The immunogenetics of viral antigen response is associated with sub-type specific glioma risk and survival

Geno Guerra, Ph.D.,^{1,*} Linda Kachuri, Ph.D.,² George Wendt, B.A.,¹ Helen M. Hansen, B.A.,¹ Steven J. Mack, Ph.D.,³ Annette M. Molinaro, Ph.D.,^{1,2} Terri Rice, M.P.H.,¹ Paige Bracci, Ph.D.,² John K. Wiencke, Ph.D.,^{1,2,4} Nori Kasahara, M.D, Ph.D.,^{1,7} Jeanette E Eckel-Passow, Ph.D.,⁵ Robert B. Jenkins, M.D., Ph.D.,⁶ Margaret Wrensch, Ph.D.,^{1,4} Stephen S. Francis, Ph.D.^{1,2,7,**}

1. Department of Neurological Surgery, University of California San Francisco, San Francisco, CA,

USA

2. Department of Epidemiology and Biostatistics, University of California San Francisco, San

Francisco, CA, USA

3. Department of Pediatrics, University of California, San Francisco, Oakland, California, USA

4. Institute of Human Genetics, University of California San Francisco, San Francisco, California

5. Department of Quantitative Health Sciences, Mayo Clinic, Rochester, Minnesota, USA

6. Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA

7. Weill Institute for Neurosciences, University of California San Francisco, San Francisco, USA

* Corresponding author: geno.guerra@ucsf.edu

** Corresponding author: stephen.francis@ucsf.edu

Abstract

Glioma is a highly fatal cancer with prognostically significant molecular subtypes and few known risk factors. Multiple studies have implicated infections in glioma susceptibility, but evidence remains inconsistent. Genetic variants in the human leukocyte antigen (HLA) region modulate host response to infection and have been linked to glioma risk. In this study we leveraged genetic predictors of antibody response to 10 viral antigens to investigate the relationship with glioma risk and survival.

Genetic reactivity scores (GRS) for each antigen were derived from genome-wide significant ($p < 5 \times 10^{-8}$) variants associated with immunoglobulin G antibody response in the UK Biobank cohort. We conducted parallel analyses of glioma risk and survival for each GRS and HLA alleles imputed at two-field resolution using data from 3418 glioma patients subtyped by somatic mutations and 8156 controls.

Genetic reactivity scores to Epstein-Barr virus (EBV) ZEBRA and EBNA antigens and Merkel cell polyomavirus (MCV) VP1 antigen were associated with glioma risk and survival (Bonferroni-corrected $p < 0.01$). GRS_{ZEBRA} and GRS_{MCV} were associated in opposite directions with risk of *IDH* wild type gliomas ($OR_{ZEBRA} = 0.91$, $p = 0.0099$ / $OR_{MCV} = 1.11$, $p = 0.0054$). GRS_{EBNA} was associated with both increased risk for *IDH* mutated gliomas ($OR = 1.09$, $p = 0.040$) and improved survival ($HR = 0.86$, $p = 0.010$). *HLA-DQA1*03:01* was significantly associated with decreased risk of glioma overall ($OR = 0.85$, $p = 3.96 \times 10^{-4}$) after multiple testing adjustment.

This first systematic investigation of the role of genetic determinants of viral antigen reactivity in glioma risk and survival provides insight into complex immunogenomic mechanisms of glioma pathogenesis. These results may inform applications of antiviral based therapies in glioma treatment.

44 **Keywords**

45 Glioma, human leukocyte antigen, Epstein-Barr virus, Merkel cell polyomavirus, polygenic risk

46 score

47

48

49

50

Introduction

Studies linking viruses and cancer date back over 100 years¹ and laid the foundation for understanding oncogenes². It has also become increasingly clear, as well evidenced by the current pandemic, that host genetics play an important role in response to viruses^{3,4}. To date, seven viruses have been accepted to be tumor-initiating in humans: Epstein-Barr virus (EBV), hepatitis B virus (HBV), human papillomavirus (HPV), human T-lymphotropic virus-1 (HTLV-1), hepatitis C (HCV), Kaposi's sarcoma herpesvirus (HHV-8) and Merkel cell polyomavirus (MCV)⁵. Recent analyses have shown that infections account for approximately 13% of human cancers worldwide⁶. However viruses have not been definitively implicated in the etiology of glioma despite decades of suggestive associations⁷⁻¹¹.

Glioma is a highly fatal brain cancer with a paucity of known risk factors, with exposure to ionizing radiation being the accepted causal factor¹². A history of previous infection with Varicella-Zoster virus (VZV) has been the only infectious agent consistently linked to adult glioma, conferring an estimated 20% decrease in risk^{10,13}. A suite of other viruses have been associated with the risk and grade of glioma, including EBV, MCV, John Cunningham virus (JCV), BK virus (BKV), human Cytomegalovirus (CMV) and human herpesvirus-6 (HHV-6), but with discordant results^{11,14-16}. The identification of key somatic molecular alterations (e.g. *IDH* mutation, 1p/19q chromosomal arm codeletion, *TERT* mutation), which drastically affect glioma prognosis, have uncovered subtype-specific risk factors^{17,18}. Recently, several studies/clinical trials have suggested a prognostic benefit of antiviral medications in the treatment of glioma¹⁹⁻²¹.

Genetic host response to viral infection could play a role in elucidating potential links between viruses and cancer incidence and prognosis. Studies have demonstrated significant heritable components (32-48%) of antibody response to many viruses and have identified genetic loci within host genes related to cell entry, cytokine production, and immune response²²⁻²⁴. Genetic variants of class I and II human leukocyte antigen (HLA) genes contribute the most important identified components of genetic determinants of response to viral antigens. Class II genes each encode half of a heterodimeric class II HLA protein, which presents extra-cellularly-derived peptides to CD4+ helper T-cells. The immune response is triggered when a CD4+ T-cell recognizes the combination of a class II HLA protein and its bound peptide. It is well studied that CD4+ T-cells have an important role in creating and sustaining effective anti-tumor immunity²⁵.

