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 ($p<5\times10^{-8}$) 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, $p=3.96\times10^{-4}$) 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.

Keywords

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

51 # Introduction

52

53 Studies linking viruses and cancer date back over 100 years[1] and laid the foundation for
54 understanding oncogenes[2]. It has also become increasingly clear, as well evidenced by the
55 current pandemic, that host genetics play an important role in response to viruses[3,4]. To date,
56 seven viruses have been accepted to be tumor-initiating in humans: Epstein-Barr virus (EBV),
57 hepatitis B virus (HBV), human papillomavirus (HPV), human T-lymphotropic virus-1 (HTLV-1),
58 hepatitis C (HCV), Kaposi's sarcoma herpesvirus (HHV-8) and Merkel cell polyomavirus (MCV)[5].
59 Recent analyses have shown that infections account for approximately 13% of human cancers
60 worldwide[6]. However viruses have not been definitively implicated in the etiology of glioma despite
61 decades of suggestive associations[7–11].

62

63 Glioma is a highly fatal brain cancer with a paucity of known risk factors, with exposure to
64 ionizing radiation being the accepted causal factor[12]. A history of previous infection with Varicella-
65 Zoster virus (VZV) has been the only infectious agent consistently linked to adult glioma,
66 conferring an estimated 20% decrease in risk[10,13]. A suite of other viruses have been associated
67 with the risk and grade of glioma, including EBV, MCV, John Cunningham virus (JCV), BK virus
68 (BKV), human Cytomegalovirus (CMV) and human herpesvirus-6 (HHV-6), but with discordant
69 results[11,14–16]. The identification of key somatic molecular alterations (e.g. *IDH* mutation, 1p/19q
70 chromosomal arm codeletion, *TERT* mutation), which drastically affect glioma prognosis, have
71 uncovered subtype-specific risk factors[17,18]. Recently, several studies/clinical trials have
72 suggested a prognostic benefit of antiviral medications in the treatment of glioma[19–21].

73

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 response[22–24]. 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 call 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 immunity[25].

The HLA region of the genome is considered the most polymorphic region of the human genetic system[26] 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 risk[29], 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 response[30] 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 panel[32]. 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.

124 Molecular subtype information (*IDH* mutation and 1p/19q codeletion status) was downloaded from
125 Ceccarelli et al.[35] supplementary table 1 for the third dataset and was provided directly from the
126 UCSF AGS study and Mayo Clinic for the first and second datasets.
127
128 **Quality Control and Imputation**
129
130 Standard quality control procedures were implemented prior to imputation. Analyses were
131 restricted to individuals of predominantly (>70%) European ancestry, determined using
132 ADMIXTURE[36] and the HapMap 3 reference populations. Within each ancestral group, we
133 removed samples with excess heterozygosity (>3 standard deviations (SD) from mean), <95%
134 call rates, and discordant self-reported and genetically inferred sex. Relatedness checks were
135 performed within each study using KING[37] (kinship >0.12), filtering out up to second-degree
136 relations and retaining the samples with higher call rate. TCGA blood samples were preferentially
137 chosen when both blood and tumor sequencing data were available for the same patient (693
138 European samples had both blood and tumor samples available). SNPs with <95% call rate were
139 removed, along with variants deviating from Hardy-Weinberg equilibrium ($p<10^{-6}$) or at a low minor
140 allele frequency (MAF<0.005). Samples genotyped on the same platform (i.e., Affymetrix 6.0 for
141 TCGA and WTCCC) were imputed together using the multi-ethnic TOPMed reference panel (ver.
142 r2).
143
144 **Statistical Analysis**
145
146 **Genetically Predicted Viral Antigen Response**
147

148 Genetic seroreactivity scores (GRS) were calculated to obtain genetically inferred viral
149 antigen response profiles in each of the glioma datasets. For each GRS, candidate variants and
150 corresponding effect sizes were obtained from genome-wide summary statistics for seroreactivity
151 to 10 viral antigens previously identified in Kachuri and Francis et al. 2020[30] in 7895 randomly
152 selected individuals of European descent from the UK Biobank (UKB) cohort using serological
153 measures. For each antigen, we preferentially selected independent SNPs (linkage
154 disequilibrium, LD, $r^2<0.01$ within 500kb) with the lowest p-value among genome-wide significant
155 variants ($p<5x10^{-8}$). LD proxies ($r^2>0.9$) were obtained for variants unavailable in the target glioma
156 datasets. For each individual, an antigen-specific GRS was calculated as a weighted sum with
157 weights (β) corresponding to a standard deviation increase in antibody response:

158 $$GRS_{Antigen}=\beta_1 x SNP_1+\ldots+\beta_k x SNP_k$$

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

161

162 *Antigen GRS Associations with Glioma Susceptibility*

163

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

171

172 *Antigen GRS Associations with Glioma Survival*

173

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

185

186 *Regional HLA Analyses of glioma risk and survival*

187

188 Classical HLA alleles were imputed for samples in all cohorts at two-field resolution using
189 SNP2HLA[38] and the Type 1 Diabetes Genetics Consortium (T1DGC) reference panel of 2,767
190 unrelated individuals. Associations were tested for 77 alleles (MAF>=0.01) with imputation quality
191 >0.4 across eight genes: *HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-*
192 *DQB1, HLA-DRB1*.

193

194 Subtype-specific associations with risk were estimated using logistic regression models.
195 HLA allele associations with mortality were assessed using SPACox[39], an extension of the Cox
196 model with improved type I error control in high-dimensional settings. Study-specific risk and

197 survival associations were combined in a meta-analysis. Associations for each HLA allele were
198 considered statistically significant if $p<6.5x10^{-4}$ (0.05/77 alleles).
199

200 **Results**

201

202 The creation of a GRS was attempted for all antigens with genome-wide significant
203 seroreactivity-associated variants ($p<5x10^{-8}$) as reported in Kachuri and Francis et al. 2020[30]. This
204 included 10 antigens: Four EBV antigens (EA-D, EBNA, p18, ZEBRA) and antigens for BKV,
205 HHV7, HSV1, JCV, MCV, and VZV. We do not include antigens for CMV and HHV6 in our study
206 as there was only one significantly associated SNP each in the previously published GWAS. An
207 antigen-specific GRS was considered successfully created if at least four independent SNPs
208 passed LD clumping and thresholding and were present (or via proxy SNP) in the three glioma
209 study datasets. GRS were successfully generated for EBV EA-D, EBV EBNA, EBV p18, EBV
210 ZEBRA, and MCV. Only one independent SNP remained for each of HSV1, BKV, JCV and VZV
211 after accounting for long-range LD in the HLA region, although having numerous SNPs meet
212 genome-wide significance in the previous seroreactivity study. Three SNPS remained for HHV7,
213 each on a different chromosome (6,11, and 17). Complete information of variants considered for
214 all 12 viral antigens, including nearest gene and corresponding glioma risk associations are
215 reported in **Supplement Tables S1 and S2**. LD correlations for all associated chromosome 6
216 variants are available in a heatmap in **Supplement Figure S1**.
217

