Impaired activation of Transposable Elements in SARS-CoV-2 infection

Matan Sorek[1,2,*], Eran Meshorer[1,2,*] and Sharon Schlesinger[3,*]

[1]Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem 9190100, Israel,

[2]Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem 9190100, Israel

[3]Department of Animal Sciences, Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot 7610001, Israel

[*]Corresponding authors (matan.sorek@mail.huji.ac.il; eran.meshorer@mail.huji.ac.il; sharon.shle@mail.huji.ac.il)

Key points:

Unlike other viruses, SARS-CoV-2 invokes a weak and inefficient transposable element (TE) response

TE induction precede and predict IFN response

Basal TE expression and viral load explain immune responses

Since TEs increase with age, we propose that 'TE desensitization' leads age-related COVID19 death

Key words:

Transposable elements; SARS-CoV-2; interferon response; lung epithelia; COVID-19; epigenetics; aging

Abstract

Transposable element (TE) transcription is induced in response to viral infections. TE induction triggers a robust and durable interferon (IFN) response, providing a host defense mechanism. Still, the connection between SARS-CoV-2 IFN response and TEs remained unexplored. Here, we analyzed TE expression changes in response to SARS-CoV-2 infection in different human cellular models. We find that compared to other viruses, which cause global upregulation of TEs, SARS-CoV-2 infection results in a significantly milder TE response in both primary lung epithelial cells and in iPSC-derived lung alveolar type 2 cells. TE activation precedes, and correlates with, the induction of IFN-related genes, suggesting that the limited activation of TEs following SARS-CoV-2 infection may be the reason for the weak IFN response. Diminished TE activation was not observed in lung cancer cell lines with very high viral load. Moreover, we identify two variables which explain most of the observed diverseness in immune responses: basal expression levels of TEs in the pre-infected cell, and the viral load. Since basal TE levels increase with age, we propose that 'TE desensitization' leads to age-related death from COVID19. Finally, analyzing the SARS-CoV-2 interactome, as well as the epigenetic landscape around the TEs that are activated following infection, we identify SARS-CoV-2 interacting proteins, which may regulate chromatin structure and TE transcription in response to a high viral load. This work provides a mechanistic explanation for SARS-CoV-2's success in its fight against the host immune system, and suggests that TEs could be used as sensors or serve as potential drug targets for COVID-19.

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 of lung function and even death [1]. The fatality of SARS-CoV-2 increases substantially with age [2]. 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; IFN-III). This is partially achieved by active suppression of dsRNA-activated early host responses [3, 4] and is supported by several lines of evidence.

First, Blanco-Melo et al. assessed the transcriptional response to SARS-CoV-2 infection in different cellular models [1] and found that SARS-CoV-2 does not elicit robust IFN expression in lung epithelial cells, while multiple pro-inflammatory cytokines were highly expressed. Second, Huang et al. 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 [5]. The authors found that SARS-CoV-2 infection in these cells results in an inflammatory phenotype with an activation of the NF-κB pathway, and a delayed IFN signaling response. Other recently published data support these conclusions [6, 7], although the outcome of IFN response on the virus and its host are not clear [8]

Transposable elements (TEs) are abundant sequences in the mammalian genome that contain multiple regulatory elements and can amplify in a short evolutionary timescale [9]. Lately, it was found that TEs have coopted to shape the transcriptional network underlying the IFN response, and some TEs have dispersed numerous IFN-inducible enhancers independently in diverse mammalian genomes [10]. TEs are

also enriched in CD8+ T lymphocytes-specific enhancers, suggesting that their upregulation might influence not only the innate but also the adaptive, immune response [11]. What's more, genome-wide induction of TEs is common during some viral infections in humans, and transposon induction is part of the first wave response to viral infection [12, 13]. For example, infection by highly pathogenic avian influenza viruses elicits TEs induction [14]. As a result, TE dsRNA is formed, recognized by the host sensors and activates both the NFkB and the IFN pathways, thus enhancing immune response [15]. Collectively, these findings suggest a link between the level of TE overexpression and the intensity of the IFN response.

This link is also evident in aging. Aging is associated with sterile inflammation which include erroneous IFN response, or the 'Inflammaging' phenomena [16]. On the other hand, at the molecular level, aging is associated with decreased heterochromatin-associated marks, e.g. H3K9me3 [17] and DNA methylation [18]. Consequently, TE basal expression increases with age [19, 20]. Interestingly, early antiviral IFN responses are impaired and delayed in aged individuals, resulting in increased risk of COVID-19 complication [21].

Here, we suggest that SARS-CoV-2 infected cells that fail to activate an immediate and effective TE response, will demonstrate a late immune response. The late TE activation, which is induced only by very high viral levels, results in a delayed IFN response linked with severe COVID-19. Since TE expression rises with age [19, 20], the induction of the TE response to viral infection becomes less effective in old individuals. This might explain the aging-associated decline in effective innate response to SARS-CoV-2 infection and increased mortality. In contrast, young people benefit from higher SARS-CoV-2 induced TE overexpression that, in turn, prompts IFN response early in the disease course [6]. Taken together, this model explains both SARS-CoV-2's efficiency, as well as the reduced capacity of old individuals to cope with SARS-CoV-2's infection.

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 [12], we first analyzed recently published datasets of primary lung epithelial cells (NHBE) infected with influenza (IAV) [22] or SARS-CoV-2 [1]. NHBE cells mimic infected human lung cells, showing cytopathic effects after SARS-CoV-2 infection [23]. In these cells IAV infection caused, as expected, a global increase in TE subfamilies expression, but SARS-CoV-2 did not (Figure 1A). Interestingly, IFNβ treated cells similarly induced TE expression levels (Figure 1A). This raised an intriguing hypothesis that SARS-CoV-2 may avoid the robust TE expression response that often follows a viral infection, potentially explaining the ineffective immune response to SARS-CoV-2 infection.

To examine this hypothesis, we expanded our data analysis to more cell types infected with 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 [5]. We first found that the viral load affects the strength of the TE response to SARS-CoV-2 infection, where the higher the viral load the stronger the TE upregulation (Figure 1B, note the green heat map in the bottom denoting the viral load of each sample). This is in line with the view that TE upregulation serves as a defense mechanism following virus detection. However, the viral load alone did not explain the observed TE response. Instead, for both low viral load and high viral load, the TE induction levels in the primary cells were mild compared to the transformed cell lines (Figure 1B-C); A549 cells had a stronger TE response compared to the primary NHBE cells although both had similar low viral load levels, while at higher viral load levels Calu3 and A549 cells expressing the ACE2 receptor had a stronger TE response compared to the iAT2 cells. This pattern was consistent among the different TE classes (Figure S1A).

Because cancer cell lines, unlike primary cells, are somewhat “ageless”, we thought that this might explain the discrepancy between the TE responses of primary and cancer cells. Interestingly**,** a recent

large-scale study of TE expression in different tissues during aging, demonstrated that TEs are gradually elevated in most tissues as a function of age [20]. As age is the most significant risk factor for COVID19-related death, we hypothesized that high basal level of TE expression desensitizes the TE induction response to viral infection, explaining the age-related decline in survival. Indeed, we observed a clear reverse-correlation between the basal TE expression in the uninfected cells and the TE activation response following infection (Figure 1C, see circles' fill color). For example, the initial TE levels prior to infection in the NHBE normal cells, which show a milder TE response to infection, were higher than those in the A549 cancer cells, which robustly respond with a general TE induction. Interestingly, the NHBE cells with the high basal TE expression were taken from a 79 years old female, supporting the link between old age and high TE expression.

