Cell Genomics Transparent Peer Review Record

Host Genetic Variants, Epstein-Barr Virus Subtypes and the Risk of Nasopharyngeal Carcinoma: An Assessment of Interaction and Mediation

Miao Xu[a,b,1], Ruimei Feng[c,1], Zhonghua Liu[d,1], Xiang Zhou[a,e,1], Yanhong Chen[a], Yulu Cao[a], Linda Valeri[d,f], Zilin Li[b,g], Zhiwei Liu[h], Su-Mei Cao[a], Qing Liu[a], Shang-Hang Xie[a], Ellen T. Chang[i], Wei-Hua Jia[a], Jincheng Shen[j], Youyuan Yao[k], Yong-Lin Cai[l], Yuming Zheng[l], Zhe Zhang[m], Guangwu Huang[m], Ingemar Ernberg[n], Minzhong Tang[m], Weimin Ye[o,p,2], Hans-Olov Adami[f,o,q,2], Yi-Xin Zeng[a,2], Xihong Lin[b,2]

Summary

Initial submission: Received : 3/30/2023

Scientific editor: Laura Zahn

First round of review: Number of reviewers: 2
Revision invited : 5/31/2023
Revision received : 9/29/2023

Second round of review: Number of reviewers: 2
Accepted : 12/6/2023

Data freely available: Yes

Code freely available: N/A

Cell Genomics Transparent Peer Review Record

Referees' reports, first round of review

Reviewer #1: Several polymorphisms in EBV have been proposed to contribute to the high incidence of NPC in S China but the epidemiological evidence supporting the 163364 SNP in BALF2 is probably the strongest and was published in ref 5 in this manuscript. This SNP is predicted to cause a Val to Met amino acid change in the EBV BALF2 protein. There is also a well-established association of certain HLA alleles with NPC disease in China.

This paper combines the authors' previous studies with new data and analysis to distinguish whether the high-risk HLA alleles simply favour increased frequency of EBV with the 163364 SNP in the population or, alternatively, the 163364 SNP and HLA alleles jointly contribute to the NPC risk. The conclusion is that they contribute jointly to NPC. This is interpreted to mean that immune surveillance of EBV involving this part of BALF2 is a determinant of NPC and that immunisation against this EBV strain might be able to prevent NPC. The genetic analysis is good and is presented clearly.

The model and the paper would be much stronger if there were any data indicating that the relevant 163364 SNP region of BALF2 actually affects recognition by the immune response. Extensive mapping of EBV T cell epitopes by other groups (not using these exact same HLA alleles) has found T cell clones that react with BALF2 but they do not map to this location.

Reviewer #2: It is well konwn that nasopharyngeal carcinoma is characterised by distinct geographical distribution and is particularly prevalent in east and southeast Asia. Early diagnosis and treatment is the key to improve overall survial. Although early studies on the genome variation and typing of Epstein-Barr virus found that the distribution of Epstein-Barr virus strains did have geographical characteristics, how to identify high-risk groups of nasopharyngeal cancer remains unclear. This large-scale genomic study conducted by the authors reveals significant additive interactions on NPC risk between the EBV subtype 277 classified by variant 163364 and host SNPs rs2860580 at HLA-A and rs2894207 at HLA-B/C loci, which may provide guidance on risk prediction chips and preventive vaccines in the future. In general, this paper is well organized and informative. However, it might be more comprehensive and beneficial to further analysize the impact of EBV variant 163364 on protein functional domain and protein three-dimensional structure if possible, which could be helpful for future development of preventive vaccines and further interpertation of underlying mechanism as to interaction between high-risk EBV subtype and host HLA SNPs rs2860580 and rs2894207.

Authors' response to the first round of review

We would like to thank the editor and reviewers for thoroughly reviewing our manuscript and for providing insightful and constructive comments and suggestions. Your suggestions have significantly helped improve our paper. We have conducted the suggested analyses and made substantial revisions to the manuscript. Herein, we present point-by-point responses and the corresponding revisions. The changes in the revised manuscript are marked in blue.

Responses to Reviewer 1's Comments:

Several polymorphisms in EBV have been proposed to contribute to the high incidence of NPC in S China but the epidemiological evidence supporting the 163364 SNP in BALF2 is probably the strongest and was published in ref 5 in this manuscript. This SNP is predicted to cause a Val to Met amino acid change in the EBV BALF2 protein. There is also a wellestablished association of certain HLA alleles with NPC disease in China.

