




Impaired activation of transposable elements in SARS-CoV-2 infection

Matan Sorek^{1,2,3,*} , Eran Meshorer^{1,2,**}  & Sharon Schlesinger^{3,***} 

Abstract

Emerging evidence shows that transposable elements (TEs) are induced in response to viral infections. This TE induction is suggested to trigger a robust and durable interferon response, providing a host defense mechanism. Here, we analyze TE expression changes in response to SARS-CoV-2 infection in different human cellular models. Unlike other viruses, SARS-CoV-2 infection does not lead to global upregulation of TEs in primary cells. We report a correlation between TEs activation and induction of interferon-related genes, suggesting that failure to activate TEs may account for the weak interferon response. Moreover, we identify two variables that explain most of the observed diverseness in immune responses: basal expression levels of TEs in the pre-infected cells and the viral load. Finally, analyzing the SARS-CoV-2 interactome and the epigenetic landscape around the TEs activated following infection, we identify SARS-CoV-2 interacting proteins, which may regulate chromatin structure and TE transcription. This work provides a possible functional explanation for SARS-CoV-2 success in its fight against the host immune system and suggests that TEs could serve as potential drug targets for COVID-19.

Keywords COVID-19; epigenetics; interferon response; SARS-CoV-2; transposable elements

Subject Categories Chromatin, Transcription, & Genomics; Immunology; Microbiology, Virology & Host Pathogen Interaction

DOI 10.15252/embr.202255101 | Received 23 March 2022 | Revised 26 July 2022 | Accepted 27 July 2022 | Published online 16 August 2022

EMBO Reports (2022) 23: e55101

Introduction

Coronaviruses are a diverse group of single-stranded positive-strand RNA viruses infecting a wide range of vertebrate hosts. These viruses are thought to generally cause mild upper respiratory tract illnesses in humans such as the common cold. However, infection with severe acute respiratory syndrome-related coronavirus 2

(SARS-CoV-2), which causes coronavirus disease-2019 (COVID-19), can result in a cytokine storm, which develops into acute respiratory distress syndrome and acute lung injury, often leading to reduction in lung function and even death (Blanco-Melo *et al*, 2020). The fatality of SARS-CoV-2 increases substantially with age. Unlike other airborne viruses, SARS-CoV-2 is unusually effective at evading the early innate immune responses, such as type I and type III interferons (IFN-I and IFN-III). This is partially achieved by viral proteins that antagonize various steps of dsRNA-activated early host responses (Domizio *et al*, 2022; Neufeldt *et al*, 2022). However, we are only beginning to understand the kinetics of IFN response in mild and severe COVID-19 patients (Paludan & Mogensen, 2022).

In one study, Blanco-Melo *et al* (2020) assessed the transcriptional response to SARS-CoV-2 infection in different cellular models and found that SARS-CoV-2 does not elicit robust IFN expression in lung epithelial cells, while multiple pro-inflammatory cytokines were highly expressed. In another study, Huang *et al* (2020) assessed the transcriptional response to SARS-CoV-2 infection in alveolar type 2 cells (iAT2s) derived from induced pluripotent stem cells (iPSCs) by *in vitro* differentiation. The authors found that SARS-CoV-2 infection in these cells results in an inflammatory phenotype with an activation of the NF- κ B pathway, and a delayed IFN signaling response. Other recently published data support these conclusions (Neufeldt *et al*, 2022), although the outcome of IFN response on the virus and its host are not clear.

Transposable elements (TEs) are abundant sequences in the mammalian genome that contain multiple regulatory elements and can amplify in a short evolutionary timescale. Lately, it was found that TEs induction can stimulate antiviral response, via both *cis* and *trans* mechanisms (Hale, 2021). First, in *cis*, TEs have coopted to shape the transcriptional network underlying the IFN response, and some TEs serve as enhancers of antiviral genes in diverse mammalian genomes (Chuong *et al*, 2016; Wang *et al*, 2021). Thus, when TEs are transcribed, they also facilitate the transcription of the nearby antiviral genes. TEs are also enriched in enhancers of CD8⁺ T lymphocytes-specific genes, suggesting that their upregulation might influence not only the innate but also the adaptive, immune response. What is more, due to the similarities between TEs and

¹ Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

² Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Jerusalem, Israel

³ Department of Animal Sciences, Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel

*Corresponding author. Tel: +972 54 5607817; E-mail: matan.sorek@mail.huji.ac.il

**Corresponding author. Tel: +972 2 6585161; E-mail: eran.meshorer@mail.huji.ac.il

***Corresponding author. Tel: +972 8 9489426; E-mail: sharon.shle@mail.huji.ac.il

viral transcripts, cells sometimes misidentify them as invading viruses and trigger the innate immune nucleic acid sensors (e.g., RIG-I, MDA-5, and cGAS) controlling IFN response. Consequently, genome-wide global induction of TEs act *in trans* by activating these sensors during some viral infections in humans, and, is part of the first wave response to viral infection, prior to the induction of IFN (Macchietto *et al*, 2020). For example, infection by highly pathogenic avian influenza viruses elicits TEs induction (Krischuns *et al*, 2018). As a result, TE dsRNA is formed, recognized by the host sensors, and activates both the NF κ B and the IFN pathways, thus enhancing immune response (Chiappinelli *et al*, 2015). These, and other (Macchietto *et al*, 2020; Badarinarayan & Sauter, 2021) results, indicate that while IFN β treatment activates TEs expression (Hung *et al*, 2015; Attig *et al*, 2017), TEs upregulation is an early response to viral infection, which can precede the IFN response and induce it in a positive feedback loop (Macchietto *et al*, 2020). Collectively, these findings suggest a causative link between TE induction and the intensity of the IFN response (Gazquez-Gutierrez *et al*, 2021).

This link is also evident in aging as TE basal expression increases with age (preprint: Bogu *et al*, 2019; Pehrsson *et al*, 2019). Consequently, aging is associated with sterile inflammation which include erroneous IFN response, or the “Inflamaging” phenomena (Franceschi *et al*, 2018). At the molecular level, aging is associated with decreased heterochromatin-associated marks, for example, H3K9me3 and DNA methylation (Franceschi *et al*, 2018). Interestingly, early antiviral IFN responses are impaired and delayed in aged individuals, resulting in increased risk of COVID-19 complication (Galani *et al*, 2020). Similarly, elevated TE expression is found in autoimmune pathologies such as arthritis and systemic lupus erythematosus (Tokuyama *et al*, 2018), which are suggested as risk factors for severe COVID-19 (Karaderi *et al*, 2020).

Epithelial cells are the primary targets and first responders of both IAV and SARS-CoV-2 infections, which initiate the immediate immune response. Viral nucleic acids are recognized by the pattern recognition receptors and culminate in the production of ISGs and pro-inflammatory cytokines. Early type I interferon response is crucial for the host, and both IAV and SARS-CoV-2 produce proteins that interfere with interferon signaling. TEs that are upregulated following IAV infection contribute to the induction of the antiviral responses (Shen *et al*, 2022). Although some studies have addressed TEs expression following SARS-CoV-2 infection (Ferrarini *et al*, 2021; Marston *et al*, 2021), and detected minor TE expression changes in primary cells (Kitsou *et al*, 2021; Tovo *et al*, 2021), none has considered the cell type and basal TE expression levels as important characteristics of their analysis. Here, we suggest that SARS-CoV-2-infected cells that fail to activate an immediate and effective TE response will be more likely to demonstrate a late immune response. Notably, in primary cells, the IFN response to SARS-CoV-2 is observed only 96 h after infection, unlike that observed in cell lines (Rebendenne *et al*, 2021). This correlation between the IFN response and TE expression levels strengthens our model and gives rise to a hypothesis that link between mild TE levels and the ineffective innate response to SARS-CoV-2 infection in some individuals (Rebendenne *et al*, 2021). Given the high correlation between TEs activation and higher IFN response, we suggest a possible use for TEs in COVID-19 prognosis.

