Cell Genomics Transparent Peer Review Record

Summary

Initial submission:	Received : 6/14/2022
	Scientific editor: Laura Zahn
First round of review:	Number of reviewers: 3 Revision invited : 7/26/2022 Revision received : 11/15/2022
Second round of review:	Number of reviewers: 3 Accepted : 3/6/2023
Data freely available:	Yes
Code freely available:	Yes

Referees' reports, first round of review

Reviewer #1: Manuscript Summary:
In this manuscript, the authors used a computational approach to study the relationship between transposable elements (TEs) and variability in the host response to Influenza A virus (IAV) infection. They used a recently published dataset of monocyte-derived macrophages before and after infection with IAV. They reanalyzed the data focusing on transcriptional and epigenetic changes in TEs after IAV infection. The main study question is whether TEs play a role in viral load and innate immunity response variability between individuals. This is an exciting hypothesis and potentially important question to study.

Major Issues:
The authors did a good job dissecting the changes in the epigenetic landscape at TEs during IAV infection, which is appreciated. However, the authors failed to demonstrate that TEs have a causal or even a contributing role in immune response (IR) variability, which greatly diminishes my enthusiasm. Instead, they hammer on the "association" between epigenetic changes at TEs and IR variability, which could merely be a side-effect of known variability in IR after any viral infection. A variable immune response is associated with variability in pioneer Transcription factors (TFs) activity, and this is the likely cause of variability in chromatin accessibility genomewide including at TEs, which harbor many binding sites for these TFs.

Also, a significant issue is the focus on analyzing TEs bundled as families. This could give some valuable insights, but IR variation could also be tied to genetic or epigenetic alteration of individual TEs, or even polymorphic TE insertions that could contribute to IR variation. The authors could utilize their comprehensive dataset (including whole-genome sequencing) to look for such loci (which could include novel L1/SINE/Alu insertions, or LTR/LTR recombination events). If such loci could be found computationally, experimental validation using reporter assays or locus-specific knock-out could be a bonus addition to strengthen the study.

On another point, the authors built a model incorporating INF signature and TEs epigenetic changes as variables. It is unclear how this model supports a role for TEs sequences in influencing IR and viral load. As discussed above, TEs epigenetic changes could directly result from variation in INF response (a complex multi-factorial phenomenon).

In conclusion, in its current state, this manuscript doesn't provide any evidence that transposable elements (as independent factors) play a role in the immune response to IAV or in determining viral load. Establishing the association alone is not a novel finding, as it is well known that significant transcriptional changes in different biological settings are associated with major changes in chromatin accessibility, including TEs, which constitute a significant fraction of the genomes. Additional work is needed to investigate the roles of TEs polymorphism and epi-variation in IAV IR. I also find some redundancy in the results and figures presented in this manuscript to show the association between TEs epigenetic changes and IAV IR. This data could be reduced into fewer figures to convey the same conclusion.

Minor Issues/Comments:
Figure 2C: What exactly is meant by epigenetic variability and variable regions in this context? It should be clarified briefly in the text and reference the methods for details. Also, the

conclusion that "these results are consistent with some variability of TE transcription post-infection" is vague. Also, the difference in variability between TE and non-TE is evident in H3K4me3 peaks regardless of infection, how does this support a role for TEs in the modulation of infection response?

Figure S3E: add a legend for color scale.

Page 9: "Among high var. families we consistently observed more commonly (≥ 25% individuals of one group) and rarely (< 25%) accessible instances that were specific to Group 3 individuals". Rephrase this sentence; it is difficult to understand.

Page10-11: "Instances from high var. families .... Overall, low var. and high var. showed distinctive chromatin patterns post-infection". How do you interpret these findings? How does support the overall study conclusions?

Fig 5A: is THE1B Low var. Or High var.? It is mentioned as high in the text but low in the figure annotation.

Reviewer #2: In this work, the authors examined epigenomic and transcriptomic data in influenza-infected primary macrophages from a panel of patients. They found the severity of infection, measured by transcriptomic viral load, to be correlated with certain patterns of transposon expression, accessibility, and epigenetic character. They found that many transposon families were upregulated upon influenza infection of the cells, and some were significantly correlated with viral load from patient to patient. Then, they found that some of these families were variable in their response to infection across patients. The authors posit that these transposon families contribute to the observed variability in viral load and severity of infection by acting as variably active cis regulatory elements, contributing both immune and non-immune-associated transcription factor motifs. Overall this work is a rigorous and comprehensive analysis of an interesting dataset and provides new but largely correlative evidence supporting the idea that TEs could contribute to variable immune responses.

Major Comments
1. The authors conduct some interesting analysis related to KZNF proteins and show there is some enrichment based on Imbeault data. Given that the ZNF repression is suggested as a major mechanism regulating variability these analyses could and should be further strengthened. For example, analyzing whether there is a correlation with the expression of specific KZNFs (all those in Supp fig 5G rather than just TRIM28/SETDB1).
2. The authors' conclusions are drawn exclusively from observed correlations between effects of infection. Empirical testing of these conclusions may be beyond the scope of this paper, but discussion of alternative explanations for these correlations is in order. For instance, the enrichment of inducibly-accessible TEs near inducibly-transcribed genes could be explained as two independent effects of regional chromatin changes, rather than causally related as the authors

imply.
3. The sequencing data used in this study were collected from a single time point during infection and were apparently performed only once per individual. Possible variations over time are not observable but not discussed as a caveat.
4. Motif analyses to identify the transcription factors responsible for the observed correlations were performed based on transposon family consensus sequences. This is very useful but cannot identify motifs that were accrued in individual instances or subsets of instances during their expansion.

Minor Comments
* On page 5, in the first paragraph of the Results section, the wording "suggesting varying capacity to infection and/or to limit viral replication across individuals" is unclear. Perhaps a word is missing after "varying," or perhaps "varying capacity for infection" was meant?
* Minor stylistic suggestion: there are many abbreviations used in the text that reduce readibility, for example the use of "var." instead of simply spelling out "variable" is a little confusing to read at times.
* The section "High var. families contribute transcription factor.." (p11 end) is not written very clearly currently, could be made more readable.
* In several instances phrases like "higher proportion" and "more likely" are used in the text, but the comparisons being made are not always easy to follow. Language could be clarified.
* On page 17 in the first sentence there appears to be a typo. "Depression" should be "de-repression?"

Reviewer #3: In this correlative study, the authors employ RNA-seq data from monocyte-derived macrophages from 39 individuals, where cells were infected with influenza A virus for 24 hours. Post-infection, they measure the percentage of the transcriptome contributed by viral transcripts as an indicator of viral load and observe considerable variation between individuals. The infection induces upregulation of some TE families (mainly LTRs) either through direct or indirect affects following changes to the global transcriptome. In line with this, there is increased enrichment of active epigenetic marks H3K4me3 and H3K27ac at TEs in the flu samples (although H3K27ac did not reach significance). Analyses at the TE family level shows that THE1B, SVAC&D, MER52, MER41B and LTR12C and a few others gain enriched chromatin accessibility (ATAC-seq) in the Flu samples. Several of these (including THE1B) also gain H3K4me3 and H3K27ac in the Flu samples. Interestingly, some LINE-1 families become less accessible with reduced H3K27ac in the Flu samples. Some TE families show high variability of expression between individuals (including MER52, LTR12C). Several instances of these LTRs are positioned proximal to genes with roles in immunity, and have binding sites for STATs/NFkB transcription factors implying that they may function as enhancers. This suggests that differential expression of these TEs between individuals may impact on expression of interferon-regulated genes and viral load. The authors propose that 'high-variance' TEs are bound by KRAB-ZFP repressors, which are differentially expressed between individuals, potentially explaining the variable response to infection. It is interesting that high expression of certain TEs

correlates with low viral load and the idea that TEs and KRAB-ZFPs contribute to the response to infection is topical. However, there is a missing link of whether expression of the high-variance TEs tracks with higher expression of interferon-stimulated genes globally and low viral load, and there is no direct evidence that these KRAB-ZFPs regulate activity of the high-variance TEs.

Comments

1.There appears no evidence that the KRAB-ZFPs assessed (ZKSCAN5 and ZNF460) regulate the LTRs that vary in expression between individuals or expression of any interferon-stimulated genes (and is the proposed mechanism through DNA methylation of LTR enhancers)? Does KO/knockdown of these KRAB-ZFPs influence the expression of the high variance TEs, the immune response and the viral load? The proposed model is a bit vague due to the data being a bit preliminary. What is meant by 'reduced TEs' in the model?

2.The higher the expression of TEs, the lower the viral load. Is this because the high expression of TEs correlates with high expression of interferon-stimulated genes (ISGs) they are potentially proximal to? ISGs function to limit viral replication and stimulate adaptive immunity. Several ISGs are mentioned but they are not interrogated systematically as a group of genes.

3.In the abstract and elsewhere, the authors state that 'TEs contribute to the activation of innate immunity'. This should be clarified to reflect what is known, i.e. 'TE expression increases upon infection' or 'some TEs (MER41) act as enhancers for genes involved in innate immunity' since it is not known if a global increase in TE expression is necessary or contributes to the establishment of innate immunity against any pathogen.

4.The term 'known immune regulators' in the abstract is quite vague. It would be clearer to refer to specific transcription factors.

5.Page 3: 'Endogenous Retroviruses (ERVs), are derived from ancient retrovirus, suggesting a potential association with infection and Immunity'. This is a bit confusing. Do you mean, they may retain viral features (the ability to reverse transcribe for example) that are recognized by nucleic acid sensors, making them able to induce IFN responses? Please clarify.

6.Page 3: 'Confirming this, an ERV family, MER41, was found to be co-opted as cisregulatory elements in the primate innate immune response'. It would be clearer to explain co-option of ERVs in terms of them already having intact promoters and enhancers, which can then be repurposed by the host to regulate host genes.

7.Page 3: 'derived from ancient retrovirus' should be 'derived from ancient retroviruses'.