This paper combines the authors' previous studies with new data and analysis toto distinguish whether the high-risk HLA alleles simply favour increased frequency of EBV with the 163364 SNP in the population or, alternatively, the 163364 SNP and HLA alleles jointly contribute to the NPC risk. The conclusion is that they contribute jointly to NPC. This is interpreted to mean that immune surveillance of EBV involving this part of BALF2 is a determinant of NPC and that immunisation against this EBV strain might be able to prevent NPC. The genetic analysis is good and is presented clearly.

Response: We greatly appreciate the reviewer's insightful summary of our findings that the high-risk EBV variant and human HLA variants jointly contribute to the increased risk of NPC. Given linkage disequilibrium (LD) patterns within the EBV genome, the mechanistic interaction could occur between HLA alleles and the high-risk EBV BALF2 variant 163364 or its correlated variants. In this context, this interplay suggests that HLA-mediated immune surveillance of EBV involving BALF2 variant 163364 or its correlated variants is a determinant of NPC risk.

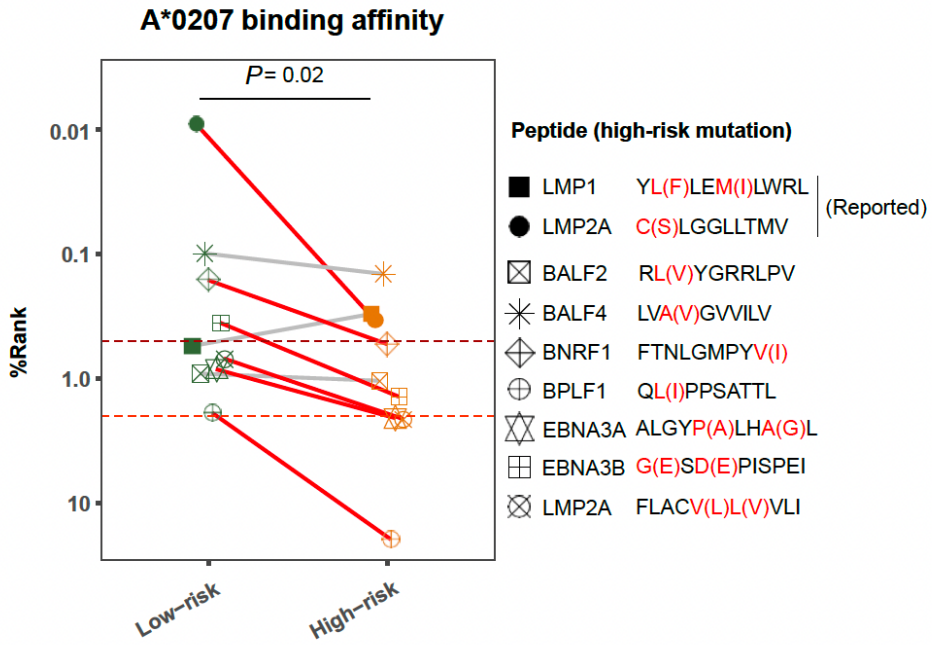
The model and the paper would be much stronger if there were any data indicating that the relevant 163364 SNP region of BALF2 actually affects recognition by the immune response. Extensive mapping of EBV T cell epitopes by other groups (not using these exact same HLA alleles) has found T cell clones that react with BALF2 but they do not map to this location.

Response: Thank you for the reference to EBV T cell epitope mapping studies, which indicate that EBV BALF2 is a prominent target for host T cell responses against EBV early antigens, eliciting T cell responses among hosts carrying HLA-B*27 and B*071-3. We agree with you that it is important to evaluate the immune recognition of BALF2 variant 163364 and its relevant variants. In response to this comment, we have conducted a comprehensive analysis of HLA recognition for the peptides encoded by high-risk EBV variants, including BALF2 variants and variants in moderate to high LD with BALF2 variant 163364.

