The impact of germline variants in DNA repair pathways on survival and temozolomide toxicity in adults with glioma

Geno Guerra, Ph.D.[1,2,*], George Wendt, B.A.[1], Lucie McCoy, M.P.H.[1], Helen M. Hansen, B.A.[1], Annette M. Molinaro, Ph.D.[1,2], Terri Rice, M.P.H.[1], Victoria Guan[3], Lianne Capistrano[3], Allison Hsieh[3], Veruna Kalsi[3], Jaimie Sallee[3], Jennie W. Taylor, M.D., M.P.H.[1,4], Jennifer L. Clarke, M.D., M.P.H.[1,4], Eduardo Rodriguez Almaraz M.D.[1,2], Linda Kachuri, Ph.D.[5,6], John K. Wiencke, Ph.D.[1,2], Jeanette E Eckel-Passow, Ph.D.[7], Robert B. Jenkins, M.D., Ph.D.[8], Margaret Wrensch, Ph.D.[1] , Stephen S. Francis, Ph.D.[1,2,9,**]

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. School of Pharmacy, University of California San Francisco, San Francisco, CA, USA
4. Department of Neurology, University of California San Francisco, San Francisco, CA, USA
5. Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA
6. Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA
7. Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA
8. Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
9. Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, USA

* Corresponding author: geno.guerra@ucsf.edu

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

Abstract:

Gliomas are highly fatal malignant brain tumors with prognostically relevant molecular subtypes. Genomic instability is a hallmark of cancer, including glioma, and DNA repair mechanisms are crucial to maintaining cell integrity. Temozolomide (TMZ), which utilizes specific

DNA repair pathways in its mechanism of action, is one of the few standard-of-care drugs with a meaningful impact on *IDH* wildtype glioblastoma survival. However, its utility in non-glioblastomas is complicated by temozolomide-induced hypermutation. This study aims to identify if germline polymorphisms in DNA repair genes are associated with overall survival within glioma subtypes and if the polymorphisms modify survival in subjects treated with temozolomide.

We utilized genotype data of 2078 adults with glioma collected through the UCSF Adult Glioma Study, Mayo Clinic, and TCGA, to study 1393 SNPs within 22 specific DNA repair genes on overall survival for all cases and for those with/without temozolomide exposure during first course of treatment. All models were fit separately for individuals grouped by major molecular subtypes.

We identified four germline SNPs (within *MLH1*, *MSH4*, *MSH3*, and *MUTYH*) with a significant ($p<0.00019$) polygenic effect on survival, where the presence of at least one SNP decreased overall survival among grade 2 and 3 glioma cases with *IDH* mutated tumors (Hazard ratio per unit increase in number of present polymorphic sites, HR-PS=1.52, 95% confidence interval, CI: 1.26-1.83, $p=1.3\times10^{-5}$). In *IDH* mutant gliomas without somatic 1p/19q chromosomal arm codeletion, we identified a group of SNPs (in *MSH4*, *ERCC2*, and *ERCC1*) which showed a stepwise significant improvement on overall survival for each present germline polymorphism (HR-PS=0.67, 95% CI: 0.56-0.80, $p=8.9\times10^{-6}$).

Within *IDH* mutant glioma cases, we also identified two mutations (rs1540354-T in *MLH1*, rs71636247-G in *MSH3*) which were significantly associated with decreased overall survival only amongst those with known temozolomide usage (HR-PS=2.23, 95% CI: 1.55-3.20, $p=1.56\times10^{-5}$), with no survival effect observed in the non-temozolomide group (HR=0.90, 95% CI: 0.63-1.28, $p=0.56$). One SNP in *MSH2* was suggestively associated ($p<0.005$) with altered *IDH* wildtype glioblastoma response to temozolomide (rs149630102-T, HR=1.74, 95% CI: 1.27-2.39, $p=0.00052$). Suggestive associations for other glioma subtypes are also reported.

We found evidence that germline polygenic alterations within DNA repair pathways may alter prognosis and potentially modulate temozolomide toxicity in adults with glioma. Further validation and functional work are warranted to assess if these markers can assist clinicians in individually tailoring therapeutic treatments for adults with glioma tumors.

Keywords:

glioma, DNA repair, base excision repair, mismatch repair, temozolomide, pharmacogenomics

Funding and acknowledgments:

Work at the University of California, San Francisco, was supported by the National Institutes of Health (grant numbers T32CA112355, R01CA52689, P50CA097257, R01CA126831, R01CA139020, and R01CA266676), as well as the loglio Collective, 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.