The HLA region of the genome is considered the most polymorphic region of the human genetic system²⁶ where polymorphisms have been shown to alter the risk and progression of disease in a variety of autoimmune (notably HLA class II) and malignant conditions^{27,28}. Certain HLA haplotypes have shown non-additive epistatic effects on glioma risk²⁹, yet no germline variants within the HLA have been directly identified as risk loci in a glioma GWAS.

In this study we leveraged previously published genome-wide SNP associations with viral antibody response³⁰ to generate genetically inferred antigen reactivity profiles in glioma cases and controls and evaluated their association with risk and survival by major glioma subtypes. We further conducted imputed HLA gene association analysis with glioma risk and survival and our findings suggest a convergence of genetic mechanisms regulating host immune response to viral challenge and glioma development and progression.

Methods

99

100 **Ethics**

101

102 Collection of patient samples and associated clinicopathological information was
103 undertaken with written informed consent and relevant ethical review board approval at the
104 respective study centers in accordance with the tenets of the Declaration of
105 Helsinki. Specifically informed consent and ethical board approval was obtained from the UCSF
106 Committee on Human Research (USA) and the Mayo Clinic Office for Human Research
107 Protection (USA). The diagnosis of glioma (ICDO-3 codes 9380-9480 or equivalent) was
108 established through histology in all cases in accordance with World Health Organization
109 guidelines.

110

111 **Study Populations**

112

113 We analyzed three glioma case-control sets assembled based on genotyping platform and
114 study population for a total sample size of 3418 cases and 8156 controls (**Figure 1, Table 1**). The
115 first set included 1973 cases from the Mayo Clinic and University of California, San Francisco
116 (UCSF) Adult Glioma Study and 1859 controls from the Glioma International Case-Control Study
117 (GICC) who were genotyped on the Illumina OncoArray, as previously described^{18,31-34}. The
118 second dataset included 659 cases and 586 controls from the UCSF Adult Glioma study
119 genotyped on the Illumina HumanHap370duo panel³². The third dataset included 786 glioma
120 cases from The Cancer Genome Atlas (TCGA) with available molecular data genotyped on the
121 Affymetrix 6.0 array. Cancer-free controls were assembled from two Wellcome Trust Case Control
122 Consortium (WTCCC) studies genotyped using the Affymetrix 6.0 array: 2,917 controls from the
123 1958 British Birth cohort and 2,794 controls from the UK Blood Service control group.

Molecular subtype information (*IDH* mutation and 1p/19q codeletion status) was downloaded from Ceccarelli et al.³⁵ supplementary table 1 for the third dataset and was provided directly from the UCSF AGS study and Mayo Clinic for the first and second datasets.

Quality Control and Imputation

Standard quality control procedures were implemented prior to imputation. Analyses were restricted to individuals of predominantly (>70%) European ancestry, determined using ADMIXTURE³⁶ and the HapMap 3 reference populations. Within each ancestral group, we removed samples with excess heterozygosity (>3 standard deviations (SD) from mean), <95% call rates, and discordant self-reported and genetically inferred sex. Relatedness checks were performed within each study using KING³⁷ (kinship >0.12), filtering out up to second-degree relations and retaining the samples with higher call rate. TCGA blood samples were preferentially chosen when both blood and tumor sequencing data were available for the same patient (693 European samples had both blood and tumor samples available). SNPs with <95% call rate were removed, along with variants deviating from Hardy-Weinberg equilibrium ($p < 10^{-6}$) or at a low minor allele frequency (MAF < 0.005). Samples genotyped on the same platform (i.e., Affymetrix 6.0 for TCGA and WTCCC) were imputed together using the multi-ethnic TOPMed reference panel (ver. r2).

Statistical Analysis

Genetically Predicted Viral Antigen Response

Genetic seroreactivity scores (GRS) were calculated to obtain genetically inferred viral antigen response profiles in each of the glioma datasets. For each GRS, candidate variants and corresponding effect sizes were obtained from genome-wide summary statistics for seroreactivity to 10 viral antigens previously identified in Kachuri and Francis et al. 2020³⁰ in 7895 randomly selected individuals of European descent from the UK Biobank (UKB) cohort using serological measures. For each antigen, we preferentially selected independent SNPs (linkage disequilibrium, LD, $r^2 < 0.01$ within 500kb) with the lowest p-value among genome-wide significant variants ($p < 5 \times 10^{-8}$). LD proxies ($r^2 > 0.9$) were obtained for variants unavailable in the target glioma datasets. For each individual, an antigen-specific GRS was calculated as a weighted sum with weights (β) corresponding to a standard deviation increase in antibody response:

$$GRS_{Antigen} = \beta_1 \times SNP_1 + \dots + \beta_k \times SNP_k$$

Each GRS was calculated using a minimum of 4 variants with $MAF \geq 0.01$ and imputation quality ($R^2 > 0.6$), and then standardized within each dataset.

Antigen GRS Associations with Glioma Susceptibility

GRS associations with disease risk were examined for glioma overall and for molecular subtypes defined by the specific tumor alterations: *IDH* mutation status and 1p/19q chromosomal arm codeletion. For each GRS, odds ratios (OR) were estimated using logistic regression models adjusted for age, sex, and the first 10 genetic ancestry principal components (PCs). Models in the UCSF-Mayo dataset were further adjusted for contributing site. GRS associations from each study were meta-analyzed using a fixed-effects inverse-variance weighted approach. Heterogeneity in study-specific GRS associations was assessed using Cochran's Q test.