8.Page 4: 'Meanwhile, loss of SETDB1 or SUMO-modified TRIM28, which are associated with histone methylation and Kruppel-associated box domain (KRAB) zinc finger proteins (ZNFs), will lead to the significant derepression of TEs in the immune response (Cuellar et al., 2017; Schmidt et al., 2019). Together, these studies suggest that TEs play a prominent role in human innate immunity'. This is a bit confusing: the SETDB1 paper cited is a cancer paper, which does

not inform us whether SETDB1 has a natural role in regulation of TEs in normal cells or upon infections. The second reference also doesn't appear to show that TEs play a prominent role in human innate immunity. Please tone down conclusions.

9.Page 7: 'That being said, we observed higher variability of H3K4me3 and lower variability of H3K27me3 mark in TEs compared to non-TE regions, respectively'. It would be helpful to include the percentages here like for the previous sentence comparing TE and non-TE regions.

10.Page 13: 'Notably, L1MA2, L1MA4, L1MA6, L1MA7, and L1MA8 were significantly enriched for MEF2 related motifs. MEF2 TFs are central developmental regulators (Potthoff and Olson, 2007), which are also required in the immune response that functions as an in vivo immune-metabolic switch' It would be helpful to explain this further and discuss why and how LINE-1 elements might be downregulated in the aftermath of a viral infection in the discussion. LINE-1 elements have been linked to inducing type I IFN responses and to being upregulated in disease settings (cancer, autoimmune diseases).

11.Page 17: 'In line with the involvement of TE transcripts in the activation of innate Immunity'. No references are cited here that relate to TE transcripts activating the innate immune response. There is a body of literature about inverted repeat Alu elements being self RNAs that are substrates for dsRNA sensing by MDA5. Some of those references would be appropriate here or other mechanistic studies. There is also a useful review on TEs and antiviral innate immunity: PMID: 33888553.

12.Figure 1a: The legend is a little confusing for the ethnicity data. The triangle and square could be unfilled rather than coloured grey since the colour changes depending on the infection status.

13.The figures were a bit big making them slow to download and view properly.

Authors' response to the first round of review

Reviewer #1:

Manuscript Summary:

In this manuscript, the authors used a computational approach to study the relationship between transposable elements (TEs) and variability in the host response to Influenza A virus (IAV) infection. They used a recently published dataset of monocyte-derived macrophages before and after infection with IAV. They reanalyzed the data focusing on transcriptional and epigenetic changes in TEs after IAV infection. The main study question is whether TEs play a role in viral load and innate immunity response variability between individuals. This is an exciting hypothesis and potentially important question to study.

Thank you for the accurate summary and positive comments.

Major Issues:
The authors did a good job dissecting the changes in the epigenetic landscape at TEs during IAV infection, which is appreciated. However, the authors failed to demonstrate that TEs have a causal or even a contributing role in immune response (IR) variability, which greatly diminishes my enthusiasm. Instead, they hammer on the "association" between epigenetic changes at TEs and IR variability, which could merely be a side-effect of known variability in IR after any viral infection. A variable immune response is associated with variability in pioneer Transcription factors (TFs) activity, and this is the likely cause of variability in chromatin accessibility genomewide including at TEs, which harbor many binding sites for these TFs.

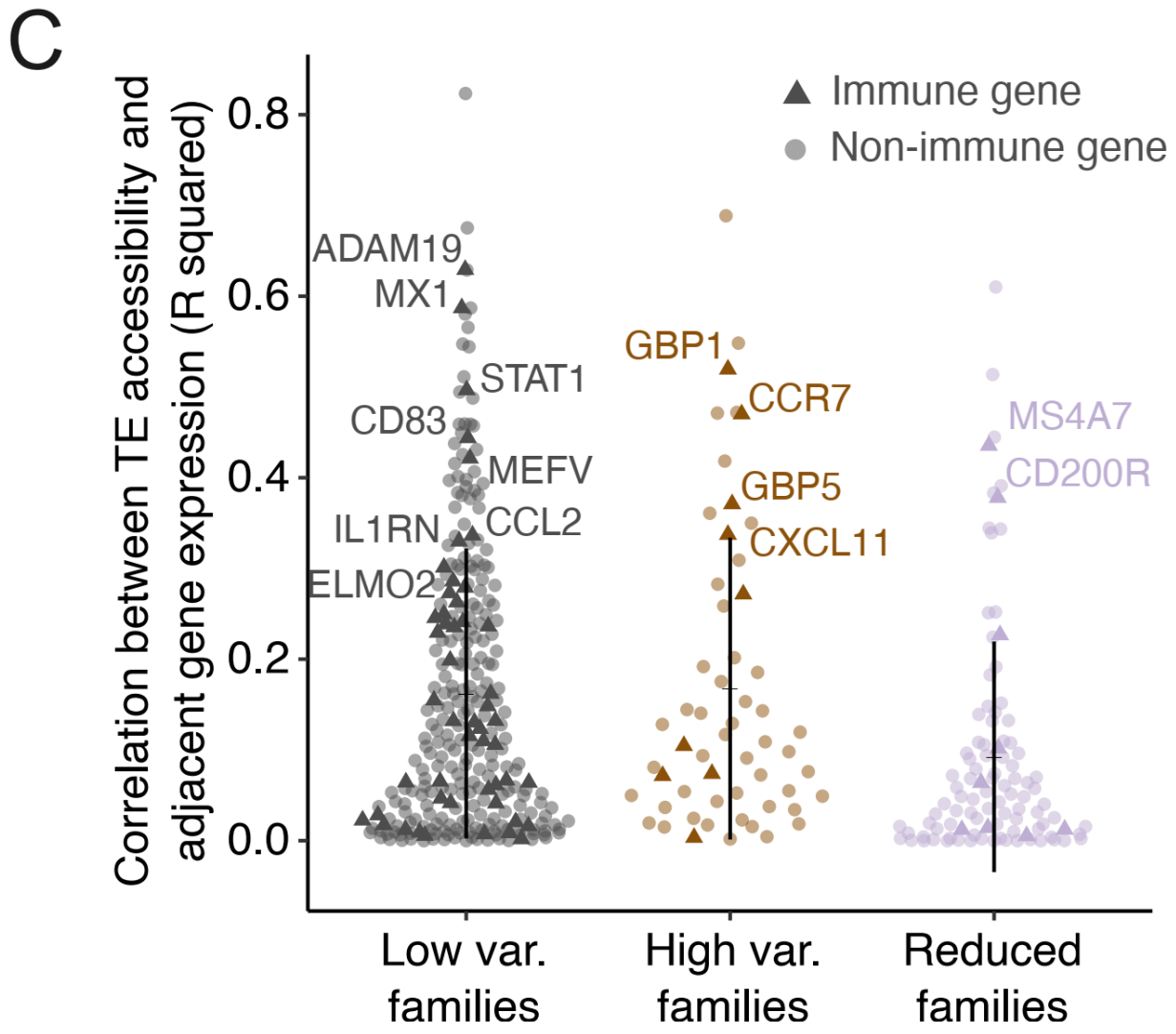
We thank you for your encouraging comments on the analysis of the TE epigenetic landscape during infection. In the text, we have been careful to say that what we observed is an association between infection and changes in the TE epigenetic landscape, which we feel was important to report since this result is novel in itself. That said, we agree with the reviewer that it would be interesting to implicate these changes to the IR itself. We do have results pointing in this direction, mainly the proximity of many of these regions to important genes known to play a role in IR. Indeed, in addition to the observation that up-regulated genes are enriched near high and low variable TE families we have identified (Figure 4A), we have now expanded our analysis and provided further evidence supporting the fact they act as variable enhancers and promoters for key immune genes. A total of 420 upregulated genes were found in proximity to repeat loci from enhanced families and 168 downregulated genes from reduced families (New Table S6). The correlation between the accessibility of some of these loci and their adjacent genes further supports coordinated regulation (New Figure 4C). Among these genes, 82 genes were near TE-loci from high variable families, including 17 IR genes. For example, we observed that two known IR genes, GBP2 and GBP5, are potentially regulated by TE-loci from the LTR12C high variable family (New Figure 4D and New Figure S5E). Moreover, a recent study using reporter assays in HEK293T and CD4+ T cell lines has independently and successfully validated that these two LTR12C instances could act as promoters regulating GBP2 and GBP5 expression (Srinivasachar Badarinarayan et al. 2020, 1). Taken together, this further suggests that high variable TE families contribute to the variable IR to IAV infection. We have expanded the corresponding section as follows:

“Finally, to further investigate which genes were potentially regulated by these TE-embedded sequences upon infection, we analyzed the list of nearby differentially expressed genes (≤ 50 kb) and observed an enrichment in various immune-related pathways (Figure S5C). Next, we selected the repeat loci from the enhanced and reduced TE families with significant changes in accessibility and active histone modifications (H3K4me1 and/or H3K27ac). A total of 420 upregulated genes were found in proximity (≤ 50 kb) to repeat loci from enhanced families and 168 downregulated genes from reduced families (Table S6). Of these, we found 17, 64, and 11 immune-related genes near instances from high variable, low variable and reduced families, respectively. The correlation between the accessibility of many of these loci and their adjacent genes further supports coordinated regulation (Figure 4C). For example, GBP5 gene is an interferon-induced gene and exhibits antiviral activity against viral infection (Tretina et al., 2019). An LTR12C instance and a MER1B instance with enhanced chromatin accessibility accompanied by an augmentation of H3K27ac and H3K4me1 upon infection can be found near this gene (Figure 4D). The accessibility of the two instances was positively correlated with GBP5 expression level post infection (Figure 4E). Furthermore, this specific LTR12C instance was previously validated to regulate GBP5 expression in cell lines (Srinivasachar Badarinarayan