Results

TE induction in response to SARS-CoV-2 infection is limited in normal lung cells

Since TEs are induced in response to many viral infections (Macchietto *et al*, 2020), we first analyzed recently published datasets of primary human lung epithelial (NHBE) cells infected with influenza (IAV) or SARS-CoV-2 (Blanco-Melo *et al*, 2020). NHBE cells mimic infected human lung cells, showing cytopathic effects after SARS-CoV-2 infection (Takayama, 2020). In these cells, IAV infection caused, as expected, a global increase in TE subfamilies expression across all TE families, but SARS-CoV-2 did not (Fig 1A). As expected, IFN β -treated cells also activated TE expression levels (Fig 1A) because many TEs have IFN-responsive sequences and are upregulated following the induction of IFN response (Tokuyama *et al*, 2018). This raised an intriguing hypothesis that SARS-CoV-2 may avoid the robust TE expression response that often follows a viral infection. To examine this hypothesis, we expanded our data analysis to more cell types infected with SARS-CoV-2 and other viruses. For SARS-CoV-2, we added lung cancer cell lines A549 and Calu3, and iAT2 primary cells, which represent a young state of normal lung cells, as well as primary human airway epithelial (HAE) cells and peripheral blood mononuclear cells (PBMCs; Huang *et al*, 2020). As expected, we found that other viruses induced a marked activation of TE (Fig EV1). Importantly, we observed that the viral load affects the strength of the TE response to SARS-CoV-2 infection: the higher the viral load the stronger the TE upregulation (Fig 1B, note green heatmap at the bottom denoting the viral load in each sample).

Remarkably, although viral load explains much of the observed TE response (Fig 1C, $R^2 = 0.66$, $P = 0.0078$), incorporating the TE basal levels into the model improved the accuracy of the prediction of the TE response (Fig 1D, $R^2 = 0.78$, $P = 0.0095$). This is because for both low viral load and high viral load, the TE induction levels in the primary cells, which have a higher TE basal level, were mild compared with the transformed cell lines (Fig 1B). A549 cells had a stronger TE response than the primary NHBE cells although both had similar low viral load levels, while at higher viral load levels, Calu3 and A549 cells expressing ACE2—the virus entry receptor—had a stronger TE response than the iAT2 cells. This pattern was consistent among the different TE classes (Fig EV1A). Therefore, viral load is not solely responsible for the upregulation of TEs expression. The other two contributing factors are the identity of the virus, where SARS-CoV-2 induces weaker activation than IAV and other viruses (Fig EV1A), and the identity of the cells, where in primary cells that have a higher basal TE level, the TEs are less induced compared with transformed cell lines.

Upregulation of TEs is positively correlated with high IFN response

Because IFN expression had previously been associated with TE expression (Chiappinelli *et al*, 2015; Roulois *et al*, 2015; Chuong *et al*, 2016), we next tested the relationship between IFN genes (defined based on gene ontology, see [Materials and Methods](#) for details) and TEs expression during SARS-CoV-2 infection. In agreement with the more robust TE induction, we found a significantly

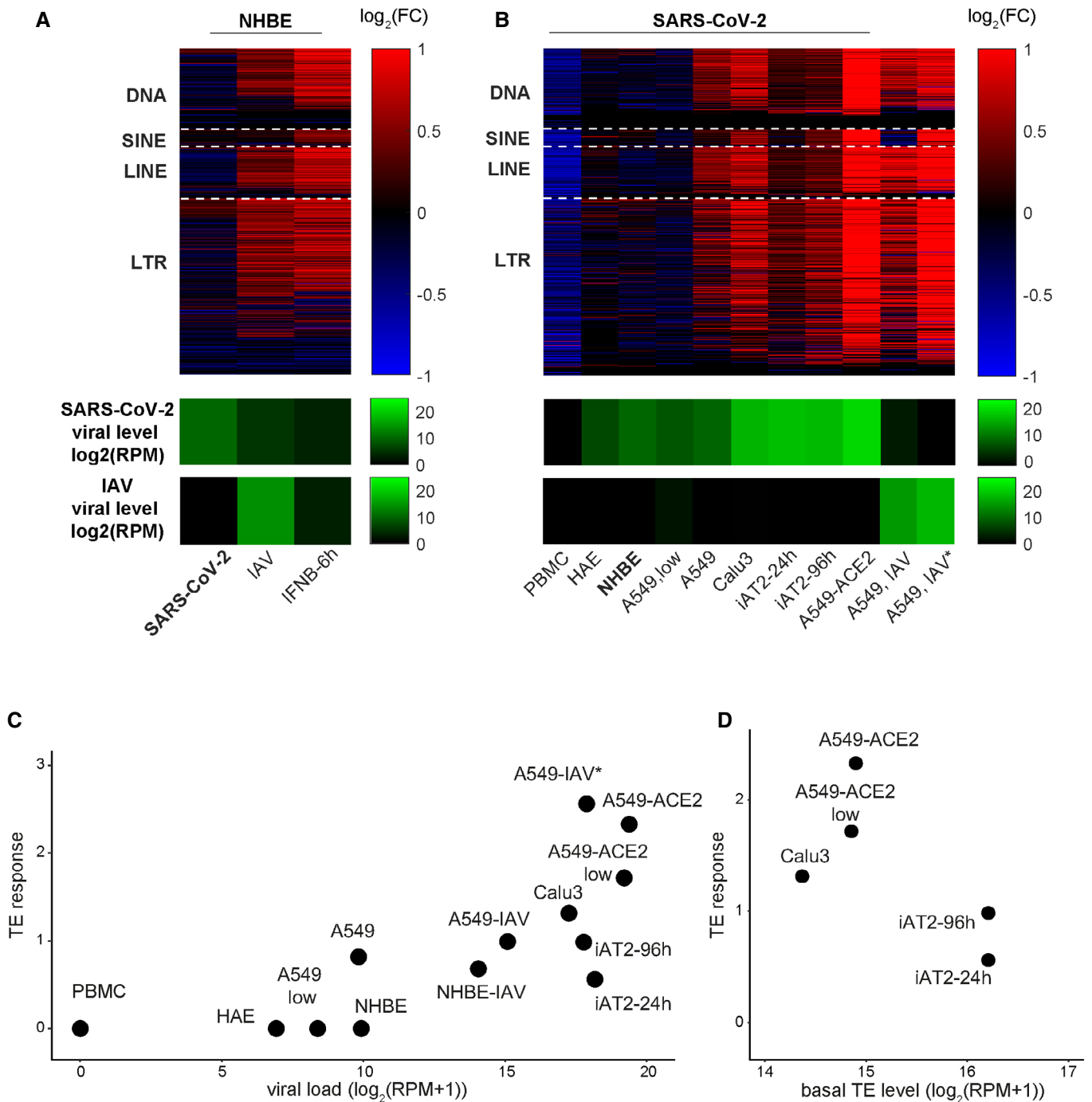


Figure 1. TEs expression changes in response to SARS-CoV-2 and IAV infections.