Our findings indicate that a substantial additional risk is associated with the concurrent presence of the high-risk EBV subtype and susceptible HLA alleles. The additional risk associated may be related to distinct HLA-mediated T cell immune responses to different EBV subtypes, suggesting an elevated potential for immune evasion by the high-risk EBV subtype from susceptible HLA alleles. Among the HLA alleles associated with NPC in southern Chinese populations, A*0207 is the most significantly associated risk allele. Therefore, we evaluated HLA-A*0207 binding affinity for the peptide pairs encoded by highrisk and low-risk EBV strains using NetMHCpan-4.14. Our analysis focused on the nonamer peptides encompassing BALF2 variant 163364, 24 non-synonymous variants associated with NPC, which were in moderate to high LD with 163364 ($R2 > 0.25$), and three nonsynonymous variants within two previously reported NPC-associated A*02 epitopes. Peptide pairs from high-risk and low-risk EBV strains were retained if either member could be assigned to A*0207 with a ranking percentile < 2%. Comprehensive MHC-I benchmarking suggests that this threshold captured approximately 90% of the epitopes that elicit T cell response in vivo (Methods, Lines 576-588).

This prediction identified nine peptide pairs as candidate A*0207 epitopes, encompassing 13 variants, including the two reported A*02 epitopes (Supplementary Table 4). Interestingly, mutations in high-risk EBV strains had a significant trend of reduced epitope binding affinity with the high-risk allele HLA-A*0207 (Supplementary Figure 6). Among the nine NPCassociated peptides, six (red lines) had variants showing substantial binding affinity decreases, possibly representing candidate loci for evading A*0207 recognition. The BALF2 mutation V700L, encoded by high-risk EBV variant 162476, caused a more than two-fold decrease in binding affinity with HLA-A*0207. Specifically, in the LMP1 protein, mutations L126F and M129I encoded by the high-risk EBV subtype entirely abolished A*02 binding with the mutated peptide YFLEILWRL, compared to the low-risk epitope YLLEMLWRL. This loss of binding impairs its capacity to stimulate functional T cell responses, as demonstrated in previous studies5,6. These findings suggest that the high-risk EBV mutations may evade immune recognition by the risk HLA allele A*0207, providing insights into the

potential interaction mechanisms underlying the synergistic effect of high-risk EBV subtype and HLA genes on NPC risk. A comprehensive study on the screening of high-risk EBVspecific epitopes and the assessment of functional T cell responses in the NPC-endemic southern Chinese population is warranted.



Supplementary Figure 6. HLA-A*0207 binding affinity with the EBV peptides of NPC-low-risk and high-risk subtypes. The 9-mer peptides are indicated on the right, and mutations in the high-risk EBV subtype are highlighted in red. The LMP1 and LMP2A peptides have been verified with functional T cell response assays in previous studies, indicating that the mutant LMP2A peptide failed to elicit T cell responses in patients with NPC. The affinity is shown as the binding ranking percentile predicted with NetMHCpan-4.1. The dark red dashed line represents a ranking percentile of 0.5%, indicative of strong binding affinity. The red dashed line represents a ranking percentile of 2%, indicative of weak binding affinity.

We have summarized the analyses and the findings described above (Lines 302-333, Supplementary Figure 6, and Supplementary Table 4) and incorporated the references you mentioned (Lines 307-308) in a new section "Potential mechanistic interaction between EBV subtypes and HLA alleles". Following your suggestion, we have expanded the Discussion Section by including the published evidence on EBV epitope changes against immune surveillance in the southern Chinese population and NPC patients. We have also added relevant discussions and highlighted the importance of functional T cell epitope mapping (Lines 354-375) as follows:

"……The strong interaction highlights that the susceptible HLA alleles increase the risk associated with high-risk EBV infection for NPC, supporting potential immune evasion by the high-risk EBV subtype from susceptible HLA alleles. The association between HLA genes and NPC risk has been confirmed consistently in both candidate-gene studies and several independent GWASs, highlighting HLA-A*1101 and A*0207 as the most significantly associated genes in these investigations. HLA-A*1101 is the protective allele, while A*0207 is the risk allele. The EBV EBNA-3B epitope IVTDFSVIK, restricted to HLA-A*11, has a high-frequency mutation (IVTDFSVIKN) among southern Chinese with a high A*11 frequency. These mutations are thought to provide selective advantage in the highly A*11-positive populations. Our findings, revealing an additional risk associated with the cooccurrence