Antigen GRS Associations with Glioma Survival

The association between each antigen-GRS and overall survival was assessed using a Cox proportional hazards regression model with follow-up time calculated from the date of first surgery to either date of death or last known contact. The latter censored at that date. Analyses were conducted for glioma overall and molecular subtypes with a minimum of 50 cases and 20 events (deaths). Proportionality assumptions were checked within each dataset via examination of Kaplan-Meier curves. Hazard ratios (HR) were estimated using Cox models adjusted for age, sex, 10 genetic PCs, and study site (if applicable). Associations with survival in each study were combined using fixed-effects meta-analysis. For each nominally significant ($p < 0.05$) glioma-GRS association, survival differences were further assessed using Kaplan-Meier curves by comparing mortality trajectories in patients with high genetically predicted immune reactivity (top 20%) to the remainder.

Regional HLA Analyses of glioma risk and survival

Classical HLA alleles were imputed for samples in all cohorts at two-field resolution using SNP2HLA³⁸ and the Type 1 Diabetes Genetics Consortium (T1DGC) reference panel of 2,767 unrelated individuals. Associations were tested for 77 alleles ($MAF \geq 0.01$) with imputation quality > 0.4 across eight genes: *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1*.

Subtype-specific associations with risk were estimated using logistic regression models. HLA allele associations with mortality were assessed using SPACox³⁹, an extension of the Cox model with improved type I error control in high-dimensional settings. Study-specific risk and

survival associations were combined in a meta-analysis. Associations for each HLA allele were considered statistically significant if $p < 6.5 \times 10^{-4}$ (0.05/77 alleles).

Results

The creation of a GRS was attempted for all antigens with genome-wide significant seroreactivity-associated variants ($p < 5 \times 10^{-8}$) as reported in Kachuri and Francis et al. 2020³⁰. This included 10 antigens: Four EBV antigens (EA-D, EBNA, p18, ZEBRA) and antigens for BKV, HHV7, HSV1, JCV, MCV, and VZV. We do not include antigens for CMV and HHV6 in our study as there was only one significantly associated SNP each in the previously published GWAS. An antigen-specific GRS was considered successfully created if at least four independent SNPs passed LD clumping and thresholding and were present (or via proxy SNP) in the three glioma study datasets. GRS were successfully generated for EBV EA-D, EBV EBNA, EBV p18, EBV ZEBRA, and MCV. Only one independent SNP remained for each of HSV1, BKV, JCV and VZV after accounting for long-range LD in the HLA region, although having numerous SNPs meet genome-wide significance in the previous seroreactivity study. Three SNPs remained for HHV7, each on a different chromosome (6, 11, and 17). Complete information of variants considered for all 12 viral antigens, including nearest gene and corresponding glioma risk associations are reported in **Supplement Tables S1 and S2**. LD correlations for all associated chromosome 6 variants are available in a heatmap in **Supplement Figure S1**.

The variants included in each of the five created GRSs were overwhelmingly located in the HLA region, with few predictors located elsewhere across the genome: rs67886110 in 3q25.1 an eQTL for *MED12L* and *P2RY12* (EBV EBNA), rs7618405 in 3p24.3 (MCV), and rs7444313 in 5q31.2 near *TMEM173* (MCV). Of the 38 SNPs included across the 5 GRSs, 2 had a significant

association (Bonferroni-corrected: $p < 1.3 \times 10^{-3}$) with overall glioma risk: rs9265517 near *HLA-B* (EBV EBNA), $p = 3.08 \times 10^{-4}$ and rs9268847 near *HLA-DRB9* (MCV), $p = 2.86 \times 10^{-4}$. **Supplement Table S2** visualizes the squared correlation between the GRS SNPs and each imputed HLA allele (two field resolution) from UCSF-Mayo cases and controls. Notably, the GRS for MCV was inversely correlated with GRSs for EBV ZEBRA (Pearson's $r = -0.28$, $p = 3.89 \times 10^{-70}$) and EBV EBNA ($r = -0.19$, $p = 6.28 \times 10^{-31}$) (**Figure 2**).

Viral antigen GRS associations with glioma risk

In the combined meta-analysis of three case-control studies, three GRS's reached at least nominal significance ($p < 0.05$), with some associations remaining statistically significant after correction for the number of antigens tested (Bonferroni: $p = 0.05/5 = 0.01$) (**Figure 3**). Genetic predisposition to an increased serological response to EBV ZEBRA was inversely associated with risk of glioma overall (per 1 SD increase in GRS: Odds Ratio, $OR_{ZEBRA} = 0.94$, 95% confidence interval 0.89-0.99, $p = 0.012$, 3418 cases). GRS_{EBNA} was associated with an increased risk of *IDH* mutated gliomas ($OR_{EBNA} = 1.09$, 1.004-1.18, $p = 0.040$, 1074 cases) and the magnitude of this effect increased for *IDH* mutated 1p/19q codeleted gliomas ($OR_{EBNA} = 1.14$, 1.012-1.28, $p = 0.031$, 396 cases).

We observed some evidence of antagonistic pleiotropy between genetic determinants of antibody response to EBV ZEBRA and MCV, which generalized across glioma subtypes. GRS_{ZEBRA} and GRS_{MCV} were associated with susceptibility to *IDH* wild type gliomas, but in opposite directions: Higher genetically predicted reactivity to EBV ZEBRA was inversely associated with *IDH* wild type glioma risk ($OR_{ZEBRA} = 0.91$, 0.85-0.98, $p = 0.0072$, 1479 cases), while increased predicted antibody response to MCV conferred an increased risk ($OR_{MCV} = 1.09$, 1.02-

1.17, $p=0.013$). This pattern persisted for *IDH* wild type 1p/19q non-codeleted gliomas ($OR_{ZEBRA}=0.91$, 0.84-0.9, $p=0.0099$; $OR_{MCV}=1.11$, 1.03-1.19, $p=0.0054$, 1280 cases). The correlation between GRS_{ZEBRA} and GRS_{MCV} (Pearson's $r=-0.28$, $p=3.9 \times 10^{-70}$ in UCSF-Mayo, **Figure 2**) suggests a possible shared underlying genetic mechanism between the response to the two antigens. GRS_{ZEBRA} and GRS_{MCV} were tested together, along with interaction term, in a single logistic model and we observed no significant interaction ($p=0.34$ in *IDH* wild type risk).