218 The variants included in each of the five created GRSs were overwhelmingly located in
219 the HLA region, with few predictors located elsewhere across the genome: rs67886110 in 3q25.1
220 an eQTL for *MED12L* and *P2RY12* (EBV EBNA), rs7618405 in 3p24.3 (MCV), and rs7444313 in
221 5q31.2 near *TMEM173* (MCV). Of the 38 SNPs included across the 5 GRSs, 2 had a significant

222 association (Bonferroni-corrected: $p<1.3x10^{-3}$) with overall glioma risk: rs9265517 near *HLA-B*
223 (EBV EBNA), $p=3.08x10^{-4}$ and rs9268847 near *HLA-DRB9* (MCV), $p=2.86x10^{-4}$. **Supplement**
224 **Table S2** visualizes the squared correlation between the GRS SNPs and each imputed HLA allele
225 (two field resolution) from UCSF-Mayo cases and controls. Notably, the GRS for MCV was
226 inversely correlated with GRSs for EBV ZEBRA (Pearson's r=-0.28, $p=3.89x10^{-70}$) and EBV EBNA
227 (r=-0.19, $p=6.28x10^{-31}$) (**Figure 2)**.
228
229 *Viral antigen GRS associations with glioma risk*
230
231 In the combined meta-analysis of three case-control studies, three GRS's reached at least
232 nominal significance (p<0.05), with some associations remaining statistically significant after
233 correction for the number of antigens tested (Bonferroni: p=0.05/5=0.01) (**Figure 3**). Genetic
234 predisposition to an increased serological response to EBV ZEBRA was inversely associated with
235 risk of glioma overall (per 1 SD increase in GRS: Odds Ratio, OR_{ZEBRA}=0.94, 95% confidence
236 interval 0.89-0.99, p=0.012, 3418 cases). GRS_{EBNA} was associated with an increased risk of *IDH*
237 mutated gliomas (OR_{EBNA}=1.09,1.004-1.18, p=0.040, 1074 cases) and the magnitude of this effect
238 increased for *IDH* mutated 1p/19q codeleted gliomas (OR_{EBNA}=1.14, 1.012-1.28, p=0.031, 396
239 cases).
240
241 We observed some evidence of antagonistic pleiotropy between genetic determinants of
242 antibody response to EBV ZEBRA and MCV, which generalized across glioma subtypes.
243 GRS_{ZEBRA} and GRS_{MCV} were associated with susceptibility to *IDH* wild type gliomas, but in
244 opposite directions: Higher genetically predicted reactivity to EBV ZEBRA was inversely
245 associated with *IDH* wild type glioma risk (OR_{ZEBRA}=0.91, 0.85-0.98, p=0.0072, 1479 cases), while
246 increased predicted antibody response to MCV conferred an increased risk (OR_{MCV}=1.09, 1.02-

247 1.17, p=0.013). This pattern persisted for *IDH* wild type 1p/19q non-codeleted gliomas
248 (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
249 correlation between GRS_{ZEBRA} and GRS_{MCV} (Pearson's r=-0.28, p=3.9×10^{-70} in UCSF-Mayo,
250 **Figure 2**) suggests a possible shared underlying genetic mechanism between the response to
251 the two antigens. GRS_{ZEBRA} and GRS_{MCV} were tested together, along with interaction term, in a
252 single logistic model and we observed no significant interaction (p=0.34 in *IDH* wild type risk).
253
254 *Viral antigen GRS-survival associations*
255
256 The number of available cases and deaths across the three studies is available in **Table**
257 **2**. Associations between genetically predicted viral antigen response profiles and survival were
258 restricted to *IDH* mutated gliomas (**Figure 4**). GRS_{EBNA} was associated with survival in 1074 *IDH*
259 mutated glioma cases (per 1 SD increase: Hazard Ratio, HR=0.86, 0.76-0.96, p=0.010, 325
260 events), suggesting that a higher genetically predicted reactivity to EBV EBNA improved duration
261 of survival.
262
263 GRS for two EBV antigens were nominally associated with survival amongst 244 *IDH*
264 mutated 1p/19q codeleted glioma cases, but in opposite directions (HR_{EBNA}=0.75, 0.57-0.99
265 p=0.048 / HR_{ZEBRA}=1.27,1.01-1.6, p=0.042, 64 events). This subtype-specific result was limited to
266 the UCSF-Mayo dataset due to insufficient number of reported events (deaths) in *IDH* mutated
267 1p/19q codeleted glioma cases from TCGA and AGS-i370.
268
269 GRS_{EBNA} was also associated with improved survival outcomes in 614 *IDH* mutated 1p/19q
270 non-codeleted glioma cases (HR=0.75, 0.74-0.997, p=0.045, 222 events), suggesting the

271 association of GRS_{EBNA} is independent of 1p/19q status. **Figure 5** visualizes the significant
272 GRS_{EBNA} associations using Kaplan-Meier survival curves.
273
274 We did not detect any significant GRS associations in the analysis of glioma overall, which
275 is consistent with the strong prognostic significance of molecular glioma subtypes[17,33]. Also, as
276 the prognosis in *IDH* wild type gliomas is the poorest, we suspect our GRS instruments are
277 underpowered to detect significant deviations over such short intervals. Full survival results are
278 available in **Supplement Figures S3 and S4**.
279
280 *HLA allele-glioma associations*
281
282 Of the 77 HLA alleles imputed at two-field resolution, only *HLA-DQA1*03:01* reached
283 Bonferroni-corrected significance ($p<6.5x10^{-4}$ after correcting for 77 HLA alleles tested) for
284 association with glioma risk in a meta-analysis. The presence of an *HLA-DQA1*03:01* allele was
285 associated with a decrease in overall glioma risk (OR=0.85, $p=3.96x10^{-4}$). This allele was also
286 nominally associated with risk of *IDH* wild type (OR=0.82, $p=1.4x10^{-3}$) and *IDH* wild type 1p/19q
287 non-codeleted gliomas (OR=0.81, $p=2.6x10^{-3}$). These subtype-specific associations and direction
288 of effect mirror those of GRS_{ZEBRA} presented above, suggesting *HLA-DQA1*03:01* as a potential
289 shared marker for both glioma risk and EBV ZEBRA seroreactivity. Results of all HLA allele (one
290 and two field resolution) associations with glioma risk tested in a meta-analysis are available in
291 **Supplement Table S2**.
292
293 # Discussion
294

