Transposable elements as sensors of 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)

Abstract

Transposable elements (TEs) are induced in response to viral infections. TEs 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. In general, we observe that TE activation correlates with, and precedes, the induction of IFN-related genes, suggesting that the failure to activate TEs may be the reason for the weak IFN response. 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. This work provides a potential 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.

Main

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 (Blanco-Melo et al., 2020). The fatality of SARS-CoV-2 increases substantially with age (Verity et al., 2020). 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 (Ancar et al., 2020; Yin et al., 2021).

Although the first cellular responses to viral infection are IFN-I; IFN-III responses, recent evidence suggests that SARS-CoV-2 does not elicit a robust IFN expression response in lung epithelial cells. First, Blanco-Melo et al. assessed the transcriptional response to SARS-CoV-2 infection in different cellular models (Blanco-Melo et al., 2020). The authors found that unlike other respiratory RNA viruses, 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 (Huang et al., 2020). 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.

Transposable elements (TEs) are abundant sequences in the mammalian genomes that contain multiple regulatory elements and can amplify in a short evolutionary timescale (Britten and Davidson, 1969). 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 (Chuong et al., 2016). 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 (Ye et al., 2020). 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 (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 NFkB and the IFN pathways, thus enhancing immune response (Chiappinelli et al., 2015). Thus, there seems to be a link between the level of TE overexpression and the intensity of the IFN response.

Aging and increased risk of COVID-19 complication are associated with impaired and delayed early IFN responses (Galani et al., 2020). On the other hand, at the molecular level, aging is associated with decreased heterochromatin-associated marks, e.g. H3K9me3 (Zhang et al., 2015) and DNA methylation (Perez et al., 2018). Consequently, TE expression increases with age (Bogu et al., 2019; Pehrsson et al., 2019).

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 (Bogu et al., 2019; Pehrsson et al., 2019), 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 (Park and Iwasaki, 2020). Taken together, this raised the "TE desensitization" hypothesis, which posits that the higher the basal expression of TEs, the less effective the early TE activation in response to viral infections and the subsequent induction of the innate anti-viral defense mechanisms.

Our "TE desensitization" model makes several predictions. First, we would anticipate 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, we anticipate 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, we suggest, 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, if TEs act as a general warning device, we should not expect to see specific TEs responding, but rather a more wide-spread, heterogeneous response, associated with active chromatin regions rather than specific locations.

Since TEs are induced in response to many viral infections (Macchietto et al., 2020), we first analyzed TE expression changes in response to SARS-CoV-2 infection in different cellular models comparing SARS-CoV-2 with other viruses. As our model predicts, we show 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, supporting both the "TE desensitization" idea, as well as the age-related decline in handling SARS-CoV-2 infection. Last, we find that a subset of TEs which are marked by active histone modifications prior to viral infection,

are poised for activation, although, save a few prominent exceptions, exactly which TEs will be activated is a stochastic process.

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

We began by analyzing a recently published datasets of primary lung epithelial cells and A549 cancer cells infected with influenza (IAV) (Schmidt et al., 2019) or SARS-CoV-2 (Blanco-Melo et al., 2020). Interestingly, while, as expected, IAV infection caused a robust global increase in TE subfamilies expression in both NHBE primary lung epithelial cells and A549 cancer cells (Figure 1A), SARS-CoV-2 caused a significant increase in TE expression in A549 cells, but not in NHBE cells, despite similar viral loads (Figure 1A, left column). This raised an intriguing hypothesis that SARS-CoV-2 may somehow avoid, at least partially, the robust TE expression response which often follows viral infection, potentially explaining the ineffective immune response to SARS-CoV-2 infection.

To test this more directly, we turned to iAT2 cells, which may represent a more natural system. We found that in the iAT2s, TEs induction in response to SARS-CoV-2 infections after 24h is, once again, relatively mild despite the extremely high viral levels (Figure 1A, compare to A549 cells expressing the ACE2 receptor). This pattern was consistent among the different TE classes (Figure S1A).

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 (Bogu et al., 2019). As age is the most significant risk factor for COVID19-related death, this raised the intriguing hypothesis 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 correlation between the basal TE expression in the uninfected cells and the TE activation response following infection (Figure 1B). 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.

Upregulation of TEs is correlated with high IFN response

Because IFN expression is associated with TE expression (Chiappinelli et al., 2015; Chuong et al., 2016; Roulois et al., 2015), we next tested the relation between IFN and TEs in 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 1C). Importantly, NHBE cells did respond to IFNβ treatment by a significant induction of IFN genes already after 4-12 hours (Figure 1C and Figure S1B), excluding the possibility that these cells simply fail to induce IFN genes in response to SARS-CoV-2 infection. Infection of iAT2 cells with SARS-CoV-2 resulted in a mild but detectable TE response compared to NHBE cells at day 1 post infection and similar, very low IFN response (Figure 1A and 1C-D). This mild TE response slightly increases after 3 days, and is accompanied by a significant IFN response (Figure 1D). This suggests that the TE response precedes the IFN response in these cells.

The idea that TE activation is inducing IFN response in response to 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 2A), while the global TE response remained the same (Figure 2B). Finally, although done on different tissues, genetic backgrounds and M.O.Is, the magnitude of IFN changes strongly correlated with TE expression changes (Figure 2C). This strong correlation ($p<0.029$, permutation test on Spearman correlation) was specific to IFNs, as other random groups of genes were not correlated with TE changes. Taken together, these data suggest that TE induction following viral infection precede expression of IFN-related genes. Therefore, TE induction is a crucial step in the activation of the cellular immune response, and a delayed induction of TE activation would enable undeterred viral replication.

Upregulated individual TEs are adjacent to key viral response related genes

To explore the potential influence of TEs on the transcriptional regulation of adjacent genes, we also analyzed individual TE expression. To avoid counting possible reads related to unspliced mRNAs, we analyzed only TEs located outside of coding genes. In agreement with our previous results, we found no upregulated TEs in NHBE cells infected with SARS-CoV-2, while in other conditions hundreds to thousands of TEs were induced, and only a dozen were suppressed (Figure S1C and Figure S1D). This is unlike differentially expressed genes, which showed similar propensities for down- and up-regulation (Figure S1E). Noticeably, however, eight TEs located within seven genes were significantly upregulated in NHBE SARS-CoV-2 infected cells (Figure S2), of which at least five (C3, S100A9, OAS1, SAA2 and IL1A) have been previously linked to SARS-CoV-1 and SARS-CoV-2 disease pathology (Gralinski et al., 2018; He et al., 2006; Li et al., 2020; Silvin et al., 2020; van de Veerdonk and Netea, 2020).