The work at Mayo was supported by National Cancer Institute (NCI) grants CA230712, P50 CA108961, and CA139020; the National Brain Tumor Society; the loglio Collective; the Mayo Clinic; and the Ting Tsung and Wei Fong Chao Foundation.

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 authors' responsibility and do not necessarily represent the official views of the NIH. The authors wish to acknowledge study participants, the clinicians, and the research staff at the participating medical centers, the UCSF Cancer Registry, and the UCSF Neurosurgery Tissue Bank.

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. All analyses, interpretations, and conclusions reached in this manuscript from the mortality data are those of the author(s) and not the State of California Department of Public Health.

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 04/26/23.

The results published here are in whole or part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov.

Declaration of interests:

The authors declare no competing interests.

Authorship:

105 GG, VG, LC, AH, VK, JS, and SSF conceived of the study. GG and SSF wrote the main
106 drafts of the manuscript. VG, LC, AG, VK, JS, GG, and SSF conducted literature searches. GG,
107 GW, and SSF conducted statistical and computational analyses. GG, SSF, JWT, JLC, AMM, JLW,
108 JEP, RBJ, LK, ERA, and MW advised on result interpretations. The primary data collection
109 involved LM, HMH, AMM, JLW, TR, JEP, RBJ, JKW, and MW. All authors contributed to,
110 reviewed, and approved the final manuscript.
111
112 **Introduction/Background:**

113 Gliomas are a histologically and molecularly diverse group of highly fatal cancers
114 originating from brain glial cells. Overall, gliomas have an incidence rate of ~6 per 100,000
115 persons[1]. Along with tumor grade, glioma prognoses can be well stratified by the identification of
116 somatic markers within the primary tumor, the most prominent features being mutations to
117 isocitrate dehydrogenase 1 and 2 (*IDH*) genes, and codeletion of the 1p and 19q chromosomal
118 arms (1p/19q)[2]. The most aggressive subtype, *IDH* wildtype glioblastoma (GBM), has a median
119 overall survival rate of roughly 1.2 years[3]. The last notable increase in survival coincided with
120 introducing the chemotherapy agent temozolomide[4].

121 Polymorphisms within DNA repair genes are strong candidates for a functional role in
122 carcinogenesis. These genes code for redundant and complex repair mechanisms that identify
123 and repair errors in our genome induced by internal or external forces to ensure proper DNA base
124 modification. Proper DNA repair plays a fundamental role in preventing genomic instability in
125 healthy cells, where instability has been shown to drive tumor development and progression in
126 almost all cancers[5,6]. Tumor cells rely heavily on DNA repair mechanisms to abate further
127 damage, such as those from oxidative and replicative stress[7].

128 Some anti-cancer drugs also rely on the overwhelming damage of tumor DNA as a form
129 of treatment, and DNA repair mechanisms can inhibit effectiveness and mediate treatment
130 resistance. Temozolomide (TMZ), a chemotherapy treatment for patients with glioma, is an

131 alkylating agent that induces DNA damage through the methylation of purine bases within tumor
132 cells, creating genomic instability upon replication. Along with the O^6-methylguanine-DNA-
133 methyltransferase (*MGMT*) gene, mismatch repair (MMR), and base-excision repair (BER),
134 pathways play critical roles in TMZ's mechanism of action by identifying and potentially removing
135 the alkylated bases, leading to either cell repair or cell death[8]. TMZ has become standard-of-care
136 in treating glioblastomas, with maximum safe surgical resection and adjuvant radiation therapy.
137 TMZ has also been used to treat lower-grade gliomas (grades 2 and 3), but its utility has been
138 controversial. Treatment with temozolomide is associated with hypermutation at the time of
139 progression in patients with primary grade 2/3 *IDH* mutant glioma, with the tumors exhibiting
140 thousands of new coding mutations, including somatic defects in DNA repair genes, that generate
141 a distinctive TMZ-induced hypermutation signature[9,10]. Cells that lack MMR function cannot
142 recognize alkylated bases, leading to unrepaired DNA alterations distributed throughout the
143 genome[11]. One study observed that amongst 82 cases, roughly 57% of *IDH* mutant LrGGs treated
144 with TMZ experienced hypermutation at recurrence, associated with shorter survival[11]. Although
145 there is still a clinical knowledge gap in characterizing which markers can predispose patients to
146 TMZ-induced hypermutation, recent findings reveal that somatic *MGMT* promoter methylation
147 levels in newly diagnosed LrGGs are predictive of hypermutation at recurrence and may serve as
148 a prospective biomarker to inform clinical decision making[12]. However, beyond *MGMT*, mention
149 the reasons why some patients develop hypermutation following TMZ treatment while others may
150 not remain unknown. The impact of germline polymorphisms in MMR and BER pathway genes
151 on TMZ efficacy/toxicity is poorly defined.