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

301 One of our main findings is that genetic predisposition to increased seroreactivity to EBV
302 ZEBRA was associated with a decreased overall glioma risk, with a significant decrease in *IDH*
303 wild type subtypes. Predicted reactivity to the MCV VP1 antigen mirrored the same *IDH* wild type
304 associations but in the opposite direction, where higher reactivity was associated with increased
305 glioma risk. The significant inverse relationship between predicted reactivity to EBV ZEBRA and
306 MCV VP1 highlights the possibility of shared genetic components of antibody response to the two
307 antigens. Understanding the underlying mechanisms guided by these genetics is left as an open
308 question. A possible link is that the class II HLA allele *DQA1*03:01* was associated with decreased
309 glioma risk in the same subtypes associated with GRSs for EBV ZEBRA. In our previous UKB
310 analysis[30] the presence of *HLA-DQA1*03:01* was positively associated with EBV ZEBRA
311 seroreactivity measurements (β=0.168, p=1.3x10^{-16}). Taken together, the associations between
312 *HLA-DQA1*03:01*, EBV ZEBRA and glioma risk suggest possible shared underlying
313 immunogenetic architecture. As HLA class II proteins present potentially antigenic peptides,
314 functional genetic alterations can result in changes to the binding affinity of specific antigens,
315 modulating the potential for recognition by CD4+ helper T-cells. It is possible that the
316 heterodimeric DQ proteins half-encoded by *DQA1*03:01* have improved binding or recognition of
317 peptides presented by both EBV ZEBRA proteins and glioma (specifically *IDH* wild type), allowing
318 for an increased immune response in both cases. As *IDH* wild type has generally been shown to
319 serve as a marker for more severe glioma cases, the exact somatic tumor alterations leading to
320 recognized peptide variants in these tumors is not clear and is an area that warrants future

321 investigation. Further analysis of viral-tumor protein homology is needed to understand if this
322 could be a possible connection.
323
324 Interestingly, a higher genetically predicted response to EBV EBNA was nominally
325 associated with increased risk in *IDH* mutated/1p/19q codeleted gliomas. Still, reactivity towards
326 EBV EBNA was more strongly associated with improved survival in *IDH* mutated gliomas. This
327 discrepancy between the disease-promoting and pro-survival associations of GRS_{EBNA} may
328 suggest the presence of different biologic pathways after initiation of disease. It may also point
329 towards EBV latency-mediated treatment effects. Previous studies show that the EBV latency
330 type predicts the relative amount of induced reactivation generated by cytotoxic chemotherapy
331 drugs[40]. Therefore, individuals who react strongly to EBV EBNA antigens may exhibit a different
332 pattern of reactivation when treated with temozolomide. Further research is required to elucidate
333 this putative association, yet it is clear that EBV lytic/latent cycling is a critical aspect of other
334 germline interactions and disease risk.
335
336 Epstein-Barr virus was the first recognized human oncovirus[41] and exists in two distinct
337 life cycles within a host: a lytic phase of active infection where new viruses are produced, and a
338 latent phase where the virus remains dormant to avoid detection by the host. Distinct sets of
339 antigens are produced during these two phases, two of such are EBNA, during the latent phase
340 of infection, and the ZEBRA protein which initiates a change from latent to lytic gene expression.
341 Past evidence suggests processes during the latent stage are responsible for the virus' oncogenic
342 properties via mechanisms promoting cell growth and preventing cell apoptosis (reviewed in
343 Akhtar et al. 2018[42]). Recent work has also implicated the ZEBRA protein as oncogenic with
344 evidence demonstrating its ability to deregulate immune surveillance and promote immune
345 escape[43].
346

347 In contrast MCV is the most recently recognized human oncovirus, first described in 2008
348 and later accepted as a causal agent for Merkel cell carcinoma[44], a neuroendocrine carcinoma.
349 Previous studies have identified the presence of MCV DNA in glioma patients and have drawn an
350 association with infection and increased risk of glioma[15,16]. It has been shown that both the large
351 T and small T antigens of MCV are oncoproteins that target tumor suppressor proteins such as
352 pRB. Although the exact latency mechanism in polyomaviruses is unknown, it has been
353 suggested that complex protein-mediated latency may be critical to the MCV lifecycle[45].
354
355 In interpreting our findings, several limitations should be acknowledged. The shared
356 genetic architecture between the viruses studied here, as seen in the correlations in **Figure 2 and**
357 **Supplement Figure S1**, may result in a lack of specificity. This limitation in our study may be
358 applicable in other viruses that share the same underlying genetic programming of antigen
359 response, particularly VZV, which is associated with a unique pattern of LD spanning a large
360 region of the HLA[30]. Furthermore, the clumping and thresholding approach to GRS development
361 may not be optimal for regions with complex LD structure, such as HLA. An approach that can
362 appropriately account for the long-range correlation structure and capture non-linear interactions,
363 such as haplotype effects, may improve future studies that use genetically inferred immune
364 responses. Another challenge is the low SNP density in the HLA region in the present study, as
365 well as reliance on imputed HLA alleles both in the glioma studies and in UKB. The HLA region
366 is highly polymorphic and imputation of alleles from SNP data is limiting and prone to error.
367 Similarly, it is difficult to disentangle the specific function of SNPs located in the HLA region. As
368 such, additional studies leveraging HLA sequencing approaches are required to provide further
369 granularity of the genetic determinants of antigen response and glioma risk/survival.
370 Additionally, we could not study the relationship between reactivity to several antigens of
371 interest (BKV, HSV-1, JCV, VZV, HHV-7, HHV-6, CMV) due to the limited number of variants
372 reaching genome-wide significance and/or meeting our GRS criteria[30]. Although the GRS-glioma

373 association results presented here require further replication in independent patient populations,
374 our findings are intriguing and suggest previous inconsistencies in observational epidemiologic
375 may be partially due to individual differences in genetic factors that affect antigen reactivity.
376 This is the first study directly examining the underlying genetic architecture of antigen
377 reactivity to common viruses and glioma risk and survival. We observed important associations
378 between programmed reactivity to viruses and glioma etiology and prognosis. This methodology
379 is not limited to the study of glioma, as the GRS instruments proposed here can readily be applied
380 to cohorts of any cancer type. This offers a unique approach for future studies to reinvestigate the
381 genetic contributions of long-running epidemiologic associations between viruses and cancer,
382 and possibly clarify effects of viral therapies in their treatment.
383

384 Supplemental Information

385 Supplemental data include five tables and five figures.
386

387 Funding and Acknowledgments

388 Work at University of California, San Francisco was supported by the National Institutes
389 of Health (grant numbers T32CA151022, R01CA52689, P50CA097257, R01CA126831,
390 R01CA139020, R01AI128775, and R25CA112355), the National Brain Tumor Foundation,
391 the Stanley D. Lewis and Virginia S. Lewis Endowed Chair in Brain Tumor Research, the Robert
392 Magnin Newman Endowed Chair in Neuro-oncology, and by donations from families and friends
393 of John Berardi, Helen Glaser, Elvera Olsen, Raymond E. Cooper, and William Martinusen.
394 This publication was supported by the National Center for Research Resources and the National
395 Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI
396 Grant Number UL1 RR024131. Its contents are solely the responsibility of the authors and do not
397 necessarily represent the official views of the NIH.