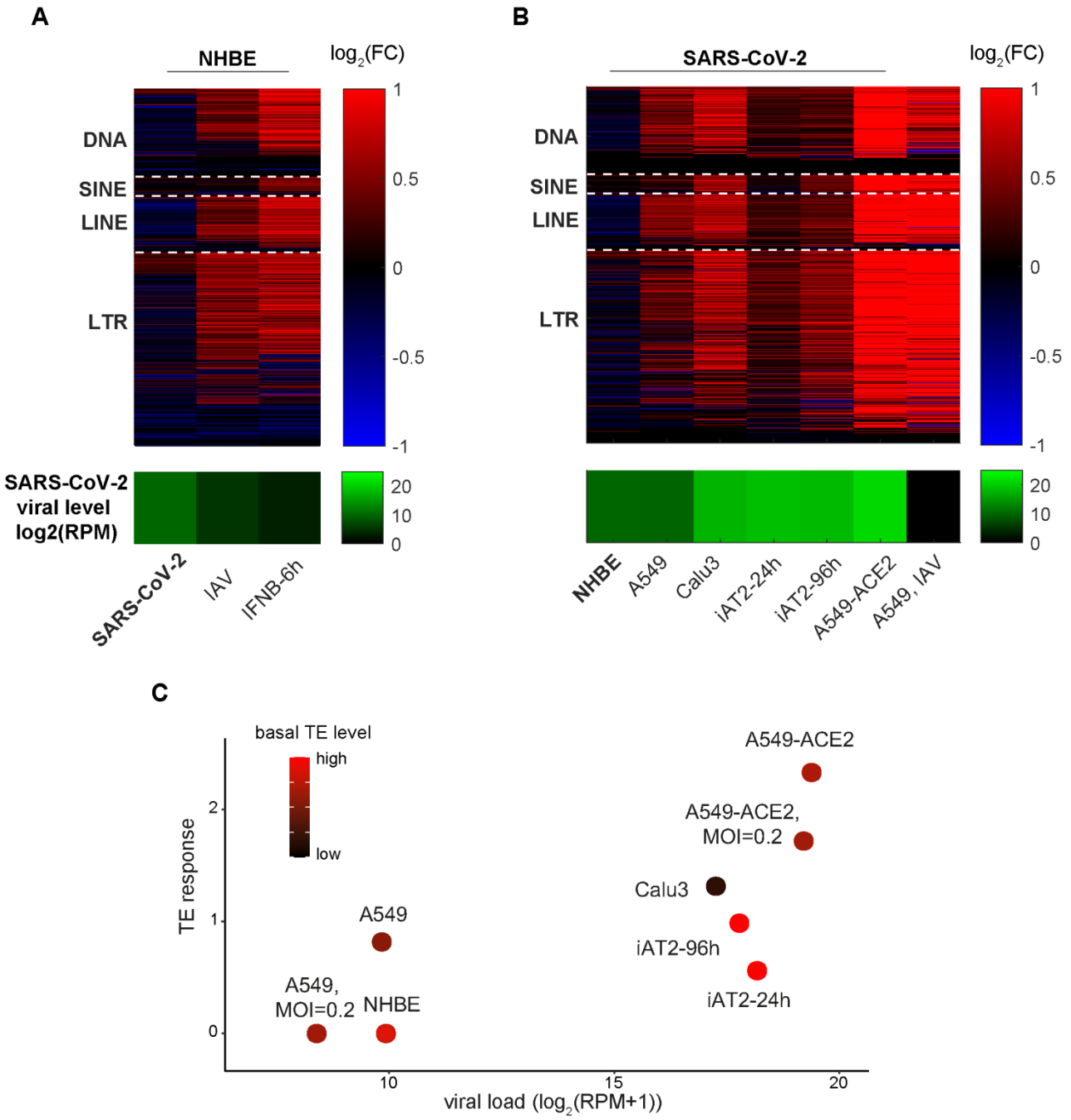

Figure 1. TEs expression changes in response to SARS-CoV-2 and IAV infections. A. Log2-fold change in expression level of TE subfamilies (DNA, SINE, LINE and LTR) in NHBE cells in response to IFNB treatment and in response to SARS-CoV-2 and IAV infections. SARS-CoV-2 viral levels (green) are depicted 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.** TE induction levels are correlated with TE basal levels pre-infection and SARS-CoV-2 viral levels. Linear regression coefficients are 0.158 and -0.57 for viral load and basal TE level, respectively (R^2=0.81).

Upregulation of TEs is correlated with high IFN response

Because IFN expression is associated with TE expression [10, 15, 24], we next tested the relation between IFN and TEs during SARS-CoV-2 infection. As expected from a more robust TE response, we found a significantly larger number of upregulated IFN genes in the cancer cell lines one day after infection with SARS-CoV-2 compared with the primary NHBE cells (Figure 2A and Table S1). Importantly, NHBE cells did respond to IFNβ treatment by a significant induction of IFN genes already after 4-12 hours (Figure 2B and Figure S1B), excluding the possibility that these cells simply fail to induce IFN genes. Infection of iAT2 cells with SARS-CoV-2, which resulted in a mild but detectable TE response at day 1 post infection, showed very low IFN response, similar to NHBE cells (Figure 1B and 2B). This mild TE response slightly increased after 4 days and was accompanied by a significant IFN response (Figure 2B). This suggests that the TE response precedes the 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 over-expressing the ACE2 receptor that were infected with SARS-CoV-2. These cells, when treated with Ruxolitinib, a drug that is known to reduce IFN response, showed reduced IFN response (Figure 2C), while the global TE response remained high (Figure 2D), demonstrating that TE response is not primarily driven by IFN response. Finally, although done on different tissues, genetic backgrounds and M.O.Is, the magnitude of IFN changes strongly correlated with TE expression changes (Figure 2E). This strong correlation ($p<0.029$, 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 enhancers for adjacent host genes encoding critical innate immune factors [10], 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 (Figure 2F). Focusing on IFN-related genes located near up-regulated TEs revealed an even stronger

bias, suggesting that the TE-adjacent gene induction relation is not affected in SARS-CoV-2 infection (Figure S1C). However, 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 (Figure 2G). This suggests that TEs that upregulate IFN related genes remain mostly silent following SARS-CoV-2 infection, and that the mild IFN response to SARS-CoV-2 is not induced for the most part by TE activation. 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, SARS-CoV-2 infection fails to upregulate the TEs coopted for immune activation. According to this model, TE induction is a crucial step in the activation of the anti-viral 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 perhaps most importantly, their specificity is altered – those TEs that induce immune response are not the ones activated by SARS-CoV-2 infection.