Viral antigen GRS-survival associations

The number of available cases and deaths across the three studies is available in **Table 2**. Associations between genetically predicted viral antigen response profiles and survival were restricted to *IDH* mutated gliomas (**Figure 4**). GRS_{EBNA} was associated with survival in 1074 *IDH* mutated glioma cases (per 1 SD increase: Hazard Ratio, $HR=0.86$, 0.76-0.96, $p=0.010$, 325 events), suggesting that a higher genetically predicted reactivity to EBV EBNA improved duration of survival.

GRS for two EBV antigens were nominally associated with survival amongst 244 *IDH* mutated 1p/19q codeleted glioma cases, but in opposite directions ($HR_{EBNA}=0.75$, 0.57-0.99 $p=0.048$ / $HR_{ZEBRA}=1.27$, 1.01-1.6, $p=0.042$, 64 events). This subtype-specific result was limited to the UCSF-Mayo dataset due to insufficient number of reported events (deaths) in *IDH* mutated 1p/19q codeleted glioma cases from TCGA and AGS-i370.

GRS_{EBNA} was also associated with improved survival outcomes in 614 *IDH* mutated 1p/19q non-codeleted glioma cases ($HR=0.75$, 0.74-0.997, $p=0.045$, 222 events), suggesting the

association of GRS_{EBNA} is independent of 1p/19q status. **Figure 5** visualizes the significant GRS_{EBNA} associations using Kaplan-Meier survival curves.

We did not detect any significant GRS associations in the analysis of glioma overall, which is consistent with the strong prognostic significance of molecular glioma subtypes^{17,33}. Also, as the prognosis in *IDH* wild type gliomas is the poorest, we suspect our GRS instruments are underpowered to detect significant deviations over such short intervals. Full survival results are available in **Supplement Figures S3 and S4**.

HLA allele-glioma associations

Of the 77 HLA alleles imputed at two-field resolution, only *HLA-DQA1*03:01* reached Bonferroni-corrected significance ($p < 6.5 \times 10^{-4}$ after correcting for 77 HLA alleles tested) for association with glioma risk in a meta-analysis. The presence of an *HLA-DQA1*03:01* allele was associated with a decrease in overall glioma risk ($OR = 0.85$, $p = 3.96 \times 10^{-4}$). This allele was also nominally associated with risk of *IDH* wild type ($OR = 0.82$, $p = 1.4 \times 10^{-3}$) and *IDH* wild type 1p/19q non-codeleted gliomas ($OR = 0.81$, $p = 2.6 \times 10^{-3}$). These subtype-specific associations and direction of effect mirror those of GRS_{ZEBRA} presented above, suggesting *HLA-DQA1*03:01* as a potential shared marker for both glioma risk and EBV ZEBRA seroreactivity. Results of all HLA allele (one and two field resolution) associations with glioma risk tested in a meta-analysis are available in **Supplement Table S2**.

Discussion

We investigated genetically predicted antibody response to ten antigens for seven viruses in relation to glioma risk and prognosis in a meta-analysis of three patient cohorts with molecular subtype information. Our analysis discovered evidence that genetic predisposition to reactivity to specific viral antigens associated with susceptibility to glioma in a subtype-specific manner. This pattern of effect modification by glioma subtype also extends into prognosis, where we observed associations between viral antigen GRSs and survival among specific glioma subtypes.

One of our main findings is that genetic predisposition to increased seroreactivity to EBV ZEBRA was associated with a decreased overall glioma risk, with a significant decrease in *IDH* wild type subtypes. Predicted reactivity to the MCV VP1 antigen mirrored the same *IDH* wild type associations but in the opposite direction, where higher reactivity was associated with increased glioma risk. The significant inverse relationship between predicted reactivity to EBV ZEBRA and MCV VP1 highlights the possibility of shared genetic components of antibody response to the two antigens. Understanding the underlying mechanisms guided by these genetics is left as an open question. A possible link is that the class II HLA allele *DQA1*03:01* was associated with decreased glioma risk in the same subtypes associated with GRSs for EBV ZEBRA. In our previous UKB analysis³⁰ the presence of *HLA-DQA1*03:01* was positively associated with EBV ZEBRA seroreactivity measurements ($\beta=0.168$, $p=1.3 \times 10^{-16}$). Taken together, the associations between *HLA-DQA1*03:01*, EBV ZEBRA and glioma risk suggest possible shared underlying immunogenetic architecture. As HLA class II proteins present potentially antigenic peptides, functional genetic alterations can result in changes to the binding affinity of specific antigens, modulating the potential for recognition by CD4+ helper T-cells. It is possible that the heterodimeric DQ proteins half-encoded by *DQA1*03:01* have improved binding or recognition of peptides presented by both EBV ZEBRA proteins and glioma (specifically *IDH* wild type), allowing for an increased immune response in both cases. As *IDH* wild type has generally been shown to serve as a marker for more severe glioma cases, the exact somatic tumor alterations leading to recognized peptide variants in these tumors is not clear and is an area that warrants future

investigation. Further analysis of viral-tumor protein homology is needed to understand if this could be a possible connection.