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

Data Availability

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

Notes

Declaration of interests

The authors declare no competing interests.

Authorship

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

428 ## References

429 1. Rous P. A Tranimissible Avian Neoplasm. (Sarcoma of the common fowl.). *J Exp Med*.
430 1910;12(5):696-705. doi:10.1084/jem.12.5.696

431 2. Duesberg PH, Vogt PK. Differences between the ribonucleic acids of transforming and
432 nontransforming avian tumor viruses. *Proc Natl Acad Sci U S A*. 1970;67(4):1673-1680.
433 doi:10.1073/pnas.67.4.1673

434 3. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19.
435 *Nature*. Published online July 8, 2021. doi:10.1038/s41586-021-03767-x

436 4. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying
437 the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med*.
438 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779

439 5. Mui UN, Haley CT, Tyring SK. Viral Oncology: Molecular Biology and Pathogenesis. *J Clin*
440 *Med*. 2017;6(12):E111. doi:10.3390/jcm6120111

441 6. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable
442 to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health*. 2020;8(2):e180-
443 e190. doi:10.1016/S2214-109X(19)30488-7

444 7. Neves AM, Thompson G, Carvalheira J, et al. Detection and quantitative analysis of human
445 herpesvirus in pilocytic astrocytoma. *Brain Res*. 2008;1221:108-114.
446 doi:10.1016/j.brainres.2008.05.009

447 8. Chauvin C, Suh M, Remy C, Benabid AL. Failure to detect viral genomic sequences of three
448 viruses (herpes simplex, simian virus 40 and adenovirus) in human and rat brain tumors. *Ital*
449 *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

511 32. Wrensch M, Jenkins RB, Chang JS, et al. Variants in the CDKN2B and RTEL1 regions are
512 associated with high-grade glioma susceptibility. *Nat Genet*. 2009;41(8):905-908.
513 doi:10.1038/ng.408

514 33. Eckel-Passow JE, Lachance DH, Molinaro AM, et al. Glioma Groups Based on 1p/19q, *IDH*
515 , and *TERT* Promoter Mutations in Tumors. *N Engl J Med*. 2015;372(26):2499-2508.
516 doi:10.1056/NEJMoa1407279

517 34. Jenkins RB, Wrensch MR, Johnson D, et al. Distinct germ line polymorphisms underlie
518 glioma morphologic heterogeneity. *Cancer Genet*. 2011;204(1):13-18.
519 doi:10.1016/j.cancergencyto.2010.10.002

520 35. Ceccarelli M, Barthel FP, Malta TM, et al. Molecular Profiling Reveals Biologically Discrete
521 Subsets and Pathways of Progression in Diffuse Glioma. *Cell*. 2016;164(3):550-563.
522 doi:10.1016/j.cell.2015.12.028

523 36. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry
524 estimation. *BMC Bioinformatics*. 2011;12(1):246. doi:10.1186/1471-2105-12-246

525 37. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship
526 inference in genome-wide association studies. *Bioinforma Oxf Engl*. 2010;26(22):2867-
527 2873. doi:10.1093/bioinformatics/btq559

528 38. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing Amino Acid Polymorphisms in Human
529 Leukocyte Antigens. Tang J, ed. *PLoS ONE*. 2013;8(6):e64683.
530 doi:10.1371/journal.pone.0064683

531 39. Bi W, Fritsche LG, Mukherjee B, Kim S, Lee S. A Fast and Accurate Method for Genome-
532 Wide Time-to-Event Data Analysis and Its Application to UK Biobank. *Am J Hum Genet*.
533 2020;107(2):222-233. doi:10.1016/j.ajhg.2020.06.003

534 40. Phan AT, Fernandez SG, Somberg JJ, Keck KM, Miranda JL. Epstein–Barr virus latency
535 type and spontaneous reactivation predict lytic induction levels. *Biochem Biophys Res*
536 *Commun*. 2016;474(1):71-75. doi:10.1016/j.bbrc.2016.04.070

537 41. Epstein MA, Achong BG, Barr YM. Virus Particles In Cultured Lymphoblasts From Burkitt's
538 Lymphoma. *The Lancet*. 1964;283(7335):702-703. doi:10.1016/S0140-6736(64)91524-7

539 42. Akhtar S, Vranic S, Cyprian FS, Al Moustafa AE. Epstein–Barr Virus in Gliomas: Cause,
540 Association, or Artifact? *Front Oncol*. 2018;8:123. doi:10.3389/fonc.2018.00123

541 43. Germini D, Sall FB, Shmakova A, et al. Oncogenic Properties of the EBV ZEBRA Protein.
542 *Cancers*. 2020;12(6):1479. doi:10.3390/cancers12061479

543 44. Chang Y, Moore PS. Merkel cell carcinoma: a virus-induced human cancer. *Annu Rev*
544 *Pathol*. 2012;7:123-144. doi:10.1146/annurev-pathol-011110-130227

545 45. Kwun HJ, Chang Y, Moore PS. Protein-mediated viral latency is a novel mechanism for
546 Merkel cell polyomavirus persistence. *Proc Natl Acad Sci U S A*. 2017;114(20):E4040-
547 E4047. doi:10.1073/pnas.1703879114

548

549

550

551

552

553 **Table 1: Clinical and molecular summary of the three case-control datasets**
554

	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

555 NA: not applicable to controls
556

557 **Table 2: Summary of available cases and events for survival analyses**
558

	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

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

561 **Figure 1: Summary of data processing and analysis**

562 Our analysis consisted of three glioma case-control datasets. The first dataset (purple)
563 included 786 glioma cases from The Cancer Genome Atlas (TCGA) genotyped on the Affymetrix
564 6.0 array and cancer-free controls assembled from two Wellcome Trust Case Control Consortium
565 (WTCCC) studies genotyped using the Affymetrix 6.0 array: 2,917 controls from the 1958 British
566 Birth cohort and 2,794 controls from the UK Blood Service control group. The second dataset
567 (yellow) included 659 cases and 586 controls from the University of California, San Francisco
568 (UCSF) Adult Glioma Study (AGS) genotyped on the Illumina HumanHap370duo panel[32]. The
569 third set (blue) included 1973 cases from the Mayo Clinic and UCSF AGS and 1859 controls from
570 the Glioma International Case-Control Study (GICC) who were genotyped on the Illumina
571 OncoArray, as previously described[18,31–34]. The three resulting case-control datasets were
572 processed through quality controls as described in the main text and imputed using the TOPMed
573 imputation server. SNP2HLA was used to impute HLA alleles from SNP data. Risk and survival
574 analysis were performed separately on each study's imputed HLA alleles, on multiple glioma
575 molecular subtypes, and a fixed-effects meta analysis was performed to aggregate results.
576 Separately, genetic reactivity scores to 5 viral antigens were created using previously published
577 GWAS data. Cases and controls across the 3 studies were then assessed a genetically predicted
578 antibody reactivity score (GRS) to each of the 5 antigens. Risk and survival analysis were
579 performed separately on each study using each GRS score as a predictor, with results aggregated
580 using a fixed effects inverse variance weighted meta analysis for each subset of glioma patients
581 and controls based on molecular subtype.