of NPC-high-risk EBV and HLA-A*0207, suggest that the high-risk EBV may carry the sequence variations correlated with reduced binding affinity for A*0207, potentially contributing to an elevated NPC risk among individuals carrying the A*0207 allele. A recent study has also reported the trend of decreased HLA-A*02 binding affinity with peptides harboring NPC-high-risk mutations. Specifically, the LMP-1 YFLEILWRL mutant peptide, which shows association with NPC risk, has been reported to evade recognition by A*02-restricted epitope(YLLEMLWRL)-specific T cells and fail to elicit T cell responses in patients with NPC. Since EBV LMP-1 protein is among the few latent antigens expressed in NPC cells, it is plausible that NPC cells infected with the high-risk EBV subtype possess an enhanced ability to evade HLA-A*0207-mediated T cell immune surveillance. This scenario could further increase the NPC risk among individuals carrying the susceptible HLA allele and the high-risk EBV subtype. Extensive mapping of T cell epitopes of the high-risk EBV subtype is important for designing EBV vaccines and T cell therapies targeting NPC."

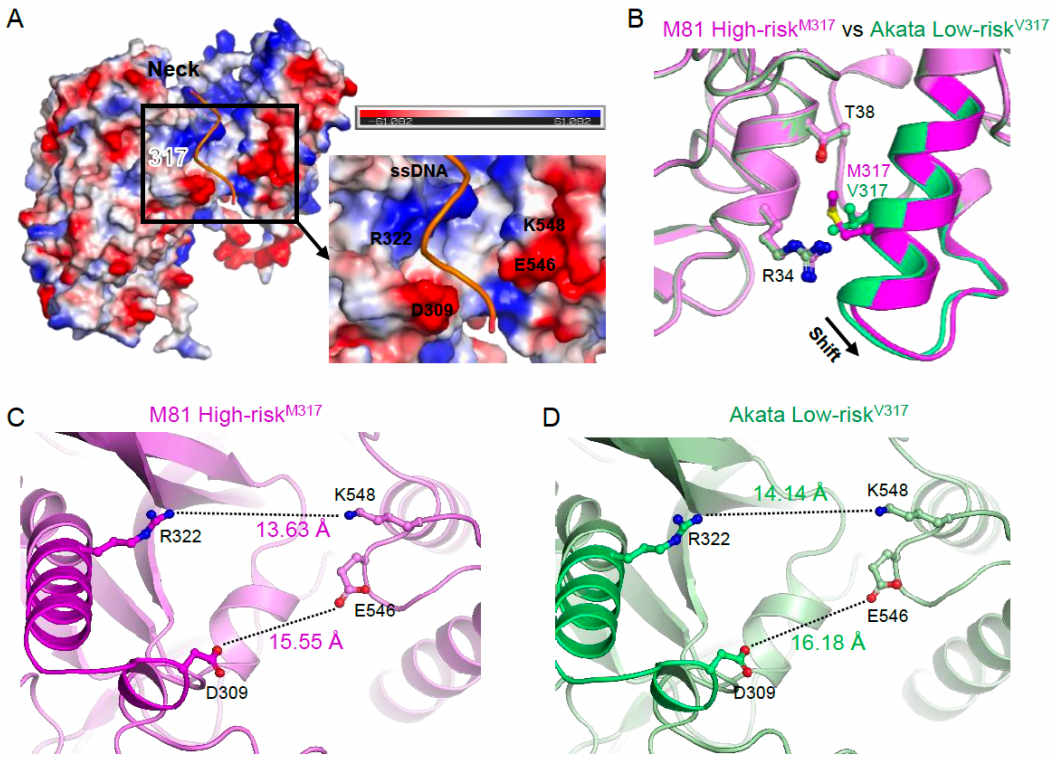
Please also see our added discussions on the amino acid changes on BALF2 protein function and its effect on viral replication in response to Reviewer 2's comments. Below please see our responses to Reviewer 2's comments.