152 Many germline polymorphisms, particularly SNPs, in genes involved in DNA repair
153 pathways, have been associated with altered genetic predispositions to cancers, such as glioma,
154 colorectal, lung, breast, and others[13–16]. Since glioma patients are susceptible to TMZ-induced
155 hypermutation, we hypothesized that polymorphisms in BER/MMR genes may be associated with
156 altered glioma survival outcomes. To test this hypothesis, we conducted a large association study

157 of 1393 SNPs across 22 BER/MMR-associated genes and overall survival in 2078 adults with
158 common glioma subtypes and further analyzed for TMZ-specific survival effects.
159
160 **Methods:**

161
162 ***Ethics:***
163 Collection of patient samples and associated clinicopathological information was
164 undertaken with written informed consent and relevant ethical review board approval at the
165 respective study centers in accordance with the tenets of the Declaration of Helsinki. Specifically
166 informed consent and ethical board approval was obtained from the UCSF Committee on Human
167 Research (USA) and the Mayo Clinic Office for Human Research Protection (USA). The diagnosis
168 of glioma (ICDO-3 codes 9380-9480 or equivalent) was established through histology and somatic
169 molecular markers in all cases in accordance with World Health Organization guidelines.
170
171 ***Study Populations:***
172 We analyzed three glioma datasets assembled based on genotyping platform and study
173 population for a total sample size of 2078 cases with available genotyping, molecular subtyping
174 (*IDH* mutation, 1p/19q codeletion status), tumor grade, treatment, and survival information. The
175 first set, "AGS-Mayo", included 875 cases from the University of California, San Francisco (UCSF)
176 Adult Glioma Study (AGS) and Mayo Clinic who were genotyped on the Illumina OncoArray, as
177 previously described[17–20]. The second dataset, "AGS-i370", included 518 cases from the UCSF
178 Adult Glioma study genotyped on the Illumina HumanHap370duo panel[19]. The third dataset
179 included 685 glioma cases from The Cancer Genome Atlas (TCGA), with available molecular data
180 genotyped on the Affymetrix 6.0 array. Subtype information for the TCGA dataset (*IDH* mutation,
181 1p/19q codeletion status, *TERT* mutation, primary diagnosis, grade) was downloaded from

182 Ceccarelli et al. 2016[21]. These datasets data quality control, imputation, and sample selection
183 details have been previously described[22].
184 For all analyses, samples were separated into major subtypes using prognostically
185 significant molecular markers (presence/absence of somatic *IDH* mutation and/or 1p/19q
186 chromosomal arm codeletion) and primary tumor grades[3]. In total, eight groups of glioma cases
187 were separately analyzed in the present study, most coarsely separated by *IDH* mutation, and
188 then further stratified by 1p/19q codeletion status and tumor grade (full description available in
189 **Figure 1**). Individuals with unknown grade or *IDH* status were excluded from all analyses.
190 Recorded temozolomide usage (yes/no) during first round of treatment was available for all glioma
191 cases from the UCSF AGS and Mayo Clinic datasets. Glioma cases included from the TCGA
192 were generally treatment naïve.
193
194 ***Study Genes:***
195 We conducted a literature review on DNA repair and temozolomide, specifically on BER
196 and MMR pathways, to identify genes known or suspected to play a role in these mechanisms.
197 We searched PubMed, Web of Science, Scopus, and Embase until April 2, 2023. Search
198 keywords included: "base-excision repair," "germline mutation," "hypermutation," "glioma,"
199 "mismatch base repair," "single nucleotide polymorphism," "temozolomide," "TMZ mechanism of
200 action," and their synonyms. Candidate genes were included regardless of any reported
201 relationship with TMZ. This search identified the following 22 candidate genes: *MSH2*, *MSH3*,
202 *MSH4*, *MSH5*, *MSH6*, *MLH1*, *MLH3*, *PMS1*, *PMS2*, *XRCC1*, *XRCC3*, *ERCC1*, *ERCC2*, *LIG1*,
203 *LIG3*, *LIG4*, *APEX1*, *ADPRT*, *MBD4*, *POLB*, *OGG1*, and *MUTYH*.
204
205 ***SNP selection:***
206 This study analyzed a total of 1393 SNPs with minor allele frequency (MAF) >0.01
207 (amongst European populations CEU, TSI, FIN, GBR, and IBS from the 1000 Genomes Project[23])