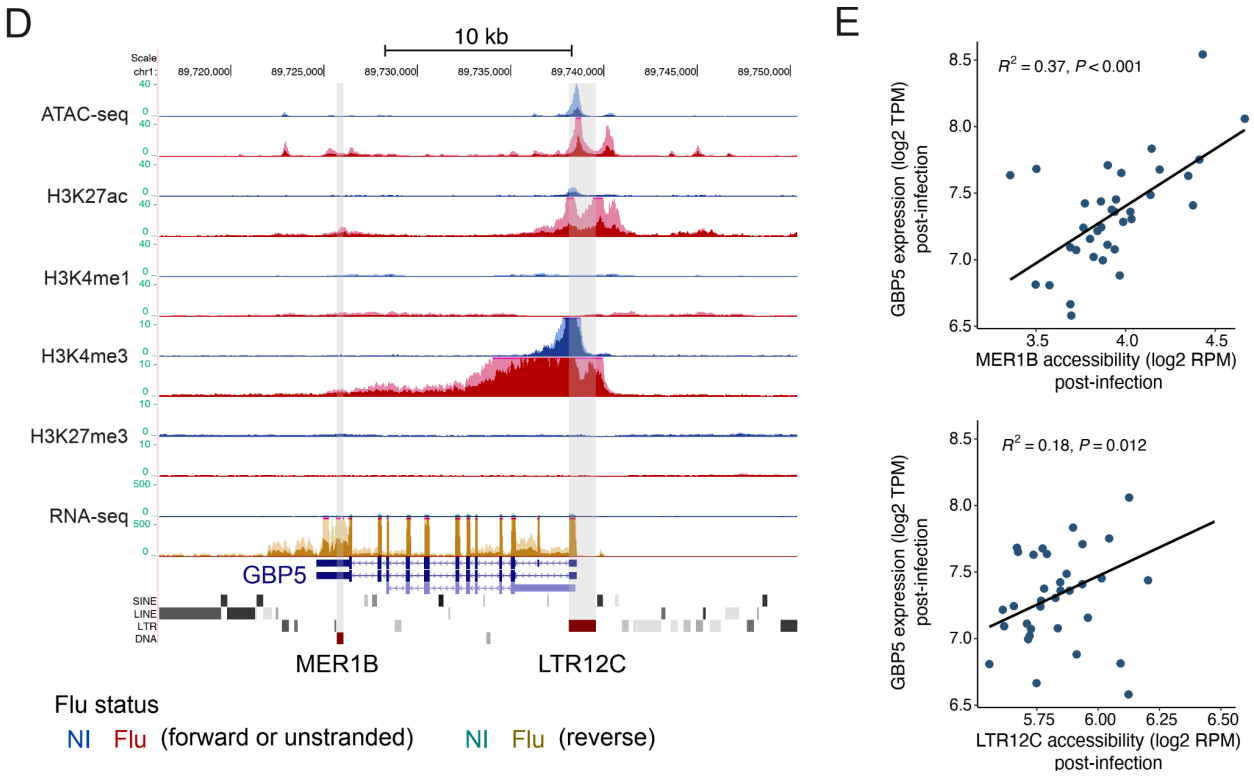
et al., 2020). In a different LTR12C instance near the up-regulated immune-related gene IL10RA, transcription was initiated at the open chromatin region within the repeat itself and was flu-specific (Figure S5D). We also confirmed the chromatin change at the LTR12C instance that was shown to be a promoter regulating GBP2 (Srinivasachar Badarinarayan et al., 2020) and a MER41 instance that was shown to be an enhancer regulating AIM2 (Figure S5E-S5F) (Chuong et al., 2016). Lastly, we identified several immune-related genes that were potentially regulated by adjacent instances from enhanced families, such as the TE gene pairs of MER52A-GBP1/3, LTR12C-TRIM22, THE1C-IFI44, THE1B-PSMA5, MLT2B3-CLEC4E, and tigger3a-ADAM19 (Figure S5G-S5L). Thus, some of the instances from the enhanced and reduced TE families behave like cis-regulatory elements regulating nearby immune genes."

We have also revised the abstract to clarify our findings:

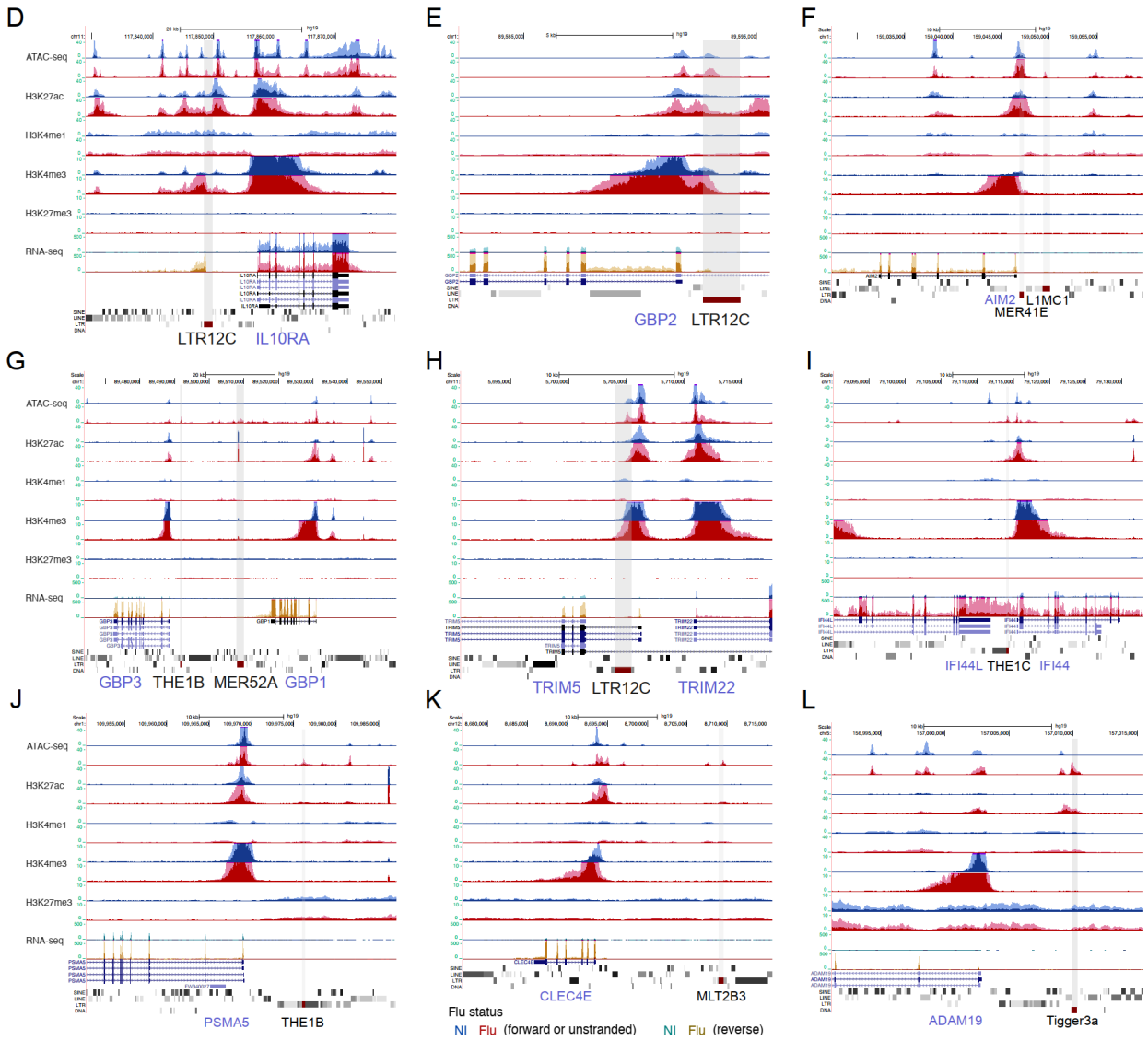
"We also observed a strong association between basal TE transcripts and viral load post infection and showed that TEs, and host factors regulating TEs, were predictive of the response. Our findings shed light on the variable transcriptional and epigenetic response to infection and the role TEs and KRAB-ZNFs may play in inter-individual variation in immunity."



New Figure 4 (C) Correlation between the accessibility of TE-loci with significant changes of both accessibility (ATAC-seq) and active histone modifications (H3K4me1 and/or H3K27ac) and adjacent gene expression (within 50 kb) post-infection (see Methods). Positively correlated upregulated genes are shown for enhanced families and downregulated genes are shown for reduced families. Strongly correlated immune genes (R squared ≥ 0.3, p value ≤ 0.05) are highlighted.



New Figure 4 (D) Example genomic view of an accessible LTR12C instance and MER1B instance potentially upregulating adjacent GBP5 gene expression post-infection. LTR12C and MER1B are highlighted as the shaded area with the increased accessibility, expression, H3K27ac, H3K4me1, and H3K4me3 activity. The dark shaded area denotes the distribution of the average RPM values and the light shaded area denotes the standard deviation. Signals of various epigenetic marks are shown in blue color for non-infected samples and red color for infected samples. For RNA-seq, forward and reverse transcripts are shown in blue and green color separately for non-infected samples; while forward and reverse transcripts are shown in red and brown color separately for infected samples. (E) Positive correlation between the accessibility of LTR12C and MER1B instances with GBP5 expression level post infection. R2 and p values computed by the linear regression model are shown.