Responses to Reviewer 2's Comments:

It is well konwn that nasopharyngeal carcinoma is characterised by distinct geographical distribution and is particularly prevalent in east and southeast Asia. Early diagnosis and treatment is the key to improve overall survival. Although early studies on the genome variation and typing of Epstein-Barr virus found that the distribution of Epstein-Barr virus strains did have geographical characteristics, how to identify high-risk groups of nasopharyngeal cancer remains unclear. This large-scale genomic study conducted by the authors reveals significant additive interactions on NPC risk between the EBV subtype 277 classified by variant 163364 and host SNPs rs2860580 at HLA-A and rs2894207 at HLA-B/C loci, which may provide guidance on risk prediction chips and preventive vaccines in the future. In general, this paper is well organized and informative. However, it might be more comprehensive and beneficial to further analyze the impact of EBV variant 163364 on protein functional domain and protein three-dimensional structure if possible, which could be helpful for future development of preventive vaccines and further interpretation of underlying mechanism as to interaction between high-risk EBV subtype and host HLA SNPs rs2860580 and rs2894207.

Response: Thank you for the positive and constructive comments. We are grateful for your valuable suggestion regarding the analysis of amino acid changes on BALF2 protein function and its effect on viral replication. Following your suggestion, we utilized AlphaFold2 to predict the three-dimensional (3D) structure of both the B95-8 wildtype BALF2 protein and the BALF2 variant carrying the high-risk SNP163364 (317 V>M mutation). The conformational prediction was based on the crystal structure of the single-stranded DNA (ssDNA)-binding protein ICP8 from Herpes Simplex Virus-1 (PDB code: 1URJ), a homologous protein of BALF2, resolved at 3.0 Å resolution7. Given ~25% of amino acid sequence similarity between BALF2 and ICP8, the predicted BALF2 structure achieved high confidence using AlphaFold2.

Interestingly, we observed that BALF2 amino acid 317, encoded by EBV SNP 163364, is located in the neck region, a presumed ssDNA binding pocket (Supplementary Figure 5A). Within this pocket, two positively-charged amino acids, D309 and E546 on the opposite side, form a gate structure, facilitating the movement of ssDNA through the binding pocket. Meanwhile, two negatively charged amino acids, R322 and K548, dock the ssDNA. With the presence of V317M mutation, encoded by the high-risk variant 163364, the long side-chain of M drives a shift of α-helix (314-328) harboring M317 and its neighboring loop (308-313), narrowing the gate between D309 and E546, as well as the docking interface between R322 and K548 (Supplementary Figure 5B-D). This narrowed gate and binding pocket due to the M317 high-risk variant may potentially affect DNA movement and viral DNA replication during the lytic cycle.



Supplementary Figure 5. Predicted structures of BALF2 protein from high-risk M81 EBV and low-risk Akata EBV. (A) Predicted protein conformation of BALF2 in complex with single-stranded DNA (ssDNA, orange). The amino acid 317, encoded by the high-risk variant 163364, and the key amino acids interacting with ssDNA are indicated. (B) The V317M mutation induces an alpha-helix shift. Magenta and green indicate regional structures of BALF2 protein from high-risk M81 EBV and low-risk Akata EBV, respectively. Other two amino acids (R34 and T38) that retain their position, in contrast to V317M, are highlighted. (C-D) Spatial distances between amino acids interacting with ssDNA are indicated for high-risk M81 EBV (C) and low-risk Akata EBV (D) BALF2 proteins, respectively.

We fully agree that genetic variations within the high-risk EBV strains may participate in shaping the viral function, which would aid in the future development of vaccines. Indeed, the NPC-derived EBV strain, M81, has been shown to exhibit an epitheliotropism and a high level of spontaneous replication in B cells8. These unique properties of M81 are thought to be related to its epithelial oncogenic potential and consistent with the observed increased viral replication preceding NPC onset. Polymorphisms within NPC-high-risk EBV subtype, particularly in the trans-activator protein BZLF1 and its promoter region, the non-coding RNA EBER2, as well as the gene structure of BALF5 have been reported to contribute to these properties8-13. Given the role of BALF2 as the ssDNA binding protein, an essential component of EBV DNA replication complex, the V317M mutation (variant 163364) in the BALF2 protein of high-risk EBV strains, could potentially influence the conformation of the ssDNA interaction surface, thereby altering its function during viral DNA replication. Functional analysis of these high-risk EBV variants would be indispensable to elucidate whether they might contribute the enhanced oncogenicity. Taken together, the distinct viral functional properties between NPC-high-risk and low-risk EBV, coupled with their interplay with HLA genetic factors, suggest that vaccine design aimed at NPC prevention should take into account the genetic variations within the high-risk EBV subtype.