Interestingly, a higher genetically predicted response to EBV EBNA was nominally associated with increased risk in *IDH* mutated/1p/19q codeleted gliomas. Still, reactivity towards EBV EBNA was more strongly associated with improved survival in *IDH* mutated gliomas. This discrepancy between the disease-promoting and pro-survival associations of GRS_{EBNA} may suggest the presence of different biologic pathways after initiation of disease. It may also point towards EBV latency-mediated treatment effects. Previous studies show that the EBV latency type predicts the relative amount of induced reactivation generated by cytotoxic chemotherapy drugs⁴⁰. Therefore, individuals who react strongly to EBV EBNA antigens may exhibit a different pattern of reactivation when treated with temozolomide. Further research is required to elucidate this putative association, yet it is clear that EBV lytic/latent cycling is a critical aspect of other germline interactions and disease risk.

Epstein-Barr virus was the first recognized human oncovirus⁴¹ and exists in two distinct life cycles within a host: a lytic phase of active infection where new viruses are produced, and a latent phase where the virus remains dormant to avoid detection by the host. Distinct sets of antigens are produced during these two phases, two of such are EBNA, during the latent phase of infection, and the ZEBRA protein which initiates a change from latent to lytic gene expression. Past evidence suggests processes during the latent stage are responsible for the virus' oncogenic properties via mechanisms promoting cell growth and preventing cell apoptosis (reviewed in Akhtar et al. 2018⁴²). Recent work has also implicated the ZEBRA protein as oncogenic with evidence demonstrating its ability to deregulate immune surveillance and promote immune escape⁴³.

In contrast MCV is the most recently recognized human oncovirus, first described in 2008 and later accepted as a causal agent for Merkel cell carcinoma⁴⁴, a neuroendocrine carcinoma. Previous studies have identified the presence of MCV DNA in glioma patients and have drawn an association with infection and increased risk of glioma^{15,16}. It has been shown that both the large T and small T antigens of MCV are oncoproteins that target tumor suppressor proteins such as pRB. Although the exact latency mechanism in polyomaviruses is unknown, it has been suggested that complex protein-mediated latency may be critical to the MCV lifecycle⁴⁵.

In interpreting our findings, several limitations should be acknowledged. The shared genetic architecture between the viruses studied here, as seen in the correlations in **Figure 2 and Supplement Figure S1**, may result in a lack of specificity. This limitation in our study may be applicable in other viruses that share the same underlying genetic programming of antigen response, particularly VZV, which is associated with a unique pattern of LD spanning a large region of the HLA³⁰. Furthermore, the clumping and thresholding approach to GRS development may not be optimal for regions with complex LD structure, such as HLA. An approach that can appropriately account for the long-range correlation structure and capture non-linear interactions, such as haplotype effects, may improve future studies that use genetically inferred immune responses. Another challenge is the low SNP density in the HLA region in the present study, as well as reliance on imputed HLA alleles both in the glioma studies and in UKB. The HLA region is highly polymorphic and imputation of alleles from SNP data is limiting and prone to error. Similarly, it is difficult to disentangle the specific function of SNPs located in the HLA region. As such, additional studies leveraging HLA sequencing approaches are required to provide further granularity of the genetic determinants of antigen response and glioma risk/survival.

Additionally, we could not study the relationship between reactivity to several antigens of interest (BKV, HSV-1, JCV, VZV, HHV-7, HHV-6, CMV) due to the limited number of variants reaching genome-wide significance and/or meeting our GRS criteria³⁰. Although the GRS-glioma

association results presented here require further replication in independent patient populations, our findings are intriguing and suggest previous inconsistencies in observational epidemiologic may be partially due to individual differences in genetic factors that affect antigen reactivity.

This is the first study directly examining the underlying genetic architecture of antigen reactivity to common viruses and glioma risk and survival. We observed important associations between programmed reactivity to viruses and glioma etiology and prognosis. This methodology is not limited to the study of glioma, as the GRS instruments proposed here can readily be applied to cohorts of any cancer type. This offers a unique approach for future studies to reinvestigate the genetic contributions of long-running epidemiologic associations between viruses and cancer, and possibly clarify effects of viral therapies in their treatment.

Supplemental Information

Supplemental data include five tables and five figures.

Funding and Acknowledgments

Work at University of California, San Francisco was supported by the National Institutes of Health (grant numbers T32CA151022, R01CA52689, P50CA097257, R01CA126831, R01CA139020, R01AI128775, and R25CA112355), the National Brain Tumor Foundation, the Stanley D. Lewis and Virginia S. Lewis Endowed Chair in Brain Tumor Research, the Robert Magnin Newman Endowed Chair in Neuro-oncology, and by donations from families and friends of John Berardi, Helen Glaser, Elvera Olsen, Raymond E. Cooper, and William Martinusen.

This publication was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 RR024131. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

398

399 The collection of cancer incidence data used in this study was supported by the California
400 Department of Public Health pursuant to California Health and Safety Code Section 103885;
401 Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries,
402 under cooperative agreement 5NU58DP006344; the National Cancer Institute's Surveillance,
403 Epidemiology and End Results Program under contract HHSN261201800032I awarded to the
404 University of California, San Francisco, contract HHSN261201800015I awarded to the University
405 of Southern California, and contract HHSN261201800009I awarded to the Public Health Institute,
406 Cancer Registry of Greater California. The ideas and opinions expressed herein are those of the
407 author(s) and do not necessarily reflect the opinions of the State of California, Department of
408 Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention
409 or their Contractors and Subcontractors.

410

411 **Data Availability**

412 Genotype data of control samples from the 1958 British Birth Cohort and UK Blood Service
413 Control Group were made available from the Wellcome Trust Case Control Consortium (WTCCC)
414 and downloaded from the European Genotype Archive under ascension numbers
415 [EGAD00000000021](#) and [EGAD00000000023](#), respectively. Genotype data of glioma cases from
416 The Cancer Genome Atlas (TCGA) were obtained from Database of Genotypes and Phenotypes
417 (dbGaP) ([phs000178](#)). Genotype data of control samples from the Glioma International Case
418 Control Study (GICC) are available from dbGaP under accession [phs001319.v1.p1](#).