New Figure S5 (D) Genomic view of an accessible LTR12C with the expression was upregulated and initiated at the open chromatin region post-infection. The LTR12C instance highlighted as the shaded area shows an upregulated accessibility, expression, and H3K4me3 activity. IL10RA gene located near the LTR12C instance is also significantly upregulated postinfection. (E-L) Example genomic views of instances with enhanced accessibility post-infection. Instances are highlighted as the shaded areas. Eight TE immune-related gene pairs are shown, i.e., LTR12C-GBP2, MER41-AIM2, MER52A/THE1B-GBP1/3, LTR12C-TRIM22, THE1C-IFI44, THE1B-PSMA5, and MLT2B3-CLEC4E, and Tigger3a-ADAM19. GBP2 has been validated to be regulated by the upstream LTR12C instance (Srinivasachar Badarinarayan et al., 2020). AIM2 has also been validated to be regulated by a MER41 instance (Chuong et al., 2016); interestingly, it may also be regulated by another TE instance. Other three TE instances reported by Chuong et al. that potentially regulate APOL1, IFI6, and SECTM1 did not show chromatin change in macrophages (https://computationalgenomics.ca/tools/epivar). The dark shaded area denotes the distribution of the average RPM values and the light shaded area denotes the standard deviation. Signals of various epigenetic marks are shown in blue color for

non-infected samples and red color for infected samples. For RNA-seq, forward and reverse transcripts are shown in blue and green color separately for non-infected samples; while forward and reverse transcripts are shown in red and brown color separately for infected samples.
Also, a significant issue is the focus on analyzing TEs bundled as families. This could give some valuable insights, but IR variation could also be tied to genetic or epigenetic alteration of individual TEs, or even polymorphic TE insertions that could contribute to IR variation. The authors could utilize their comprehensive dataset (including whole-genome sequencing) to look for such loci (which could include novel L1/SINE/Alu insertions, or LTR/LTR recombination events). If such loci could be found computationally, experimental validation using reporter assays or locus-specific knock-out could be a bonus addition to strengthen the study.

Thank you for this question. Due to the repetitive nature of sequences, instances of a TE family often share similar epigenetic states. That is why we reported some of the chromatin patterns following infection at the TE family level as was done in other studies (Chuong, Elde, and Feschotte 2016; Bogdan, Barreiro, and Bourque 2020). That being said, in this work, most of the analyses were performed at the instance-level. For example, we analyzed the variability of individual instances between individuals (Figure S3E) and also summarized the proportion of instances per family that were accessible in different individual groups (Figure S4A); the association with nearby genes was done at the instance-level (Figure 4 and Figure S5); we determined the accessible regions at each instance (Figure 5A-5B) and also grouped instances based on the accessible regions into different "TE peak regions" per family (Figure 5C); we performed the motif analysis of each individual instance and also computed the enrichment of motifs among instances with each "TE peak region" per family (Figure 5D). We have made minor modifications throughout the text to clarify this.

We also agree with you that the potential role of polymorphic insertion is interesting and that is the topic of a separate paper that uses the same dataset given the complexity of the question (Groza et al. 2022, https://www.biorxiv.org/content/10.1101/2021.09.29.462206v2).

On another point, the authors built a model incorporating INF signature and TEs epigenetic changes as variables. It is unclear how this model supports a role for TEs sequences in influencing IR and viral load. As discussed above, TEs epigenetic changes could directly result from variation in INF response (a complex multi-factorial phenomenon).

We agree with you that the model suggests a link between TEs and viral load but does not confirm that TEs influence viral load. That said, what we show in this section is that the TErelated features that we have identified at the basal state (before infection) are predictive of viral load post infection. We have also expanded our analysis of KRAB-ZNFs and showed that the basal expression of top candidates, i.e., ZNF519, ZNF566, and ZNF611, provided additional predictive value beyond previously known IFN response factors (see the detailed response to reviewer #2 below).

In conclusion, in its current state, this manuscript doesn't provide any evidence that transposable elements (as independent factors) play a role in the immune response to IAV or in determining viral load. Establishing the association alone is not a novel finding, as it is well known that significant transcriptional changes in different biological settings are associated with major changes in chromatin accessibility, including TEs, which constitute a significant fraction of the genomes. Additional work is needed to investigate the roles of TEs polymorphism and epivariation

in IAV IR. I also find some redundancy in the results and figures presented in this manuscript to show the association between TEs epigenetic changes and IAV IR. This data could be reduced into fewer figures to convey the same conclusion.

In this work we provide an in-depth characterization of the epigenetic landscape in TEs before and after infection. We believe that the association we found between TE epigenetic state preinfection and viral load post infection is novel. Moreover, through these analyses, we were able to identify new host factors (e.g., KRAB-ZNFs) likely associated with the response to infection. One of the strongest evidence supporting that TE-derived sequences impact the IAV IR is the proximity of many of the TE-loci we identified to dozens of IR genes (see the detailed response to point 1 above). Moreover, two of the sequences we identified to be variable between individuals were previously shown in a cell line to act as an enhancer for a gene known to be associated with IR (Srinivasachar Badarinarayan et al., 2020). We agree with the Reviewer that our study opens a number of new questions that will require further investigation as we also state in the discussion.

Minor Issues/Comments:

Figure 2C: What exactly is meant by epigenetic variability and variable regions in this context? It should be clarified briefly in the text and reference the methods for details. Also, the conclusion that "these results are consistent with some variability of TE transcription post-infection" is vague. Also, the difference in variability between TE and non-TE is evident in H3K4me3 peaks regardless of infection, how does this support a role for TEs in the modulation of infection response?

We have revised the sentence in the text to say:

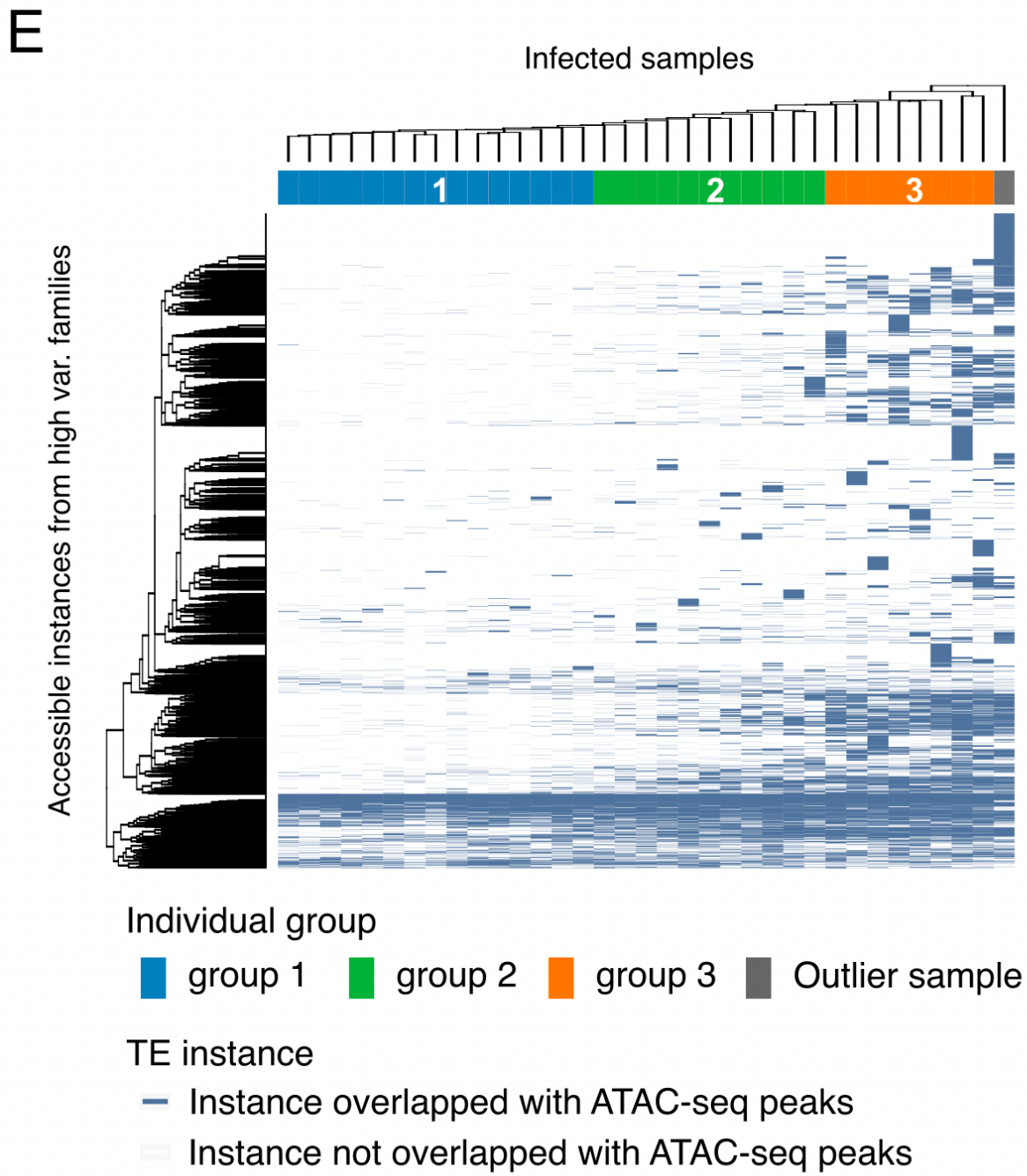
“To determine which regions were epigenetically variable between individuals, we measured the coefficients of variation (cv) in consensus peak regions (Aracena et al., 2022) and identified similar proportions of variable regions in TE and non-TE regions for most marks (0.4% to 6.4%, cv ≥ 0.5, Figure 2C, see Methods).”

You are also correct that the pattern observed in H3K4me3 is unique. We have rephrased the concluding sentence to:

“Given that H3K4me3 is typically associated with transcription, these results suggest variability of TE transcription before and after infection.”

Figure S3E: add a legend for color scale.

Thank you for the suggestion. We have revised Figure S3E and added a legend for the colors.



Page 9: "Among high var. families we consistently observed more commonly (≥ 25% individuals of one group) and rarely (< 25%) accessible instances that were specific to Group 3 individuals". Rephrase this sentence; it is difficult to understand.

We have rephrased the sentence and pushed the result to the supplements in the revised version as below:

"Among more commonly (≥ 25% individuals of one group) and rarely (< 25%) accessible instances from high variable families, we observed that they were often from Group 3 individuals (Figure S4A and Methods)."

Page10-11: "Instances from high var. families .... Overall, low var. and high var. Showed distinctive chromatin patterns post-infection". How do you interpret these findings? How does support the overall study conclusions?