A \log_2 fold-change in expression level of TE subfamilies (DNA, SINE, LINE, and LTR) in NHBE cells in response to IFN β treatment and in response to SARS-CoV-2 and IAV infections. SARS-CoV-2 viral levels (green) are depicted in the bottom panel.
 B Same as (A) for SARS-CoV-2 infection in different cellular systems and IAV infection in A549 cell line.
 C, D TE induction levels are correlated with SARS-CoV-2 viral levels (C) and also with TE basal levels preinfection (D). Linear regression coefficients are 0.17 and -0.39 for viral load and basal TE level, respectively ($R^2 = 0.78$). PBMC were removed from the regression because they had essentially zero viral load. IAV* marks an independent dataset from Schmidt *et al.* The number of replicates for each sample is $n = 3$, except for IFN β treatment for which $n = 2$.

larger number of upregulated IFN genes (Dataset EV1) in the cancer cell lines 1 day after infection with SARS-CoV-2 compared with primary cells, including NHBE cells (Fig 2A). Importantly, NHBE cells

did respond to IAV infection and to IFN β treatment by a significant induction of IFN genes (Figs 2B and EV1B) and TE expression levels (Fig 1A) already after 4–12 h, excluding the possibility that these

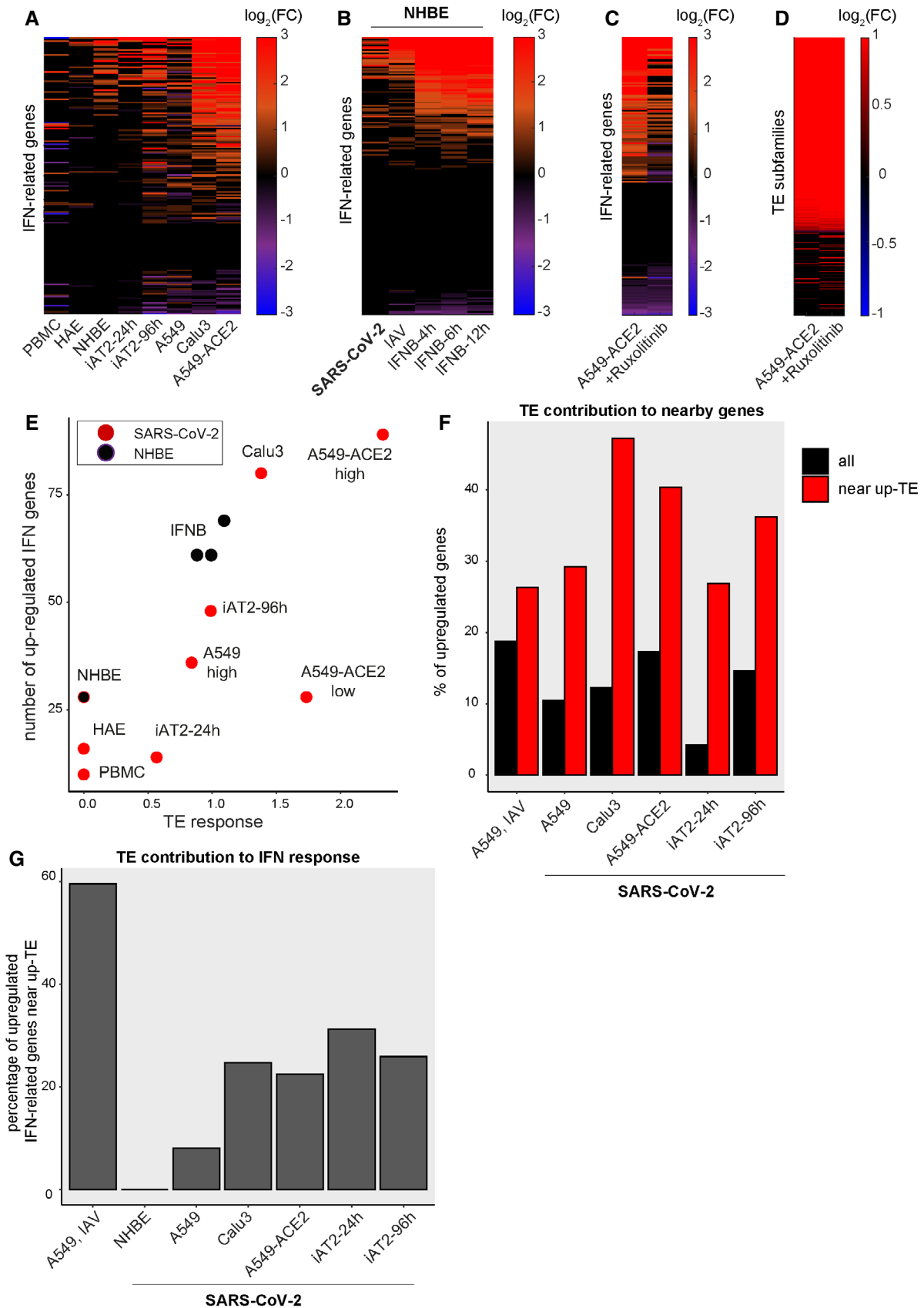


Figure 2.

Figure 2. TE induction precedes and predicts IFN response.

- A The IFN transcriptional response for SARS-CoV-2 infection in different cellular systems.
- B Same as (A) for NHBE and iAT2 cells. For A549 IAV there is one experiment from Blanco-Melo *et al* and one from Schmidt *et al* (marked by an asterisk).
- C The IFN transcriptional response of A549 cells overexpressing the ACE2 receptor dramatically decreases upon Ruxolitinib treatment.
- D TE upregulation persists even with Ruxolitinib treatment.
- E IFN transcriptional changes correlate with the TE induction levels among SARS-CoV-2-infected cells (red) and among NHBE cells (black). TE response is calculated as the 95-percentile of log₂ fold-change of all TE subfamilies.
- F The percentages of upregulated genes out of all genes and the percentages of upregulated genes that are near upregulated TEs in response to SARS-CoV-2 infection in different cellular systems.
- G The fraction of upregulated IFN-related genes that are located near upregulated TEs.

cells simply fail to activate TEs or induce IFN response. While infection of iAT2 cells with SARS-CoV-2 resulted in a mild but detectable TE response (Fig 1B, larger TE log fold-change [LFC] compared with NHBE cells, $P < 10^{-6}$ Kolmogorov–Smirnov test) 1 day postinfection, it showed almost no IFN response, even weaker than infected NHBE cells (smaller LFC of IFN genes, $P < 0.0001$ Kolmogorov–Smirnov test, Fig 2A). This mild TE response slightly increased after 4 days and was accompanied by a significant IFN response (Fig 2A). This suggests that the early TE response precedes the late IFN response in these cells.

The idea that TE activation is inducing IFN response following SARS-CoV-2 infection was also supported by A549 cells overexpressing the ACE2 receptor that were infected with SARS-CoV-2. These cells, when treated with Ruxolitinib, a JAK1/2 inhibitor that is known to reduce inflammatory response (Davis *et al*, 2011), showed reduced IFN response (Fig 2C), while the global TE response remains high (Fig 2D). This shows that although IFN response upregulates TE expression, TE overexpression is not solely dependent on the IFN response. Finally, although done on different tissues, genetic backgrounds, and multiplicity of infection (M.O.I.s), the magnitude of IFN changes strongly correlated with TE expression changes (Fig 2E). This correlation ($P < 0.05$, permutation test on Spearman correlation) was specific to IFN-related genes (and epifactors, see below), as other random groups of genes were not correlated with TE changes.