Examining the identity of the genes which are adjacent to individual up-regulated TEs, we found enrichment for genes related to IFN-I response and to cytokine response in the cellular systems that showed evident IFN response (i.e. iAT2 d4, A549-ACE2 and Calu3). In iAT2 cells at d1, the number of individual induced TEs was relatively large, and larger than in NHBE cells, in contrast to the subfamily response, which was extremely mild. This suggests that the initial TE response begins on day 1 with specific subfamilies (from the LTR class of retrotransposons, Figure S1A) but is not global yet at this early stage. In addition, also supporting the model where TE induction precedes IFN response, individual TEs up-regulated in these cells were adjacent to key viral response related genes such as ISG20, IRS2 and RELB, despite low IFN response in iAT2 cells at day 1 post-infection.

The transcriptional signature of TE over-expressing cells

Looking for a possible explanation for the almost complete absence of TE response in NHBE and the mild response in iAT2 cells at day 1 after SARS-CoV-2 infection, we searched for genes that are correlated or inversely correlated with TE over-expression. We focused on genes that had a high

correlation with TE expression 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, consistent with previous reports (Chung et al., 2019). By contrast, genes that were positively correlated with TE response were enriched for histone acetyltransferase, RNA Pol-II binding, pre-mRNA processing, regulation of H4 acetylation and histone deubiquitination, demonstrating a clear epigenetic and chromatin-related signature. Indeed, intersecting the positively correlated genes with all genes that encode for chromatin binding proteins, or epifactors (Medvedeva et al., 2015), showed highly significant enrichment (Figure 2D, green. 95 genes, fold-enrichment=3.22, $P = 2.1x10^{-24}$). Interestingly, SETD2, the human H3K36 lysine trimethylase, was among the epifactors that were positively correlated with TE response. H3K36m3 marks gene bodies of active genes. In addition, recent evidence shows that SETD2 is essential for microsatellite stability, implicating its role in non-genic transcriptional regulation (Li et al., 2013)

In addition to SETD2, the two epifactor genes that code for BRD2 and BRD4, two proteins which were recently shown to interact with the SARS-CoV-2 E protein (Gordon et al., 2020), were also positively correlated with TE response. We therefore crossed the list of TE-correlated genes with all of the recently published human interactome of SARS-CoV-2 (Gordon et al., 2020). In contrast to BRD2 and BRD4, genes that encode for proteins that interact with proteins produced by SARS-CoV-2 were inversely correlated with TE response (Figure 2E, red. 42 genes, fold-enrichment=2.38, $P = 1.6x10^{-7}$). Approximately 10% of the genes that encode for proteins that interact with SARS-CoV-2 were inversely correlated with the TE response, suggesting that SARS-CoV-2 is actively working to counteract the host TE-response to infection.

The epigenetic signature of SARS-CoV-2 induced TEs

The high correlation of SETD2 with TE response, as well as the general observation that genes which were positively correlated with TE response were enriched for histone H3 methylation and acetylation,

led us investigate whether histone modifications may mitigate the upregulation in TEs following SARS-CoV-2 infection.

To this end, we used a large-scale dataset including multiple ChIP-seq profiles for transcription factors (TFs) and histone modifications (HMs) in A549 cells (Schreiber et al., 2020). 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 TFs and HMs) around the TEs that are up-regulated post-infection of SARS-CoV-2 and IAV (Figure 3 and S3).

We found that the TEs that were upregulated in response to both SARS-CoV-2 and IAV infections in A549 cells of all different classes were enriched for active histone marks, with a subset of TEs highly enriched in H3K36m3 as wells as the combination of H3K27Ac, H3K4m3 and H3K9Ac (Figure 3A-B). Upregulated SINEs and DNA elements were especially enriched in response to SARS-CoV-2 infection compared to both random TEs and IAV-induced TEs. By contrast, all classes of SARS-CoV-2-induced TEs were depleted for H3K27me3 prior to infection (Figure 3C). This is in stark contrast to IAV-induced TEs, which show no such depletion (Figure 3C).

As IAV infection results in HERV-W induction following H3K9me3 demethylation, (Li et al., 2014), H3K9me3 was enriched on LTRs which were induced in response to IAV infection (Figure 3D). In contrast, H3K9m3 was mainly enriched on LINEs that were responsive to SARS-CoV-2 infection. A closer examination revealed that many of the upregulated LINEs and SINEs (but not LTRs) marked by H3K9me3, were also marked by the active HM signatures (Figure 3E-F). This suggests that these TEs are in a poised state ready for infection-induced expression.

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 H3K27m3, and are comprised of subsets of TEs, including SINEs and LINEs with a highly

active chromatin profile, LINEs uniquely marked by H3K9m3, and a bivalent groups of LINEs and SINEs marked by both repressive and active marks.

Re-analyzing data published by Huang et al. and Blanco-Melo et al. we provide evidence that unlike other viruses, which can strongly induce TEs following infection (Macchietto et al., 2020), SARS-CoV-2 infection of an 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 (Burns, 2020). Our analysis further showed that SARS-CoV-2 can 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 (Burns, 2020; Chiappinelli et al., 2015; Roulois et al., 2015). Finally, since TE expression increases globally across most tissues with age, our model explains the aging-related severity to SARS-CoV-2 infection.

Discussion

As more than 40% of the human genome is composed of TEs, proper epigenetic silencing of these elements is fundamental to maintaining genome stability. In differentiated cells, TE expression is inhibited via DNA methylation, suppressive histone modifications, and RNA-mediated mechanisms (Pehrsson et al., 2019). In this study we observed that TEs are globally upregulated in response to SARS-CoV-2 infection in cancer cells, but not in primary lung epithelial cells. In cancer cells, global epigenetic dysregulation of TEs is often observed, rendering the TEs poised for activation by viral infection. This is supported by our results, which show that both SARS-CoV-2 and IAV infections induce higher TE expression in A594 but not in NHBE or iAT2 cells. For all samples examined, we observed a direct correlation between TE expression and IFN levels, supporting a role of TEs upregulation in the activation of the host immune response.

Impaired IFN response is a hallmark of severe COVID-19 in humans (Hadjadj et al., 2020) and in mouse models infected with either SARS-CoV-1 or SARS-CoV-2 (Channappanavar et al., 2016; Israelow et al., 2020). This is possibly explained by evasion strategies to dampen innate immune surveillance developed by SARS-CoV-2, which delay the kinetics of IFN production in response to viral challenges (Yin et al., 2021). For example, SARS-CoV-2 was shown to inhibit IFN signaling by blocking STAT1 nuclear import (Miorin et al., 2020; Xia et al., 2020). Several ERVs contain STAT1 binding sites and act as IFN-inducible enhancers for a subset of IFN stimulated genes, which are specifically important for promoting the inflammatory response (Chuong et al., 2016). Our results also point to the potential involvement of some epigenetic modifiers, which show distinct patterns of expression in response to infection with SARS-CoV-2. Their expression is strongly correlates with TE expression, hinting at possible beneficial advantage to the virus. Additionally, a significant number of the host-viral interacting proteins are inversely correlated with the TE response, again suggesting an active evasion mechanism activated by the SARS-CoV-2. Taken together, this might explain the low TE response to SARS-CoV-2 compared to other viruses, when infecting the same cells in a similar MOI.

