiments and analyzed data. A.D.E. and G.J.F. performed bioinformatic analysis. J.F., I.W.D., A.A.K., D.W., and G.J.F. provided resources. G.J.F. designed the project and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Ahl, V., Keller, H., Schmidt, S., and Weichenrieder, O. (2015). Retrotransposition and Crystal Structure of an Alu RNP in the Ribosome-Stalling Conformation. Mol. Cell 60, 715–727.

Amarilla, A.A., Modhiran, N., Setoh, Y.X., Peng, N.Y.G., Sng, J.D.J., Liang, B., McMillan, C.L.D., Freney, M.E., Cheung, S.T.M., Chappell, K.J., et al. (2021). An optimized high-throughput immuno-plaque assay for SARS-CoV-2. Front. Microbiol. 12, 625136.

Belyi, V.A., Levine, A.J., and Skalka, A.M. (2010). Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. PLoS Pathog. 6, e1001030.

Bill, C.A., and Summers, J. (2004). Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. Proc. Natl. Acad. Sci. USA 101, 11135–11140.

Briggs, E., Ward, W., Rey, A., Law, D., Nelson, K., Bois, M., Ostrov, N., Lee, H.H., Laurent, J.M., and Mita, P. (2021). Assessment of potential SARS-CoV-2 virus N gene integration into human genome reveals no significant impact on RT-qPCR COVID-19 diagnostic testing. medRxiv. https://doi.org/10.1101/2021.06.21.21258023.

Deniz, Ö., Frost, J.M., and Branco, M.R. (2019). Regulation of transposable elements by DNA modifications. Nat. Rev. Genet. 20, 417–431.

Dewannieux, M., Esnault, C., and Heidmann, T. (2000). LINE-mediated retrotransposition. Mol. Cell 6, 728–741.

D dellin, O., Maestre, J., and Heidmann, T. (1997). Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription. EMBO J. 16, 6590–6602.

Dombroski, B.A., Scott, A.F., and Kazazian, H.H., Jr. (1993). Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. Proc. Natl. Acad. Sci. USA 90, 6513–6517.

Doucet, A.J., Wilusz, J.E., Miyoshi, T., Liu, Y., and Moran, J.V. (2015). A 3′ Poly(A) Tract Is Required for LINE-1 Retrotransposition. Mol. Cell 60, 728–741.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. Nat. Genet. 24, 363–367.
Evrony, G.D., Lee, E., Mehta, B.K., Benjamini, Y., Johnson, R.M., Cai, X., Yang, L., Haseley, P., Lehmann, H.S., Park, P.J., and Walsh, C.A. (2015). Cell lineage analysis in human brain using endogenous retroelements. Neuron 85, 49–59.

Ewing, A.D., Smits, N., Sanchez-Luque, F.J., Fairev, J., Brennan, P.M., Richardson, S.R., Cheetham, S.W., and Faulkner, G.J. (2020). Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. Mol. Cell 80, 915–928.e5.

Faulkner, G.J., and Billon, V. (2018). L1 retrotransposition in the soma: a field jumping ahead. Mob. DNA 9, 22.

Feschotte, C., and Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. 13, 283–296.

Feusier, J., Watkins, W.S., Thomas, J., Farrell, A., Witherspoon, D.J., Baird, L., Ha, H., Xing, J., and Jorde, L.B. (2019). Pedigree-based estimation of human mobile element retrotransposition rates. Genome Res.

Ha, H., Xing, J., and Jorde, L.B. (2019). Novel mechanism for the creation of inversions in L1 retrotransposition. Genome Res.

Jern, P., Gojobori, T., Coffin, J.M., and Tomonaga, K. (2010). Endogenous human endogenous retroviruses: From evolution to impact on host biology. Nat. Rev. Genet.

Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (2011). Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum. Mol. Genet. 20, 3386–3400.

Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nat. Genet. 7, 143–148.

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daigo, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M., and Tomonaga, K. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 463, 84–87.

Hrecka, K., Hao, C., Gierszewskia, M., Swanson, S.K., Kesik-Brodacka, M., Srivastava, S., Floren, L., Washburn, M.P., and Skowronski, J.B. (2011). Vpx represses inhibition of HIV-1 infection by macrophages mediated by the SAMHD1 protein. Nature 474, 658–661.

Jiang, Z., Jhunjhunwala, S., Liu, J., Haverty, P.M., Kenneren, M.I., Guan, Y., Lee, W., Carnevali, P., Stinson, J., Johnson, S., et al. (2012). The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. Genome Res. 22, 593–601.

Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retroelements. Proc. Natl. Acad. Sci. USA 94, 1872–1877.

Karki, R., Sharma, B.R., Tuladhar, S., Williams, E.P., Zaldoaundo, L., Samir, P., Zheng, M., Sundaram, B., Banot, B., Mailredi, R.K.S., et al. (2021). Synergism of TNF-α and IFN-γ Triggers Inflammatory Cell Death, Tissue Damage, and Mortality in SARS-CoV-2 Infection and Cytokine Shock Syndromes. Cell 184, 149–168.

Kawasaki, J., Kojima, S., Mukai, Y., Tomonaga, K., and Horie, M. (2021). 100-Year History of Bornavirus Infections Hidden in Vertebrate Genomes. Proc. Natl. Acad. Sci. USA 118, e202635118.

Kazachekina, A., and Kassiotis, G. (2021). SARS-CoV-2-Host Chimeric RNA-Sequencing Reads Do Not Necessarily Arise From Virus Integration Into the Host DNA. Front. Microbiol. 12, 676693.

Kazazian, H.H., Jr., and Moran, J.V. (2017). Mobile DNA in Health and Disease. N. Engl. J. Med. 377, 361–370.

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332, 164–166.

Klawitter, S., Fuchs, N.V., Upton, K.R., Muñoz-Lopez, M., Shukla, R., Wang, J., García-Canahíbas, M., Lopez-Ruiz, C., Gerhardt, D.J., Sebe, A., et al. (2016). Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. Nat. Commun. 7, 10286.

Klenerman, P., Hengartner, H., and Zinkernagel, R.M. (1997). A non-retroviral RNA virus persists in DNA form. Nature 390, 298–301.

Kubo, S., Seleme, M.C., Soifer, H.S., Perez, J.L.G., Moran, J.V., Kazazian, H.H., Jr., and Kasahara, N. (2006). L1 retrotransposition in nondividing and primary human somatic cells. Proc. Natl. Acad. Sci. USA 103, 8036–8041.

Kulpa, D.A., and Moran, J.V. (2006). cis- Preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat. Struct. Mol. Biol. 13, 655–660.

Lagouette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Ségréal, E., Yatim, A., Emiliani, S., Schwartz, O., and Benkirane, M. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature 474, 654–657.

Lauer, G.M., and Walker, B.D. (2001). Hepatitis C virus infection. N. Engl. J. Med. 345, 41–52.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Mason, W.S., Gill, U.S., Litwin, S., Zhou, Y., Peri, S., Pop, O., Hong, M.L.W., Naik, S., Quaglia, A., Bertolatti, A., and Kennedy, P.T. (2016). HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant. Gastroenterology 151, 986–998.

Monot, C., Kuciak, M., Violet, S., Mir, A.A., Gabus, C., Darlix, J.-L., and Cristofari, G. (2013). The specificity and flexibility of HIV reverse transcription priming at imperfect T-tracts. PLoS Genet. 9, e1003499.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell 87, 917–927.

Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. Science 283, 1530–1534.

Nagaya, T., Nakamura, T., Tokino, T., Tsurimoto, T., Imai, M., Mayumi, T., Kamino, K., Yamamura, K., and Matsubara, K. (1987). The mode of hepatitis B virus persists in DNA form. Nature 328, 10286.

Nguyen, T.H.M., Carreira, P.E., Sanchez-Luque, F.J., Schauer, S.N., Fagg, A.C., Richardson, S.R., Davies, C.M., Jesudian, J.S., Kempen, M.H.C., Troskie, R.-L., et al. (2018). L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution. Cell Rep. 23, 3730–3740.

Niewiadomska, A.M., Tian, C., Tan, L., Wang, T., Sarkis, P.T.N., and Yu, X.-F. (2007). Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. J. Virol. 81, 9577–9583.

Osterberg, E.M., and Kazazian, H.H., Jr. (2001). Twin priming: a proposed mechanism for the creation of insertions in L1 retrotransposition. Genome Res. 11, 2059–2065.
Philippe, C., Vargas-Landin, D.B., Doucet, A.J., van Essen, D., Vera-Otario, J., Kuciak, M., Corbin, A., Nigmann, P., and Cristofari, G. (2016). Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. eLife 5, e13926.

Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Harndorf, M., Löwer, J., Strätling, W.H., Löwer, R., and Schumann, G.G. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. 40, 1666–1683.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

Rodriguez-Martin, B., Alvarez, E.G., Baez-Ortega, A., Zamora, J., Supef, K., Demeulemeester, J., Santamarina, M., Ju, Y.S., Temes, J., Garcia-Souto, D., et al.; PCAWG Structural Variation Working Group; PCAWG Consortium (2020). Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat. Genet. 52, 306–319.

Sanchez-Luque, F.J., Kempen, M.H.C., Gerdes, P., Vargas-Landin, D.B., Richardson, S.R., Troskie, R.-L., Jesuadian, J.S., Cheetham, S.W., Carreira, P.E., Salvador-Palomeque, C., et al. (2019). LINE-1 Evasion of Epigenetic Repression in Humans. Mol. Cell 75, 590–604.

Schauer, S.N., Carreira, P.E., Shukla, R., Gerhardt, D.J., Gerdes, P., Sanchez-Luque, F.J., Nicoli, P., Kindlова, M., Ghisletti, S., Santos, A.D., et al. (2019). L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. Genome Res. 28, 453–463.

Scott, E.C., Gardner, E.J., Masood, A., Chuang, N.T., Vertino, P.M., and Devine, S.E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res. 26, 745–755.

Shukla, R., Upton, K.R., Muñoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell 153, 101–111.

Sludeka, K., van den Beek, M., Riddiford, N., Boumand, B., Wurmser, A., Stefanutti, N., Lameiras, S., and Bardin, A.J. (2021). Unraveling the features of somatic transposition in the Drosophila intestine. EMBO J. 40, e106386.

Sultana, T., van Essen, D., Siol, O., Bailly-Bechet, M., Philippe, C., Zine El Aabidine, A., Poger, L., Nigmann, P., Saccani, S., Andraud, J.-C., et al. (2019). The Landscape of L1 Retrotransposons in the Human Genome is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. Mol. Cell 74, 555–570.

Sungnak, W., Huang, N., Bécauví, C., Berg, M., Queen, R., Litvinukova, M., Tavares-López, C., Maatz, H., Reichart, D., Sampaiozitis, F., et al.; HCA Lung Biological Network (2020). SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat. Med. 26, 681–687.

Treiber, C.D., and Waddell, S. (2017). Resolving the prevalence of somatic transposition in Drosophila. eLife 6, e28297.

Tu, T., Budzinska, M.A., Vondran, F.W.R., Shackel, N.A., and Urban, S. (2018). Hepatitis B virus DNA integration occurs early in the viral life cycle in an in vitro infection model via sodium taurocholate cotransporting polypeptide-dependent uptake of enveloped virus particles. J. Virol. 92, e02007–e02017.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. Nucleic Acids Res. 40, e115.

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. Nat. Rev. Microbiol. 19, 155–170.

Varga, Z., Flammer, A.J., Steiger, P., Haberecker, M., Andermatt, R., Zinken, A.S., Mehra, M.R., Schuepbach, R.A., Rutschitzka, F., and Moch, H. (2020). Endothelial cell infection and endotheliitis in COVID-19. Lancet 395, 1417–1418.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. Mol. Cell. Biol. 21, 1429–1439.

Wiersinga, W.J., Rhodes, A., Cheng, A.C., Peacock, S.J., and Prescott, H.C. (2020). Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA 324, 782–793.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269.

Yan, B., Chakravorty, S., Mirabelli, C., Wang, L., Trujillo-Ochoa, J.L., Chaus, D., Kumar, D., Lionakis, M.S., Olson, M.R., Wobus, C.E., et al. (2021). Host-Virus Chimeric Events in SARS-CoV-2-Infected Cells Are Infrequent and Artificial. J. Virol. 95, e0029421.

Zhang, L., Richards, A., Barrasa, M.I., Hughes, S.H., Young, R.A., and Jahnisch, R. (2021). 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. USA 118, e2105968118.

Zhao, K., Du, J., Han, X., Goodier, J.L., Li, P., Zhou, X., Wei, W., Evans, S.L., Li, L., Zhang, W., et al. (2013). Modulation of LINE-1 and Alu/SVA retrotransposition by Acardia-Goutieres syndrome-related SAMHD1. Cell Rep. 4, 1108–1115.
## STAR METHODS