We found that high variable families have higher DNA methylation, are histone repressed and have lower enrichment of active marks, as compared to low variable families (Figure 4A-B). It suggests different regulatory activity at 24-hr infection post infection. We have added the interpretation in the conclusive sentence:

"Overall, low variable and high variable showed distinctive chromatin patterns following infection suggesting different activation patterns and potential regulatory impact."

It supports our conclusions regarding the enrichment of KRAB-ZNFs found in high variable families.

Fig 5A: is THE1B Low var. Or High var.? It is mentioned as high in the text but low in the figure annotation.

Thank you for picking up this error. THE1B is a low variable family. We have corrected it in the text:

"For example, we can visualize the peak centroids identified along the consensus sequences for THE1B, a low variable family (Figure 5A), and LTR12C, a high variable family (Figure 5B)."

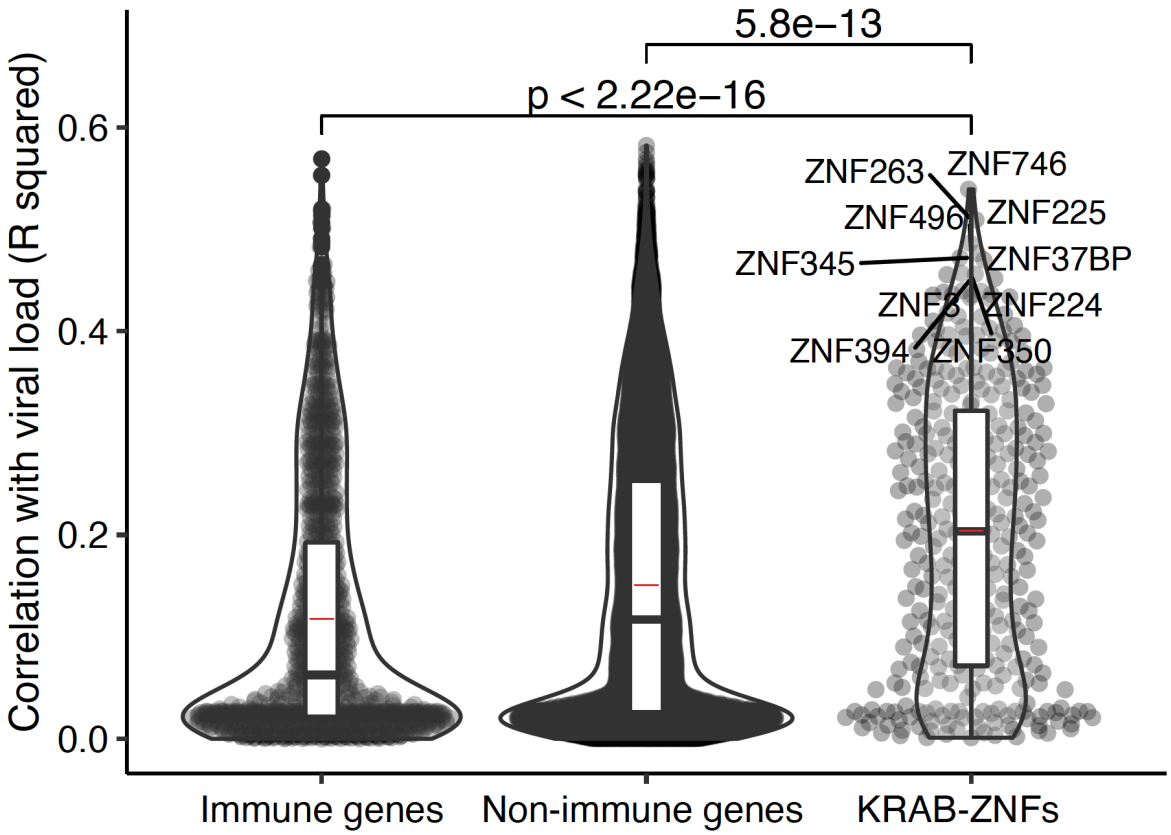
Reviewer #2
In this work, the authors examined epigenomic and transcriptomic data in influenza-infected primary macrophages from a panel of patients. They found the severity of infection, measured by transcriptomic viral load, to be correlated with certain patterns of transposon expression, accessibility, and epigenetic character. They found that many transposon families were upregulated upon influenza infection of the cells, and some were significantly correlated with viral load from patient to patient. Then, they found that some of these families were variable in their response to infection across patients. The authors posit that these transposon families contribute to the observed variability in viral load and severity of infection by acting as variably active cis-regulatory elements, contributing both immune and non-immune-associated transcription factor motifs. Overall this work is a rigorous and comprehensive analysis of an interesting dataset and provides new but largely correlative evidence supporting the idea that TEs could contribute to variable immune responses.

Thank you for this accurate summary.

Major Comments
1. The authors conduct some interesting analysis related to KZNF proteins and show there is some enrichment based on Imbeault data. Given that the ZNF repression is suggested as a major mechanism regulating variability these analyses could and should be further strengthened. For example, analyzing whether there is a correlation with the expression of specific KZNFs (all those in Supp fig 5G rather than just TRIM28/SETDB1).

Thank you for the suggestion. We have now performed a more in-depth analysis between the basal expression of human KRAB-ZNFs and viral load post-infection. Notably, we found that the correlation with KRAB-ZNFs was significantly stronger than with other genes, including both immune and non-immune genes (New Figure 6F). We have now added to the results:
"We then examined the basal expression levels of all KRAB-ZNFs and observed a significantly higher correlation with viral load compared to immune and non-immune related genes (Figure 6F and Table S7)."

F



New Figure 6 (F) Violin plot of the correlation coefficients between the basal expression (TPM) of KRAB-ZNFs and other genes with viral load post-infection. A list of human KRAB-ZNFs was obtained from Imbeault et al., 2017 and immune genes were obtained from the InnateDB database (Breuer et al., 2013). The top 10 most correlated KRAB-ZNFs are highlighted. Furthermore, we redid the enrichment analysis for all KRAB-ZNF binding sites in TEs using Imbeault data and observed the enrichment of various KRAB-ZNFs in multiple high variable families (New Figure S7A). We further tested whether the binding sites were located in the accessible regions upon IAV infection (New Figure S7B). As expected, the proportions of KRAB-ZNFs binding sites in the accessible regions were observed to be significantly higher in high variable families compared to low variable families. We have added the following sentences to the results:

"Supporting the potential role of KRAB-ZNFs in high variable families, we observed that the binding sites for KAP1 and multiple ZNF TFs (Imbeault et al., 2017) were enriched in some high

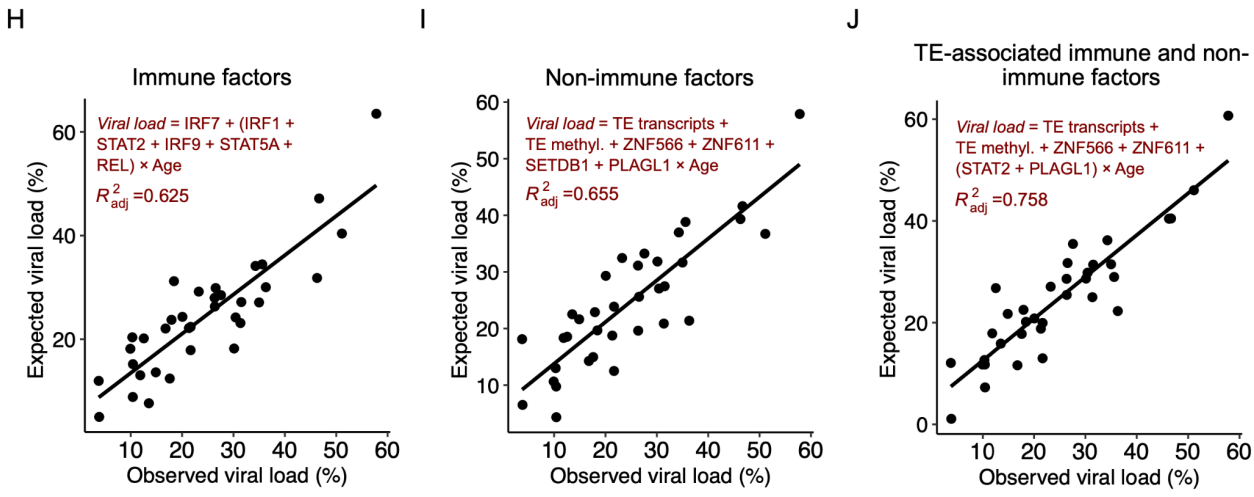
variable families (Figure S7A and Table S7); Moreover, the binding regions significantly overlapped the open chromatin regions in some high variable families post-infection (Figure S7B)."

Next, because the JASPAR motif database contains less than 30 KRAB-ZNF motifs, we expanded our motif analysis by using another source of KRAB-ZNF motifs (Barazandeh et al. 2018). We have added the following sentences to the results:

"Due to the limited number of KRAB-ZNF motifs in the JASPAR database, we used another source of KRAB-ZNF motifs (Barazandeh et al., 2018) to identify motifs across the accessible instances from enhanced families. We observed enrichment of KRAB-ZNF motifs in high variable families but not in low variable ones (Figure S7C and Table S7). KRAB-ZNFs are commonly found to interact with the KAP1/TRIM28 machinery to repress TEs through DNA and histone repression (Helleboid et al., 2019; Iyengar and Farnham, 2011), thus the enrichment of KRAB-ZNF binding sites and motifs in high variable families is also consistent with the high DNA and histone repression observed in these families (Figure 4B)."