419 **Notes**

420 **Declaration of interests**

421 The authors declare no competing interests.

422 **Authorship**

423 GG, LK, and SSF conceived of the study and wrote main drafts of the manuscript. GG,
424 LK, GW, SJM, NK, and SSF conducted and advised on informatic and statistical analyses along
425 with result interpretations. HMH, AMM, TR, PB, JKW, JEE, RBJ, and MW were involved in primary
426 data collection. All authors contributed and reviewed the final manuscript.

427

References

1. Rous P. A Transmissible Avian Neoplasm. (Sarcoma of the common fowl.). *J Exp Med*. 1910;12(5):696-705. doi:10.1084/jem.12.5.696
2. Duesberg PH, Vogt PK. Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses. *Proc Natl Acad Sci U S A*. 1970;67(4):1673-1680. doi:10.1073/pnas.67.4.1673
3. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature*. Published online July 8, 2021. doi:10.1038/s41586-021-03767-x
4. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
5. Mui UN, Haley CT, Tying SK. Viral Oncology: Molecular Biology and Pathogenesis. *J Clin Med*. 2017;6(12):E111. doi:10.3390/jcm6120111
6. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health*. 2020;8(2):e180-e190. doi:10.1016/S2214-109X(19)30488-7
7. Neves AM, Thompson G, Carneiro J, et al. Detection and quantitative analysis of human herpesvirus in pilocytic astrocytoma. *Brain Res*. 2008;1221:108-114. doi:10.1016/j.brainres.2008.05.009
8. Chauvin C, Suh M, Remy C, Benabid AL. Failure to detect viral genomic sequences of three viruses (herpes simplex, simian virus 40 and adenovirus) in human and rat brain tumors. *Ital J Neurol Sci*. 1990;11(4):345-357. doi:10.1007/BF02335937

- 450 9. Cobbs CS, Harkins L, Samanta M, et al. Human cytomegalovirus infection and expression
451 in human malignant glioma. *Cancer Res.* 2002;62(12):3347-3350.
- 452 10. Wrensch M, Weinberg A, Wiencke J, et al. Does Prior Infection with Varicella-Zoster Virus
453 Influence Risk of Adult Glioma? *Am J Epidemiol.* 1997;145(7):594-597.
454 doi:10.1093/oxfordjournals.aje.a009155
- 455 11. Strojnik T, Duh D, Lah TT. Prevalence of Neurotropic Viruses in Malignant Glioma and Their
456 Onco-Modulatory Potential. *In Vivo.* 2017;31(2):221-230. doi:10.21873/invivo.11049
- 457 12. Bondy ML, Scheurer ME, Malmer B, et al. Brain tumor epidemiology: Consensus from the
458 Brain Tumor Epidemiology Consortium. *Cancer.* 2008;113(S7):1953-1968.
459 doi:10.1002/cncr.23741
- 460 13. Amirian ES, Scheurer ME, Zhou R, et al. History of chickenpox in glioma risk: a report from
461 the glioma international case-control study (GICC). *Cancer Med.* 2016;5(6):1352-1358.
462 doi:10.1002/cam4.682
- 463 14. Kofman A, Marcinkiewicz L, Dupart E, et al. The roles of viruses in brain tumor initiation and
464 oncomodulation. *J Neurooncol.* 2011;105(3):451-466. doi:10.1007/s11060-011-0658-6
- 465 15. Limam S, Missaoui N, Bdioui A, et al. Investigation of simian virus 40 (SV40) and human JC,
466 BK, MC, KI, and WU polyomaviruses in glioma. *J Neurovirol.* 2020;26(3):347-357.
467 doi:10.1007/s13365-020-00833-4
- 468 16. Egan KM, Kim Y, Bender N, et al. Prospective investigation of polyomavirus infection and
469 the risk of adult glioma. *Sci Rep.* 2021;11(1):9642. doi:10.1038/s41598-021-89133-3

- 470 17. Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR. Genetic and molecular epidemiology
471 of adult diffuse glioma. *Nat Rev Neurol*. 2019;15(7):405-417. doi:10.1038/s41582-019-0220-
472 2
- 473 18. Eckel-Passow JE, Drucker KL, Kollmeyer TM, et al. Adult diffuse glioma GWAS by molecular
474 subtype identifies variants in *D2HGDH* and *FAM20C*. *Neuro-Oncol*. 2020;22(11):1602-1613.
475 doi:10.1093/neuonc/noaa117
- 476 19. Volpin F, Casaos J, Sesen J, et al. Use of an anti-viral drug, Ribavirin, as an anti-
477 glioblastoma therapeutic. *Oncogene*. 2017;36(21):3037-3047. doi:10.1038/onc.2016.457
- 478 20. Stragliotto G, Pantalone MR, Rahbar A, Bartek J, Söderberg-Naucler C. Valganciclovir as
479 Add-on to Standard Therapy in Glioblastoma Patients. *Clin Cancer Res*. 2020;26(15):4031-
480 4039. doi:10.1158/1078-0432.CCR-20-0369
- 481 21. Piper K, Foster H, Gabel B, Nabors B, Cobbs C. Glioblastoma Mimicking Viral Encephalitis
482 Responds to Acyclovir: A Case Series and Literature Review. *Front Oncol*. 2019;9:8.
483 doi:10.3389/fonc.2019.00008
- 484 22. Besson C, Amiel C, Le-Pendeven C, et al. Strong Correlations of Anti-Viral Capsid Antigen
485 Antibody Levels in First-Degree Relatives from Families with Epstein-Barr Virus-Related
486 Lymphomas. *J Infect Dis*. 2009;199(8):1121-1127. doi:10.1086/597424
- 487 23. Kenney AD, Dowdle JA, Bozzacco L, et al. Human Genetic Determinants of Viral Diseases.
488 *Annu Rev Genet*. 2017;51(1):241-263. doi:10.1146/annurev-genet-120116-023425
- 489 24. Venkataraman T, Valencia C, Mangino M, et al. *Antiviral Antibody Epitope Selection Is a*
490 *Heritable Trait*. *Genetics*; 2021. doi:10.1101/2021.03.25.436790