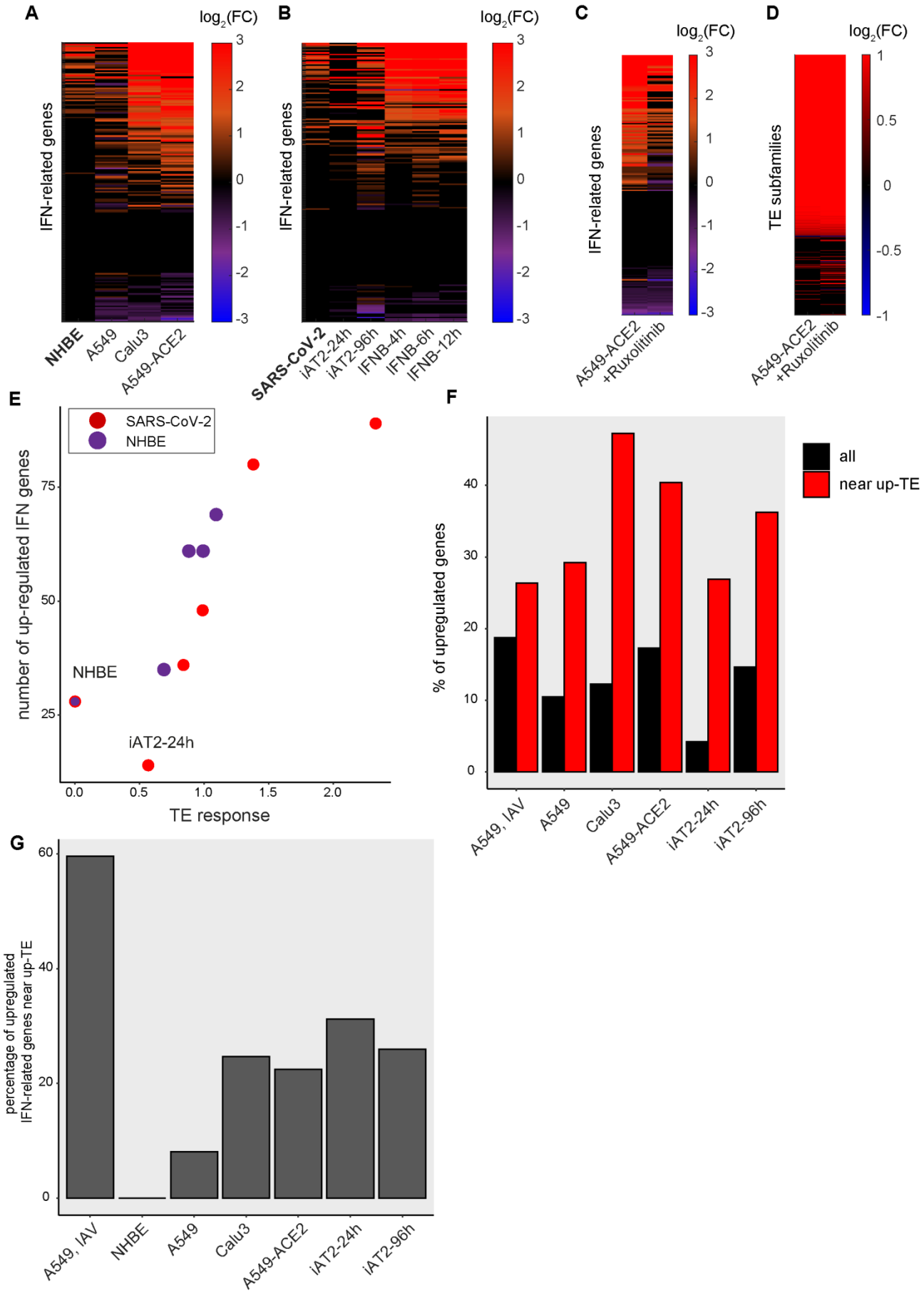
A
log2(FC)
IFN-related genes
NHBE
A549
Calu3
A549-ACE2
B
log2(FC)
IFN-related genes
SARS-CoV-2
iAT2-24h
iAT2-96h
IFNB-4h
IFNB-6h
IFNB-12h
C
log2(FC)
IFN-related genes
A549-ACE2 +Ruxolitinib
D
log2(FC)
TE subfamilies
A549-ACE2 +Ruxolitinib
E
SARS-CoV-2
NHBE
number of up-regulated IFN genes
NHBE
iAT2-24h
TE response
F
all
near up-TE
% of upregulated genes
A549, IAV
A549
Calu3
A549-ACE2
iAT2-24h
iAT2-96h
G
percentage of upregulated IFN-related genes near up-TE
A549, IAV
NHBE
A549
Calu3
A549-ACE2
iAT2-24h
iAT2-96h

Figure 2: TE induction preceds and predicts IFN response. A. The IFN transcriptional response among for SARS-CoV-2 infection in difference cellular systems. **B.** Same as (A) for NHBE and iAT2 cells. **C.** The IFN transcriptional response of A549 cells over-expressing the ACE2 receptor dramatically decreases upon Ruxolitinib treatment. **D.** TE up-regulation 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 (purple). 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 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.

The epigenetic signature of SARS-CoV-2 induced TEs

Regulation of TEs transcription is largely achieved through epigenetic silencing [25]. To understand the nature of the distinctive TE regulation following SARS-CoV-2 infection, we investigated which specific histone modifications are found on the upregulated TEs before SARS-CoV-2 infection (Table S2).

To this end, we used a large-scale dataset including multiple ChIP-seq profiles for histone modifications (HMs) in A549 cells [26]. Since for A549 cells we also have pre- and post- SARS-CoV-2 infection data, it allowed us to examine the 'epigenetic signature' [27] (relative enrichment of HMs) around the TEs that are upregulated post-infection of SARS-CoV-2 and IAV (Figure 3, Figure S2 and Table S3).

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, with a subset of TEs marked by H3K36me3 as well as the combination of H3K27Ac, H3K4me3, H3K79me2 and H3K9Ac (Figure 3A-E and Figure S2). This was consistent among different ChIP-seq experiments for the same histone modification (Figure S3). SINEs and DNA elements upregulated in response to SARS-CoV-2 infection were especially enriched compared to 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 (Figure 3G and Figure S4). Interestingly, all classes of SARS-CoV-2-induced TEs were depleted for H3K27me3 prior to infection (Figure 3F). This is in stark contrast to IAV-induced TEs, which show no such depletion (Figure 3F).

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 subsets of TEs, including SINEs and DNA elements with a highly active chromatin profile and LINEs marked by both repressive and active marks uniquely marked by H3K9me3, and a bivalent group of LINEs marked by both repressive and active marks, keeping them in a poised state ready for infection-induced expression. This attests at the relative failure of the cells to activate a general TE response, which includes the induction of TEs located in an open active chromatin state as well as those that are in a repressive chromatin state prior to SARS-CoV infection.

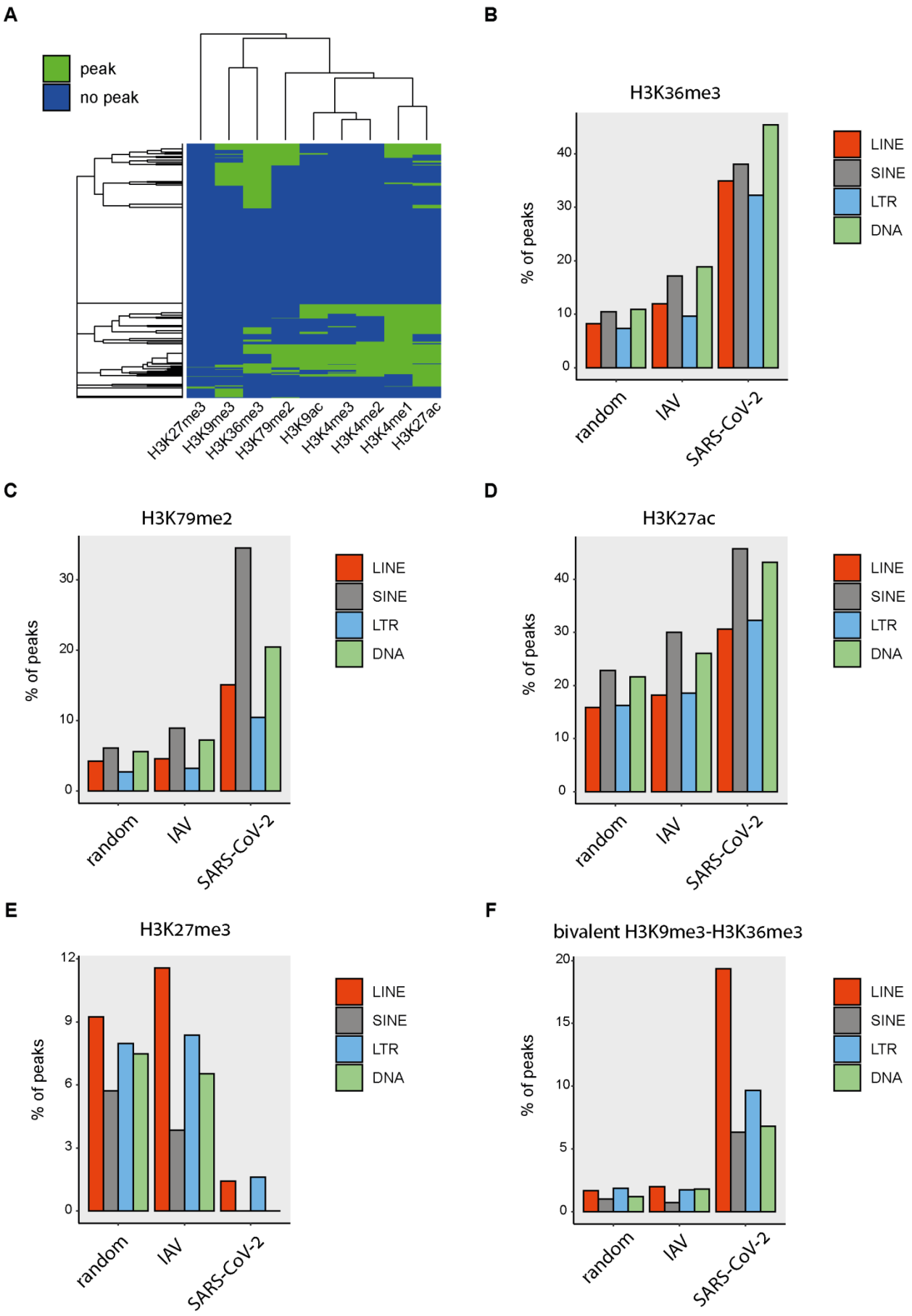
A
peak
no peak
H3K27me3
H3K9me3
H3K36me3
H3K79me2
H3K9ac
H3K4me3
H3K4me2
H3K4me1
H3K27ac
B
H3K36me3
% of peaks
LINE
SINE
LTR
DNA
random
IAV
SARS-CoV-2
C
H3K79me2
D
H3K27ac
E
H3K27me3
F
bivalent H3K9me3-H3K36me3