The TEs that are upregulated in response to SARS-CoV-2 infection share a unique epigenetic pattern: they are frequently marked by the active histone modifications H3K4me3 and H3K36me3, but curiously, some of them also carry the heterochromatin-related mark, H3K9me3. We suggest that these TEs are caught in a poised state which renders them ready for activation as the first line of host defense in response to viral infection.

Severe COVID-19 is much more frequently observed in old individuals. Aging is strongly associated with global changes in several epigenetic marks, including DNA demethylation and loss of the heterochromatin specific H3K9me3 mark (Perez et al., 2018; Zhang et al., 2015). Consequently, up-regulation of TE expression is positively correlated with aging (Bogu et al., 2019). We examined the basal expression levels of TEs in the different cell types. In line with our 'TE desensitization' hypothesis,

we observed a striking reverse correlation between basal TE expression and TE induction / IFN response. In two samples infected with a similar SARS-CoV-2 viral load, the 'old' normal cells (from a 79-year old female) show about 10 times less TE induction compared with the A549 cancer cells. We suggest that the gradual elevation in cytoplasmic TE transcripts that occur during aging cause desensitization of the cellular vRNA sensors, which are known to play a role in the Coronavirus induced IFN response (Channappanavar et al., 2016; Channappanavar et al., 2019). This reduction in the sensitivity of dsRNA-activated antiviral responses, together with the coronavirus specific EndoU cleavage of viral RNA (Ancar et al., 2020; Hackbart et al., 2020), likely explains the evasion from the dsRNA-induced IFN response in aged individuals.

In sum, according to our proposed desensitization model, the impaired TE response to SARS-CoV-2 infection in human lung epithelial cells is a result of both the active evasion mechanisms of the virus and the desensitization of the host cell caused by the normal epigenetic dysregulation observed with aging. The outcome of the impaired TE upregulation is a reduced dysfunctional early IFN response, which may result in an exacerbated inflammatory response leading possibly to a cytokine storm, manifested clinically by severe acute respiratory distress syndrome (ARDS) with systemic consequences, such as disseminated intravascular coagulation (Garcia, 2020; Park and Iwasaki, 2020). This is in line with previous works which demonstrate the effectiveness of reverse transcriptase inhibitors as a treatment to reduce IFN activation (De Cecco et al., 2019; Tunbak et al., 2020). Therefore, we propose that TE activation could be used as a therapeutic tool to generate higher cellular IFN response against 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.

ChIP-Seq data from ENCODE project (Schreiber et al., 2020):

ChIP Name (in this manuscript)	Experiment identifier (ENCODE)	GSE accession (GEO)	GSM accessions (GEO)
H3K4m3_1	ENCSR000DPD	GSE35583	GSM945244
H3K4m3_2	ENCSR203XPU	GSE91218	GSM2421502-GSM2421504
H3K9Ac	ENCSR000ASV	GSE29611	GSM1003544
H3K27Ac	ENCSR783SNV	GSE91337	GSM2421872- GSM2421874
H3K9m3	ENCSR775TAI	GSE91335	GSM2421866-GSM2421868
H3K27m3_1	ENCSR000AUK	GSE29611	GSM1003455
H3K27m3_2	ENCSR000AUJ	GSE29611	GSM1003577

Preprocessing and alignment

Raw reads from GSE147507 and GSE133329 were trimmed to remove Illumina adapters using the Trimmomatic software (Bolger et al., 2014) version 0.39. We used the STAR aligner version 2.7.1a (Dobin et al., 2013) 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 (Jin et al., 2015). 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 (Love et al., 2014). 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 (Jin et al., 2015; Liao et al., 2014) 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. *IFN* 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/ (Medvedeva et al., 2015).

ChIP-Seq signal quantification

To quantify the strength of ChIP-signals, we first calculated the number of reads aligned to flanking regions around all individual TEs located outside genes. Flanking regions were defined as the TE location and surrounding 1500 bp up- and down-stream, except for H3K27Ac for which only 500 bp were used. Next, raw count were normalized to regions lengths, followed by Z-score normalization. Z-scores were averaged among biological replicates. ChIP-signal clustering was carried out using R heatmap.2 function with the default parameters. Up-regulated TEs were defined based on a cutoff of at least 2-fold change with p-value < 0.05. Random controls (Figure S3) are based on 10,000 randomly selected TEs.

Percentages of TEs with active and repressive epigenetic signature of H3K36m3 and H3K9m3, respectively, were based on a cutoff of Z>2. Trivalent signature was defined as TEs that had high H3K36m3 and H3K9m3 levels (Z>2) and in addition had a general high active ChIP signal of H3K27Ac, H3K9Ac and H3K3m4 (Z>1 for all). For H3K27m3 depletion signature, a cutoff of Z=1 was used.

TE-gene correlation analysis

For TE response-gene correlation, we used Spearman rank correlation. Genes were defined as highly correlated and inversely-correlated if the correlated was larger than 0.8 and smaller than -0.8, respectively. This cutoffs corresponded to approximately the top and bottom 5%.

References

Ancar, R., Li, Y., Kindler, E., Cooper, D.A., Ransom, M., Thiel, V., Weiss, S.R., Hesselberth, J.R., and Barton, D.J. (2020). Physiologic RNA targets and refined sequence specificity of coronavirus EndoU. RNA *26*, 1976-1999.

Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.C., Uhl, S., Hoagland, D., Moller, R., Jordan, T.X., 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 e1039.

Bogu, G.K., Reverter, F., Marti-Renom, M.A., Snyder, M.P., and Guigó, R. (2019). Atlas of transcriptionally active transposable elements in human adult tissues. BioRxiv, 714212.

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

Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. Science (New York, NY) *165*, 349-357.

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

Channappanavar, R., Fehr, A.R., Vijay, R., Mack, M., Zhao, J., Meyerholz, D.K., and Perlman, S. (2016). Dysregulated Type I Interferon and Inflammatory Monocyte-Macrophage Responses Cause Lethal Pneumonia in SARS-CoV-Infected Mice. Cell Host Microbe *19*, 181-193.

Channappanavar, R., Fehr, A.R., Zheng, J., Wohlford-Lenane, C., Abrahante, J.E., Mack, M., Sompallae, R., McCray, P.B., Jr., Meyerholz, D.K., and Perlman, S. (2019). IFN-I response timing relative to virus replication determines MERS coronavirus infection outcomes. The Journal of clinical investigation *129*, 3625-3639.

Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., 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, G.M., Quinton, S., Ross, A., Sexton, C.E., Alberto, A., Clymer, C., Churchill, D., Navarro Leija, O., and Han, M.V. (2019). Transcriptome analyses of tumor-adjacent somatic tissues reveal genes co-expressed with transposable elements. Mob DNA *10*, 39.

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

De Cecco, M., Ito, T., Petrashen, A.P., Elias, A.E., Skvir, N.J., Criscione, S.W., Caligiana, A., Brocculi, G., Adney, E.M., Boeke, J.D., *et al.* (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature *566*, 73-78.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Galani, I.-E., Rovina, N., Lampropoulou, V., Triantafyllia, V., Manioudaki, M., Pavlos, E., Koukaki, E., Fragkou, P.C., Panou, V., and Rapti, V. (2020). Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. Nat Immunol, 1-9.

Garcia, L.F. (2020). Immune Response, Inflammation, and the Clinical Spectrum of COVID-19. Front Immunol *11*, 1441.

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., *et al.* (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459-468.

Gralinski, L.E., Sheahan, T.P., Morrison, T.E., Menachery, V.D., Jensen, K., Leist, S.R., Whitmore, A., Heise, M.T., and Baric, R.S. (2018). Complement Activation Contributes to Severe Acute Respiratory Syndrome Coronavirus Pathogenesis. mBio *9*.

Hackbart, M., Deng, X., and Baker, S.C. (2020). Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. Proc Natl Acad Sci USA *117*, 8094-8103.

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Pere, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., *et al.* (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science *369*, 718-724.

He, J., Feng, D., de Vlas, S.J., Wang, H., Fontanet, A., Zhang, P., Plancoulaine, S., Tang, F., Zhan, L., Yang, H., *et al.* (2006). Association of SARS susceptibility with single nucleic acid polymorphisms of OAS1 and MxA genes: a case-control study. BMC Infect Dis *6*, 106.

Huang, J., Hume, A.J., Abo, K.M., Werder, R.B., Villacorta-Martin, C., Alysandratos, K.D., Beermann, M.L., 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 *Sep 18*, S1934-5909.

Israelow, B., Song, E., Mao, T., Lu, P., Meir, A., Liu, F., Alfajaro, M.M., Wei, J., Dong, H., Homer, R.J., *et al.* (2020). Mouse model of SARS-CoV-2 reveals inflammatory role of type I interferon signaling. J Exp Med *217*.

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

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*.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. Cell *153*, 590-600.

Li, F., Nellaker, C., Sabunciyan, S., Yolken, R.H., Jones-Brando, L., Johansson, A.S., Owe-Larsson, B., and Karlsson, H. (2014). Transcriptional derepression of the ERVWE1 locus following influenza A virus infection. J Virol *88*, 4328-4337.

Li, H., Xiang, X., Ren, H., Xu, L., Zhao, L., Chen, X., Long, H., Wang, Q., and Wu, Q. (2020). Serum Amyloid A is a biomarker of severe Coronavirus Disease and poor prognosis. J Infect *80*, 646-655.

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

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

Macchietto, M.G., Langlois, R.A., and Shen, S.S. (2020). Virus-induced transposable element expression up-regulation in human and mouse host cells. Life science alliance *3*.

Medvedeva, Y.A., Lennartsson, A., Ehsani, R., Kulakovskiy, I.V., Vorontsov, I.E., Panahandeh, P., Khimulya, G., Kasukawa, T., Consortium, F., and Drablos, F. (2015). EpiFactors: a comprehensive database of human epigenetic factors and complexes. Database (Oxford) *2015*, bav067.

Miorin, L., Kehrer, T., Sanchez-Aparicio, M.T., Zhang, K., Cohen, P., Patel, R.S., Cupic, A., Makio, T., Mei, M., and Moreno, E. (2020). SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. Proc Natl Acad Sci USA *117*, 28344-28354.

Park, A., and 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, E.C., Choudhary, M.N., Sundaram, V., and Wang, T. (2019). The epigenomic landscape of transposable elements across normal human development and anatomy. Nature Communications *10*, 1-16.

Perez, R.F., Tejedor, J.R., Bayon, G.F., Fernández, A.F., and Fraga, M.F. (2018). Distinct chromatin signatures of DNA hypomethylation in aging and cancer. Aging Cell *17*, e12744.

Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang, G., Jones, P.A., Pugh, T.J., *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, M.H., Hay, R.T., and Hale, B.G. (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.

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

Silvin, A., Chapuis, N., Dunsmore, G., Goubet, A.G., Dubuisson, A., Derosa, L., Almire, C., Henon, C., Kosmider, O., Droin, N., *et al.* (2020). Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. Cell *182*, 1401-1418 e1418.

Sorek, M., Cohen, L.R.Z., and 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.

Tunbak, H., Enriquez-Gasca, R., Tie, C.H.C., Gould, P.A., Mlcochova, P., Gupta, R.K., Fernandes, L., Holt, J., van der Veen, A.G., Giampazolias, E., *et al.* (2020). The HUSH complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s. Nat Commun *11*, 5387.

van de Veerdonk, F.L., and Netea, M.G. (2020). Blocking IL-1 to prevent respiratory failure in COVID-19. Crit Care *24*, 445.

Verity, R., Okell, L.C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P.G., and Fu, H. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet infectious diseases *20*, 669-677.

Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J.Y.-C., Wang, H., Menachery, V.D., Rajsbaum, R., and Shi, P.-Y. (2020). Evasion of type I interferon by SARS-CoV-2. Cell Reports *33*, 108234.

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

Yin, X., Riva, L., Pu, Y., Martin-Sancho, L., Kanamune, J., Yamamoto, Y., Sakai, K., Gotoh, S., Miorin, L., De Jesus, P.D., *et al.* (2021). MDA5 Governs the Innate Immune Response to SARS-CoV-2 in Lung Epithelial Cells. Cell Rep *34*, 108628.

Zhang, W., Li, J., Suzuki, K., Qu, J., Wang, P., Zhou, J., Liu, X., Ren, R., Xu, X., and Ocampo, A. (2015). A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. Science *348*, 1160-1163.