208 which were available on all three arrays within the regions 1kb upstream or downstream of any of
209 the 22 candidate DNA repair genes. To properly adjust for multiple testing, the quantity of
210 putatively independent SNP signals was estimated using a linkage disequilibrium (LD) r^2>=0.10
211 (within European populations) filter for SNPs within each chromosome, resulting in 259 effectively
212 independent SNPs across the 22 genes. This study therefore used a Bonferroni-adjusted p-value
213 threshold of 0.05/259 = 0.00019 to determine strict statistical significance for all survival analyses.
214 A less stringent threshold of p<0.005 was used to highlight potentially important associations.
215
216 ***Statistical Analysis*:**
217
218 *Overall survival SNP analysis:*
219 We evaluated the relationship between germline genetic variants within DNA repair genes
220 and overall survival outcomes of adults with subtype-specific gliomas. Single SNP survival
221 analyses were conducted using SPACox[24] using Cox proportional hazards (PH) models with
222 adjustments for age at diagnosis, sex, primary tumor grade, history of radiotherapy (binary),
223 history of chemotherapy (binary), year of primary diagnosis, and the top 10 genetic ancestry
224 principal components (PCs). No treatment or year of diagnosis information was available for the
225 TCGA dataset. SNPs were modeled as continuous dosage values. Analyses were conducted
226 separately for each of the three glioma datasets (AGS-Mayo, AGS-i370, TCGA) and results were
227 combined in a fixed-effects meta-analysis using the software METAL[25] to account for study-
228 specific biases. SNP associations with p<0.00019 were considered statistically significant after
229 multiple testing adjustments, and associations with p<0.005 were considered suggestively
230 important. Within each subtype, SNPs where less than 10 cases carried any copy of the minor
231 allele across the three studies were removed from consideration.
232
233 *SNP-Temozolomide interaction analysis:*

234 We aimed to identify germline genetic polymorphisms in the nominated MMR/BER
235 pathway genes described above. These may interact with temozolomide and alter survival
236 outcomes only in adults with glioma treated with the drug. Cox PH regression models with an
237 included SNP-TMZ interaction term were fit using gwasurvivr[26] with further adjustments for age
238 at diagnosis, sex, tumor grade, radiotherapy, diagnosis year, data source (UCSF AGS or Mayo
239 Clinic), and the top 10 genetic PCs:

240 $$\lambda(t, X) = \lambda_0(t)\exp(\beta_1 \times \text{SNP} + \beta_2 \times \text{TMZ} + \beta_3 \times \text{TMZ} \times \text{SNP} + \text{adjustment covariates})$$

241 Associations were conducted separately for samples typed on the i370 and Oncoarray chips, and
242 overall effect sizes were estimated in a fixed-effects meta-analysis.

243 For each suggestively important SNP-TMZ interaction (p<0.005 for β_3), we partitioned
244 individuals into four groups according to temozolomide usage and presence of the minor allele,
245 and performed unadjusted Kaplan-Meier (KM) survival analysis on these groups to further confirm
246 if the candidate SNP was explicitly associated with altered survival outcomes only in those who
247 received temozolomide, with no survival effect between the two non-temozolomide groups. Only
248 SNPs with at least ten individuals in each group were included for KM analysis. SNPs with a KM
249 p-value p<0.05 in the temozolomide users' group were retained. Amongst this set, SNPs with a
250 KM p-value p<0.05 in the non-temozolomide users' group were discarded, as there was
251 insufficient evidence to conclude that survival effects of the polymorphism were temozolomide-
252 specific.

253

254 *DNA repair polymorphism burden analysis:*

255 DNA repair mutational burden scores (DRBS) were constructed to obtain estimates of the
256 cumulative effect of polymorphisms on survival in each of the glioma subtypes. For each subtype
257 with more than one suggestively important SNP association, highly correlated SNPs were filtered
258 out (linkage disequilibrium, LD, r^2 >0.5, within 500kb), preferentially selecting those with the lowest
259 p-value amongst suggestive variants (p<0.005). For each individual in the subtype, a DRBS was