Finally, we also correlated the expression of KRAB-ZNFs and the aggregated accessibility of high variable families post infection (New Table S7). After we integrated these results, we were able to identify top candidate host factors. We have revised the corresponding sentences and updated our predictive models in the new Figure 6H-6J and new Figure S9G, S9I-S9J as below:

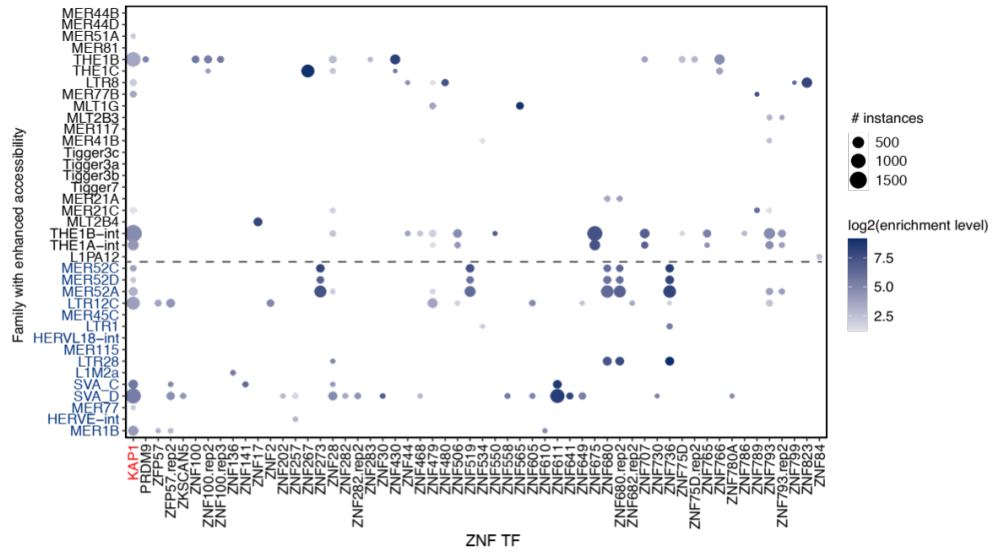
"We further found that the expressions of ten KRAB-ZNF genes were strongly correlated with the aggregated accessibility of high variable families post infection (Table S7, R squared ≥ 0.3, p value ≤ 0.05). After integrating these results, we identified PLAGL1 and three KRAB-ZNFs, i.e., ZNF519, ZNF566, and ZNF611 as top candidate host factors (Figure S9G and Table S7)."
"Notably, when we included six non-immune factors associated with TEs and age in our model, we obtained a slightly better fit with a model that includes TE transcripts and the new factors including ZNF566, ZNF611, and PLAGL1 (adjusted R2 = 0.655) (Figure 6I). Adding the top correlated immune TF, i.e., STAT2, further increased the accuracy of the model (adjusted R2 = 0.758) (Figure 6J)."



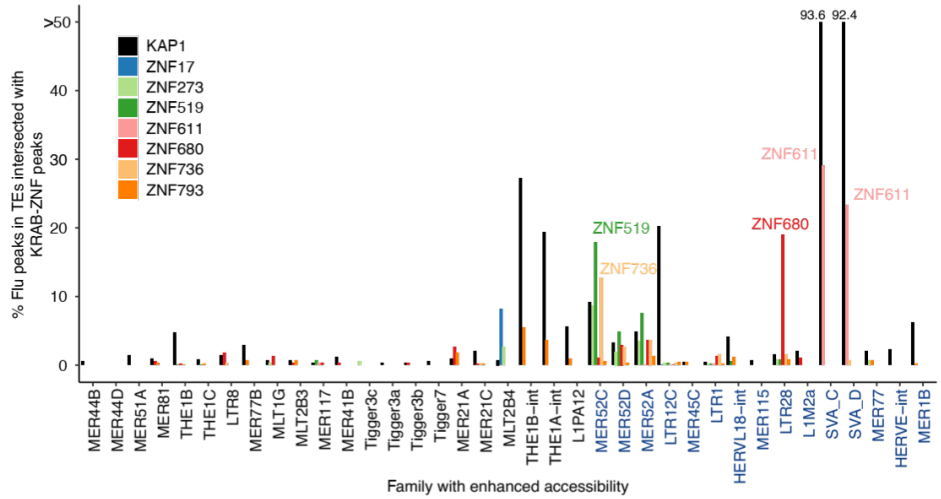
New Figure 6 (H) Multivariable regression model developed for the prediction of viral load using the expression levels of immune TFs in the basal state. The top six correlated TFs to viral load that are also associated with TEs were used. The model was generated as we described in the Methods. The formula and variables and adjusted R2 are shown. (I) Multivariable regression model developed for the predictive of viral load using the TE-associated non-immune (novel) host factors in the basal state. Using the same approach (see Methods), a subset of features were selected among the age and six non-immune factors, including SETDB1, TE transcripts, TE methylation, ZNF566, ZNF611, and PLAGL1. (J) Multivariable regression model developed for the predictive of viral load using the TE-associated immune and non-immune host factors in the basal state. We included all the non-immune factors as well as STAT2 to generate the model. STAT2 was selected based on the correlation to viral load.

Altogether, our new data support the role of KRAB-ZNFs in regulating these high variable families during infection.

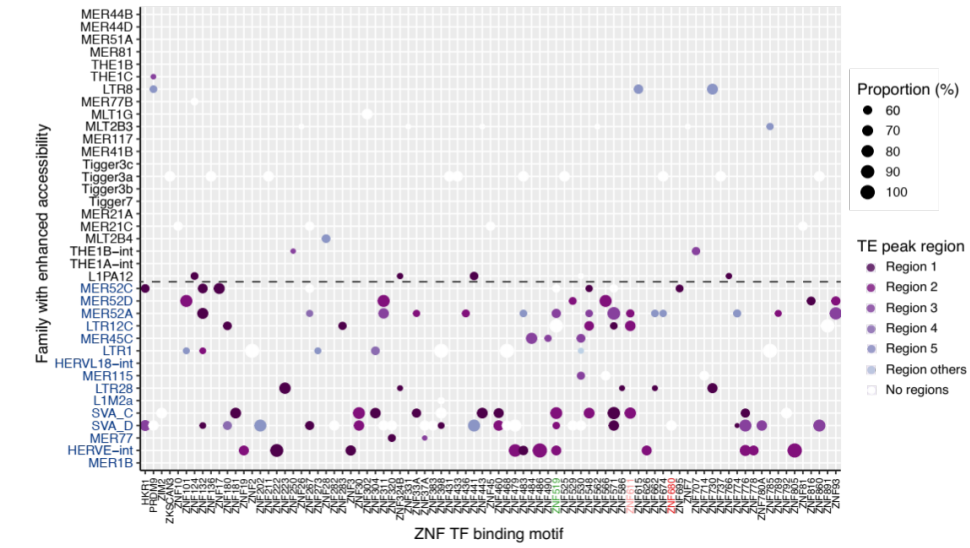
A


Family with enhanced accessibility
ZNF TF
instances
500
1000
1500
log2(enrichment level)
7.5
5.0
2.5

B


% Flu peaks in TEs intersected with KRAB-ZNF peaks
Family with enhanced accessibility
KAP1
ZNF17
ZNF273
ZNF519
ZNF611
ZNF680
ZNF736
ZNF793
93.6
92.4
ZNF611
ZNF611
ZNF680
ZNF519
ZNF736

C


Family with enhanced accessibility
ZNF TF binding motif
Proportion (%)
60
70
80
90
100
TE peak region
Region 1
Region 2
Region 3
Region 4
Region 5
Region others
No regions

New Figure S7. KAP1 and KRAB-ZNFs are associated with high variability in chromatin accessibility in high variable families. (A) Enrichment levels of KRAB-ZNF binding sites in high variable and low variable families in 257 HEK293T cell lines (Imbeault et al., 2017). High variable families MER52s, SVAs, LTR12C, and LTR28 are shown to be enriched for KRAB-ZNF binding sites. Color intensity refers to the fold enrichment relative to the random distribution (see Methods). (B) Proportion of KAP1 and KRAB-ZNF binding sites that overlap with accessible regions in TEs post-infection. A 100-bp of genomic regions centered at the ATACseq peak centroids was used for this analysis. KRAB-ZNFs with a minimum of 5% across enhanced families were visualized. (C) KRAB-ZNF binding motifs enriched in enhanced families. Motifs were obtained from Barazandeh et al. 2018. The same motifs enriched across TE peak regions were aggregated. TE peak regions with the most number of instances are shown as representatives. KRAB-ZNFs with their enrichment of binding sites in high variable families are highlighted.

2. The authors' conclusions are drawn exclusively from observed correlations between effects of infection. Empirical testing of these conclusions may be beyond the scope of this paper, but discussion of alternative explanations for these correlations is in order. For instance, the enrichment of inducibly-accessible TEs near inducibly-transcribed genes could be explained as two independent effects of regional chromatin changes, rather than causally related as the authors imply.

This is related to the first point of reviewer #1 and our response there. As you pointed out, what we observed is a correlation between basal TE transcription, accessibility of high variable TE families post-infection and viral load post-infection. We have added the following to the Discussion:

“Altogether, our data depict major epigenetic shifts in TEs in human macrophages upon infection -- opening mostly in LTR/ERVs and closing in LINEs. The proximity of these variable TE-loci to important immune genes suggest that they may contribute to the variable response to influenza infection, although further work will be needed to demonstrate a causal link between variation in TE activity and viral control.”

3. The sequencing data used in this study were collected from a single time point during infection and were apparently performed only once per individual. Possible variations over time are not observable but not discussed as a caveat.

We thank you for this excellent point. We have added the following sentence to the discussion: “Another aspect that would be interesting to dissect is whether the variation observed is consistent over time or a consequence of the fact that we looked at a specific time-point.”

4. Motif analyses to identify the transcription factors responsible for the observed correlations were performed based on transposon family consensus sequences. This is very useful but cannot identify motifs that were accrued in individual instances or subsets of instances during their expansion.

We apologize for not clearly describing our approach. Indeed, we first screened for motifs across all the accessible instances and then group them in different “TE peak regions” (subsets)

to identify shared motifs in each family. As you also pointed out in your minor comment below, we have revised the corresponding results section "High var. families contribute transcription factor…" to clearly describe how we performed the instance-level motif analysis.

"To further investigate the molecular mechanism underlying the enhanced families, we examined the TF binding motifs that were enriched in each TE peak region (Figure 5D and Figure S6C)."