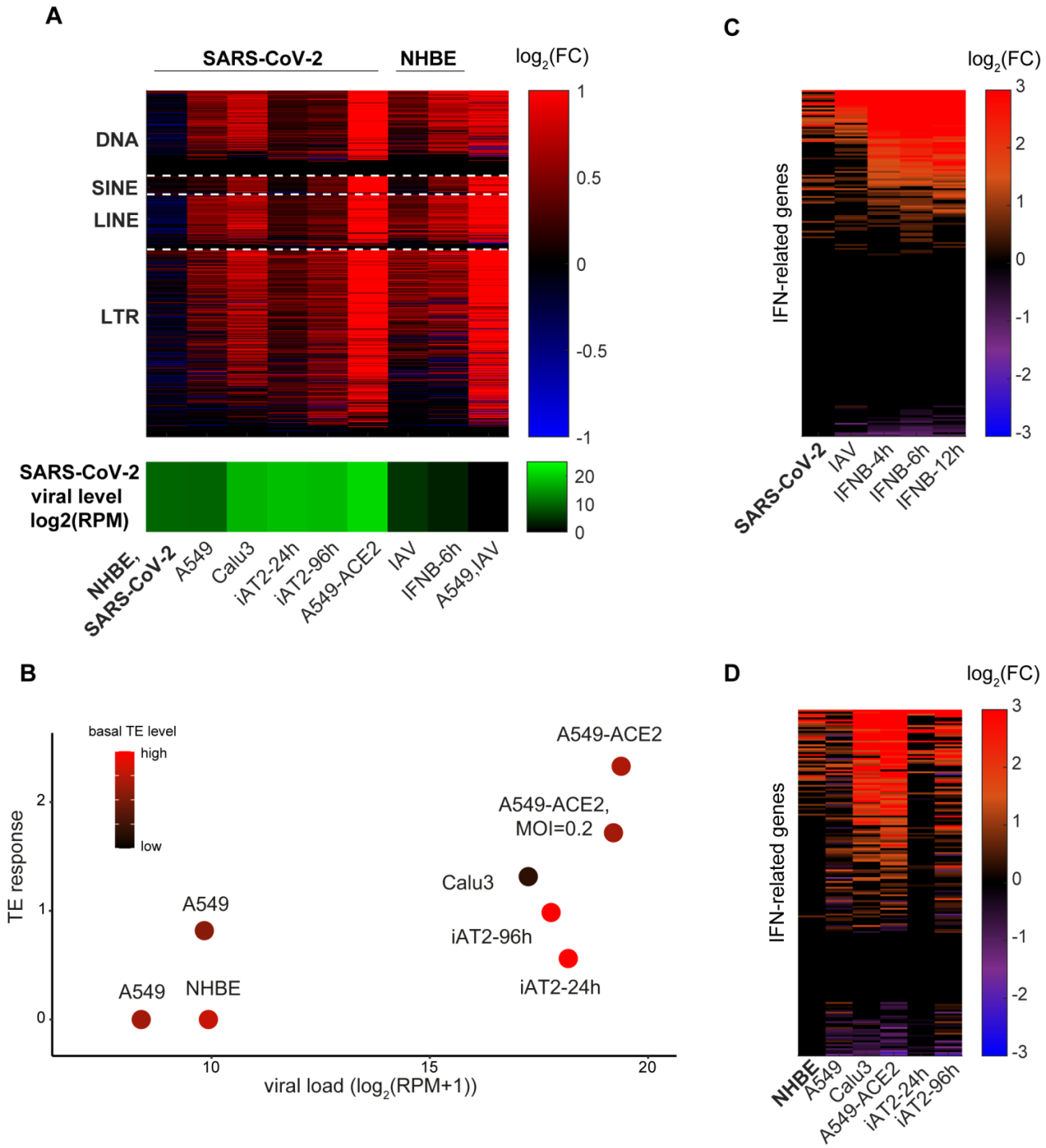

Figure 1. TEs expression changes and IFN response to SARS-CoV-2 and IAV infections. A. Log2-fold change in expression level of TE subfamilies (DNA, SINE, LINE and LTR) in different cell lines and in the different conditions indicated. SARS-CoV-2 viral levels (green) are depicted in the bottom panel. **B.** TE induction levels is 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). **C-D.** The IFN transcriptional response among NHBE cells (C) and among SARS-CoV-2-infected cells (D).