Figure 3: individual TEs induced by SARS-CoV-2 in A549 cells have a unique epigenetic profile. A. Hierarchical clustering of histone modifications signal in non-infected A549 cells around all upregulated TEs in response to SARS-CoV-2 infection in A549 cells. **B-E.** Percentage of TEs with peaks of H3K36me3 (B), H3K79me2 (C), H3K27ac (D), H3K27me3 (E) on SARS-CoV-2 induced TEs, IAV induced TEs and on all expressed TEs outside genes. **F.** Same as B-E for bivalent mark of H3K9me3 and H3K36me3 mark.

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 (see Methods) were enriched for mitochondrial-related genes and processes (Figure 4A and Table S4), consistent with previous reports [28]. 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 (Figure 4B and Table S4).

Indeed, intersecting the positively correlated genes with all genes that encode for chromatin binding proteins, or epifactors [29], showed highly significant enrichment (Figure 3C, green. 44 genes, fold-enrichment=2.33, $p < 10^{-7}$). Interestingly, SETD2, the human H3K36 lysine trimethylase, was among the epifactors that were positively correlated with TE response. H3K36me3 marks gene bodies of active genes. In addition, SETD2 methylation of STAT1 is crucial for interferon response [30] 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 non-genic transcriptional regulation [31, 32]. Interestingly, SETD2 is also among the interacting proteins of SARS-CoV-2, and comparing with other interactomes of coronaviruses as well as of different IAV strains, we found that SETD2 is specific for SARS-CoV-2 (Table S5).

We therefore searched for more epifactors in the SARS-CoV-2 interacting proteins. Among the ten epifactors that interact with SARS-CoV-2 we found the histone acetylation related proteins BRD2, BRD4 and HDAC2, all of which are, once again, SARS-CoV-2-specific (Table S5). We also found DDX21, a DNA damage and dsRNA sensing protein, as a SARS-CoV-2-specific interacting epifactor. Lastly, MOV10, an RNA helicase that also restricts LINE expression [33], interacts with SARS-CoV-2, although it also interacts with IAV. These observations suggest that SARS-CoV-2 may affect TE expression through interaction with a subset of specific epifactors.

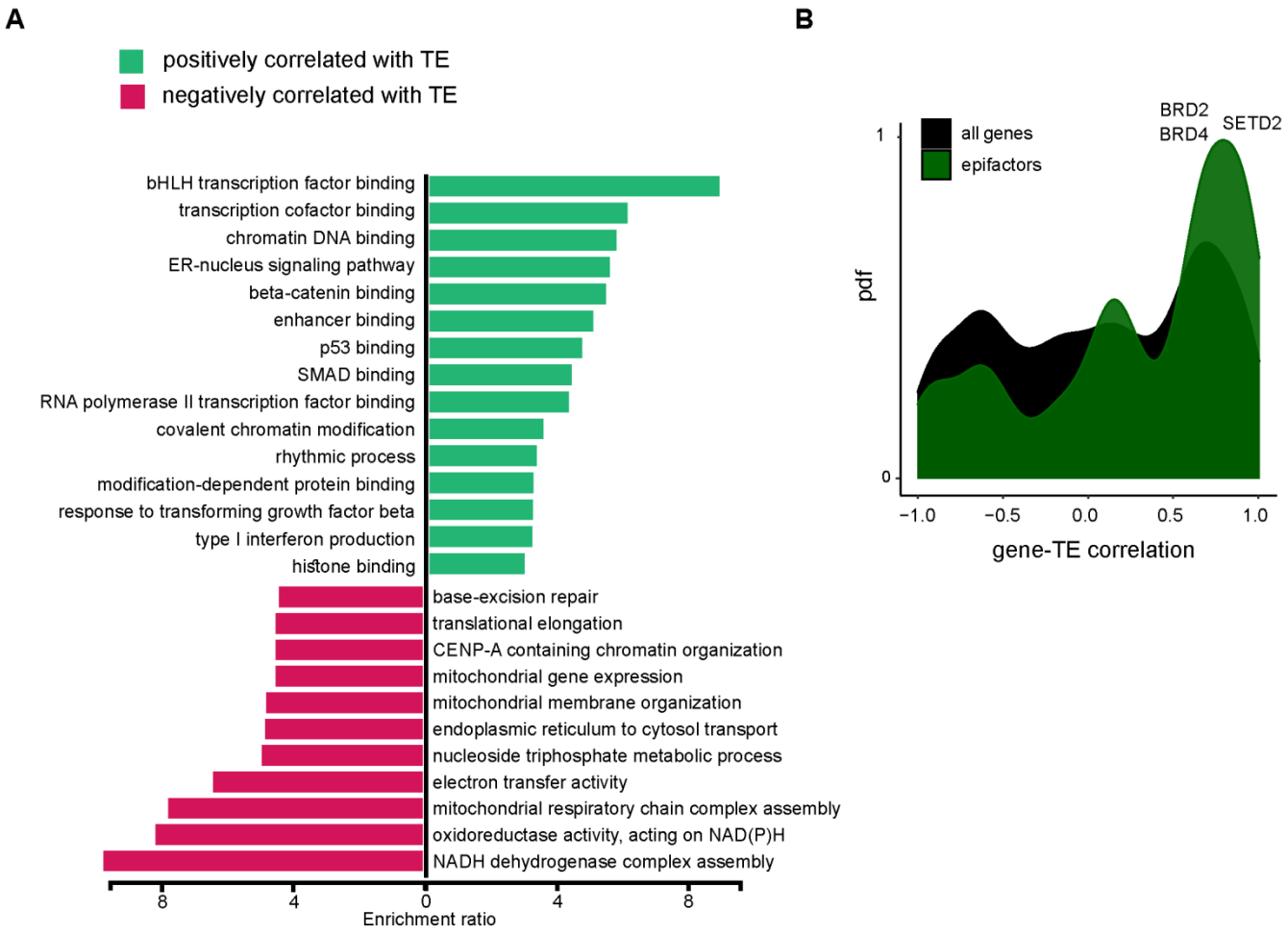

Figure 4: TE response is correlated with epigenetic and mitochondrial gene expression changes. A. Gene ontology analysis of genes correlated (green) and inversely-correlated (red) with TE response to SARS-CoV-2 infection. **B.** Distribution of TE response-gene correlations for genes that increase with TE expression in iAT2 cells. Black represents correlation distribution of all genes. Correlation distribution of epifactors are in green.

Discussion

In this study, we re-analyzed published data to examine the link between transcriptional activation of Transposable Elements (TEs) and SARS-CoV-2 infection. 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 (Figure 5). This phenomenon is even more prominent in cells with high basal TE levels, which are gradually increased with old age [16, 34]. This raised the "TE desensitization" hypothesis, which posits that the higher the basal expression of TEs, the less effective the TE activation in response to viral infections at early stage and the subsequent induction of the innate anti-viral defense mechanisms.