We have added these results (Lines 283-300 and Supplementary Figure 5) and methods (Lines 565-574) into the new result section "Potential mechanistic interaction between EBV subtypes and HLA alleles". We have also added the above discussions on the biological

impact of EBV genetic variation on viral functions (Lines 377-391) and highlighted the necessity for functional investigations into NPC-high-risk EBV variations (Lines 388-391). Following both your and the first reviewer's comments, we have further included the implication of our findings for vaccine development (Lines 374-375, 388-391, and 434-436) in the Discussion section of the revised manuscript.

Finally, we thank the editor and both reviewers for their valuable inputs, which have significantly improved our manuscript. By incorporating your suggested analyses and expanding the discussion, we have strengthened the scientific contribution of our research. This refinement helps enhance the understanding of the interplay between EBV subtypes and HLA alleles in NPC risk and the potential impact on vaccine strategies.

References :

1. Pudney, V.A., Leese, A.M., Rickinson, A.B., and Hislop, A.D. (2005). CD8+ immunodominance among Epstein-Barr virus lytic cycle antigens directly reflects the efficiency of antigen presentation in lytically infected cells. J Exp Med 201, 349-360. 10.1084/jem.20041542.
2. Turcanová, V., and Höllsberg, P. (2004). Sustained CD8+ T-cell immune response to a novel immunodominant HLA-B*0702-associated epitope derived from an Epstein-Barr virus helicaseprimase-associated protein. J Med Virol 72, 635-645. 10.1002/jmv.20023.
3. Forrest, C., Hislop, A.D., Rickinson, A.B., and Zuo, J. (2018). Proteome-wide analysis of CD8+ T cell responses to EBV reveals differences between primary and persistent infection. PLOS Pathogens 14, e1007110. 10.1371/journal.ppat.1007110.
4. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Research 48, W449-W454. 10.1093/nar/gkaa379.
5. Duraiswamy, J., Burrows, J.M., Bharadwaj, M., Burrows, S.R., Cooper, L., Pimtanothai, N., and Khanna, R. (2003). Ex vivo analysis of T-cell responses to Epstein-Barr virus-encoded oncogene latent membrane protein 1 reveals highly conserved epitope sequences in virus isolates from diverse geographic regions. J Virol 77, 7401-7410. 10.1128/jvi.77.13.7401-7410.2003.
6. Lin, J.C., Cherng, J.M., Lin, H.J., Tsang, C.W., Liu, Y.X., and Lee, S.P. (2004). Amino acid changes in functional domains of latent membrane protein 1 of Epstein-Barr virus in nasopharyngeal carcinoma of southern China and Taiwan: prevalence of an HLA A2-restricted 'epitope-loss variant'. J Gen Virol 85, 2023-2034. 10.1099/vir.0.19696-0.
7. Mapelli, M., Panjikar, S., and Tucker, P.A. (2005). The crystal structure of the herpes simplex virus 1 ssDNA-binding protein suggests the structural basis for flexible, cooperative singlestranded DNA binding. J Biol Chem 280, 2990-2997. 10.1074/jbc.M406780200.
8. Tsai, M.H., Raykova, A., Klinke, O., Bernhardt, K., Gärtner, K., Leung, C.S., Geletneky, K., Sertel, S., Münz, C., Feederle, R., and Delecluse, H.J. (2013). Spontaneous lytic replication and epitheliotropism define an Epstein-Barr virus strain found in carcinomas. Cell Rep 5, 458-470. 10.1016/j.celrep.2013.09.012.
9. Bristol, J.A., Djavadian, R., Albright, E.R., Coleman, C.B., Ohashi, M., Hayes, M., Romero-Masters, J.C., Barlow, E.A., Farrell, P.J., Rochford, R., et al. (2018). A cancer-associated Epstein-Barr virus BZLF1 promoter variant enhances lytic infection. PLoS Pathog 14, e1007179. 10.1371/journal.ppat.1007179.
10. Li, Z., Baccianti, F., Delecluse, S., Tsai, M.-H., Shumilov, A., Cheng, X., Ma, S., Hoffmann, I., Poirey, R., and Delecluse, H.-J. (2021). The Epstein–Barr virus noncoding RNA EBER2 transactivates the UCHL1 deubiquitinase to accelerate cell growth. Proceedings of the National Academy of Sciences 118, e2115508118. doi:10.1073/pnas.2115508118.
11. Li, Z., Tsai, M.-H., Shumilov, A., Baccianti, F., Tsao, S.W., Poirey, R., and Delecluse, H.-J. (2019). Epstein–Barr virus ncRNA from a nasopharyngeal carcinoma induces an inflammatory response that promotes virus production. Nature Microbiology 4, 2475-2486. 10.1038/s41564-019-0546-y.
12. Wang, Y., Ungerleider, N., Hoffman, B.A., Kara, M., Farrell, P.J., Flemington, E.K., Lee, N., and Tibbetts, S.A. (2022). A Polymorphism in the Epstein-Barr Virus EBER2 Noncoding RNA Drives In Vivo Expansion of Latently Infected B Cells. mBio 13, e0083622. 10.1128/mbio.00836-22.
13. Church, T.M., Verma, D., Thompson, J., and Swaminathan, S. (2018). Efficient Translation of Epstein-Barr Virus (EBV) DNA Polymerase Contributes to the Enhanced Lytic Replication