Since TEs were shown to function as regulatory elements, or enhancers, for adjacent host genes encoding critical innate immune factors (Chuong *et al*, 2016), we also tested the relation between expression changes in individual TEs and their neighboring genes. In general, genes that were adjacent to upregulated TEs, were prone to be upregulated as well (Fig 2F). Focusing on IFN-related genes (Dataset EV2) located near upregulated TEs revealed an even stronger effect, suggesting that this IFN genes induction by an adjacent TE also occurs following SARS-CoV-2 infection (Fig EV1C). This suggests that the TEs that are induced following SARS-CoV-2 infection have the capacity to serve as *cis*-regulatory enhancers to nearby genes including IFN-related genes.

To test whether the IFN genes that are induced in response to SARS-CoV-2 are in fact the result of TEs acting in *cis* or the result of TEs acting in *trans* or other cellular pathways, we tested the TEs located near the upregulated IFN genes. While ~60% of the upregulated IFN-related genes were located near upregulated TEs in response to IAV infection, only 10–30% of the upregulated IFN-related genes were located near upregulated TEs in response to SARS-CoV-2 infection (Fig 2G). This suggests that although following SARS-CoV-2 infection TEs can induce nearby IFN genes, the IFN response is not the result of TEs acting in *cis*. In concordance, gene

ontology (GO) analysis revealed that while the upregulated genes adjacent to upregulated TEs in response to IAV infection were enriched for cytokine receptor binding genes, no immune-related function was enriched among those in response to SARS-CoV-2 infection.

Taken together, these data suggest that while TE induction following IAV infection precedes and contributes to the expression of IFN-related genes (Schmidt *et al*, 2019), SARS-CoV-2 infection fails to upregulate the TEs coopted for immune activation. According to this hypothesis, TE induction is a crucial step in the activation of the antiviral immune response against RNA viruses. However, TE response to SARS-CoV-2 infection is limited: considerably less TEs are activated, the level of their upregulation is lower, and most importantly, their specificity is altered: those TEs that induce immune response are not the ones activated by SARS-CoV-2 infection.

The epigenetic signature of SARS-CoV-2-induced TEs

Regulation of TEs transcription is largely achieved through epigenetic silencing (Chuong *et al*, 2017). To understand the nature of the distinctive TE regulation of those TEs that do go up following SARS-CoV-2 infection, we investigated which specific histone modifications are found on the upregulated TEs before SARS-CoV-2 infection (Dataset EV2).

To this end, we used a large-scale dataset including multiple ChIP-seq profiles for histone modifications (HMs) in uninfected A549 cells (<http://www.encodeproject.org>). Since for A549 cells we also have pre- and post-SARS-CoV-2 infection data, it allowed us to examine the “epigenetic signature” (Sorek *et al*, 2019; relative enrichment of HMs) around the TEs that are upregulated postinfection of SARS-CoV-2 and IAV (Figs 3 and EV2A–C, Appendix Figs S1A–F and S2A–F, and Dataset EV3).

We found that the TEs that were upregulated in response to SARS-CoV-2 infection in A549 cells of all different classes were enriched for active histone marks in uninfected cells, with a subset of TEs marked by H3K36me3 as well as the combination of H3K27ac, H3K4me3, H3K79me2, and H3K9ac (Fig 3A–E and Appendix Fig S1A–F). This was consistent among different ChIP-seq experiments for the same histone modification (Appendix Fig S2A–F). SINES and DNA elements upregulated in response to SARS-CoV-2 infection were especially enriched for active marks compared with both random TEs and IAV-induced TEs. By contrast, LINES that were upregulated in response to SARS-CoV-2 infection were highly enriched for a bivalent signature of the repressive H3K9me3 mark together with the active H3K36me3 mark spread along the repeat and its flanking sequences (Figs 3F and EV2A–C). Strikingly, all classes of SARS-CoV-2-induced TEs were depleted for H3K27me3 in the uninfected cells (Fig 3E).

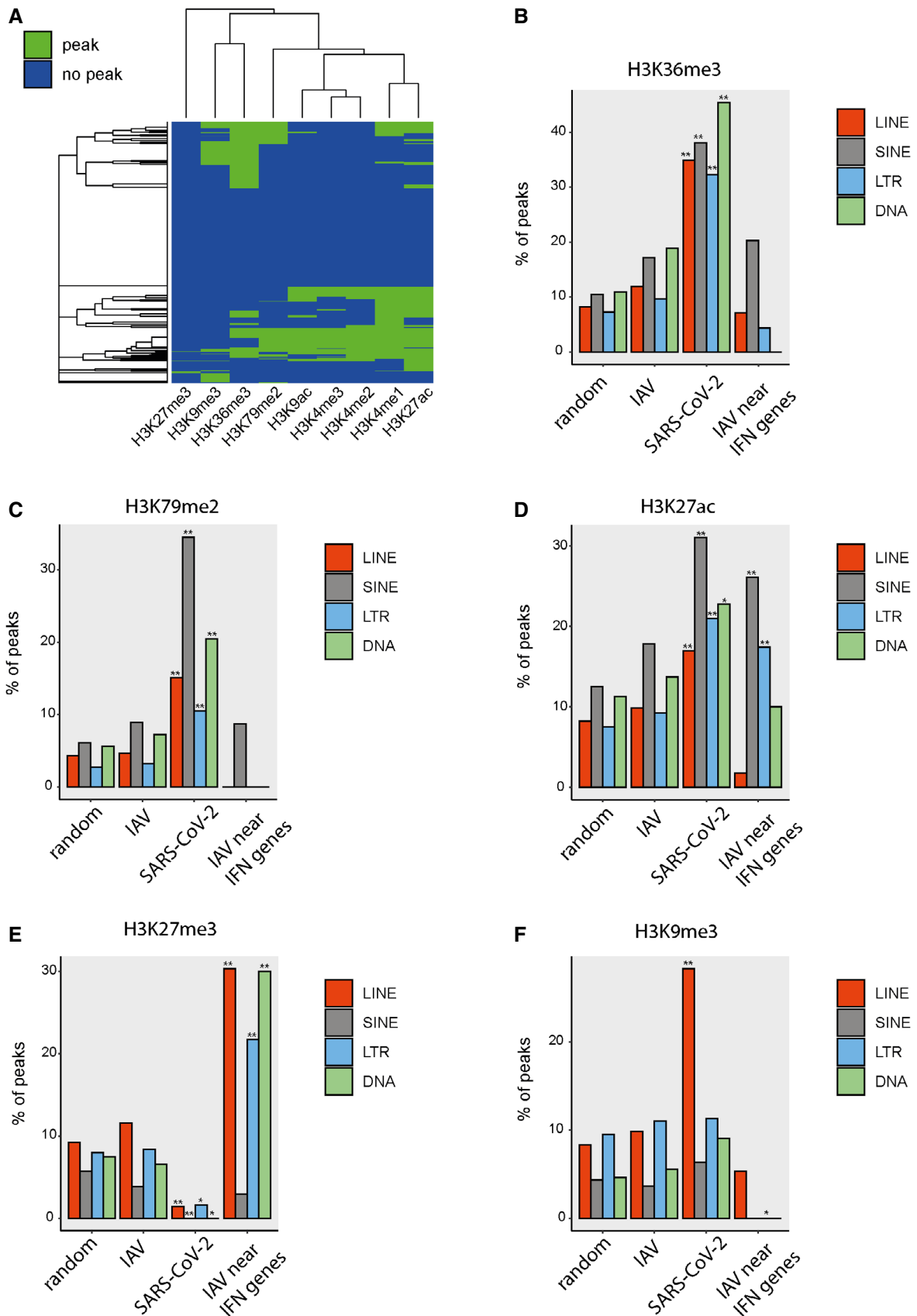


Figure 3.