260 calculated as the integer count of effect polymorphisms amongst the set suggestive SNPs. For a
261 subtype with k suggestive SNP associations, the score takes integer values in {0,1,...,k}. The
262 effect alleles were chosen such that the direction of effect was consistent across all SNPs. Cox
263 proportional hazard models adjusting for age, sex, grade, data source, and the top 10 genetic
264 PCs were used to determine the statistical significance of both the continuous DRBS and the
265 binary >0 versus 0 DRBS on survival outcomes. Significant score associations were determined
266 using the above multiple testing adjusted p-value threshold of p<0.00019.
267
268 **Results**:
269 ***Sample Population:***
270 **Table 1** provides a descriptive summary of the 2078 included samples with available
271 genotyping/covariate information. Sample summaries for each analyzed subgroup can be found
272 in **Supplementary Tables S1-S8**.
273
274 ***DNA repair SNP associations with overall survival:***
275
276 *IDH mutant glioma:*
277 The median survival time for the 813 included individuals with an *IDH* mutant glioma
278 (grades 2,3, or 4, and regardless of 1p/19q codeletion) was 11.3 years from primary diagnosis.
279 Of the 1393 SNPs analyzed for association with *IDH* mutant survival, none reached strict
280 statistical significance (p<0.00019), however polymorphism in *MLH1* had suggestively important
281 (meta p<0.005) associations with survival. The top tagging SNP of *MLH1*, based on p-value was
282 an intron variant, rs4647218-G (Hazard ratio, HR=0.81, 95% confidence interval: 0.61-0.95,
283 p=0.0035). Carriers of two copies of the G allele had a median survival time of 14.3 years,
284 compared to 11.2 and 10.5 years for one or zero copies of G, respectively.
285

286 *Grades 2/3 IDH mutant glioma:*

287 We next measured the survival associations of the MMR/BER DNA repair variant SNPs
288 within 724 grade 2/3 *IDH* mutant gliomas (regardless of 1p/19q codeletion). The median survival
289 for individuals with this tumor type was 12.2 years from initial surgery-confirming diagnosis. The
290 analysis of 1393 SNPs resulted in no significant associations after Bonferroni multiple testing
291 correction (meta p<0.00019), but SNPs in 5 genes had suggestively (meta p<0.005) important
292 associations (**Table 2, Supplementary File 1**). Like the overall *IDH* mutant group, the strongest
293 association was rs4647218-G in *MLH1* (HR=0.73, 95% CI: 0.61-0.88, p=0.0014), where carriers
294 of two copies of the G allele had a median survival of 16.1 years, compared to 13.3 and 11.4 for
295 carriers of one or zero copies, respectively. A set of SNPs tagged by rs4949677-A in *MSH4* on
296 chromosome 1 were associated with increased risk of early death (HR=1.33, 95% CI: 1.07-1.65,
297 p=0.0018). Carriers of the A/A pair had a median survival time of only 9.2 years, compared to
298 12.8 and 14.4 for A/T and T/T carriers. A single rare SNP in *MSH3* was also associated with poor
299 survival outcomes (rs146901049-A, HR=3.28, 95% CI: 1.70-6.33, p=0.0029), with carriers of the
300 A/G pair (n=21) having a dismal median survival time of only 4.7 years, compared to 13.3 for
301 carriers of the G/G pair (n=703). However, the low number of individuals carrying the allele should
302 be considered. Lastly, the G allele of rs3219489, a missense variant in *MUTYH* on chromosome
303 1, was suggestively associated with increased survival times (HR=0.69, 95% CI: 0.54-0.88,
304 p=0.0034). Carriers of any copy of the G allele (n=304) had median survival times of 16.9 years,
305 compared to 11.3 years for those with the C/C genotype (n=420).

306 To evaluate the effect of the co-occurrence of these polymorphisms, we constructed a
307 DRBS, where individuals were grouped based on the quantity of effect polymorphisms (k=4
308 presented above) they harbor (0 to 4 sites with the 'survival-decreasing' polymorphism). We
309 defined the four effect polymorphisms as: rs4647218-AA, rs4848677-AA, rs146901049-A,
310 rs3219489-CC. In a multivariate cox model adjusting for age, sex, grade, data source, and the
311 top 10 genetic PCs, the per-unit increase in the DRBS was significantly associated with decreased