Minor Comments
* On page 5, in the first paragraph of the Results section, the wording "suggesting varying capacity to infection and/or to limit viral replication across individuals" is unclear. Perhaps a word is missing after "varying," or perhaps "varying capacity for infection" was meant?

We thank you for picking up this mistake. We have corrected it as you suggested:

"Even though all samples engaged a strong transcriptional response to infection, we noticed extensive variation in the levels of viral reads (from 3.77% to 65.7%, Figure 1B), suggesting varying capacity for infection and/or to limit viral replication across individuals."

* Minor stylistic suggestion: there are many abbreviations used in the text that reduce readibility, for example the use of "var." instead of simply spelling out "variable" is a little confusing to read at times.

We thank you for this suggestion. We have replaced "low var." with "low variable" and "high var." with "high variable" throughout the text. We only use the "var." abbreviation in the figures due to the limited space.

* The section "High var. families contribute transcription factor.." (p11 end) is not written very clearly currently, could be made more readable.

We have revised the section as follows:

"To look for regulatory proteins associated with enhanced and reduced families, we aggregated the reads in open chromatin regions across samples to fine-map the actual peak summit on each TE instance, which was termed a "centroid". After the removal of instances with inaccurate or inconsistent annotations (Figure S6A), we re-mapped the reads from each TE instance to its TE family consensus sequence. For example, we can visualize the peak centroids identified along the consensus sequences for THE1B, a low variable family (Figure 5A), and LTR12C, a high variable family (Figure 5B). We observed a higher complexity of open chromatin regions for LTR12C compared to THE1B. Centroids were mainly detected at around 180 bp for THE1B and were scattered between 150 to 600 bp for LTR12C. Next, we defined a "TE peak region" as a location on the consensus sequence containing peak centroids from five or more instances, starting with the region with the largest number of instances, named Region 1, and so on. For most families, more than 80% of instances were accessible in one of the top 5 TE peak regions (Figure 5C, inset). The location of these TE peak regions can be shown on their consensus sequence and reveals that they are quite dispersed (Figure 5C). For example, 52% MER41B instances were accessible in Region 1 located around 380 bp, while another 18% and 11% of them were accessible in Region 2 (around 170 bp) and Region 3 (around 570 bp) separately.

Notably, compared to low variable families, high variable families had significantly more TE peak regions (student's t test, p value = 0.022) and lower proportions of accessible instances in the top TE peak region (student's t test, p value = 0.0037) (Figure S6B). This is consistent with the longer length of high variable families (Figure 3G)."

* In several instances phrases like "higher proportion" and "more likely" are used in the text, but the comparisons being made are not always easy to follow. Language could be clarified.

Thank you for pointing this out. We have gone through the text carefully and made a number of adjustments as the track changes.

* On page 17 in the first sentence there appears to be a typo. "Depression" should be "derepression?"

We have corrected this typographical error.

Reviewer #3:
In this correlative study, the authors employ RNA-seq data from monocyte-derived macrophages from 39 individuals, where cells were infected with influenza A virus for 24 hours. Post-infection, they measure the percentage of the transcriptome contributed by viral transcripts as an indicator of viral load and observe considerable variation between individuals. The infection induces upregulation of some TE families (mainly LTRs) either through direct or indirect affects following changes to the global transcriptome. In line with this, there is increased enrichment of active epigenetic marks H3K4me3 and H3K27ac at TEs in the flu samples (although H3K27ac did not reach significance). Analyses at the TE family level shows that THE1B, SVAC&D, MER52, MER41B and LTR12C and a few others gain enriched chromatin accessibility (ATAC-seq) in the Flu samples. Several of these (including THE1B) also gain H3K4me3 and H3K27ac in the Flu samples. Interestingly, some LINE-1 families become less accessible with reduced H3K27ac in the Flu samples. Some TE families show high variability of expression between individuals (including MER52, LTR12C). Several instances of these LTRs are positioned proximal to genes with roles in immunity, and have binding sites for STATs/NFkB transcription factors implying that they may function as enhancers. This suggests that differential expression of these TEs between individuals may impact on expression of interferon-regulated genes and viral load. The authors propose that 'high-variance' TEs are bound by KRAB-ZFP repressors, which are differentially expressed between individuals, potentially explaining the variable response to infection. It is interesting that high expression of certain TEs correlates with low viral load and the idea that TEs and KRAB-ZFPs contribute to the response to infection is topical. However, there is a missing link of whether expression of the highvariance TEs tracks with higher expression of interferon-stimulated genes globally and low viral load, and there is no direct evidence that these KRAB-ZFPs regulate activity of the highvariance TEs.

Thank you for your accurate summary and helpful comments.

Comments
1.There appears no evidence that the KRAB-ZFPs assessed (ZKSCAN5 and ZNF460) regulate the LTRs that vary in expression between individuals or expression of any interferon-stimulated genes (and is the proposed mechanism through DNA methylation of LTR enhancers)? Does

KO/knockdown of these KRAB-ZFPs influence the expression of the high variance TEs, the immune response and the viral load? The proposed model is a bit vague due to the data being a bit preliminary. What is meant by 'reduced TEs' in the model?

We did observe differences in methylation between high and low variable families, which is consistent with the role of KRAB-ZFPs in the high variable families. Since KO/knockdown of KRAB-ZFPs in these primary macrophage cells is challenging and would probably lead to many indirect effects, we have now performed a number of additional analyses including an in-depth exploration of KRAB-ZNF binding sites and motifs. These new results also support the role of KRAB-ZNFs in high variable families. In brief, we first observed that the correlation between KRAB-ZNFs and viral load is among the highest compared to other genes (New Figure 6F). We also observed the enrichment of KRAB-ZNF binding sites and motifs that are enriched in high variable families (New Figure S7). We have also toned down to highlight the roles of KRABZNFs in high variable families rather than specific ZNFs in the text and the Figure 7. See response to reviewer #1 and #2 for more details.

As for “reduced TEs”, they are defined as TE families with reduced accessibility post infection. We have now clarified this in the text:

“One of the advantages of comparing two conditions is that we could also look for TE families showing reduced accessibility upon infection. We identified 39 such “reduced families”.”
“In contrast, families with reduced accessibility, also called reduced families, are accessible and bound by a distinct set of known immune-related (IR) TFs, including MEF2s and SPIs.”

2.The higher the expression of TEs, the lower the viral load. Is this because the high expression of TEs correlates with high expression of interferon-stimulated genes (ISGs) they are potentially proximal to? ISGs function to limit viral replication and stimulate adaptive immunity. Several ISGs are mentioned but they are not interrogated systematically as a group of genes.

We thank you for this interesting question. We did observe the inverse correlation between the basal TE transcripts and viral load post-infection (Figure 6A-6B and Figure S9B); however, the TE expression changes were not correlated with viral load (Figure 6A, top). Before infection, most TEs remain repressed and expressed at a low level; similarly, the ISGs are also not activated yet. Thus, we hypothesized that the global TE transcripts but not the expression of specific TE instances may be involved in the basal stimulation of innate immunity through the antiviral sensors, such as TLR3, RIG-I (Gázquez-Gutiérrez et al. 2021; Hale 2022). Meanwhile, many TEs may express with nearby genes, but it is hard to validate the causal role of these expressed TEs to their correlated adjacent genes, which could also be because of the genomic regional co-expression effects (see reviewer #2 comment above). Moreover, at the epigenetic level, we did not observe individual differences in chromatin accessibility in TE families with enhanced accessibility before infection.

Instead, as also suggested by review #1, we focused on the TEs with chromatin changes to understand their roles during infection. Previously, we have observed enrichment in various immune-related pathways among adjacent differentially expressed genes to TE families with enhanced/reduced accessibility upon infection (Figure S5C). We now newly identified 420 significantly up-regulated genes that are potentially regulated by enhancers/promoters derived from high variable and low variable families (Table S6). More importantly, we identified 17

immune-related genes that are proximal to high variable families, e.g. antiviral GBP genes. See details in the response to reviewer #1.

3.In the abstract and elsewhere, the authors state that 'TEs contribute to the activation of innate immunity'. This should be clarified to reflect what is known, i.e. 'TE expression increases upon infection' or 'some TEs (MER41) act as enhancers for genes involved in innate immunity' since it is not known if a global increase in TE expression is necessary or contributes to the establishment of innate immunity against any pathogen.

Good point, we have revised the text as follows:

"Given that the regulation of transposable elements (TEs) contributes to the activation of innate immunity, we wanted to explore their potential role in this variability."

Later in the text we have also modified:

"Some polymorphic TEs were also found to be eQTLs for genes upon infection, such as TRIM25 (Groza et al. 2021), thus we speculate that polymorphic TEs may act as enhancers and further contribute to the variable response to infection."

4.The term 'known immune regulators' in the abstract is quite vague. It would be clearer to refer to specific transcription factors.

We have now clarified in the abstract:

"Motif analysis showed an association with known immune regulators (e.g., BATFs, FOSs/JUNs, IRFs, STATs, NFkBs, NFYs, and RELs) in stably enriched TE families and with other factors in variable families, including KRAB-ZNFs."

5.Page 3: 'Endogenous Retroviruses (ERVs), are derived from ancient retrovirus, suggesting a potential association with infection and Immunity'. This is a bit confusing. Do you mean, they may retain viral features (the ability to reverse transcribe for example) that are recognized by nucleic acid sensors, making them able to induce IFN responses? Please clarify.

We have revised the sentences as follows:

"Notably, a particular subclass of TEs, endogenous retroviruses (ERVs), are derived from ancient retroviruses and retain virus-like features that could stimulate the innate immunity, suggesting a potential association with infection and immunity (Buttler and Chuong, 2021; Kassiotis and Stoye, 2016; Srinivasachar Badarinarayan and Sauter, 2021)."

6.Page 3: 'Confirming this, an ERV family, MER41, was found to be co-opted as cisregulatory elements in the primate innate immune response'. It would be clearer to explain co-option of ERVs in terms of them already having intact promoters and enhancers, which can then be repurposed by the host to regulate host genes.