### KEY RESOURCES TABLE

| REAGENT OR RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Bacterial and virus strains** | | |
| SARS-CoV-2 isolate hCoV-19/Australia/QLD02/2020 | Queensland Health Forensic and Scientific Services | EPI_ISL_407896 |
| **Biological samples** | | |
| Snap frozen hepatocellular carcinoma and matched non-tumor liver tissue | Centre Hépatobiliaire, Paul-Brousse Hospital | HCC32 |
| **Chemicals, peptides, and recombinant proteins** | | |
| SYBR Safe DNA Gel Stain | Invitrogen | S33102 |
| Agarose | Bioline | BIO-41026 |
| Carboxymethylcellulose | Sigma-Aldrich | C4888 |
| KPL Milk Diluent/Blocking Solution Concentrate | SeraCare | 5140-0011 |
| IRDye 800CW Goat anti-Mouse IgG antibody | LI-COR | 926-32210 |
| Anti-SARS-CoV-2-spike antibody | Amarilla et al., 2021 | CR3022 |
| **Critical commercial assays** | | |
| Ligation Sequencing Kit | Oxford Nanopore Technologies | SQK-LSK110 |
| Qubit dsDNA HS Assay Kit | Invitrogen | Q32851 |
| Nanobind CBB Big DNA Kit | Circulomics | NB-900-001-01 |
| Short Read Eliminator XS Kit | Circulomics | SS-100-121-01 |
| MyTaq DNA Polymerase | Bioline | BIO-21105 |
| **Deposited data** | | |
| Nanopore WGS of mock and SARS-CoV-2 infected HEK293T cells, and HBV infected samples from liver cancer patient HCC32 | This paper | ENA: PRJEB44816 |
| Nanopore WGS of SARS-CoV-2 infected HEK293T cells overexpressing L1 | Zhang et al., 2021 | SRA: PRJNA721333 |
| Nanopore WGS of normal liver and HCV infected samples from liver cancer patient HCC33 | Ewing et al., 2020 | SRA: PRJNA629858 |
| **Experimental models: Cell lines** | | |
| Human embryonic kidney 293T (HEK293T) cells | ATCC | CRL-1568 |
| Cercopithecus aethiops Vero E6 cells | ATCC | CRL-3216 |
| **Software and algorithms** | | |
| TLDR | https://github.com/adamewing/tldr | Ewing et al., 2020 |
| Minimap2 | https://github.com/lh3/minimap2 | Li, 2018 |
| SAMtools | https://github.com/samtools/ | Li et al., 2009 |
| IGV | https://software.broadinstitute.org/software/igv/ | Robinson et al., 2011 |
| Primer3 | https://bioinfo.ut.ee/primer3-0.4.0/ | Untergasser et al., 2012 |

Report

Cell Reports 36, 109530, August 17, 2021 e1

OPEN ACCESS
RESOURCE AVAILABILITY

Lead contact
Requests for further information and for resources and reagents should be directed to and will be fulfilled by the lead contact, Geoffrey J. Faulkner (faulknergj@gmail.com).

Materials availability
This study did not generate new unique reagents.

Data and code availability
Oxford Nanopore Technologies sequencing data generated by this study were deposited in the European Nucleotide Archive (ENA) under project PRJEB44816. TLDR and instructions for its use and application are available at https://github.com/adamewing/TLD.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Liver tumor and non-tumor tissue were previously obtained from a HBV-positive patient (HCC32, male, 73yrs) who underwent surgical resection at the Centre Hepatobiliaire, Paul-Brousse Hospital, and made available for research purposes with approval from the French Institute of Medical Research and Health (Reference: 11-047). Further ethics approvals were provided by the Mater Health Services Human Research Ethics Committee (Reference: HREC-15-MHS-52) and the University of Queensland Medical Research Review Committee (Reference: 2014000221). HEK293T and Vero E6 cells were obtained from the American Type Culture Collection (ATCC).

METHOD DETAILS

SARS-CoV-2 infection of HEK293T cells
HEK293T cells and African green monkey kidney cells (Vero E6) were maintained in standard Dulbecco’s Modified Eagle Medium (DMEM). Culture media were supplemented with sodium pyruvate (11mg/L), penicillin (100U/mL), streptomycin (100 μg/mL) (P/S) and 10% fetal calf serum (FCS) (Bovogen, USA). Cells were maintained at 37°C with 5% CO2.

An early Australian SARS-CoV-2 isolate (hCoV-19/Australia/QLD02/2020; GISAID Accession EPI_ISL_407896) was sampled from patient nasopharyngeal aspirates by Queensland Health Forensic and Scientific Services and used to inoculate Vero E6 African green monkey kidney cells (passage 2). A viral stock (passage 3) was then generated on Vero E6 cells and stored at −80°C. Viral titration was determined by immuno-plaque assay (iPA), as previously described (Amarilla et al., 2021). To verify viral replication in HEK293T cells, a growth kinetic was assessed using a MOI of 0.01, 0.1 or 1.0, and showed efficient SARS-CoV-2 replication (Figure S1).