312 survival outcomes (HR=1.52, 95% CI: 1.26-1.83, $p=1.3x10^{-5}$). Individuals carrying 3 of the
313 reported polymorphisms (n=11) had a median survival time of 3.4 years, compared to 9.3 years
314 for those carrying 2 polymorphisms (n=146), 12.8 years for those carrying only 1 polymorphism
315 (n=374), and 20.4 years for those carrying none of the four (n=193) (**Figure 2**). In a similarly
316 parameterized cox model, individuals having a polymorphism at any of the four sites described
317 above (n=531) had decreased survival times (median 11.3 years) as compared to those carrying
318 none of the four polymorphisms (HR=1.80, 95% CI: 1.26-2.59, p=0.0014). No substantial
319 differences between WHO grade or age at diagnosis were observed between DRBS groups.
320
321 *Grades 2/3 IDH mutant 1p/19q codeleted glioma:*
322 The median survival of the included 279 individuals with lower grade *IDH* mutant 1p/19q
323 codeleted gliomas was 17.5 years from diagnosis. A single SNP in *MSH2* had suggestive effects
324 (p<0.005) on survival outcomes. The intron variant rs12615967-G in *MSH2* was associated with
325 poor prognosis (HR=2.04, 95% CI: 1.22-3.33, p=0.0032), where carriers of any copy of the G
326 allele (n=69) had a median survival of 11.3 years, compared to 17.9 years for those with the
327 common A/A genotype (n=210).
328
329 *Grades 2/3 IDH mutant 1p/19q non-codeleted glioma:*
330 The median survival of the 418 individuals with grade 2/3 *IDH* mutant gliomas without
331 somatic 1p/19q codeletion was 10.3 years from diagnosis. SNPs in *MSH4*, *ERCC2*, and *ERCC1*
332 were suggestively associated (p<0.005) with overall survival. A set of SNPs in an LD block tagged
333 by rs5745366-T in *MSH4* were suggestively associated with improved survival (HR=0.33, 95%
334 CI: 0.15-0.72, p=0.0010) where carriers of the C/T genotype (n=26) had a median survival of 13.6
335 years compared to 9.5 years for those with the common C/C genotype (n=392). Two SNPs in
336 *ERCC2*, rs3916804-C and rs3916788-C (LD r^2=0.24) were associated with altered survival. The
337 first, rs3916804-C, an intron variant was suggestively associated with improved survival

338 (HR=0.44, 95% CI: 0.28-0.70, p=0.0027), where individuals with the C/T or C/C genotype (n=98)
339 lived a median 13.0 years post-surgery compared to 8.5 years for those with the T/T genotype
340 (n=320). The second *ERCC2* SNP, rs3916788-C was suggestively associated with improved
341 survival times (HR=0.62, 95% CI: 0.47-0.83, p=0.0033), where carriers of the C/C genotype
342 (n=66) had the longest survival, with 13.0 years post diagnosis, carriers of the C/A genotype
343 (n=208) had a survival of 11.1 years, and carriers of the A/A genotype (n=144) had a median
344 survival time of only 6.9 years. Lastly, a set of four SNPs in *ERCC1*, tagged by intron SNP
345 rs144057986-C, were associated with improved survival outcomes (HR=0.58, 95% CI: 0.42-0.82,
346 p=0.0045). Individuals carrying any copy of the C polymorphism (n=155) had a median survival
347 of 13.3 years (95% CI: 11.1-NA), compared to only 8.0 years (95% CI: 6.6-10.5) for those with
348 the T/T genotype (n=263).

349 To construct a DRBS, individuals were grouped based on the quantity of polymorphisms
350 (k=4 presented above) they harbor (0 to 4 sites with 'survival-increasing' polymorphism).
351 Individuals carrying 3 or 4 of the reported polymorphisms (n=40) had a median survival time of
352 14.5 years, compared to 11.5 years for those carrying 2 polymorphisms (n=147), 9.7 years for
353 those carrying only 1 polymorphism (n=135), and 6.4 years for those carrying none of the four
354 (n=96) (**Figure 3**). In a multivariate cox model adjusting for age, sex, grade, data source, and the
355 top 10 genetic PCs, individuals having a polymorphism at any of the four sites described above
356 (n=322) had significantly improved survival times as compared to those carrying none of the four
357 polymorphisms (n=96) (HR=0.47, 95% CI: 0.33-0.69, $p=8.07 \times 10^{-5}$). In a similarly parameterized
358 cox model, the per-unit increase in the continuous DRBS was also significantly associated with
359 improved survival outcomes (HR=0.67, 95% CI: 0.56-0.80, $p=8.9 \times 10^{-6}$). No differences were
360 observed between DRBS groups and age of diagnosis or WHO group.