- 491 25. Tay RE, Richardson EK, Toh HC. Revisiting the role of CD4+ T cells in cancer
492 immunotherapy—new insights into old paradigms. *Cancer Gene Ther.* 2021;28(1-2):5-17.
493 doi:10.1038/s41417-020-0183-x
- 494 26. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression,
495 interaction, diversity and disease. *J Hum Genet.* 2009;54(1):15-39. doi:10.1038/jhg.2008.5
- 496 27. Bateman AC, Howell WM. Human leukocyte antigens and cancer: is it in our genes? *J*
497 *Pathol.* 1999;188(3):231-236. doi:10.1002/(SICI)1096-9896(199907)188:3<231::AID-
498 PATH325>3.0.CO;2-A
- 499 28. Gebe JA, Swanson E, Kwok WW. HLA Class II peptide-binding and autoimmunity: Gebe et
500 al: HLA Class II peptide-binding and autoimmunity. *Tissue Antigens.* 2002;59(2):78-87.
501 doi:10.1034/j.1399-0039.2002.590202.x
- 502 29. Zhang C, de Smith AJ, Smirnov IV, et al. Non-additive and epistatic effects of HLA
503 polymorphisms contributing to risk of adult glioma. *J Neurooncol.* 2017;135(2):237-244.
504 doi:10.1007/s11060-017-2569-7
- 505 30. Kachuri L, Francis SS, Morrison ML, et al. The landscape of host genetic factors involved in
506 immune response to common viral infections. *Genome Med.* 2020;12(1):93.
507 doi:10.1186/s13073-020-00790-x
- 508 31. Melin BS, Barnholtz-Sloan JS, Wrensch MR, et al. Genome-wide association study of glioma
509 subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-
510 glioblastoma tumors. *Nat Genet.* 2017;49(5):789-794. doi:10.1038/ng.3823

32. Wrensch M, Jenkins RB, Chang JS, et al. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat Genet.* 2009;41(8):905-908. doi:10.1038/ng.408
33. Eckel-Passow JE, Lachance DH, Molinaro AM, et al. Glioma Groups Based on 1p/19q, *IDH*, and *TERT* Promoter Mutations in Tumors. *N Engl J Med.* 2015;372(26):2499-2508. doi:10.1056/NEJMoa1407279
34. Jenkins RB, Wrensch MR, Johnson D, et al. Distinct germ line polymorphisms underlie glioma morphologic heterogeneity. *Cancer Genet.* 2011;204(1):13-18. doi:10.1016/j.cancergencyto.2010.10.002
35. Ceccarelli M, Barthel FP, Malta TM, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell.* 2016;164(3):550-563. doi:10.1016/j.cell.2015.12.028
36. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics.* 2011;12(1):246. doi:10.1186/1471-2105-12-246
37. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinforma Oxf Engl.* 2010;26(22):2867-2873. doi:10.1093/bioinformatics/btq559
38. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. Tang J, ed. *PLoS ONE.* 2013;8(6):e64683. doi:10.1371/journal.pone.0064683

39. Bi W, Fritsche LG, Mukherjee B, Kim S, Lee S. A Fast and Accurate Method for Genome-Wide Time-to-Event Data Analysis and Its Application to UK Biobank. *Am J Hum Genet.* 2020;107(2):222-233. doi:10.1016/j.ajhg.2020.06.003
40. Phan AT, Fernandez SG, Somberg JJ, Keck KM, Miranda JL. Epstein–Barr virus latency type and spontaneous reactivation predict lytic induction levels. *Biochem Biophys Res Commun.* 2016;474(1):71-75. doi:10.1016/j.bbrc.2016.04.070
41. Epstein MA, Achong BG, Barr YM. Virus Particles In Cultured Lymphoblasts From Burkitt's Lymphoma. *The Lancet.* 1964;283(7335):702-703. doi:10.1016/S0140-6736(64)91524-7
42. Akhtar S, Vranic S, Cyprian FS, Al Moustafa AE. Epstein–Barr Virus in Gliomas: Cause, Association, or Artifact? *Front Oncol.* 2018;8:123. doi:10.3389/fonc.2018.00123
43. Germini D, Sall FB, Shmakova A, et al. Oncogenic Properties of the EBV ZEBRA Protein. *Cancers.* 2020;12(6):1479. doi:10.3390/cancers12061479
44. Chang Y, Moore PS. Merkel cell carcinoma: a virus-induced human cancer. *Annu Rev Pathol.* 2012;7:123-144. doi:10.1146/annurev-pathol-011110-130227
45. Kwun HJ, Chang Y, Moore PS. Protein-mediated viral latency is a novel mechanism for Merkel cell polyomavirus persistence. *Proc Natl Acad Sci U S A.* 2017;114(20):E4040-E4047. doi:10.1073/pnas.1703879114