HEK293T viral infection was undertaken as follows: 3 × 10^6 HEK293T cells were seeded onto 6-well plates pre-coated with polylysine one day before infection. Cells were infected at MOI of 1 in 200μL of DMEM (2% FCS and P/S) and incubated for 30min at 37°C. Plates were rocked every 5min to ensure the monolayer remained covered with inoculum. The inoculum was then removed, and the monolayer washed five times with 1mL of additive-free DMEM. Finally, cells were maintained with 3mL of DMEM (supplemented with 2% fetal bovine serum and P/S) and incubated at 37°C with 5% CO2. Cell supernatant was harvested 0, 1, 2 and 3 days post-infection. The mock infected control differed only in that virus was not added to the inoculum media.

Genomic DNA was extracted from mock and SARS-CoV-2 infected (MOI 1.0) HEK293T cells sampled 2 days post-infection, using a Nanobind CBB Big DNA Kit (Circulomics) following the manufacturer’s instructions for high molecular weight (HMW) DNA extraction. DNA was eluted in elution buffer (10mM Tris-Cl, pH 8.5) and concentration measured by Qubit dsDNA High-Sensitivity Assay Kit on a Qubit Fluorometer (Life Technologies).

Hepatocellular carcinoma sample processing
DNA was extracted from the HCC32 tissues in our earlier study (Shukla et al., 2013) with a DNeasy Blood and Tissue Kit (QIAGEN, Germany) and stored at −80°C. To enrich for HMW DNA, 4.5 μg of DNA from the patient HCC32 tumor and non-tumor liver samples was diluted to 75ng/μL in a 1.5mL Eppendorf DNA LoBind tube and processed with a Short Read Eliminator XS Kit (Circulomics) following the manufacturer’s instructions.

ONT sequencing
DNA libraries were prepared at the Kinghorn Centre for Clinical Genomics (KCCG) using 3-4 μg HMW input DNA, without shearing, and a SQK-LSK110 ligation sequencing kit. 350-500ng of each prepared library was sequenced separately on one PromethION (Oxford Nanopore Technologies) flow cell (FLO-PR0002, R9.4.1 chemistry) (Table S1). SARS-CoV-2 infected HEK293T DNA was sequenced on two flow cells. Flow cells were washed (nuclease flush) and reloaded at 24hr and 48hr with 350-500ng of additional library to maximize output. Bases were called with guppy 4.0.11 (Oxford Nanopore Technologies).
ONT bioinformatic analyses
To call non-reference insertions with TLDR (Ewing et al., 2020), ONT reads generated here, by Zhang et al. (2021), and by our previous ONT study of human tissues (Ewing et al., 2020; Table S1) were aligned to the human reference genome build hg38 using minimap2 (Li, 2018) version 2.17 (index parameter: -x map-ont; alignment parameters: -ax map-ont -L -t 32) and SAMtools (Li et al., 2009) version 1.12. BAM files were then processed as a group with TLDR (Ewing et al., 2020) version 1.1 (parameters -e virus.fa -p 128 -m 1–max_te_len 100–min_te_len 100–keep_pickles -n nonref.collection.hg38.chr.bed.gz). The file virus.fa was composed of: representative HBV and HCV isolate genomes (Table S1), the SARS-CoV-2 isolate used here (GISAID Accession EPI_ISL_407896), the L1HS sequence L1.3 (Dombroski et al., 1993) (GenBank Accession L19088), several Alu and SVA subfamily consensus sequences, and a consensus sequence for human endogenous retrovirus K (HERVK), the youngest human long terminal repeat (LTR) retrotransposon family. The file nonref.collection.hg38.chr.bed.gz is a collection of known non-reference retrotransposon insertions available from https://github.com/adamewing/tldr/. The TLDR output table was further processed to remove calls not passing all TLDR filters, representing homopolymer insertions, where MedianMapQ < 50 or family = “NA” or remappable = “FALSE” or UnmapCover < 0.75 or Lenghtths < 100 or EndTE-StartTE < 100 or strand = “None” or SpanReads < 1. As 3’ truncation is rarely encountered for L1-mediated insertions, calls where EndTE was more than 10bp less than the consensus length were filtered, as were Alu insertions 5’ truncated by more than 1bp. The filtered TLDR output table is provided as Table S2. Insertions detected in only our mock or SARS-CoV-2 infected HEK293T datasets, but not in both experiments, and not matching a known non-reference insertion, were designated as putative cultivar-specific insertions (Table S2). Many if not most of these insertions were likely to have occurred in cell culture prior to the cultivars being separated.