582

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

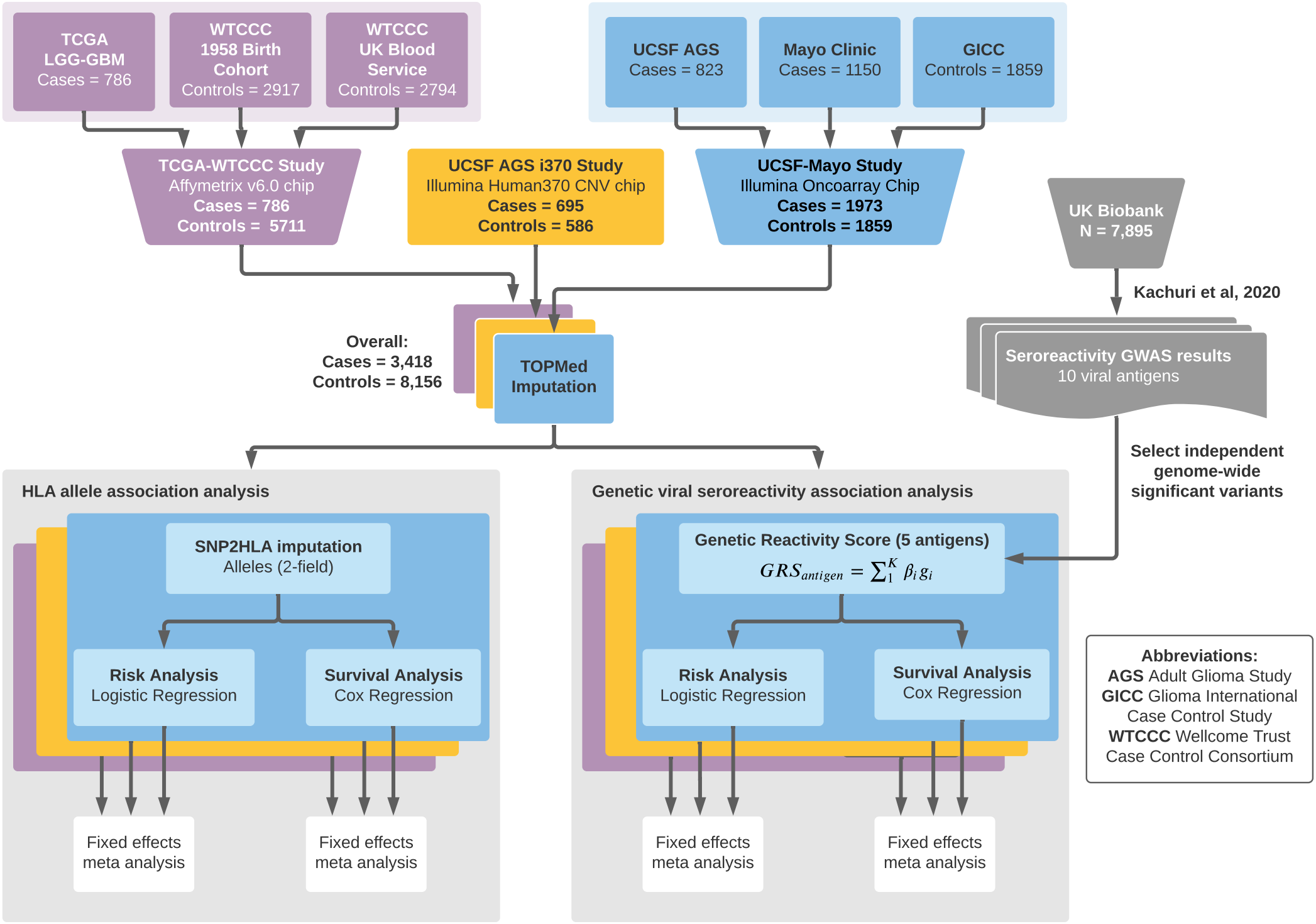
584 Pearson correlations between genetically predicted antigen responses (via GRS) as
585 computed in the UCSF-Mayo glioma cases and controls. Values were printed in each block if and
586 only if the associated correlation test p-value was less than 0.01.