361

362 *Grade 4 IDH mutant glioma:*

363 The median survival time for the 83 included individuals with grade 4 *IDH* mutated glioma
364 tumors was 2.5 years. There were no significant or suggestively important SNP associations with
365 survival in this group, likely due to small sample size.
366
367 *IDH wildtype glioma:*
368 The median survival for the 1265 included individuals with an *IDH* wildtype glioma was
369 16.1 months from time of primary diagnosis. None of the 1393 SNPs reached strict significance
370 after multiple testing correction, however 3 SNPs in LD in *MSH3* had $p<0.0005$ associations with
371 *IDH* wildtype survival. The top polymorphism, rs6151863-A, an intron variant in *MSH3* on
372 chromosome 5, was associated with an increase in overall survival (HR=0.67, 95% CI: 0.50-0.89,
373 p=0.00093), where individuals with any copy of the A allele (n=56) had a median survival time of
374 20.5 months compared to 15.7 for those with the common G/G genotype (n=1209).
375
376 *Grade 2/3 IDH wildtype glioma:*
377 A total of 222 individuals with grade 2/3 IDH wildtype gliomas were included in this
378 analysis, with a median overall survival time of 22.4 months. A single SNP in *MSH6*, rs62139913-
379 A, was suggestively associated with improved overall survival times (HR=0.26, 95% 0.09-0.76,
380 p=0.0046). Carriers of any copy of the A allele (n=11) had a median survival of 6.5 years ((5% Ci:
381 2.0-NA), compared to 1.8 years (95% CI: 1.6-2.2) for the 211 individuals with the G/G genotype.
382
383 *Grade 4 IDH wildtype glioma:*
384 The median survival for the 1043 included adults with grade 4 *IDH* wildtype gliomas was
385 15.4 months from initial diagnosis. There were no significant or suggestively important SNP
386 associations with survival in this group, however the MSH3 association observed in *IDH* wildtype
387 overall gliomas is only moderately attenuated when subsetting to grade 4, (rs6151863-A,

388 HR=0.73, 95% CI: 0.54-0.98, p=0.0064), where carriers of any copy of the A allele had a median
389 survival of 19.1 months, compared to only 15.2 for those with the G/G genotype.
390
391 ***Quantitative Trait Loci analysis:***
392 Using the Genotype-Tissue Expression (GTEx) project (Analysis release v8) and paired
393 genotype/RNA-seq data from the TCGA, we investigated expression quantitative trait loci (eQTL)
394 associations for each SNP from **Table 2**. Within GTEx, the rs4949677-A polymorphism was
395 associated with increased *MSH4* expression in the cerebellum (p=2.53e-11) amongst many other
396 tissues (**Figure 4A**). We observed that the eQTL effect remained within tumor samples of 323
397 grade 2/3 *IDH* mutant glioma cases using paired tumor RNAseq and germline genotype data
398 (**Figure 4B**). The rs146901049-A polymorphism, while located within the *MSH3* gene, was
399 associated with increased *DHFR* expression in the cortex (p=3.37e-5), an adjacent gene which
400 shares a promoter region (**Figure 4C**). The association was directionally consistent, but not
401 significant (p=0.36), amongst tumor samples of grade 2/3 IDH mutant glioma cases from the
402 TCGA (not reported). The rs3219489-G SNP was a strong splicing QTL for *MUTYH* within the
403 brain (p=1.51e-12), where the G allele was associated with significantly decreased intron excision
404 ratios (**Figure 4D**). Bulk tissue gene expression for *MUTYH* across 54 tissue types revealed that
405 the gene is most highly expressed in the brain (cerebellum) (**Figure 4E**). Amongst the remaining
406 SNPs in **Table 2,** there were no significant QTL associations within the brain within GTEX, or
407 within associated tumor types in TCGA.
408
409 ***SNP-TMZ Interaction Analysis:***
410 A total of 1393 glioma cases (480 *IDH* mutant, 913 *IDH* wildtype) from the UCSF AGS and
411 Mayo Clinic datasets with known temozolomide exposure during first round of treatment were
412 included. **Table 3** provides a demographic breakdown of each group separated by temozolomide
413 usage.