Table 1: Clinical and molecular summary of the three case-control datasets

	UCSF-Mayo Cases N=1973	GICC Controls N=1859	AGS i370 Cases N=659	AGS i370 Controls N=586	TCGA Cases N=786	WTCCC Controls N=5711
Age						
<40	983 (50%)	538 (29%)	101 (15%)	85 (15%)	213 (27%)	0 (0%)
40-59	476 (24%)	555 (30%)	329 (50%)	252 (43%)	279 (36%)	0 (0%)
>= 60	514 (26%)	766 (41%)	229 (35%)	249 (42%)	245 (31%)	0 (0%)
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)	49 (6%)	5711 (100%)
Sex						
Female	822 (42%)	744 (40%)	229 (35%)	280 (48%)	328 (42%)	2819 (49%)
Male	1151 (58%)	1115 (60%)	430 (65%)	306 (52%)	458 (58%)	2892 (51%)
IDH mutation status						
Mutant	588 (30%)	NA	111 (17%)	NA	375 (48%)	NA
Wild type	699 (35%)	NA	416 (63%)	NA	364 (46%)	NA
Missing	686 (35%)	NA	132 (20%)	NA	47 (6%)	NA
Molecular subtype based on IDH mutation and 1p/19q codeletion						
MT-codel	244 (12%)	NA	9 (1%)	NA	143 (18%)	NA
MT-noncodel	291 (15%)	NA	94 (14%)	NA	230 (29%)	NA
WT-noncodel	507 (26%)	NA	416 (63%)	NA	357 (46%)	NA
Missing/other	905 (47%)	NA	140 (22%)	NA	56 (7%)	NA

NA: not applicable to controls

Table 2: Summary of available cases and events for survival analyses

	UCSF-Mayo		TCGA		AGS i370	
Molecular Subtype	Cases	Events	Cases	Events	Cases	Events
Glioma	1973	1218	786	310	659	592
IDH mutated	588	201	375	50	111	74
1p/19q codeleted	244	64	143	13	9	1
1p/19q non-codeleted	291	117	230	36	94	69
IDH wild type	699	594	364	228	416	402
1p/19q non-codeleted	507	458	357	224	416	402

Study/subtype combinations with less than 50 cases or 20 events were not utilized in the meta-analysis of survival associations.
Indentation indicates subtype is further subset of the above

Figure 1: Summary of data processing and analysis

Our analysis consisted of three glioma case-control datasets. The first dataset (purple) included 786 glioma cases from The Cancer Genome Atlas (TCGA) genotyped on the Affymetrix 6.0 array and cancer-free controls assembled from two Wellcome Trust Case Control Consortium (WTCCC) studies genotyped using the Affymetrix 6.0 array: 2,917 controls from the 1958 British Birth cohort and 2,794 controls from the UK Blood Service control group. The second dataset (yellow) included 659 cases and 586 controls from the University of California, San Francisco (UCSF) Adult Glioma Study (AGS) genotyped on the Illumina HumanHap370duo panel³². The third set (blue) included 1973 cases from the Mayo Clinic and UCSF AGS and 1859 controls from the Glioma International Case-Control Study (GICC) who were genotyped on the Illumina OncoArray, as previously described^{18,31-34}. The three resulting case-control datasets were processed through quality controls as described in the main text and imputed using the TOPMed imputation server. SNP2HLA was used to impute HLA alleles from SNP data. Risk and survival analysis were performed separately on each study's imputed HLA alleles, on multiple glioma molecular subtypes, and a fixed-effects meta analysis was performed to aggregate results. Separately, genetic reactivity scores to 5 viral antigens were created using previously published GWAS data. Cases and controls across the 3 studies were then assessed a genetically predicted antibody reactivity score (GRS) to each of the 5 antigens. Risk and survival analysis were performed separately on each study using each GRS score as a predictor, with results aggregated using a fixed effects inverse variance weighted meta analysis for each subset of glioma patients and controls based on molecular subtype.

Figure 2: GRS correlations within the UCSF-Mayo cases and controls

Pearson correlations between genetically predicted antigen responses (via GRS) as computed in the UCSF-Mayo glioma cases and controls. Values were printed in each block if and only if the associated correlation test p-value was less than 0.01.

Figure 3: Significant GRS-glioma subtype risk association meta-analysis forest plots

Forest plot meta-analysis results of GRS-glioma risk associations which were at least nominally statistically significant ($p < 0.05$). Response to antigens EBV ZEBRA, MCV and EBV EBNA had associations which reached this threshold. Results are reported as odds ratios, along with 95% confidence intervals. Briefly, each header indicates the studied viral antigen GRS, within are its association with molecular glioma subtypes reported with $p < 0.05$, and the 95% confidence interval of each study-specific effect. The diamond visualizes the 95% confidence interval for the fixed effect (FE) meta-analysis across all 3 studies. Each meta-analysis was tested for between-study heterogeneity (Q statistic), with $p < 0.05$ indicating evidence of study-specific associations.

Figure 4: Nominal GRS-glioma subtype survival association meta-analysis forest plots

Forest plot meta-analysis results of GRS-glioma survival associations which were at least nominally statistically significant ($p < 0.05$). Genetically inferred response to antigens EBV ZEBRA, MCV and EBV EBNA had associations which reached this threshold. Results are reported as hazard ratios, along with 95% confidence intervals. Briefly, each header indicates the studied viral antigen GRS, within are its association with molecular glioma subtypes reported with $p < 0.05$, and the 95% confidence interval of each study-specific effect. The diamond visualizes the 95% confidence interval for the fixed effect (FE) meta-analysis across included studies. Studies which had an insufficient number of cases/events in a subtype were not included in the meta-analysis. Each meta-analysis (where more than one study was included) was tested for between-study heterogeneity (Q statistic), with $p < 0.05$ indicating evidence of study-specific associations.

Figure 5: Kaplan Meier curves for significant GRS_{EBNA} - glioma molecular subtype associations

Kaplan-Meier curves for subtypes where GRS_{EBNA} was nominally associated with subtype specific glioma survival differences ($p < 0.05$ via Cox regression). The second and third plots are distinct partitions of the IDH mutated subgroup. To visualize, unnormalized GRS scores across the included studies were binned based on the case-specific 80th percentile score in the UCSF-Mayo dataset. P-values included on each plot are results of a log-rank test for difference between the two curves. Below each set of curves provides the number of cases surviving beyond that time point, for each of the two GRS groups. In all cases, the glioma cases with higher GRS score for EBV EBNA had visually improved survival outcomes compared to the bottom 80%.