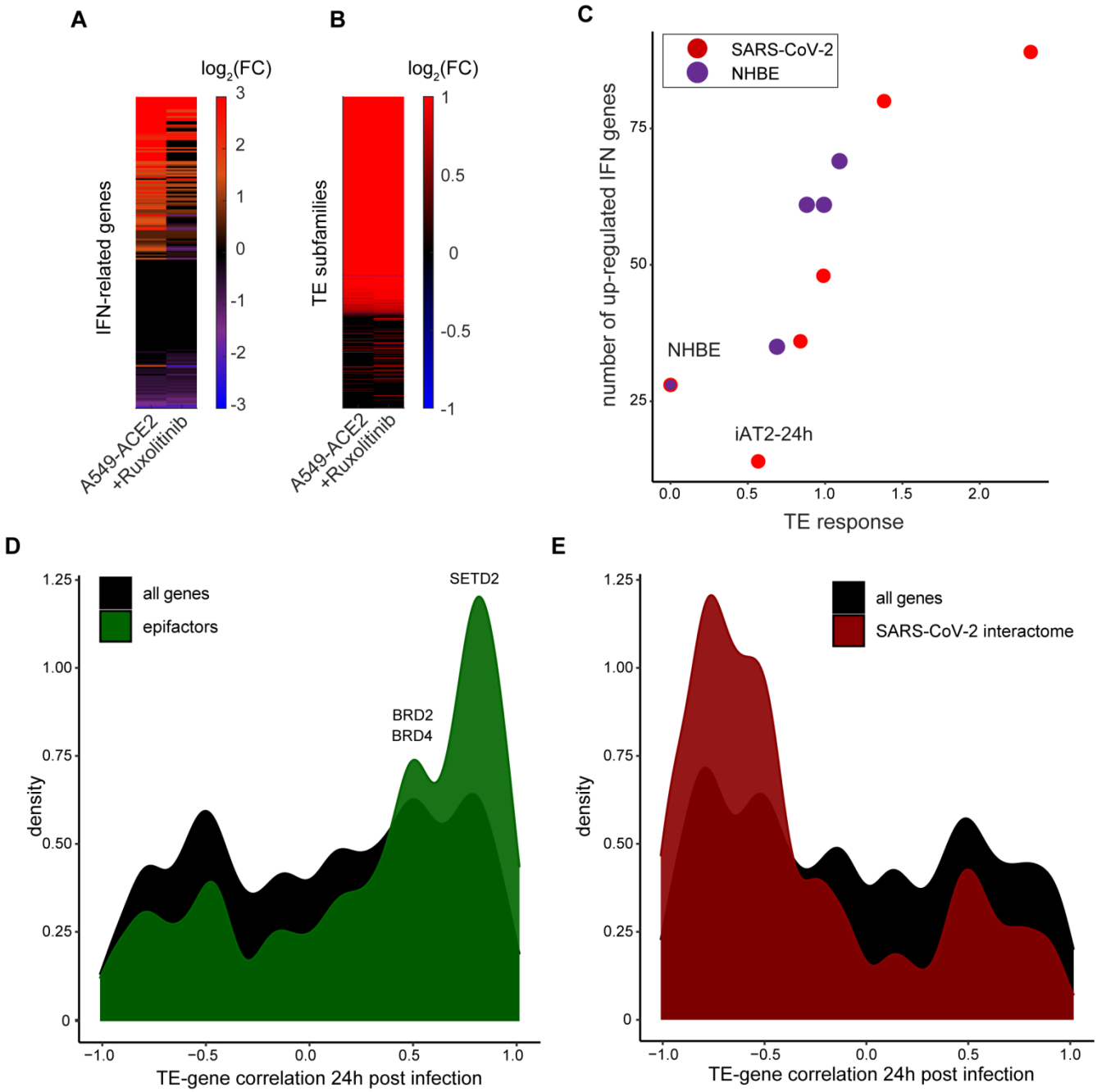

Figure 2: TE response correlates. A. The IFN transcriptional response of A549 cells over-expressing the ACE2 receptor dramatically decreases upon Ruxolitinib treatment. **B.** TE up-regulation persists even with Ruxolitinib treatment. **C.** IFN transcriptional changes correlate with the TE induction levels among SARS-CoV-2-infected cells (red) and among NHBE cells (purple). The circle sizes are proportional to the TE induction level. TE response is calculated as the 95-percentile of log2 fold-change of all TE subfamilies. **D-E**. Distribution of TE response-gene correlations in Blanco-Melo et al. for genes that increase (D) or decrease (E) with TE expression in iAT2 cells. Black represents correlation distribution of all genes. Correlation distribution of epifactors and SARS-CoV-2 interactome genes are in green and red, respectively.

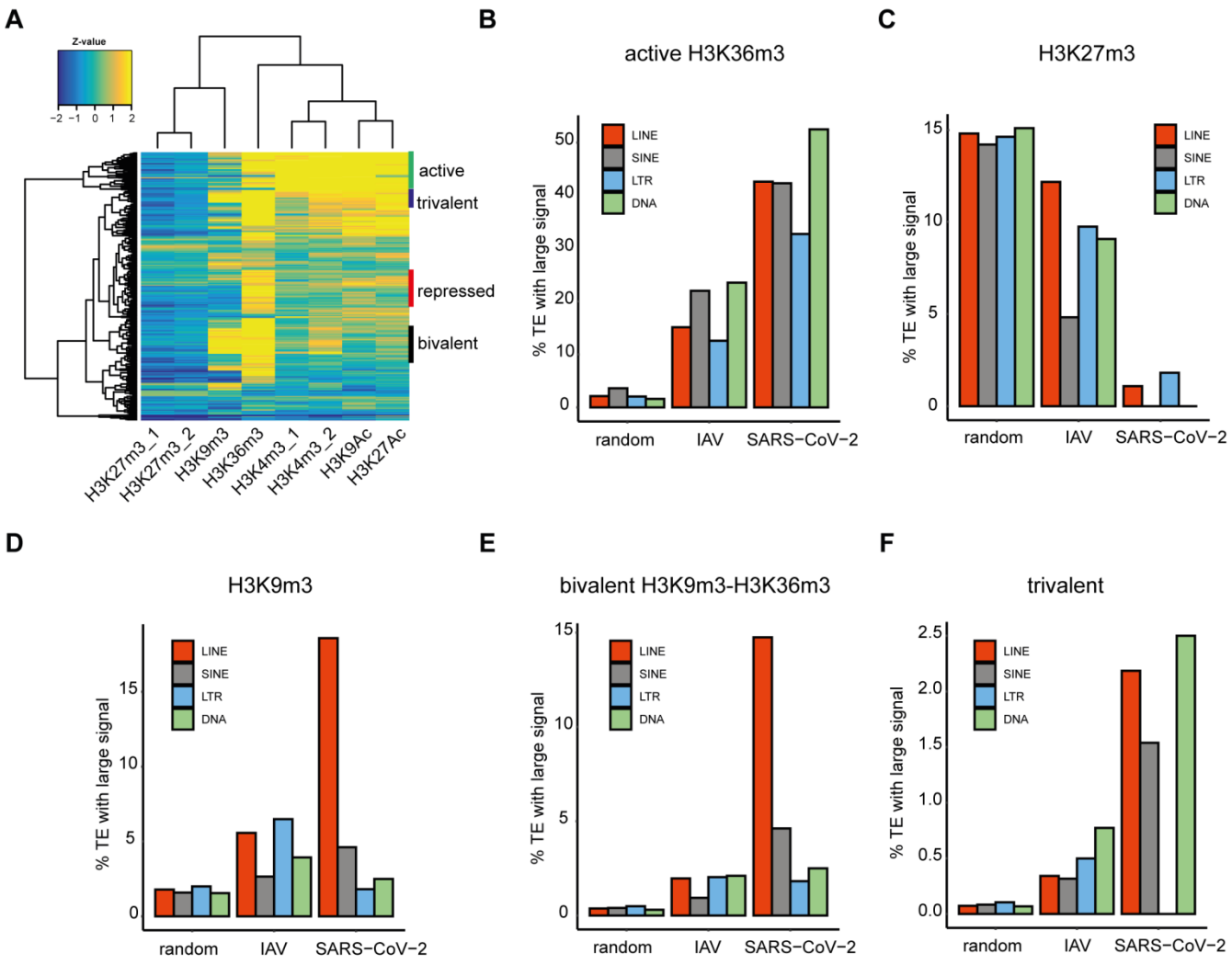

Figure 3: individual TEs induced by SARS-CoV-2 in A549 cells have a unique epigenetic profile. A. Hierarchical clustering of histone modifications signal in non-infected A549 cells around all up-regulated TEs in response to SARS-CoV-2 infection in A549 cells. **B.** Percentage of TEs with high normalized ChIP signal of H3K36m3 on SARS-CoV-2 induced TEs, IAV induced TEs and on all TEs outside genes. **C-D.** Same as B for H3K27m3 and H3K9m3, respectively. **E.** Percentage of TEs with high normalized ChIP signal of both H3K9m3 and H3K36m3. **F.** Percentage of TEs with high normalized ChIP signal of H3K9m3 and all active histone modifications (H3K4m3, H3K9Ac and H3K27Ac). See Methods for more details.