We have revised the sentence as follows:

"Confirming this, an ERV family, MER41, contains regulatory sequences that are repurposed by the host to regulate host genes in the primate innate immune response (Bogdan et al., 2020; Chuong et al., 2016)."

7.Page 3: 'derived from ancient retrovirus' should be 'derived from ancient retroviruses'.

We have corrected this typographical error in the text.

8.Page 4: 'Meanwhile, loss of SETDB1 or SUMO-modified TRIM28, which are associated with histone methylation and Kruppel-associated box domain (KRAB) zinc finger proteins (ZNFs), will lead to the significant derepression of TEs in the immune response (Cuellar et al., 2017; Schmidt et al., 2019). Together, these studies suggest that TEs play a prominent role in human innate immunity'. This is a bit confusing: the SETDB1 paper cited is a cancer paper, which does not inform us whether SETDB1 has a natural role in regulation of TEs in normal cells or upon infections. The second reference also doesn't appear to show that TEs play a prominent role in human innate immunity. Please tone down conclusions.

Thank you for your suggestion. We have toned down our conclusion and added the following review papers:

"Meanwhile, loss of SETDB1 or SUMO-modified TRIM28, which are associated with histone methylation and Kruppel-associated box domain (KRAB) zinc finger proteins (ZNFs), leads to the de-repression of TEs (Cuellar et al., 2017; Schmidt et al., 2019). Several studies have also suggested that upregulated TE transcripts may play a role in human innate immunity (Gázquez-Gutiérrez et al., 2021; Hale, 2022)."

9.Page 7: 'That being said, we observed higher variability of H3K4me3 and lower variability of H3K27me3 mark in TEs compared to non-TE regions, respectively'. It would be helpful to include the percentages here like for the previous sentence comparing TE and non-TE regions.
We have added the percentages to the corresponding sentence in the revised version:
"Compared to non-TE regions, we observed higher variability of H3K4me3 (an average of 7.3% for TE and 3.6% for non-TE regions) and lower variability of H3K27me3 mark (0.3% for TE and 1.3% for non-TE regions) in TEs, respectively."

10.Page 13: 'Notably, L1MA2, L1MA4, L1MA6, L1MA7, and L1MA8 were significantly enriched for MEF2 related motifs. MEF2 TFs are central developmental regulators (Potthoff and Olson, 2007), which are also required in the immune response that functions as an in vivo immunemetabolic switch' It would be helpful to explain this further and discuss why and how LINE-1 elements might be downregulated in the aftermath of a viral infection in the discussion. LINE-1 elements have been linked to inducing type I IFN responses and to being upregulated in disease settings (cancer, autoimmune diseases).

We have added the following to the Discussion:

"On the other hand, the observed epigenetic changes in the LINE families with reduced accessibility may not affect their transcription which were slightly upregulated post infection."

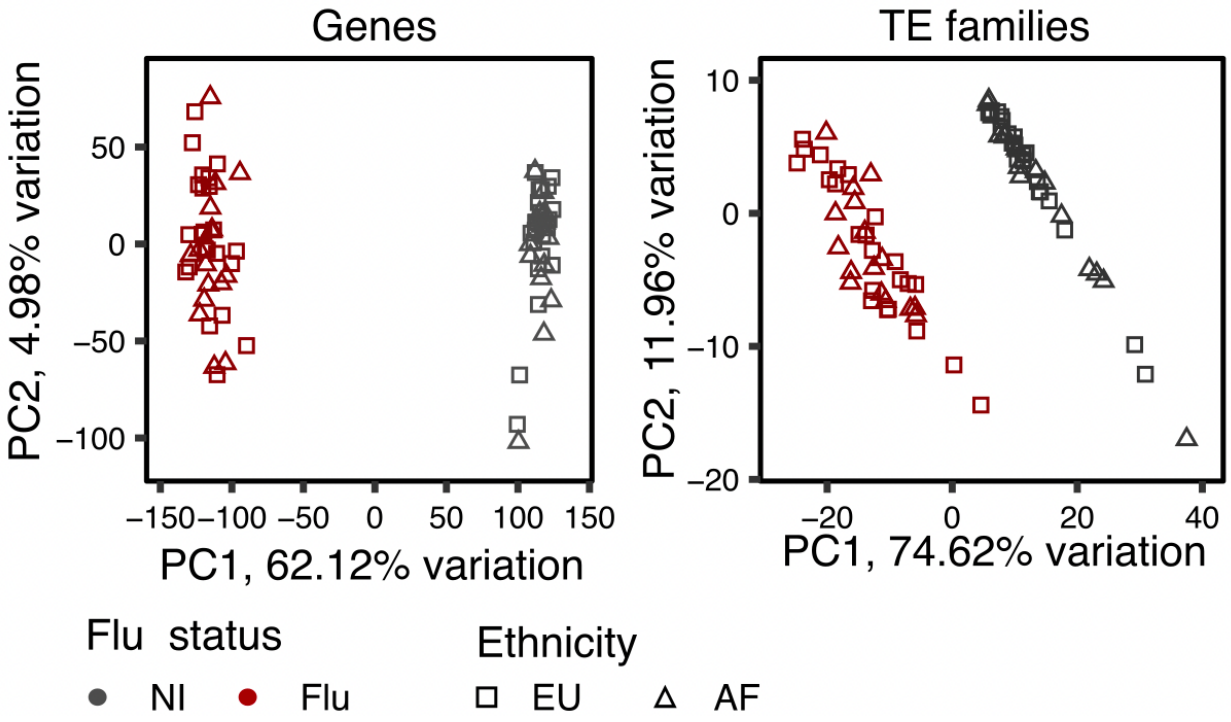
11.Page 17: 'In line with the involvement of TE transcripts in the activation of innate Immunity'. No references are cited here that relate to TE transcripts activating the innate immune response. There is a body of literature about inverted repeat Alu elements being self RNAs that are substrates for dsRNA sensing by MDA5. Some of those references would be appropriate here or other mechanistic studies. There is also a useful review on TEs and antiviral innate immunity: PMID: 33888553.

Thank you for the great suggestion. We have removed the original literature and added the suggested paper as well as another latest review as the new citations:

"In line with the involvement of TE transcripts in the activation of innate immunity (Gázquez-Gutiérrez et al., 2021; Hale, 2022)"

12.Figure 1a: The legend is a little confusing for the ethnicity data. The triangle and square could be unfilled rather than coloured grey since the colour changes depending on the infection status.

We appreciate your constructive comment. We have changed them to unfilled shapes as we shown below.



13.The figures were a bit big making them slow to download and view properly.

We have resized all the main figures to a smaller size.

REFERENCE

Aracena, Katherine A., Yen-Lung Lin, Kaixuan Luo, Alain Pacis, Saideep Gona, Zepeng Mu,

Vania Yotova, et al. 2022. “Epigenetic Variation Impacts Ancestry-Associated Differences in the Transcriptional Response to Influenza Infection.” bioRxiv. https://doi.org/10.1101/2022.05.10.491413.

Barazandeh, Marjan, Samuel A Lambert, Mihai Albu, and Timothy R Hughes. 2018. “Comparison of ChIP-Seq Data and a Reference Motif Set for Human KRAB C2H2 Zinc Finger Proteins.” G3 Genes|Genomes|Genetics 8 (1): 219–29. https://doi.org/10.1534/g3.117.300296.

Bogdan, Lucia, Luis Barreiro, and Guillaume Bourque. 2020. “Transposable Elements Have Contributed Human Regulatory Regions That Are Activated upon Bacterial Infection.” Philosophical Transactions of the Royal Society B: Biological Sciences 375 (1795): 20190332. https://doi.org/10.1098/rstb.2019.0332.

Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2016. “Regulatory Evolution of Innate Immunity through Co-Option of Endogenous Retroviruses.” Science 351 (6277): 1083–87. https://doi.org/10.1126/science.aad5497.

Gázquez-Gutiérrez, Ana, Jeroen Witteveldt, Sara R. Heras, and Sara Macias. 2021. “Sensing of Transposable Elements by the Antiviral Innate Immune System.” RNA 27 (7): 735–52. https://doi.org/10.1261/rna.078721.121.

Groza, Cristian, Xun Chen, Alain Pacis, Marie-Michelle Simon, Albena Pramatarova, Katherine A. Aracena, Tomi Pastinen, Luis B. Barreiro, and Guillaume Bourque. 2021. “Genome Graphs Detect Human Polymorphisms in Active Epigenomic States during Influenza Infection.” bioRxiv. https://doi.org/10.1101/2021.09.29.462206.

Hale, Benjamin G. 2022. “Antiviral Immunity Triggered by Infection-Induced Host Transposable Elements.” Current Opinion in Virology 52 (February): 211–16. https://doi.org/10.1016/j.coviro.2021.12.006.

Srinivasachar Badarinarayan, Smitha, Irina Shcherbakova, Simon Langer, Lennart Koepke, Andrea Preising, Dominik Hotter, Frank Kirchhoff, Konstantin M J Sparrer, Gunnar Schotta, and Daniel Sauter. 2020. “HIV-1 Infection Activates Endogenous Retroviral Promoters Regulating Antiviral Gene Expression.” Nucleic Acids Research 48 (19): 10890–908. https://doi.org/10.1093/nar/gkaa832.

Referees’ report, second round of review

Reviewer #1: The authors have made a number of changes that have improved the manuscript. They also carefully and thoughtfully wrote the manuscript so as not overstate their conclusions, acknowledging the (not directly addressed) major limitation of the study (as noted by all reviewers) that they are assessing correlations between TEs and immune response to influenza. But the paper still has value, and the predictive nature of their observations is quite interesting and worthy of publication.

Reviewer #2: The revised manuscript is significantly improved in clarity and satisfactorily addresses my comments as well as those of other reviewers. In particular, the new analysis of the ZNF proteins has yielded interesting results that further strengthen the paper.

Reviewer #3: The authors have answered my questions thank you.

Authors' response to the second round of review

No further changes were requested by the reviewers.