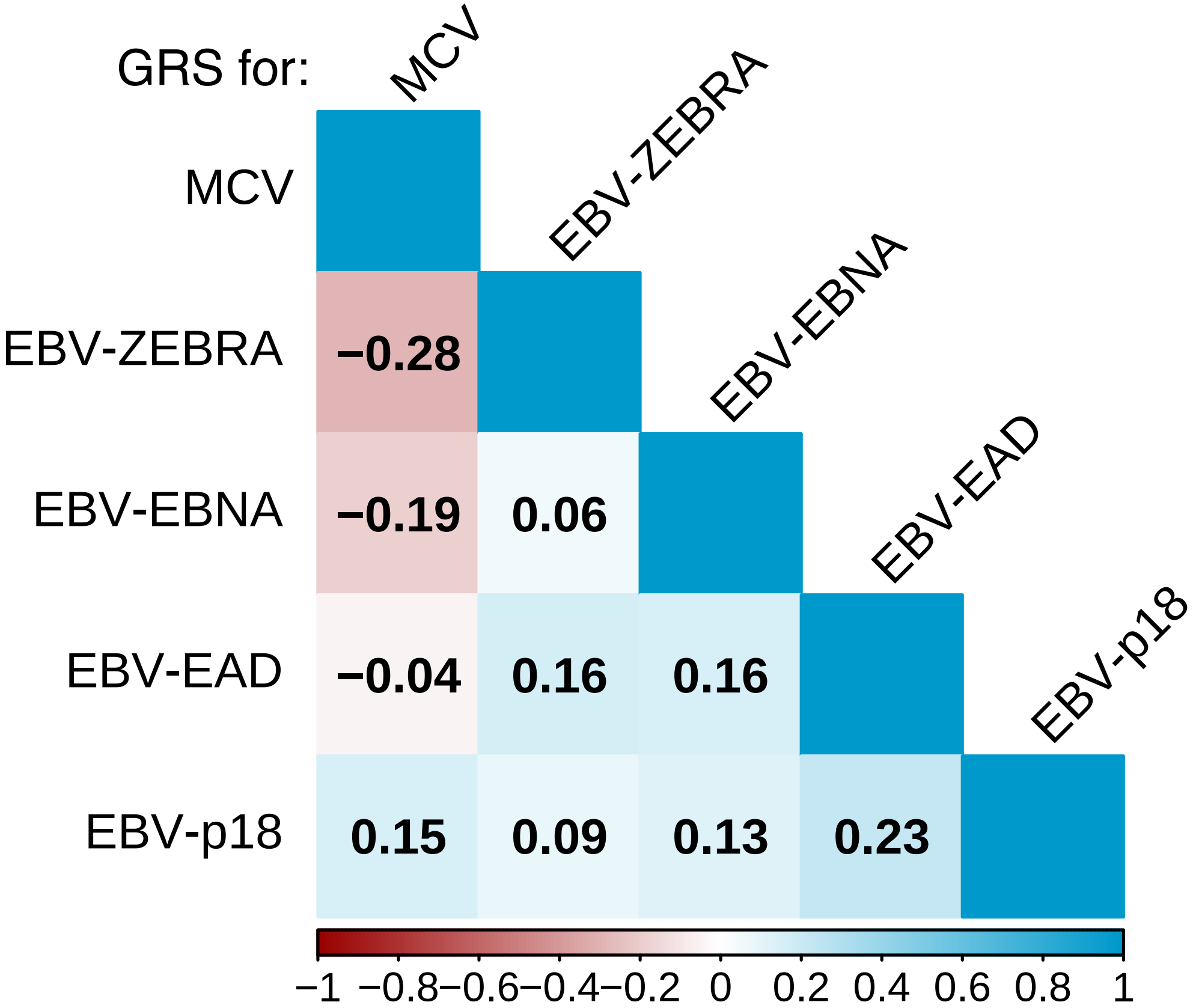
587

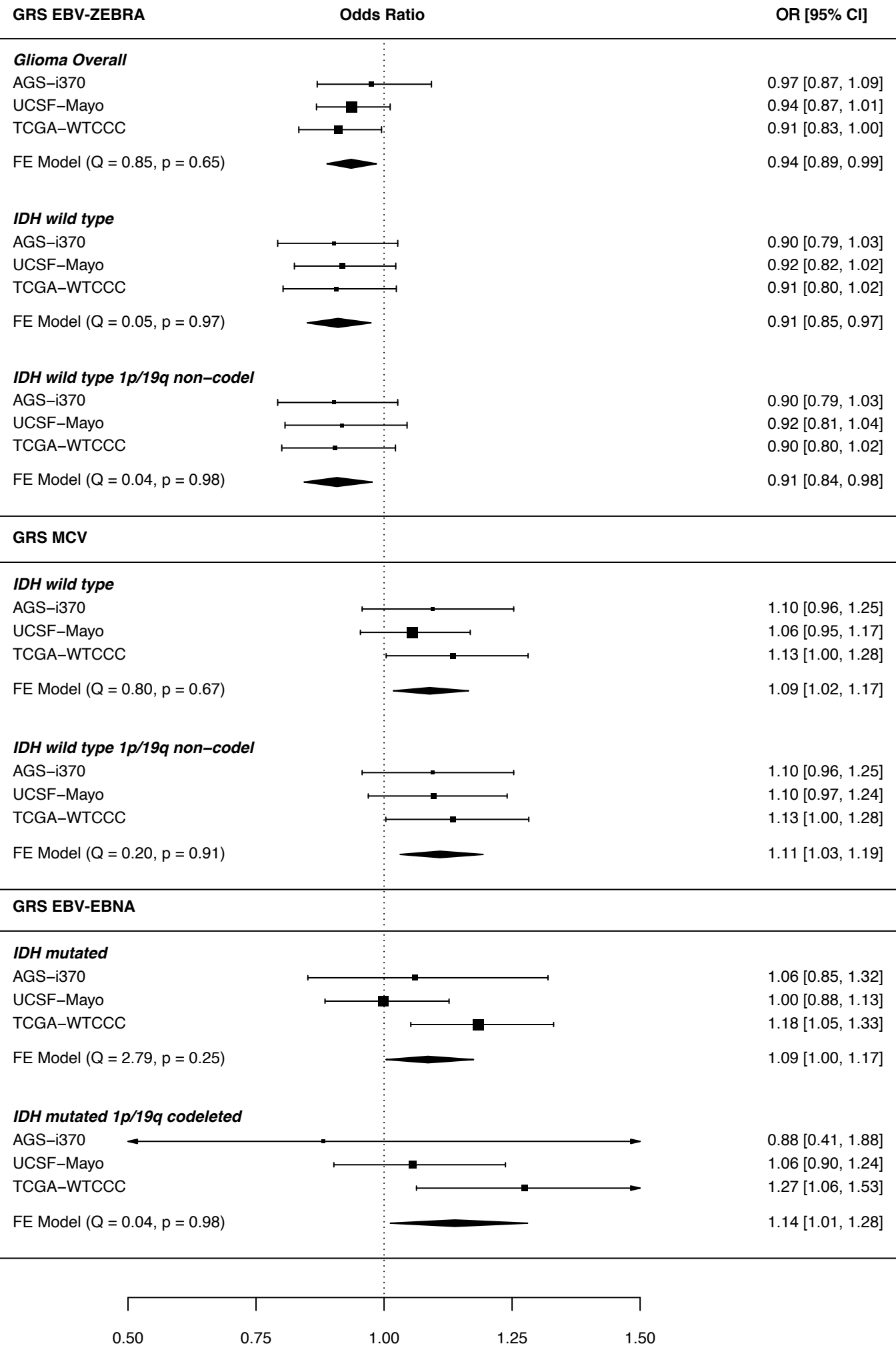
588 **Figure 3: Significant GRS-glioma subtype risk association meta-analysis**
589 **forest plots**