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 TE expression, where lower basal levels predict stronger cellular responses. As such, and thirdly, 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, fourthly, 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 very high levels of viral load, 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 [15, 24], as well as studies

that show suppression of dsRNA-activated early host responses following SARS-CoV-2 infection [3, 4]. Our own analysis also shows a highly significant correlation between TE activation and the induction of IFN response. TEs may have additional mechanisms of inducing immune responses. Second, some TEs act as enhancers for immune genes [25], however in our study no enrichment of TEs near IFN genes activation was observed following SARS-CoV-2 infection. These results suggest that TEs initiate inflammatory responses to viral infection through both dsRNA sensing and gene-regulatory roles. Therefore, the failed TE activation following SARS-CoV-2 infection have a double impact, yielding the delayed IFN response seen in COVID-19 patients.

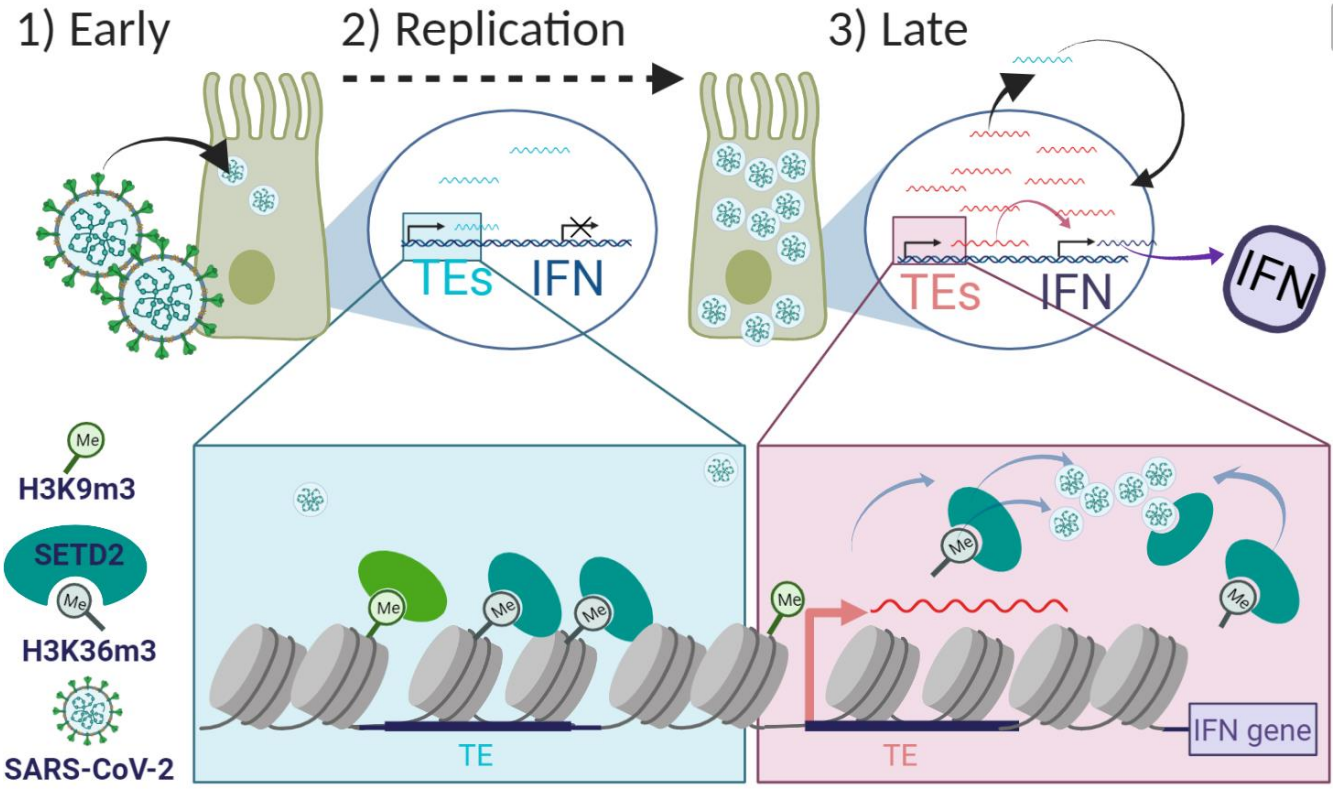

Figure 5: A model for SARS-CoV-2 impaired TE activation. Early following SARS-CoV-2 infections of normal lung epithelial cells, viral load is low and IFN response is still absent. SETD2 is normally localized on the H3K36me3 enriched chromatin. Next, viral replication (2) increases viral load and induces TE activation and IFN response (3). The hypothesis (in the pink box) is that specific SARS-CoV-2 interacting proteins sequester chromatin factors, such as SETD2, away from the H3K36me3

enriched TEs, which are then transcribed, inducing IFN response that might be too weak or too late for the host defense

In addition, SARS-CoV-2 viral load seems to be linked with TE expression level. 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. This 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.

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. 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 re-analyzing data published by Huang et al. and Blanco-Melo et al. we provide evidence that unlike other viruses, which strongly induce TEs following infection [12], SARS-CoV-2 infection of a 79 years old normal lung cells has a relatively weak effect on TE expression. 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 [35]. Our analysis further showed that SARS-CoV-2 could elicit strong TE induction in lung cancer cells, Calu3 and A549. This could potentially be explained by epigenetic aberrations often observed in transformed cells, rendering them prone to TEs dysregulation [15, 24, 35]. Finally, since TE expression increases globally across most tissues with age, our model explains the aging-related severity to SARS-CoV-2 infection.