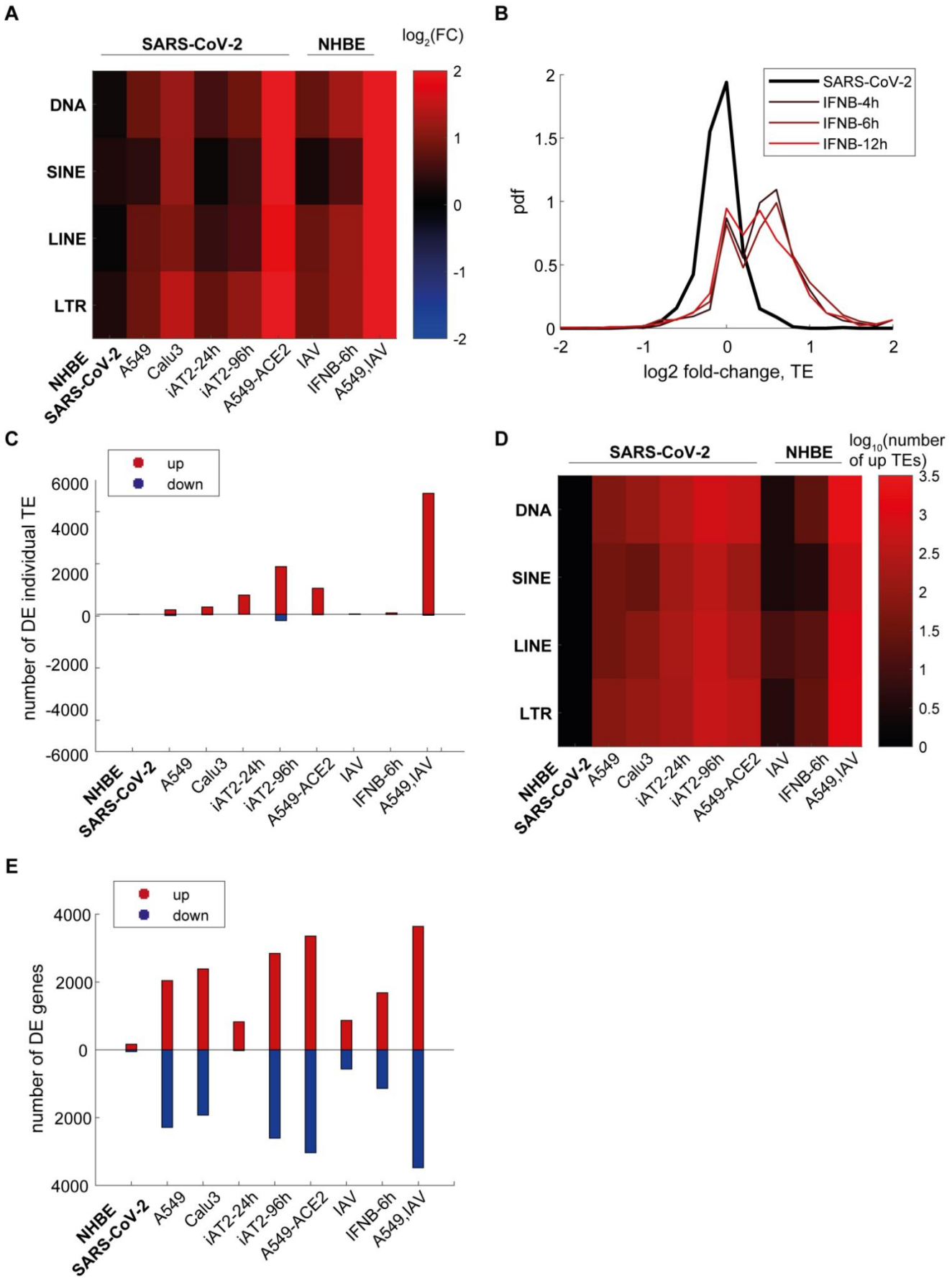
A
SARS-CoV-2
NHBE
log2(FC)
DNA
SINE
LINE
LTR
NHBE
SARS-CoV-2
A549
Calu3
iAT2-24h
iAT2-96h
A549-ACE2
IAV
IFNB-6h
A549,IAV
B
SARS-CoV-2
IFNB-4h
IFNB-6h
IFNB-12h
pdf
log2 fold-change, TE
C
up
down
number of DE individual TE
D
SARS-CoV-2
NHBE
log10(number of up TEs)
E
up
down
number of DE genes

Figure S1. TEs of all class types are not induced in normal human bronchial epithelial (NHBE) cells infected with SARS-CoV-2. A. Heatmap of the 95-percentile of the log2-fold expression change of TE subfamilies in different conditions among the four main TE classes. **B**. The distribution of log2-fold expression change of TE subfamilies in NHBE cells treated with IFNB (different shades of red) compared to cells infected with SARS-CoV-2 (black). **C.** The number of significantly up-regulated (red) and down-regulated (blue) individual TEs in the different indicated cell lines and conditions. **D.** same as (A) for the number of significantly up-regulated individual TEs. **E.** The number of significantly up- and down-regulated genes.