Figure 3. TE classes that are induced by SARS-CoV-2 in A549 cells have a unique epigenetic profile.

A Hierarchical clustering of histone modifications signal in noninfected A549 cells around all upregulated TEs in response to SARS-CoV-2 infection in A549 cells.
 B–F Percentage of TEs with peaks of H3K36me3 (B), H3K79me2 (C), H3K27ac (D), H3K27me3 (E) and H3K9me3 (F) on SARS-CoV-2-induced TEs, IAV induced TEs, IAV induced TEs that reside in the proximity of upregulated IFN response genes and on all expressed TEs outside genes. Asterisks mark significance level of the difference compared with all expressed TE outside genes: one asterisk marks FDR-adjusted P -value < 0.05 and two asterisks mark $P < 0.001$, requiring at least a twofold difference in Fisher's exact test. $n = 39,528$ DNA elements, $n = 94,474$ LTR elements, $n = 151,223$ LINE elements, and $n = 214,879$ SINE elements. For full statistics see Dataset EV3.

This is in stark contrast to IAV-induced TEs, which show no such depletion (Fig 3E). Furthermore, the upregulated TEs that reside close to the upregulated IFN response genes following IAV infection were significantly enriched for H3K27me3 (Fig 3E). This suggests that SARS-CoV-2 failure to activate IFN response is the result of failure to evacuate H3K27me3 mark from TEs near IFN response genes.

Repeating the analysis at the TE family level, we found that the largest enrichment was in the specific L1 and L2 LINE families and in the MIR and Alu SINE families (Fig EV3A–D). These families showed a strong enrichment especially in H3K36me3 and in H3K79me2 and were depleted for H3K27me3. The H3K36me3 and H3K79me2 were also significantly enriched on the main ERV/LTR families. Overall, these results suggest that the TEs that are upregulated in response to SARS-CoV-2 in A549 cells have a distinct epigenetic profile, which differs from that of TEs upregulated by IAV, as well as from the general epigenetic profile of TEs in the human genome. Specifically, SARS-CoV-2-induced TEs are devoid of H3K27me3 and are comprised of two major subsets of TEs: (i) SINEs and DNA elements marked by a highly active chromatin profile, and (ii) a bivalent group of LINEs marked by both repressive and active marks, which keeps them in a poised state ready for infection-induced expression. This attests at the relative failure of the SARS-CoV-2-infected cells to activate TEs in a repressive chromatin state as IAV does.

The transcriptional signature of infected cells

Seeking a possible mechanism for the activation of TEs with this distinctive chromatin modification pattern, we searched for genes, expression of which changes in correlation with TE response. We focused on genes that had a high correlation with TE response both in the SARS-CoV-2 infection from Blanco-Melo *et al* and in the iAT2-infected cells from Huang *et al*. We found that the inversely correlated genes were enriched for mitochondrial-related genes and processes (Fig 4A and Dataset EV4, and see [Materials and Methods](#)), consistent with previous reports (Chung *et al*, 2019). By contrast, genes that were positively correlated with TE response among all samples were enriched, in addition to type I interferon production, for chromatin, DNA and enhancer binding, RNA Pol-II binding, transcription factor and cofactor binding as well as histone binding, demonstrating a clear epigenetic and chromatin-related signature (Fig 4A and Dataset EV4).

Indeed, intersecting the positively correlated genes with all genes that encode for chromatin binding proteins, or epifactors (<https://epifactors.autosome.ru/>), showed highly significant enrichment (Fig 4B, green, 44 genes, fold-enrichment = 2.33, $P < 10^{-7}$, Fisher's exact test). Interestingly, SETD2, the human H3K36 lysine trimethylase, was among the epifactors that were positively correlated with TE response (Fig 4C). H3K36me3 marks gene bodies of active

genes. In addition, SETD2 methylation of STAT1 is crucial for interferon response (Chen *et al*, 2017), and its H3K36 methylation contributes to ISG activation, pointing at its role in the cellular response to viral infection. Finally, recent evidence shows that SETD2 is essential for microsatellite stability, implicating its role in nongenic transcriptional regulation (Li *et al*, 2013a). Interestingly, SETD2 is also among the interacting proteins of SARS-CoV-2, and compared with other interactomes of coronaviruses as well as of different IAV strains, we found that SETD2 is specific for SARS-CoV-2 (Dataset EV5).

We therefore searched for more epifactors in the SARS-CoV-2 interacting proteins. Among the 10 epifactors that interact with SARS-CoV-2, we found the histone acetylation-related proteins BRD2, BRD4, which were highly correlated with the TE response (Fig 4C), and HDAC2, all of which are, once again, SARS-CoV-2-specific (Dataset EV5). We also found two TE-related epifactors: the SARS-CoV-2-specific interacting epifactor DDX21, a DNA damage and dsRNA sensing protein, and MOV10, an RNA helicase that also restricts LINE expression (Li *et al*, 2013b), which interacts with SARS-CoV-2, as well as with IAV. These observations suggest that SARS-CoV-2 may affect TE expression through interaction with a subset of specific epifactors.

Discussion

In this study, we reanalyzed published data to examine the link between SARS-CoV-2 infection and transcriptional activation of transposable elements (TEs). We find that in normal lung epithelial cells, SARS-CoV-2 does not induce global upregulation of TEs, as observed for IAV and other RNA viruses (Fig 4D). This phenomenon is in correlation with the viral load and with the intensity of IFN response in the infected cells. The low IFN response after infection is also correlated with high basal TE expression in the uninfected cells (Franceschi *et al*, 2018). Interestingly, a recent large-scale study of TE expression in different tissues during aging demonstrated that TE expression levels are gradually elevated in most tissues as a function of age (preprint: Bogu *et al*, 2019).

Since TE expression rises with age (preprint: Bogu *et al*, 2019; Pehrsson *et al*, 2019), and age is the most significant risk factor for COVID-19-related death, we hypothesize that high basal level of TE expression desensitizes the TE induction response to viral infection, explaining the age-related decline in survival. In contrast, young people should benefit from higher SARS-CoV-2-induced TE overexpression that, in turn, prompts IFN response early in the disease course (Park & Iwasaki, 2020). Our “TE desensitization” model makes several predictions. First, it anticipates that induction of TEs would precede the IFN response. Second, the cellular TE activation response to SARS-CoV-2 should be associated with basal levels of