To identify L1HS and viral sequences, we directly aligned all reads to the virus.fa file with minimap2 (index parameter: -x map-ont; alignment parameters: -ax map-ont -L -t 32). Reads containing alignments of ≥ 100bp to a sequence present in virus.fa were counted with SAMtools idxstats. Alignments to HBV, HCV or SARS-CoV-2 were excluded if they overlapped by ≥ 10bp with a genomic alignment of ≥ 100bp. Read alignments were visualized with SAMtools view and the Integrative Genomics Viewer (Robinson et al., 2011) version 2.8.6.

PCR validation
We used Primer3 (Untergasser et al., 2012) to design PCR primers for L1 insertions found by a single spanning ONT read, using the reference genome and L1HS sequences as inputs (Table S2). These validation experiments were conducted in three phases. First, we performed an “empty/filled site” PCR using primers positioned on either side of the L1, where the filled site is the L1 allele, and the empty site is the remaining allele(s). Each empty/filled reaction was performed using a DNA Engine Tetrad 2 Thermal Cycler (Bio-Rad) and Expand Long Range Enzyme Mix, with 1X Expand Long Range Buffer with MgCl2, 50pmol of each primer, 500μM of dNTPs, 2% DMSO, 100ng of template DNA and 1.75U of enzyme, in a 25 μL final volume. PCR cycling conditions were as follows: (92°C, 3min) × 1; (92°C, 30sec; 54-57°C, 30sec; 68°C, 7min) × 10; (92°C, 30sec; 52-55°C, 30sec; 68°C, 7min + 20sec/cycle) × 30; (68°C, 10min; 4°C, hold) × 1. Amplicons were visualized on a 1% agarose gel stained with SYBR SAFE (Invitrogen). GeneRulerTM 1kb plus (Thermo Scientific) was used as the ladder. Second, we combined each empty/filled primer with a primer positioned within the L1 sequence, to amplify the 5’ and 3’ L1-genome junctions. These reactions were undertaken on a DNA Engine Tetrad 2 Thermal Cycler (Bio-Rad), with MyTaq HS DNA polymerase, 1X MyTaq Reaction Buffer, 10pmol of each primer, 10ng of template DNA, and 2.5U of enzyme, in a 25 μL final volume. PCR cycling conditions were as follows: (95°C, 1min) × 1; (95°C, 15sec; 53-55°C, 15sec; 72°C, 15sec) × 35; (72°C, 5min; 4°C, hold) × 1. Amplicons were visualized on a 1.5% agarose gel stained with SYBR SAFE (Invitrogen). Third, we repeated the 5’ L1-genome junction-specific PCR using 200ng template DNA. All PCRs were performed with non-template control, as well as DNA extracted from the same HEK293T cells (SARS-CoV-2 and mock) subjected to genomic analysis. Notably, L1 insertions that did not amplify in either cultivar are likely to be genuine events as they carried all of the relevant sequence hallmarks of L1-mediated retrotransposition.

PCR primers for the HBV insertion 3’ junction (Figure 3B; Table S2) were designed with Primer3 using the reference genome and closest match HBV sequence (GenBank accession AB602818) as inputs. PCR amplification and capillary sequencing was conducted as per the L1 insertions, except using Expand Long Range polymerase (Roche) with 1X Expand Long Range buffer with MgCl2, 10pmol of each primer, 100ng of template DNA, 500μM of PCR Nucleotide Mix, and 3.5U of enzyme, in a 25 μL final volume. PCR cycling conditions were as follows: (92°C, 2min) × 1; (92°C, 15sec; 65°C, 15sec; 68°C, 7:30min) × 10; (92°C, 15sec; 65°C, 15sec; 68°C, 7min+ 20sec per cycle) × 35 (68°C, 10min; 4°C, hold) × 1. Amplicons were visualized on a 1.2% agarose gel.

Amplicons in each experiment were visualized using a GelDoc (Bio-Rad) and, if of the correct size, gel extracted using a QIAGEN MinElute Gel Extraction Kit and capillary sequenced by the Australian Genomics Research Facility (Brisbane).

QUANTIFICATION AND STATISTICAL ANALYSIS
Error bars and replicate values are defined in figure legends, where appropriate. No statistical tests for significance were conducted.