Phenotype of M81 EBV. J Virol 92. 10.1128/jvi.01794-17.

Referees' reports, second round of review

Reviewer #1: Authors have addressed the previous comments by adding a computer-based prediction of the BALF2 structure (alphafold, Fig S5) and NetMHCpan-4.1 predictions of HLA binding affinity for EBV peptides likely to contribute to the T cell response to EBV in NPC patients, comparing the sequences from the high-risk strain marked by the 163364 SNP. There is no difference in the predicted binding to BALF2 (Fig S6) but some other epitopes have a lower binding affinity in the high-risk strain.
I am slightly puzzled why the authors do not simply conclude that the 163364 SNP is a marker linked to the still unidentified important variation in the NPC high risk strain, but the relevant information is now there in the complicated description of the genetics that they present.

Reviewer #2: Thank you for asking me to review this study on "Host Genetic Variants, Epstein-Barr Virus Subtypes and the Risk of Nasopharyngeal Carcinoma: An Assessment of Interaction and Mediation". The major strengths of this large-scale genomic study conducted by the authors reveals significant additive interactions on NPC risk between the EBV subtype 277 classified by variant 163364 and host SNPs rs2860580 at HLA-A and rs2894207 at HLA-B/C loci, which may provide guidance on risk prediction chips and preventive vaccines in the future. In general, this paper is well organized and informative. The authors have made many useful improvements to the article, which are generally in line with the requirements and the answers to the reviewer's comments.

Authors' response to the second round of review

We thank the editor and reviewers for accepting our study in principle and providing useful comments and suggestions. We have revised our manuscript according to the reviewers' and editorial comments. Herein, we present point-by-point responses and the corresponding revisions. The changes in the revised manuscript are marked in blue.

Reviewers' Comments:
Reviewer #1: Authors have addressed the previous comments by adding a computer-based prediction of the BALF2 structure (alphafold, Fig S5) and NetMHCpan-4.1 predictions of HLA binding affinity for EBV peptides likely to contribute to the T cell response to EBV in NPC patients, comparing the sequences from the high-risk strain marked by the 163364 SNP. There is no difference in the predicted binding to BALF2 (Fig S6) but some other epitopes have a lower binding affinity in the high-risk strain.

I am slightly puzzled why the authors do not simply conclude that the 163364 SNP is a marker linked to the still unidentified important variation in the NPC high risk strain, but the relevant information is now there in the complicated description of the genetics that they present.

Response: We highly appreciate your suggestion and fully agree with it. Therefore, we have taken your advice and restructured the results to make the conclusion clear and concise in the revised manuscript (Lines 292-304).

Reviewer #2: Thank you for asking me to review this study on "Host Genetic Variants, Epstein-Barr Virus Subtypes and the Risk of Nasopharyngeal Carcinoma: An Assessment of Interaction and Mediation". The major strengths of this large-scale genomic study conducted by the authors reveals significant additive interactions on NPC risk between the EBV subtype 277 classified by variant 163364 and host SNPs rs2860580 at HLA-A and rs2894207 at HLA-B/C loci, which may provide guidance on risk prediction chips and preventive vaccines in the future. In general, this paper is well organized and informative. The authors have made many useful improvements to the article, which are generally in line with the requirements and the answers to the reviewer's comments.

Response: We greatly appreciate your comments.