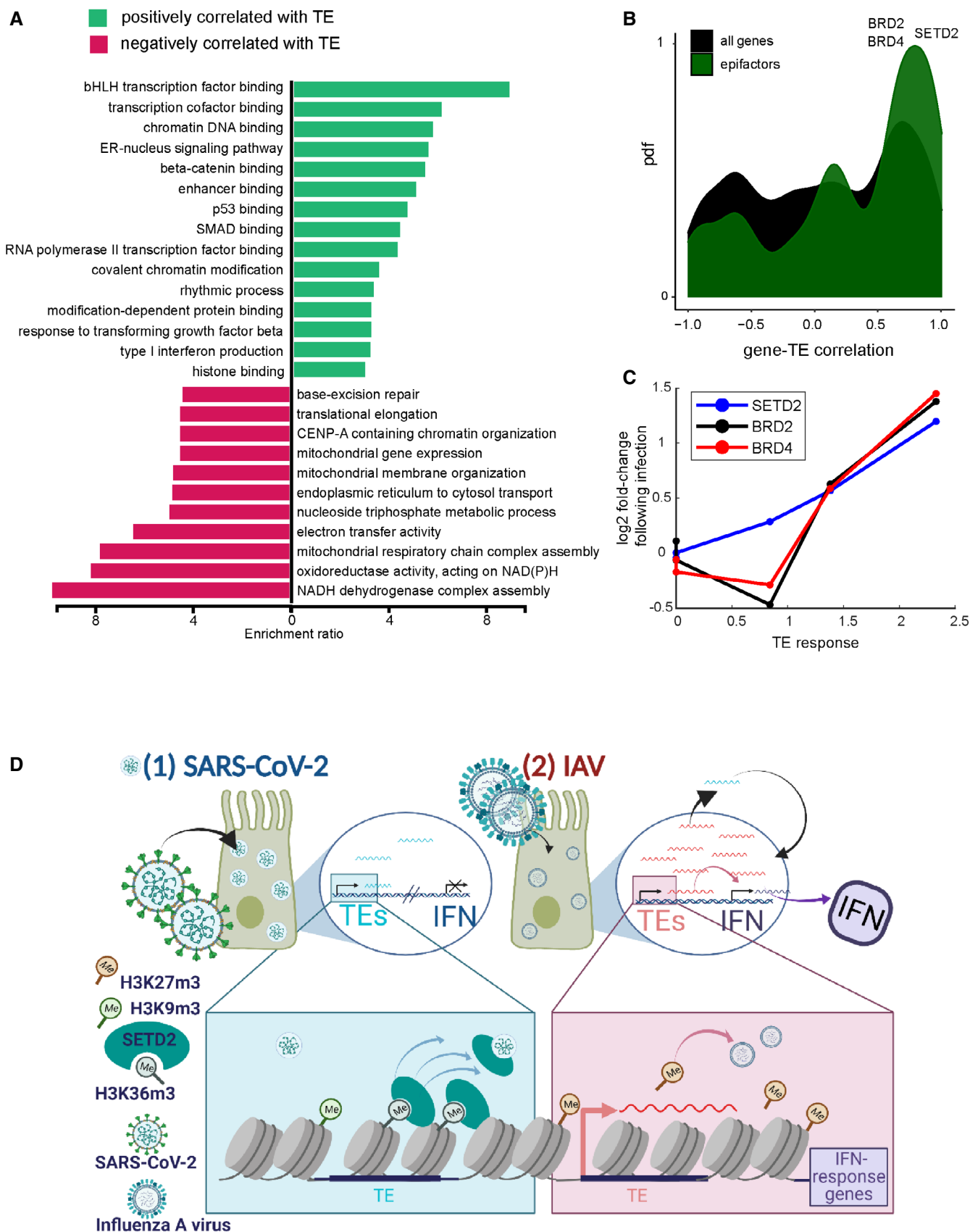


Figure 4.

Figure 4. TE response is correlated with epigenetic and mitochondrial gene expression changes.

- A Gene ontology analysis of genes correlated (green) and inversely correlated (red) with TE response to SARS-CoV-2 infection.
- B Distribution of TE response-gene correlations for genes that increase with TE expression in iAT2 cells. Black represents correlation distribution of all genes. Correlation distribution of epifactors are in green.
- C The LFC of gene expression following SARS-CoV-2 infection vs the TE response in different systems for SETD2, BRD2 and BRD4, three SARS-CoV-2 interactome-related epigenetic factors. $n = 3$ replicates for each of mock and infected cells.
- D A model for SARS-CoV-2 impaired TE activation. (1) SARS-CoV-2 infection of normal lung epithelial cells results in low viral load, weak IFN response, and modest TE upregulation, predominantly of TEs not located in the proximity of IFN response genes, and which are enriched with H3K9me3 and H3K36me3. A specific SARS-CoV-2 interacting protein may sequester SETD2 away from the otherwise H3K36me3-enriched TEs, enabling their transcription. (2) IAV infection induces TEs activation and IFN response. H3K27me3 marked TEs are prone to be upregulated, inducing near IFN gene activation. The illustration was created with BioRender.com.

TE expression, where lower basal levels predict stronger cellular responses. As such, and third, it anticipates that in older cells, which would have a higher basal level of TEs and would be “TE desensitized,” the TE activation response will be significantly milder. This would lead, fourth, to a less effective IFN induction response, allowing SARS-CoV-2 to operate “under the radar” at early disease stages when the viral load is still low. Finally, at later stages, as the viral load accumulates, selected TEs are induced. Those TEs are characterized by histone modifications, regulators of which are found to interact with the SARS-CoV-2 proteome.

As our model predicts, we showed that SARS-CoV-2 elicits a weaker TE activation response compared with other viruses, and a weaker TE activation response in primary cells compared with cancer cells. These primary cells have higher initial levels of TEs. We speculate that in old normal cells, low TE expression fails to induce viral mimicry and thus fails to initiate early innate immune reaction. This is in line with previously published data that link TE expression to immune reaction (Chiappinelli *et al*, 2015; Roulois *et al*, 2015; Kitsou *et al*, 2021; Marston *et al*, 2021; Tovo *et al*, 2021), as well as studies that show suppression of dsRNA-activated early host responses following SARS-CoV-2 infection (Domizio *et al*, 2022; Neufeldt *et al*, 2022). In later stages of the disease and especially in severe cases, when the viral load significantly increases, TEs are eventually induced and generate the delayed IFN response (Balestrieri *et al*, 2021). Although it is not known whether TE activation contribute to IFN response and therefore to immune system activation in viral infections, our own analysis shows a highly significant correlation between TE activation and the induction of IFN response to viral infections.

Cells that showed a large TE response also had a strong IFN response to SARS-CoV-2 infection, but this was not the result of upregulation of adjacent TEs. This is because the immune response can be induced by other mechanisms besides TEs acting as enhancers for immune genes (Chuong *et al*, 2017). For example, the TEs themselves can also act in *trans*, where their dsDNA and dsRNA products are sensed by the cells. Therefore, the impaired TE activation following SARS-CoV-2 infection may have a double impact, yielding the delayed IFN response seen in COVID-19 patients.

In addition, SARS-CoV-2 viral load seems to be linked with TE expression level changes. Accordingly, we were able to compare TEs induced following a rigorous SARS-CoV-2 infection to that of IAV infection in A549 cells. Interestingly, the TEs that are upregulated in response to SARS-CoV-2 in A549 cells showed a distinct epigenetic profile, which differed from that of TEs upregulated by IAV, as well as from the general epigenetic profile of TEs in the human genome. Specifically, SARS-CoV-2-induced TEs are devoid of H3K27me3 and enriched for H3K36me3, a subset of which is

bivalently marked also with H3K9me3 (Fig EV2A–C). This atypical pattern was identified as a mark for poised enhancers that control surrounding gene expression (Barral *et al*, 2022). Here, the activation is potentially mediated by the H3K36 trimethylase SETD2. SETD2 expression is closely correlated with TE induction, and it specifically interacts with the SARS-CoV-2 NSP9 protein. One caveat of our analysis is that as we analyze only uniquely mapped reads, we may potentially miss data on highly repetitive TEs which may be relevant to the regulation of the IFN response.

Together, these data suggest a model where SARS-CoV-2 entry modifies SETD2 deposition or activity, leading to aberrant H3K36me3 enrichment on a subset of TEs (Fig 4D). Consequently, these TEs are transcribed, leading to the induction of the IFN response. In older age, where basal TE levels are high, the changes in chromatin structure and histone modifications will have minor effects, either because the TEs are promiscuously expressed or because of the compromised TE-induced IFN response.