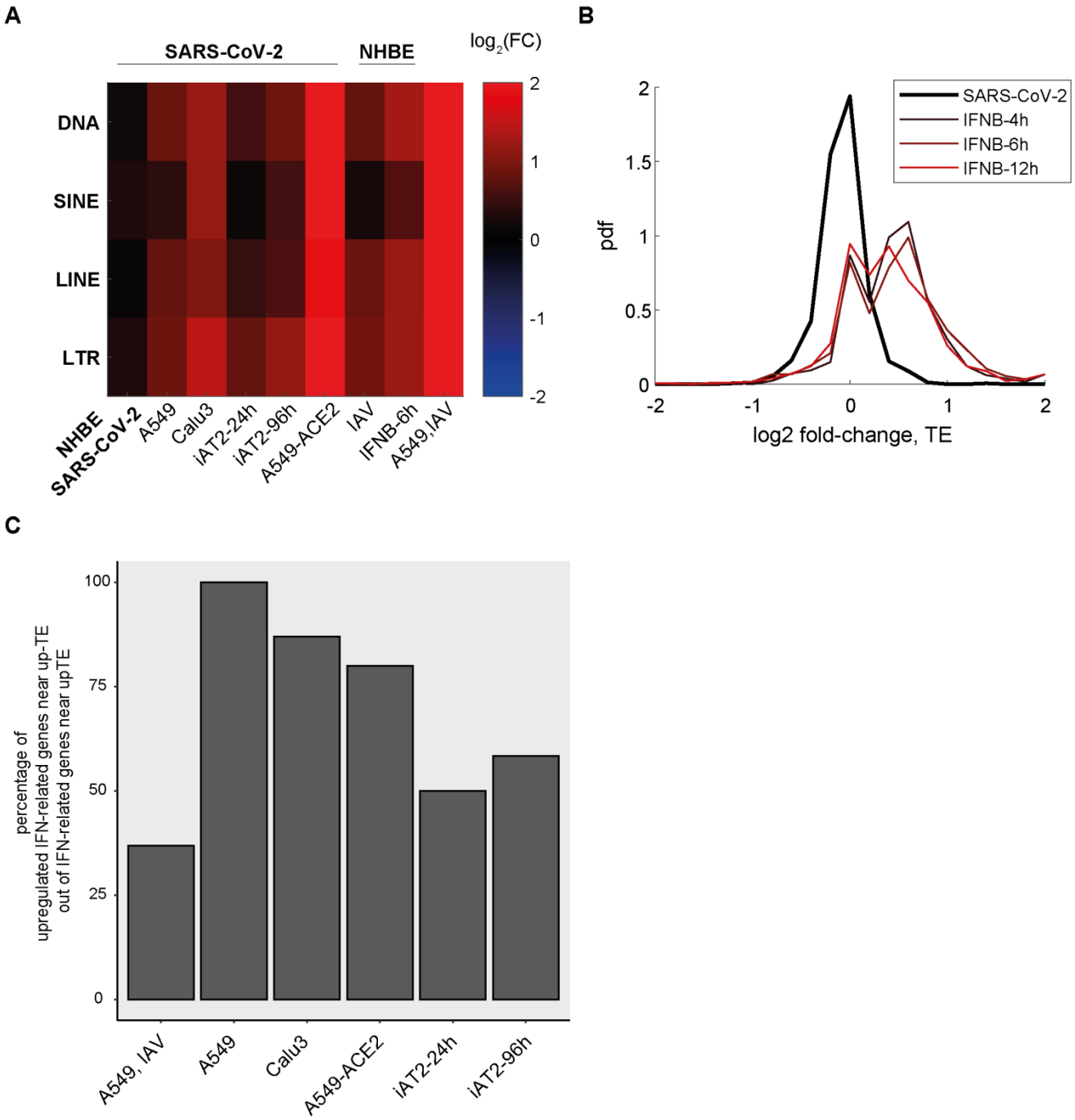

Figure S1: impaired TE activation in response to SARS-CoV-2 infection. A. The 95-percentile of TE subfamilies log2 fold changes in response IFNB treatment, and to SARS-CoV-2 and IAV infection in different cellular systems by class. **B.** The distribution of log2 fold-changes of TE subfamilies in response to SARS-CoV-2 infection and IFNB treatment at different times in NHBE cells. **C.** The

percentage of up-regulated IFN-related genes located near TEs that are up-regulated following SARS-CoV-2 and IAV infections out of all IFN-related genes located near up-regulated TEs.

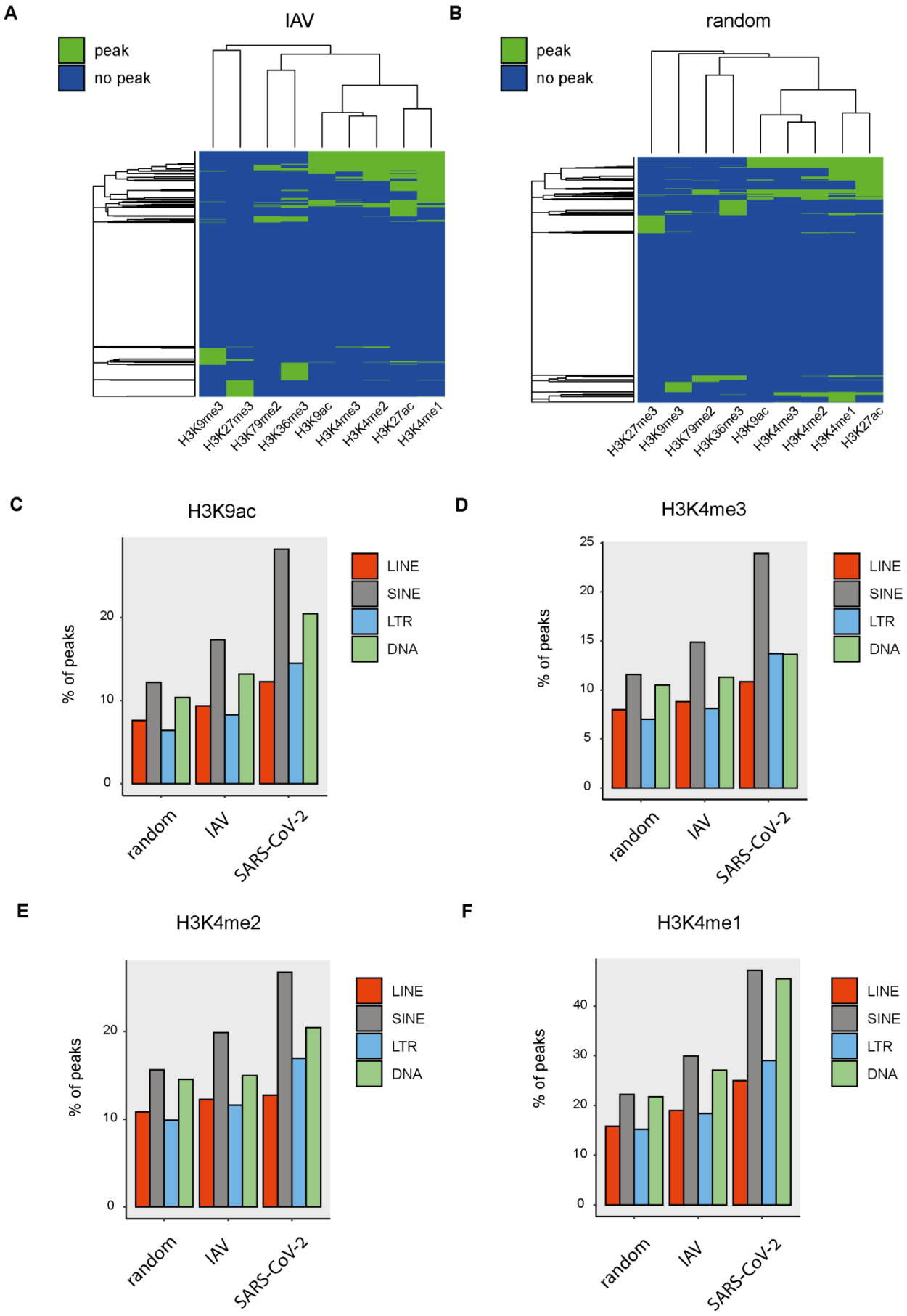
A
IAV
peak
no peak
H3K9me3
H3K27me3
H3K79me2
H3K36me3
H3K9ac
H3K4me3
H3K4me2
H3K27ac
H3K4me1
B
random
peak
no peak
H3K27me3
H3K9me3
H3K79me2
H3K36me3
H3K9ac
H3K4me3
H3K4me2
H3K4me1
H3K27ac
C
H3K9ac
% of peaks
0
10
20
random
IAV
SARS-CoV-2
LINE
SINE
LTR
DNA
D
H3K4me3
% of peaks
0
5
10
15
20
25
random
IAV
SARS-CoV-2
LINE
SINE
LTR
DNA
E
H3K4me2
% of peaks
0
10
20
random
IAV
SARS-CoV-2
LINE
SINE
LTR
DNA
F
H3K4me1
% of peaks
0
10
20
30
40
random
IAV
SARS-CoV-2
LINE
SINE
LTR
DNA

Figure S2: epigenetic signature of up-regulated TEs in response to SARS-CoV-2 and IAV infections compared to all expressed TEs. A-B. Hierarchical clustering of histone modifications signal in non-infected A549 cells around all up-regulated TEs in response to IAV infection (A) and 5000 random expressed TEs (B) in A549 cells. C-F. Percentage of TEs with peaks of H3K9Ac (C), H3K4me3 (D),

H3K4me2 (E), H3K4me1 (F) on SARS-CoV-2 induced TEs, IAV induced TEs and on all expressed TEs outside genes.