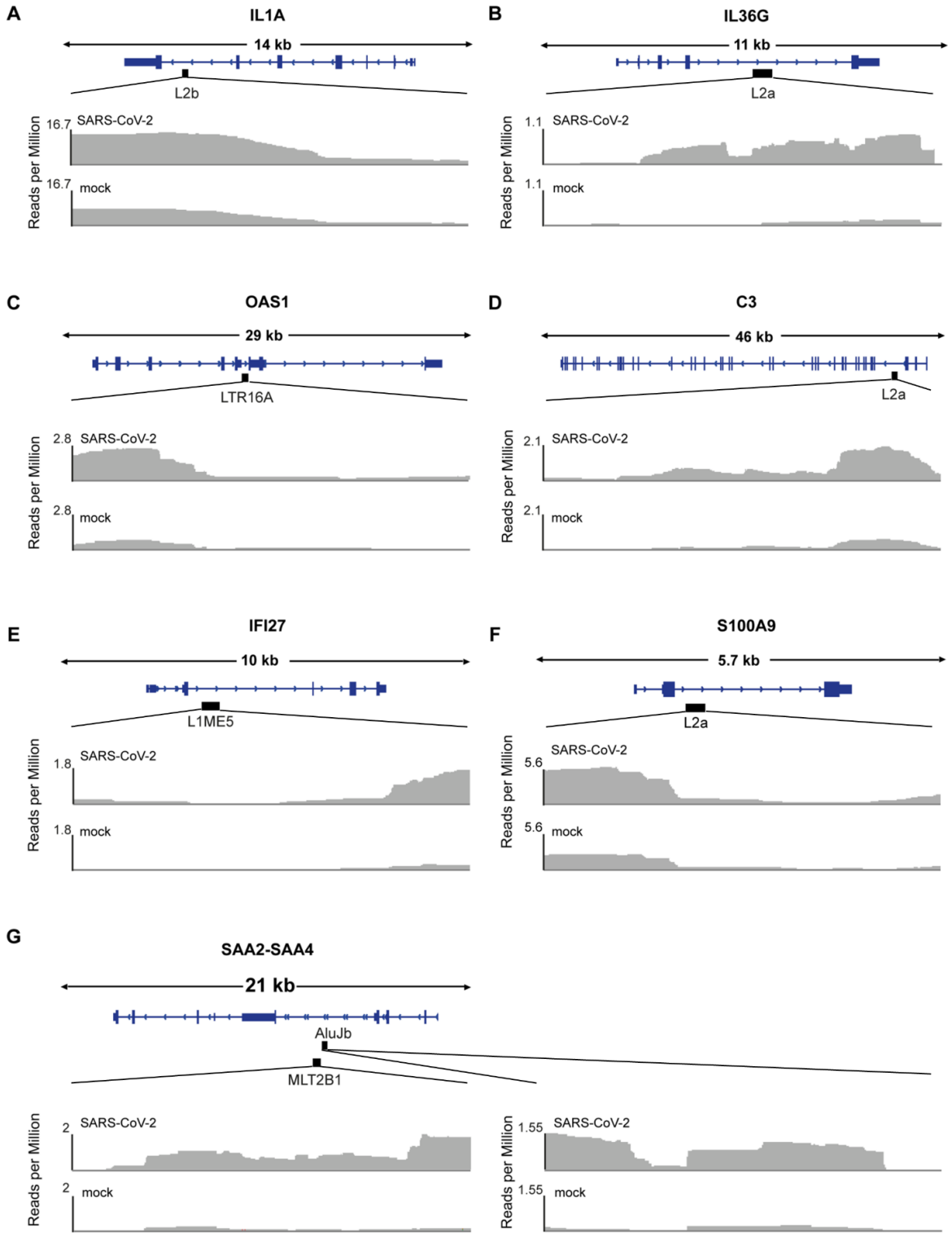

Figure S2. IGV screen shots of the TEs up-regulated in SARS-CoV-2 infected NHBE cells. Shown are the normalized counts (Reads per Million) in infected cells vs control of the eight significantly up-regulated TEs located within the seven genes shown: A. IL1A B. IL36G C. OAS1 D. C3 E. IFI27 F. S100A9 G. SAA2-SAA4.

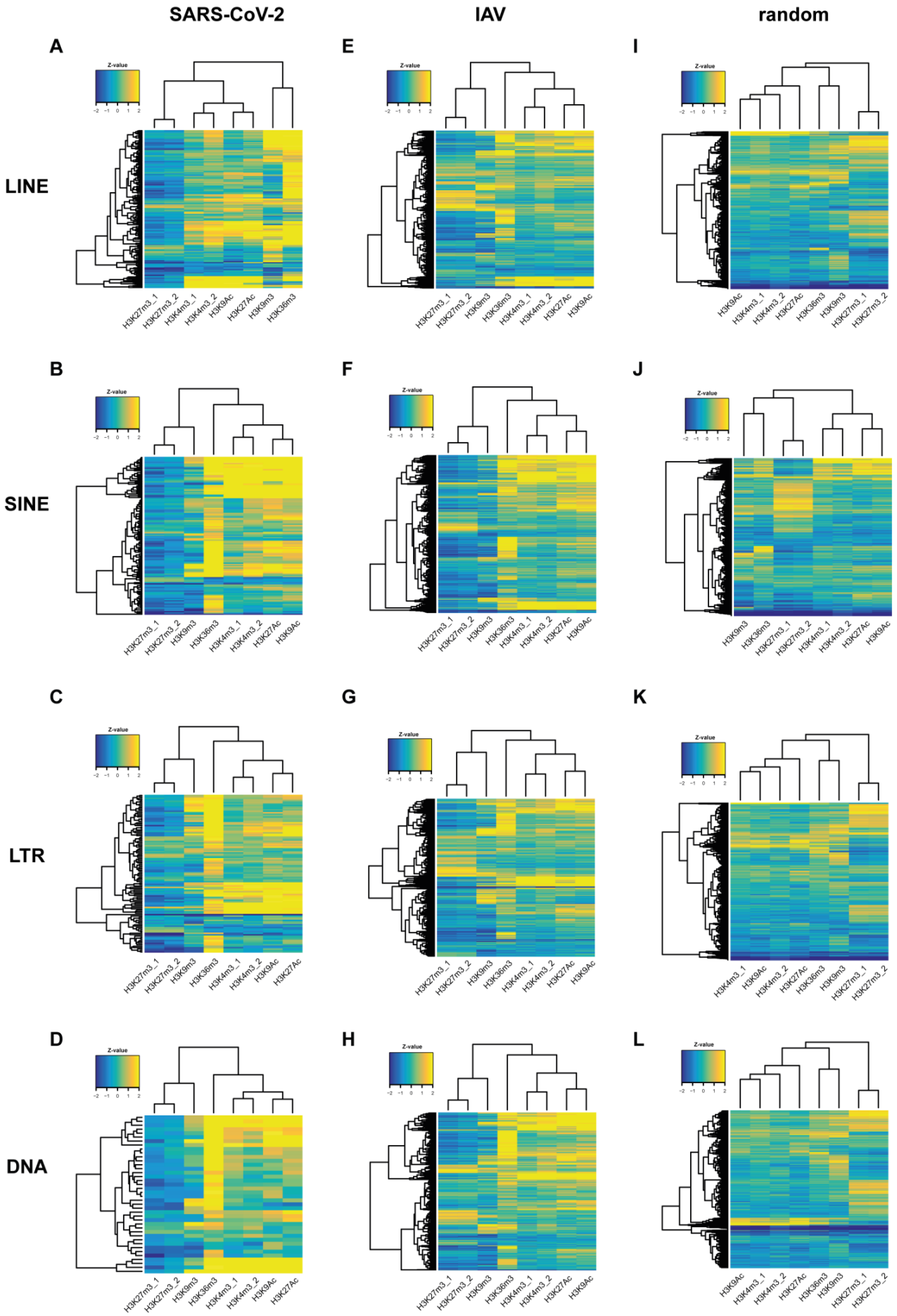
SARS-CoV-2
IAV
random
A
B
C
D
E
F
G
H
I
J
K
L
LINE
SINE
LTR
DNA
Z-value
-2 -1 0 1 2
H3K27m3_1
H3K27m3_2
H3K4m3_1
H3K4m3_2
H3K9Ac
H3K27Ac
H3K9m3
H3K36m3

Figure S3: The epigenetic signature TEs induced by SARS-CoV-2 and IAV infections in A549 cells. Hierarchical clustering of the ChIP signal of the TEs that are up-regulated in response to SARS-CoV-2 infection (A-D), IAV infection (E-H) or random 10,000 TEs (I-L). TEs are divide by the four main classes: LINE (A,E,I), SINE (B,F,J), LTR (C,G,K) and DNA (D,H,L).