Overall, by reanalyzing data published by Huang *et al*, Blanco-Melo *et al*, and others, we provide evidence that unlike other viruses, which strongly induce TEs following infection (Macchietto *et al*, 2020), SARS-CoV-2 infection has a relatively weak effect on TE expression in primary lung epithelia. One potential consequence of this reduced TE induction in normal cells is weak activation of the IFN response which usually responds to TEs via a positive feedback loop mechanism (Burns, 2020).

Materials and Methods

Datasets

The original sequencing datasets for IAV infection in A549 cells and the Blanco-Melo *et al* datasets can be found on the NCBI Gene Expression Omnibus (GEO) server under the accession numbers GSE133329 and GSE147507, respectively. The data produced by Huang *et al* can be found at GSE153277. The PBMC and HAE data can be found at the Genome Sequence Archive in BIG Data Center under the accession number CRA002390 and on GEO under the accession number GSE153970, respectively. Details of the ChIP-Seq datasets from ENCODE project used in this study are in Table EV1.

Preprocessing and alignment

Raw reads from GSE147507 and GSE133329 were trimmed to remove Illumina adapters using the Trimmomatic software version 0.39. We used the STAR aligner version 2.7.1a to align raw reads to human RefSeq reference genome (GRCh38). We used the

parameters `--outFilterMultimapNmax 100` and `--winAnchor MultimapNmax 200` to allow for large numbers of multi-mapped reads for downstream analysis of TEs.

TE and gene expression quantification

Gene and TE expression were analyzed separately. Genes were considered significantly up-(down-) regulated if they had at least 1.5-fold difference and FDR-corrected P -value < 0.05 . We quantified TE expression changes both at the level of TE subfamilies and at the level of individual repeat loci. For TE subfamily quantification, we used Tetranscripts from the TEToolKit. The Tetranscript algorithm quantifies TE subfamilies and genes simultaneously by assigning together multi-mapped reads which are associated with the same TE subfamily. Human repeat annotations for hg38 were downloaded from Tetranscripts site. Tetranscripts was run using `--mode multi` and `-n TC`. This was followed by differential expression analysis using DESeq2. TE response was quantified using the 95-percentile of the log₂ fold-change of all TE subfamilies.

To quantify gene expression and to determine the locations of individual TEs that change in expression, we used featureCounts v2.0.0 from the Subread package which uses only uniquely mapped reads. Simple repeat elements were removed prior to the analysis. The TE and gene count matrices were combined, followed by DESeq2 to compare between mock and infected cells. Individual TEs were considered significantly up- (down-) regulated if they had at least 1.5-fold difference and P -value < 0.05 . To robustly calculate the IFN response, we used all genes associated with the following GO terms: GO_0035458, GO_0035457, GO_0035456, GO_0035455, GO_0034340, as well as genes associated with the following pathways in pathcards (<https://pathcards.genecards.org/Pathway>): Immune response IFN alpha/beta signaling super-pathway and pathways 2,747, 2,388, 213 (Dataset EV1). IFN-related genes were considered significantly up- (down-) regulated if they had at least 1.5-fold difference and P -value < 0.05 . The list of epifactors was downloaded from <https://epifactors.autosome.ru/>.

ChIP-Seq signal quantification

To calculate the ChIP profile of different histone modification on TEs, we used the processed output files from the ENCODE project, which are filtered for the ENCODE blacklist regions (see Table EV1). In cases where the data were based on more than one replicate, we used the filtered file based on the IDR method. For each TE, we defined the flanking region as the TE location and its surrounding 500 bp up- and downstream. If the flanking region intersection with ChIP peak locations was nonempty, then this TE was considered as TE with peak. For histone modification clustering (Figs 2,3 and Appendix Fig S1A–F), we used the Jaccard metric. Enrichment of peaks on upregulated TEs was calculated using hypergeometric test and was then FDR-corrected. Genome tracks were produced using the ggbio package in R.

TE-gene correlation analysis

For TE response-gene correlation, we used Spearman rank correlation between the 95-percentile TE subfamily log fold-change and the log fold-change of genes in the SARS-CoV-2-infected cells. We

used NHBE, A549, Calu3, and A549-ACE2 cells infected using M.O.I. = 2 in addition to A549 cells infected with M.O.I. = 0.2. Genes were defined as highly correlated and inversely correlated if the correlated was larger than 0.97 and smaller than -0.97 among these samples, respectively, and if, in addition, the direction of change aligned with that in the iAT2 cell between d1 and d4 after infection. These cutoffs corresponded to approximately the top and bottom 5%.

Interactome analysis

For interactome analysis, we downloaded publicly available interactome lists (for details, see Table EV2). A gene was considered in SARS-CoV-2 interactome if it was included in at least three of five sources, and as a SARS-CoV-1 interacting partner if it was included in both relevant sources. Dataset All the software and resources used for data analysis are listed in Table EV3.

Data availability

This study has not produced novel datasets.

Expanded View for this article is available online.

Acknowledgements

We thank Prof. Sagiv Shifman, Prof. Lior David, and Dr. Asaf Marko for critical reading of the manuscript, and all EM and SS laboratory members for discussions. MS was supported by Edmond de Rothschild and HUJI internal funding for COVID-19 research. SS was supported by the Israel Science Foundation grant No. 761/17. The illustration in the synopsis Fig 4 image has been created with BioRender.com.

Author contributions

Matan Sorek: Conceptualization; data curation; software; formal analysis; validation; investigation; visualization; writing – original draft. **Eran Meshorer:** supervision; funding acquisition; visualization; writing – review and editing. **Sharon Schlesinger:** Conceptualization; supervision; funding acquisition; project administration; writing – review and editing.

In addition to the CRediT author contributions listed above, the contributions in detail are:

MS, EM, and SS conceptualized the study, designed the analyses, revised, and edited the manuscript. MS performed the bioinformatic analyses, interpreted the results, and wrote the manuscript. EM and SS provided guidance throughout the studies, assisted in writing the manuscript, and provided the funding for the study.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

References

- Attig J, Young GR, Stoye JP, Kassiotis G (2017) Physiological and pathological transcriptional activation of endogenous retroelements assessed by RNA-sequencing of B lymphocytes. *Front Microbiol* 8: 2489
- Badarinarayan SS, Sauter D (2021) Switching sides: how endogenous retroviruses protect us from viral infections. *J Virol* 95: e02299-20