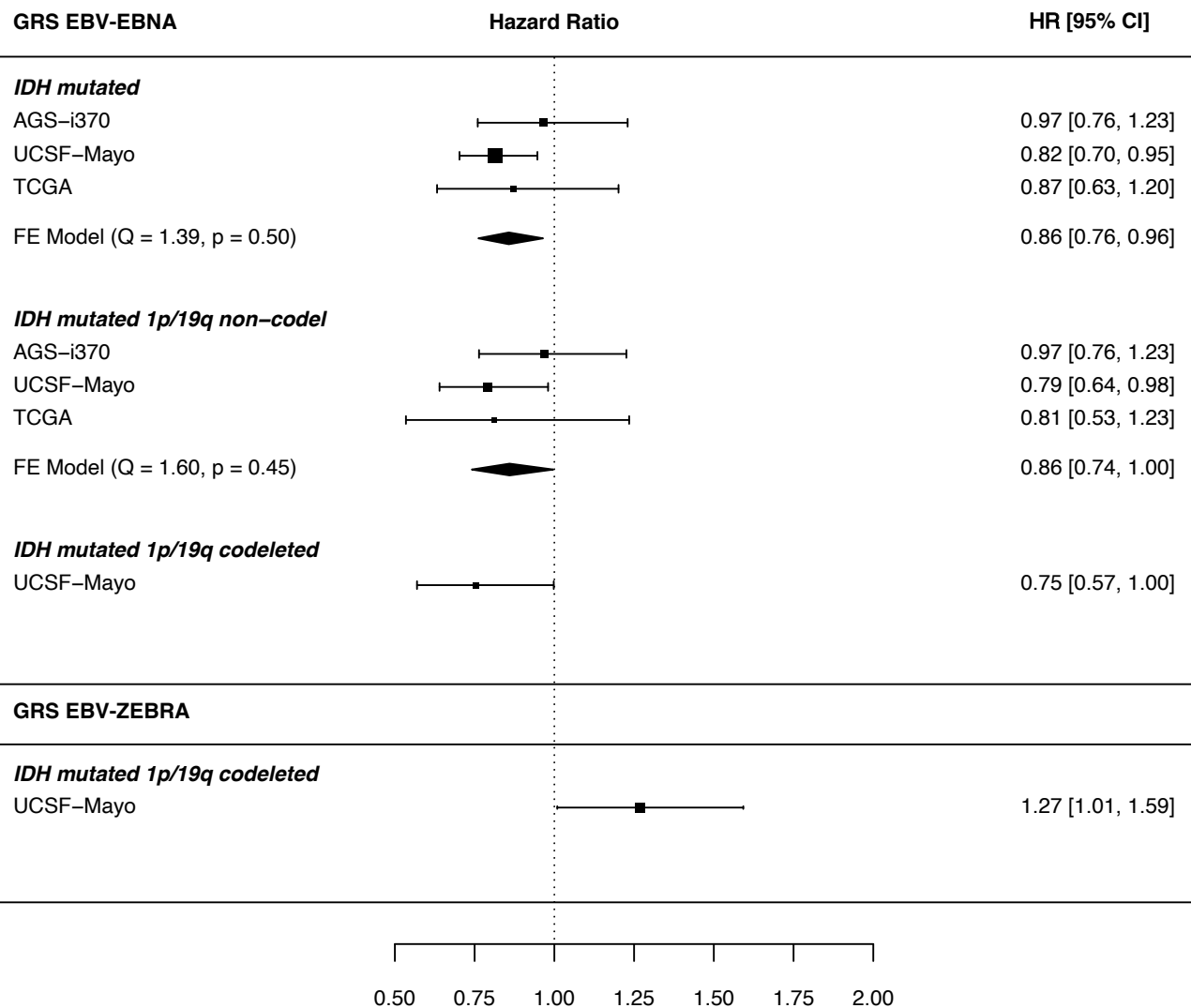
590 Forest plot meta-analysis results of GRS-glioma risk associations which were at least
591 nominally statistically significant ($p < 0.05$). Response to antigens EBV ZEBRA, MCV and EBV
592 EBNA had associations which reached this threshold. Results are reported as odds ratios, along
593 with 95% confidence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

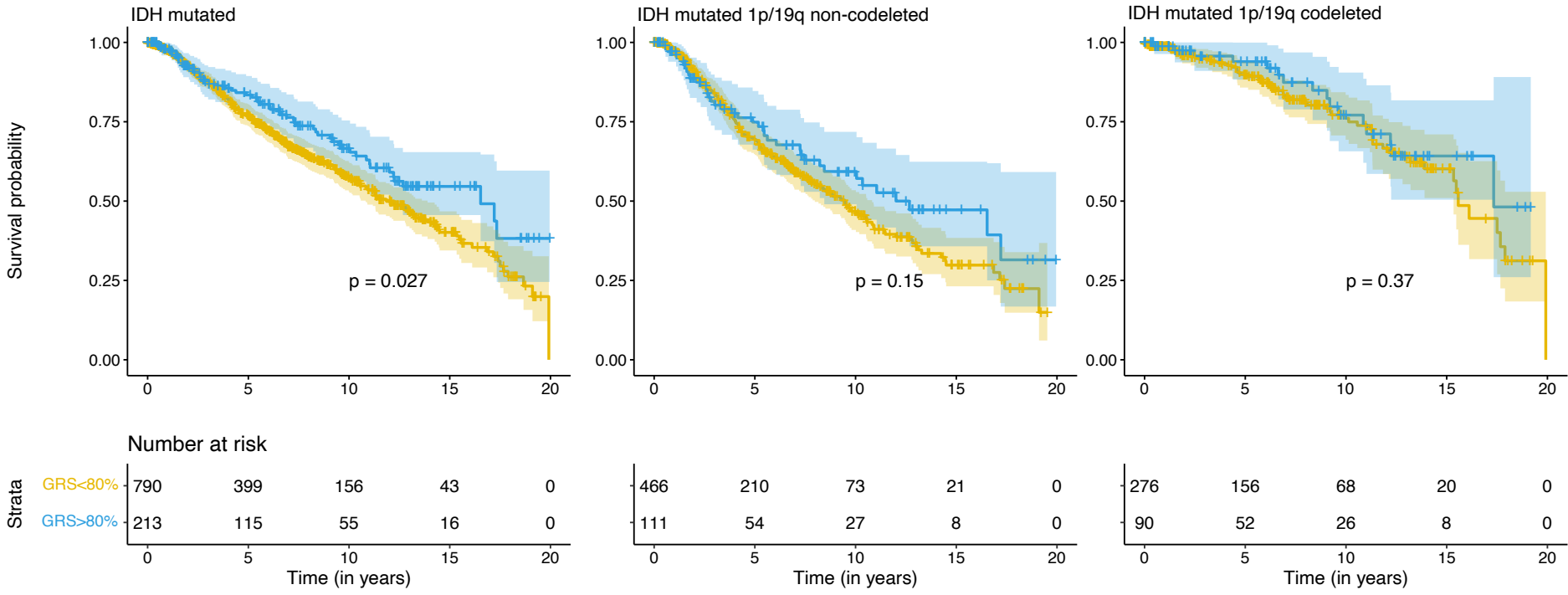
598 **Figure 4: Nominal GRS-glioma subtype survival association meta-analysis**
599 **forest plots**
600 Forest plot meta-analysis results of GRS-glioma survival associations which were at least
601 nominally statistically significant ($p < 0.05$). Genetically inferred response to antigens EBV ZEBRA,
602 MCV and EBV EBNA had associations which reached this threshold. Results are reported as
603 hazard ratios, along with 95% confidence intervals. Briefly, each header indicates the studied viral
604 antigen GRS, within are its association with molecular glioma subtypes reported with $p<0.05$, and
605 the 95% confidence interval of each study-specific effect. The diamond visualizes the 95%
606 confidence interval for the fixed effect (FE) meta-analysis across included studies. Studies which
607 had an insufficient number of cases/events in a subtype were not included in the meta-analysis.
608 Each meta-analysis (where more than one study was included) was tested for between-study
609 heterogeneity (Q statistic), with $p<0.05$ indicating evidence of study-specific associations.
610
611 **Figure 5: Kaplan Meier curves for significant GRS_{EBNA} - glioma molecular**
612 **subtype associations**
613 Kaplan-Meier curves for subtypes where GRS_{EBNA} was nominally associated with subtype
614 specific glioma survival differences ($p < 0.05$ via Cox regression). The second and third plots are
615 distinct partitions of the IDH mutated subgroup. To visualize, unnormalized GRS scores across
616 the included studies were binned based on the case-specific 80th percentile score in the UCSF-
617 Mayo dataset. P-values included on each plot are results of a log-rank test for difference between
618 the two curves. Below each set of curves provides the number of cases surviving beyond that
619 time point, for each of the two GRS groups. In all cases, the glioma cases with higher GRS score
620 for EBV EBNA had visually improved survival outcomes compared to the bottom 80%.