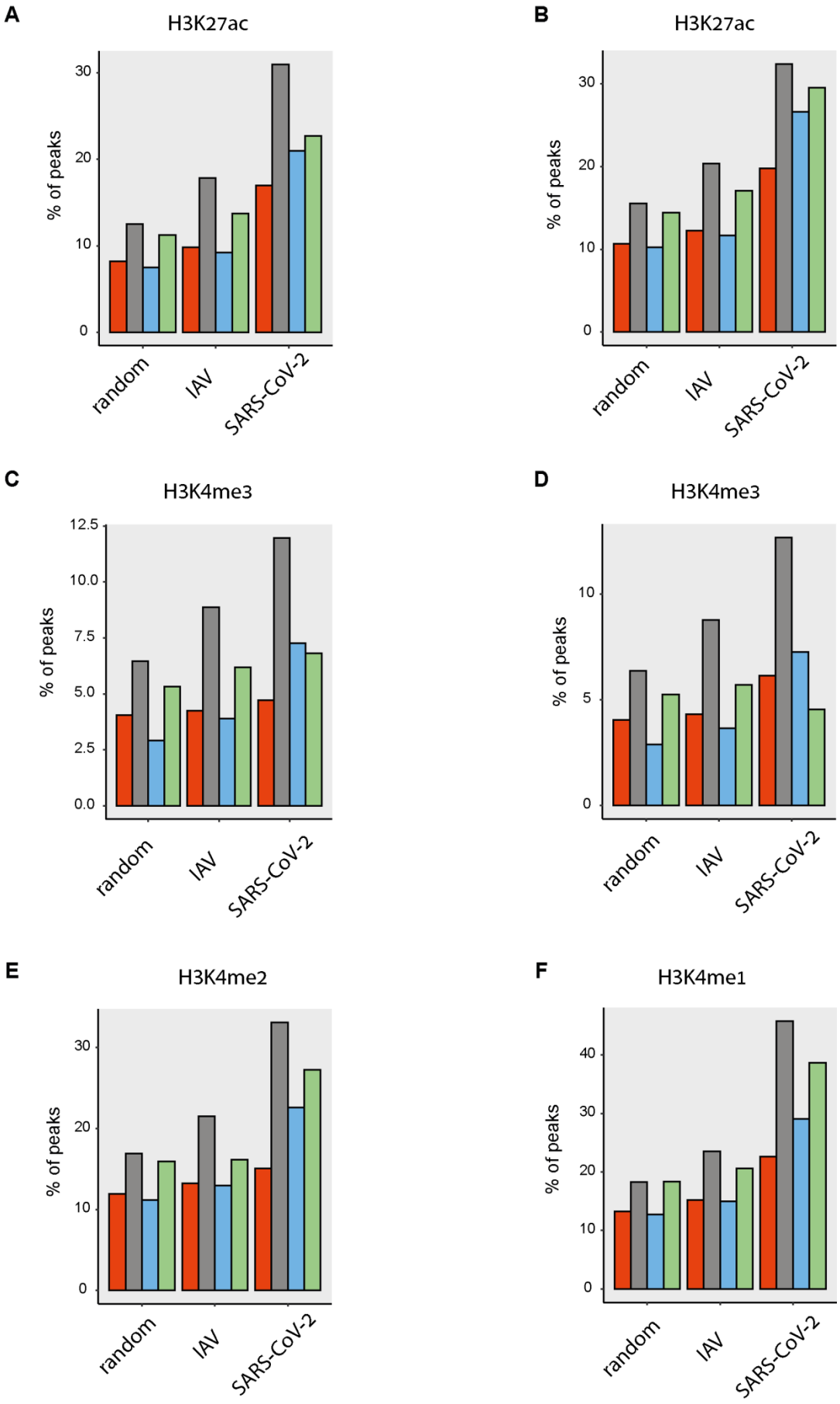
A
H3K27ac
% of peaks
0
10
20
30
random
IAV
SARS-CoV-2
B
H3K27ac
% of peaks
0
10
20
30
random
IAV
SARS-CoV-2
C
H3K4me3
% of peaks
0.0
2.5
5.0
7.5
10.0
12.5
random
IAV
SARS-CoV-2
D
H3K4me3
% of peaks
0
5
10
random
IAV
SARS-CoV-2
E
H3K4me2
% of peaks
0
10
20
30
random
IAV
SARS-CoV-2
F
H3K4me1
% of peaks
0
10
20
30
40
random
IAV
SARS-CoV-2

Figure S3: epigenetic profile of SARS-CoV-2 up-regulated TEs is consistent among different experiments. A-F. Percentage of TEs with peaks additional H3K27ac (A-B), H3K4me3 (C-D), H3K4me2 (E) and H3K4me1 (F) datasets on SARS-CoV-2 induced TEs, IAV induced TEs and on all expressed TEs outside genes.

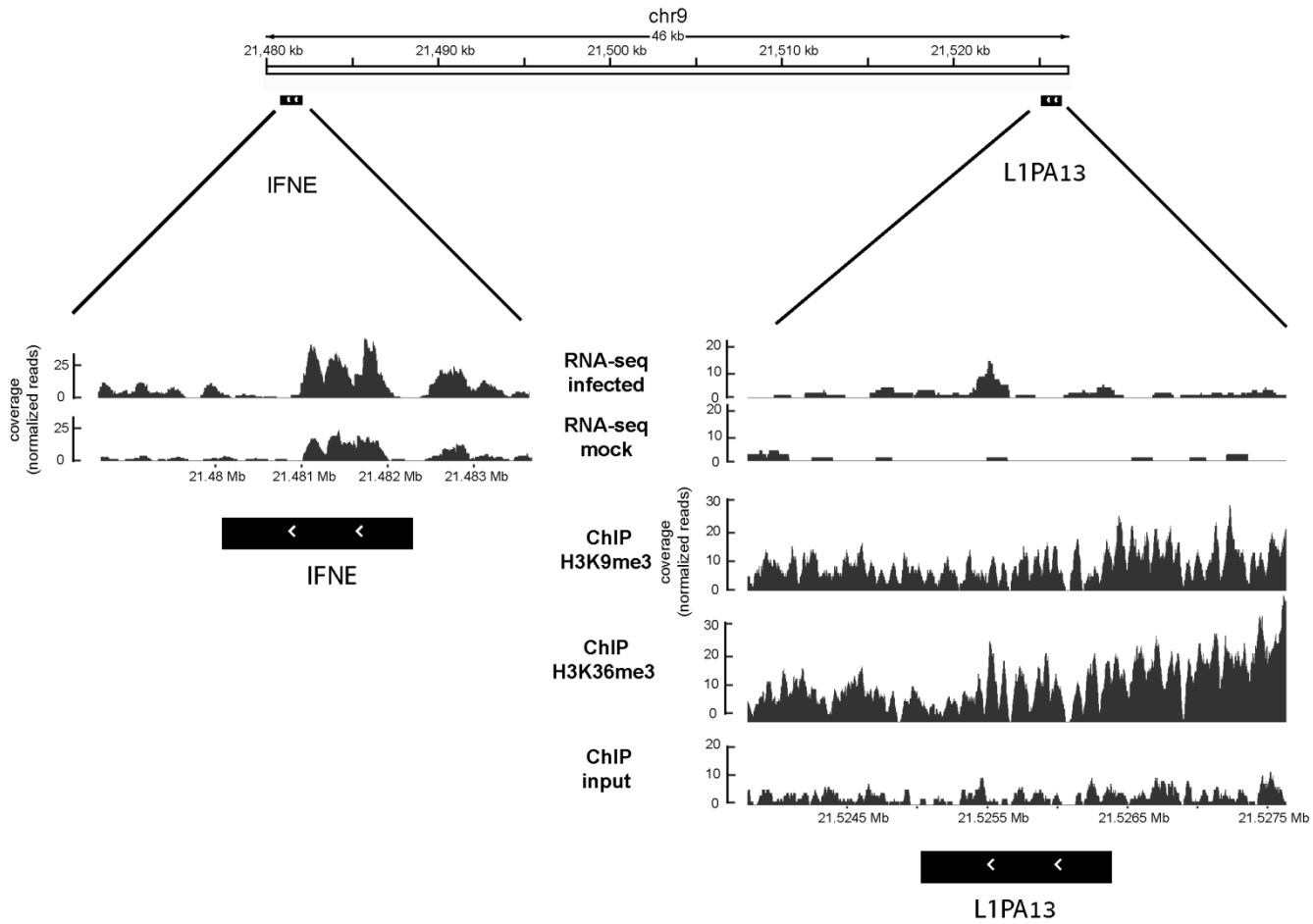

Figure S4: up-regulated TE with unique epigenetic signature contributes to gene expression response to SARS-CoV-2 infection. Shown is UCSC track for the bivalently H3K9me3-H3K36me3 marked LINE copy L1PA13 which is up-regulated following SARS-CoV-2 infection in A549 cells. The TE is located near the IFN-related gene IFNE which is also up-regulated in these cells in response to SARS-CoV-2 infection.