- Balestrieri E, Minutolo A, Petrone V, Fanelli M, Iannetta M, Malagnino V, Zordan M, Vitale P, Charvet B, Horvat B et al (2021) Evidence of the pathogenic HERV-W envelope expression in T lymphocytes in association with the respiratory outcome of COVID-19 patients. *EBioMedicine* 66: 103341
- Barral A, Pozo G, Ducrot L, Papadopoulos GL, Sauzet S, Oldfield AJ, Cavalli G, Dejardin J (2022) SETDB1/NSD-dependent H3K9me3/H3K36me3 dual heterochromatin maintains gene expression profiles by bookmarking poised enhancers. *Mol Cell* 82: 816–832
- Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Moller R, Jordan TX, Oishi K, Panis M, Sachs D et al (2020) Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* 181: 1036–1045
- Bogu GK, Reverter F, Marti-Renom MA, Snyder MP, Guigó R (2019) Atlas of transcriptionally active transposable elements in human adult tissues. *BioRxiv* <https://doi.org/10.1101/714212> [PREPRINT]
- Burns KH (2020) Our conflict with transposable elements and its implications for human disease. *Annu Rev Pathol* 15: 51–70
- Chen K, Liu J, Liu S, Xia M, Zhang X, Han D, Jiang Y, Wang C, Cao X (2017) Methyltransferase SETD2-mediated methylation of STAT1 is critical for interferon antiviral activity. *Cell* 170: 492–506
- Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM, Snyder A et al (2015) Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* 162: 974–986
- Chung N, Jonaid GM, Quinton S, Ross A, Sexton CE, Alberto A, Clymer C, Churchill D, Navarro Leija O, Han MV (2019) Transcriptome analyses of tumor-adjacent somatic tissues reveal genes co-expressed with transposable elements. *Mob DNA* 10: 39
- Chuang EB, Elde NC, Feschotte C (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351: 1083–1087
- Chuang EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: From conflicts to benefits. *Nat Rev Genet* 18: 71–86
- Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 29: 1046–1051
- Domizio JD, Gulen MF, Saidoune F, Thacker VV, Yatim A, Sharma K, Nass T, Guenova E, Schaller M, Conrad C et al (2022) The cGAS-STING pathway drives type I IFN immunopathology in COVID-19. *Nature* 603: 145–151
- Ferrarini MG, Lal A, Rebollo R, Gruber AJ, Guarracino A, Gonzalez IM, Floyd T, de Oliveira DS, Shanklin J, Beausoleil E et al (2021) Genome-wide bioinformatic analyses predict key host and viral factors in SARS-CoV-2 pathogenesis. *Commun Biol* 4: 590
- Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A (2018) Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat Rev Endocrinol* 14: 576–590
- Galani I-E, Rovina N, Lampropoulou V, Triantafyllia V, Manioudaki M, Pavlos E, Koukaki E, Fragkou PC, Panou V, Rapti V (2020) Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. *Nat Immunol* 22: 32–40
- Gazquez-Gutierrez A, Witteveldt J, Heras SR, Macias S (2021) Sensing of transposable elements by the antiviral innate immune system. *RNA* 27: 735–752
- Hale BG (2021) Antiviral immunity triggered by infection-induced host transposable elements. *Curr Opin Virol* 52: 211–216
- Huang J, Hume AJ, Abo KM, Werder RB, Villacorta-Martin C, Alysandratos KD, Beermann ML, Simone-Roach C, Lindstrom-Vautrin J, Olejnik J et al (2020) SARS-CoV-2 infection of pluripotent stem cell-derived human lung alveolar type 2 cells elicits a rapid epithelial-intrinsic inflammatory response. *Cell Stem Cell* 18: S1934–S5909
- Hung T, Pratt GA, Sundararaman B, Townsend MJ, Chaivorapol C, Bhangale T, Graham RR, Ortman W, Criswell LA, Yeo GW et al (2015) The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science* 350: 455–459
- Karaderi T, Bareke H, Kunter I, Seytanoglu A, Cagnan I, Balci D, Barin B, Hocaoglu MB, Rahmioglu N, Asilmaz E et al (2020) Host genetics at the intersection of autoimmunity and COVID-19: a potential key for heterogeneous COVID-19 severity. *Front Immunol* 11: 586111
- Kitsou K, Kotanidou A, Paraskevis D, Karamitros T, Katzourakis A, Tedder R, Hurst T, Sapounas S, Kotsinas A, Gorgoulis V et al (2021) Upregulation of human endogenous retroviruses in bronchoalveolar lavage fluid of COVID-19 patients. *Microbiol Spectr* 9: e0126021
- Krischuns T, Günl F, Henschel L, Binder M, Willemsen J, Schloer S, Rescher U, Gerlt V, Zimmer G, Nordhoff C et al (2018) Phosphorylation of TRIM28 enhances the expression of IFN- β and proinflammatory cytokines during HPAIV infection of human lung epithelial cells. *Front Immunol* 9: 2229
- Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li GM (2013a) The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. *Cell* 153: 590–600
- Li X, Zhang J, Jia R, Cheng V, Xu X, Qiao W, Guo F, Liang C, Cen S (2013b) The MOV10 helicase inhibits LINE-1 mobility. *J Biol Chem* 288: 21148–21160
- Macchietto MG, Langlois RA, Shen SS (2020) Virus-induced transposable element expression up-regulation in human and mouse host cells. *Life Sci Alliance* 3: e201900536
- Marston JL, Greenig M, Singh M, Bendall ML, Duarte RR, Feschotte C, Iniguez LP, Nixon DF (2021) SARS-CoV-2 infection mediates differential expression of human endogenous retroviruses and long interspersed nuclear elements. *JCI Insight* 6: e147170
- Neufeldt CJ, Cerikan B, Cortese M, Frankish J, Lee JY, Plociennikowska A, Heigwer F, Prasad V, Joecks S, Burkart SS et al (2022) SARS-CoV-2 infection induces a pro-inflammatory cytokine response through cGAS-STING and NF-kappaB. *Commun Biol* 5: 45
- Paludan SR, Mogensen TH (2022) Innate immunological pathways in COVID-19 pathogenesis. *Sci Immunol* 7: eabm5505
- Park A, Iwasaki A (2020) Type I and type III interferons – induction, signaling, evasion, and application to combat COVID-19. *Cell Host Microbe* 27: 870–878
- Pehrsson EC, Choudhary MN, Sundaram V, Wang T (2019) The epigenomic landscape of transposable elements across normal human development and anatomy. *Nat Commun* 10: 1–16
- Rebendenne A, Valadao ALC, Tauziet M, Maarifi G, Bonaventure B, McKellar J, Planes R, Nisole S, Arnaud-Arnould M, Moncorge O et al (2021) SARS-CoV-2 triggers an MDA-5-dependent interferon response which is unable to control replication in lung epithelial cells. *J Virol* 95: e02415-20
- Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, Han H, Liang G, Jones PA, Pugh TJ et al (2015) DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* 162: 961–973
- Schmidt N, Domingues P, Golebiowski F, Patzina C, Tatham MH, Hay RT, Hale BG (2019) An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *Proc Natl Acad Sci USA* 116: 17399–17408

Shen SS, Nanda H, Aliferis C, Langlois RA (2022) Characterization of influenza a virus induced transposons reveals a subgroup of transposons likely possessing the regulatory role as eRNAs. *Sci Rep* 12: 2188

Sorek M, Cohen LRZ, Meshorer E (2019) Open chromatin structure in PolyQ disease-related genes: a potential mechanism for CAG repeat expansion in the normal human population. *NAR Genom Bioinform* 1: e3

Takayama K (2020) In vitro and animal models for SARS-CoV-2 research. *Trends Pharmacol Sci* 41: 513–517

Tokuyama M, Kong Y, Song E, Jayewickreme T, Kang I, Iwasaki A (2018) ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc Natl Acad Sci USA* 115: 12565–12572

Tovo PA, Garazzino S, Dapra V, Pruccoli G, Calvi C, Mignone F, Alliaudi C, Denina M, Scolfaro C, Zoppo M et al (2021) COVID-19 in

children: expressions of type I/II/III interferons, TRIM28, SETDB1, and endogenous retroviruses in mild and severe cases. *Int J Mol Sci* 22: 7481

Wang M, Wang LY, Liu HZ, Chen JJ, Liu D (2021) Transcriptome analyses implicate endogenous retroviruses involved in the host antiviral immune system through the interferon pathway. *Viral Sin* 36: 1315–1326



License: This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.