TCGA
LGG-GBM
Cases = 786
WTCCC
1958 Birth
Cohort
Controls = 2917
WTCCC
UK Blood
Service
Controls = 2794
UCSF AGS
Cases = 823
Mayo Clinic
Cases = 1150
GICC
Controls = 1859
TCGA-WTCCC Study
Affymetrix v6.0 chip
Cases = 786
Controls = 5711
UCSF AGS i370 Study
Illumina Human370 CNV chip
Cases = 695
Controls = 586
UCSF-Mayo Study
Illumina Oncoarray Chip
Cases = 1973
Controls = 1859
UK Biobank
N = 7,895
Kachuri et al, 2020
Overall:
Cases = 3,418
Controls = 8,156
TOPMed
Imputation
Seroreactivity GWAS results
10 viral antigens
Select independent
genome-wide
significant variants
HLA allele association analysis
SNP2HLA imputation
Alleles (2-field)
Risk Analysis
Logistic Regression
Survival Analysis
Cox Regression
Fixed effects
meta analysis
Fixed effects
meta analysis
Genetic viral seroreactivity association analysis
Genetic Reactivity Score (5 antigens)
$GRS_{antigen} = \sum_1^K \beta_i g_i$
Risk Analysis
Logistic Regression
Survival Analysis
Cox Regression
Fixed effects
meta analysis
Fixed effects
meta analysis
Abbreviations:
AGS Adult Glioma Study
GICC Glioma International
Case Control Study
WTCCC Wellcome Trust
Case Control Consortium


GRS for:
MCV
EBV-ZEBRA
EBV-EBNA
EBV-EAD
EBV-p18
MCV
EBV-ZEBRA
EBV-EBNA
EBV-EAD
EBV-p18
−0.28
−0.19
0.06
−0.04
0.16
0.16
0.15
0.09
0.13
0.23
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1


GRS EBV-ZEBRA
Odds Ratio
OR [95% CI]
Glioma Overall
AGS–i370
0.97 [0.87, 1.09]
UCSF–Mayo
0.94 [0.87, 1.01]
TCGA–WTCCC
0.91 [0.83, 1.00]
FE Model (Q = 0.85, p = 0.65)
0.94 [0.89, 0.99]
IDH wild type
AGS–i370
0.90 [0.79, 1.03]
UCSF–Mayo
0.92 [0.82, 1.02]
TCGA–WTCCC
0.91 [0.80, 1.02]
FE Model (Q = 0.05, p = 0.97)
0.91 [0.85, 0.97]
IDH wild type 1p/19q non–codel
AGS–i370
0.90 [0.79, 1.03]
UCSF–Mayo
0.92 [0.81, 1.04]
TCGA–WTCCC
0.90 [0.80, 1.02]
FE Model (Q = 0.04, p = 0.98)
0.91 [0.84, 0.98]
GRS MCV
IDH wild type
AGS–i370
1.10 [0.96, 1.25]
UCSF–Mayo
1.06 [0.95, 1.17]
TCGA–WTCCC
1.13 [1.00, 1.28]
FE Model (Q = 0.80, p = 0.67)
1.09 [1.02, 1.17]
IDH wild type 1p/19q non–codel
AGS–i370
1.10 [0.96, 1.25]
UCSF–Mayo
1.10 [0.97, 1.24]
TCGA–WTCCC
1.13 [1.00, 1.28]
FE Model (Q = 0.20, p = 0.91)
1.11 [1.03, 1.19]
GRS EBV-EBNA
IDH mutated
AGS–i370
1.06 [0.85, 1.32]
UCSF–Mayo
1.00 [0.88, 1.13]
TCGA–WTCCC
1.18 [1.05, 1.33]
FE Model (Q = 2.79, p = 0.25)
1.09 [1.00, 1.17]
IDH mutated 1p/19q codeleted
AGS–i370
0.88 [0.41, 1.88]
UCSF–Mayo
1.06 [0.90, 1.24]
TCGA–WTCCC
1.27 [1.06, 1.53]
FE Model (Q = 0.04, p = 0.98)
1.14 [1.01, 1.28]
0.50
0.75
1.00
1.25
1.50


GRS EBV-EBNA
Hazard Ratio
HR [95% CI]
IDH mutated
AGS−i370
0.97 [0.76, 1.23]
UCSF−Mayo
0.82 [0.70, 0.95]
TCGA
0.87 [0.63, 1.20]
FE Model (Q = 1.39, p = 0.50)
0.86 [0.76, 0.96]
IDH mutated 1p/19q non−codel
AGS−i370
0.97 [0.76, 1.23]
UCSF−Mayo
0.79 [0.64, 0.98]
TCGA
0.81 [0.53, 1.23]
FE Model (Q = 1.60, p = 0.45)
0.86 [0.74, 1.00]
IDH mutated 1p/19q codeleted
UCSF−Mayo
0.75 [0.57, 1.00]
GRS EBV-ZEBRA
IDH mutated 1p/19q codeleted
UCSF−Mayo
1.27 [1.01, 1.59]
0.50
0.75
1.00
1.25
1.50
1.75
2.00


IDH mutated
IDH mutated 1p/19q non-codeleted
IDH mutated 1p/19q codeleted
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.027
p = 0.15
p = 0.37
0
5
10
15
20
Number at risk
Strata
GRS<80%
GRS>80%
790 399 156 43 0
213 115 55 16 0
466 210 73 21 0
111 54 27 8 0
276 156 68 20 0
90 52 26 8 0
Time (in years)