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.

Table 1: ChIP-Seq data from ENCODE project [26]:

ChIP Name (in this manuscript)	Experiment identifier (ENCODE)
H3K27Ac	ENCFF443MVY ENCFF918FET ENCFF747IZX
H3K9ac	ENCFF350BNS
H3K79me2	ENCFF705EEZ
H3K36me3	ENCFF584FSY
H3K4me3	ENCFF832EOL ENCFF361TGO ENCFF535EYL
H3K9me3	ENCFF287WGC ENCFF164FDB
H3K27me3	ENCFF684ZZH
H3K4me2	ENCFF536TYP ENCFF794FQH
H3K4me1	ENCFF116HIQ ENCFF594YDK

Preprocessing and alignment

Raw reads from GSE147507 and GSE133329 were trimmed to remove Illumina adapters using the Trimmomatic software [36] version 0.39. We used the STAR aligner version 2.7.1a [37] to align raw reads to human RefSeq reference genome (GRCh38). We used the parameters --outFilterMultimapNmax 100 and --winAnchorMultimapNmax 200 to allow for large numbers of multi-mapped reads for downstream analysis of TEs. For ChIP-Seq data, alignment with STAR was followed by filtering out of non-uniquely mapped reads.

TE and gene expression quantification

Gene expression was 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 [38]. 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 [39]. 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 [38, 40] 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 DE-Seq2 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. In order 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 2747 , 2388, 213 (Table S1). 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/ [29].

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 1). For each TE we defined the flanking region as the TE location and its surrounding 500 bp up- and down-stream. If the flanking region intersection with ChIP peak locations was non-empty then this TE was considered as TE with peak. For histone modification clustering (Figure 3 and Figure S3) we used the Jaccard metric. Enrichment of peaks on up-regulated TEs was calculated using hypergeometric test. Genome tracks were produced using the ggbio package in R [41].

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 publically available interactome lists (see Table 2). A gene was considered in SARS-CoV-2 interactome if it was included in at least 3 out of 5 sources, and as a SARS-CoV-1 interacting partner if it was included in both relevant sources.

Table 2: interactome sources

virus	Interactome papers
SARS-CoV-2	42, 43, 44, 45, 46
MERS-Cov	45
SARS-CoV-1	42, 45
HCovNL63	42
HCoV229E	42
IAV (different strains)	47
EBV	48
WNV	49

References

1. Blanco-Melo D, *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045 e1039 (2020).

2. Verity R, *et al.* Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases* **20**, 669-677 (2020).

3. Yin X, *et al.* MDA5 Governs the Innate Immune Response to SARS-CoV-2 in Lung Epithelial Cells. *Cell Rep* **34**, 108628 (2021).

4. Ancar R, *et al.* Physiologic RNA targets and refined sequence specificity of coronavirus EndoU. *RNA* **26**, 1976-1999 (2020).

5. Huang J, *et al.* 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* **Sep 18**, S1934-5909 (2020).

6. Park A, Iwasaki A. Type I and Type III Interferons - Induction, Signaling, Evasion, and Application to Combat COVID-19. *Cell Host Microbe* **27**, 870-878 (2020).

7. King C, Sprent J. Dual Nature of Type I Interferons in SARS-CoV-2-Induced Inflammation. *Trends Immunol* **42**, 312-322 (2021).

8. Ziegler CGK, *et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* **181**, 1016-1035 e1019 (2020).

9. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. *Science (New York, NY)* **165**, 349-357 (1969).

10. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science (New York, NY)* **351**, 1083-1087 (2016).

11. Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proc Natl Acad Sci USA* **117**, 7905-7916 (2020).

12. Macchietto MG, Langlois RA, Shen SS. Virus-induced transposable element expression up-regulation in human and mouse host cells. *Life science alliance* **3**, (2020).

13. Tunbak H, *et al.* The HUSH complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s. *Nat Commun* **11**, 5387 (2020).

14. Krischuns T, *et al.* Phosphorylation of TRIM28 Enhances the Expression of IFN-β and Proinflammatory Cytokines During HPAIV Infection of Human Lung Epithelial Cells. *Front Immunol* **9**, (2018).

15. Chiappinelli KB, *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162**, 974-986 (2015).

16. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat Rev Endocrinol* **14**, 576-590 (2018).

17. Zhang W, *et al.* A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science* **348**, 1160-1163 (2015).

18. Perez RF, Tejedor JR, Bayon GF, Fernández AF, Fraga MF. Distinct chromatin signatures of DNA hypomethylation in aging and cancer. *Aging Cell* **17**, e12744 (2018).

19. Pehrsson EC, Choudhary MN, Sundaram V, Wang T. The epigenomic landscape of transposable elements across normal human development and anatomy. *Nature Communications* **10**, 1-16 (2019).

20. Bogu GK, Reverter F, Marti-Renom MA, Snyder MP, Guigó R. Atlas of transcriptionally active transposable elements in human adult tissues. *BioRxiv*, 714212 (2019).

21. Galani I-E, *et al.* Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. *Nat Immunol*, 1-9 (2020).

22. Schmidt N, *et al.* An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *Proc Natl Acad Sci USA* **116**, 17399-17408 (2019).

23. Takayama K. In Vitro and Animal Models for SARS-CoV-2 research. *Trends Pharmacol Sci* **41**, 513-517 (2020).

24. Roulois D, *et al.* DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162**, 961-973 (2015).

25. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18**, 71-86 (2017).

26. Schreiber J, Bilmes J, Noble WS. Completing the ENCODE3 compendium yields accurate imputations across a variety of assays and human biosamples. *Genome Biol* **21**, 82 (2020).

27. Sorek M, Cohen LRZ, Meshorer E. 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 (2019).

28. Chung N, *et al.* Transcriptome analyses of tumor-adjacent somatic tissues reveal genes co-expressed with transposable elements. *Mob DNA* **10**, 39 (2019).

29. Medvedeva YA, *et al.* EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database (Oxford)* **2015**, bav067 (2015).

30. Chen K, *et al.* Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity. *Cell* **170**, 492-506 e414 (2017).

31. Li F, *et al.* The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. *Cell* **153**, 590-600 (2013).

32. Dhayalan A, *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* **285**, 26114-26120 (2010).

33. Li X, *et al.* The MOV10 helicase inhibits LINE-1 mobility. *J Biol Chem* **288**, 21148-21160 (2013).

34. De Cecco M, *et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73-78 (2019).

35. Burns KH. Our Conflict with Transposable Elements and Its Implications for Human Disease. *Annu Rev Pathol* **15**, 51-70 (2020).

36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

37. Dobin A, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

38. Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593-3599 (2015).

39. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

40. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

41. Yin T, Cook D, Lawrence M. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol* **13**, R77 (2012).

42. Stukalov A, *et al.* Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594**, 246-252 (2021).

43. Samavarchi-Tehrani P, *et al.* A SARS-CoV-2 – host proximity interactome. *bioRxiv*, 2020.2009.2003.282103 (2020).

44. Laurent EMN, *et al.* Global BioID-based SARS-CoV-2 proteins proximal interactome unveils novel ties between viral polypeptides and host factors involved in multiple COVID19-associated mechanisms. *bioRxiv*, 2020.2008.2028.272955 (2020).

45. Gordon DE, *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, (2020).

46. Gordon DE, *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459-468 (2020).

47. Wang L, *et al.* Comparative influenza protein interactomes identify the role of plakophilin 2 in virus restriction. *Nat Commun* **8**, 13876 (2017).

48. Calderwood MA, *et al.* Epstein-Barr virus and virus human protein interaction maps. *Proc Natl Acad Sci U S A* **104**, 7606-7611 (2007).

49. Li M, *et al.* Identification of antiviral roles for the exon-junction complex and nonsense-mediated decay in flaviviral infection. *Nat Microbiol* **4**, 985-995 (2019).