414 In this section, we applied a two-stage analysis: For stage 1, we performed a Cox PH
415 survival analysis on each of the 1393 SNPs with an included TMZ x SNP interaction term
416 (significance threshold $p<0.005$). For stage 2, for all SNPs with significant TMZ-survival
417 interaction, we calculated Kaplan Meier survival curves for the SNP of interest, separately for
418 those with TMZ treatment and those without, to filter for TMZ-specific survival associations, using
419 the log-rank p-values. All results which passed the 2-stage analysis are summarized in **Table 4**,
420 along with temozolomide and non-temozolomide specific Cox model associations. The full results
421 set for all considered glioma subtypes are available in **Supplementary File 2**.
422
423 *IDH mutant gliomas:*
424 The median overall survival time for the 263 included IDH mutant glioma cases with record
425 of temozolomide during first course of treatment was 10.9 years (95% CI: 9.2-NA), compared to
426 13.1 years (95% CI: 11.3-17.5) for the 217 without temozolomide. 9 SNPs across four genes
427 reached the threshold of $p<0.005$ for an interactive survival effect with temozolomide. Of those
428 11, two SNPs had significant (KM log rank $p<0.05$) associations with survival outcomes within
429 case treated with temozolomide, and no association ($p>0.05$) with survival for those without
430 temozolomide (**Table 4**).
431 The first, rs1540354-T, an intron variant in *MLH1*, was associated with decreased overall
432 survival only for cases receiving temozolomide (interaction p-value = 0.0044, TMZ KM p-value =
433 0.00047, non-TMZ KM p-value = 0.71, TMZ-only Cox model: HR=1.70, 95% CI: 1.17-2.47,
434 p=0.0058 / non-TMZ-only Cox model: HR=0.88, 95% CI:0.55-1.42, p=0.60). Among the 263
435 temozolomide treated individuals, those with T/T or A/T genotypes (n=84) had an overall survival
436 time of 6.3 years (95% CI: 5.5-9.2), contrasting to 13.7 years (95% CI:11.3-17.3) for the 179 with
437 the A/A genotype. The untreated group had indistinguishable survival times: carriers of the T
438 polymorphism (n=63) lived a median 13.8 years after diagnosis, compared to 13.1 years for those
439 with the A/A genotype (n=153) (**Figure 5A**).

440 The second variant, rs71636247-G, an intron in *MSH3* was also associated with
441 decreased survival amongst cases treated with temozolomide (interaction p-value =0.0045, TMZ
442 KM p-value = 0.025, non-TMZ KM p-value = 0.73, TMZ-only Cox model: HR=1.93, 95% CI: 1.14-
443 3.27, p=0.015 / non-TMZ-only Cox model: HR=0.74, 95% CI: 0.40-1.38, p=0.35). The median
444 survival for individuals treated with temozolomide carrying any copy of the G allele (n=28) was
445 6.2 years, compared to 12.2 years for the 235 with the A/A genotype. For those without
446 temozolomide treatment, the median survival times were 12.2 and 13.2 years for those with
447 (n=24) and without (n=193) any copy of the G allele (**Figure 5C**).
448 In constructing a DRBS using the two SNPs, individuals were grouped based on the
449 quantity of polymorphisms (k=2 presented above) present in their germline DNA (0 to 2 sites with
450 polymorphism). Amongst individuals with temozolomide, in a multivariate cox model adjusting for
451 age, sex, grade, and the top 10 genetic PCs, a per unit increase in the DRBS was associated with
452 decreased survival probability (HR=2.23, 95% CI: 1.55-3.20, $p=1.56x10^{-5}$). Those with both
453 polymorphisms (n=7) had a median survival time of 4.3 years, those with only one (n=98) survived
454 a median 7.0 years after diagnosis, and of the n=158 with neither polymorphism, there have not
455 been enough events to determine a median time (NA, 95% CI: 11.3-NA). For those not treated
456 with temozolomide during first course of treatment, in an identically parameterized cox model, the
457 per-unit increase in the DRBS was not associated with survival outcomes (HR=0.90, 95% CI:
458 0.63-1.28, p=0.56). Median survival times were NA, 11.8, and 13.6 years for those with DRBS
459 scores of two, one, and zero, respectively (**Figure 5E**).
460
461 *Grades 2/3 IDH mutant gliomas:*
462 Amongst the grade 2/3 *IDH* mutant cohort of 410 cases, 213 had known temozolomide
463 treatment, and 197 had confirmed absence of temozolomide treatment. The unadjusted median
464 survival times were 13.4 years for those with treatment (95% CI: 10.9-17.3) and 15.2 years for
465 those without (95% CI: 12.8-18.2). 2 SNPs in *XRCC1* reached significance (p<0.005) in stage 1.

466 Neither passed stage 2 filters in separate KM analyses for cases with and without TMZ treatment
467 (Supplemental Table 2).
468
469 *Grades 2/3 IDH mutant 1p/19q codeleted gliomas:*
470 The median survival time for grade 2/3 *IDH* mutant gliomas was further 1p/19q codeletion
471 was 17.5 years for those treated with temozolomide (n=83), and 17.9 for those without the
472 treatment (n=69). Nine SNPs across four genes had a suggestively important TMZ-SNP
473 interaction term (**Supplementary File 2**). Of those, four SNPs in *MSH2*, tagged by lead variant
474 rs6739909-G (Table 4), passed stage 2 Kaplan-Meier filters. Polymorphism rs6739909-G was
475 associated with decreased survival amongst individuals treated with temozolomide (interaction p-
476 value = 0.0045, TMZ KM p-value = 0.030, non-TMZ KM p-value = 0.13, TMZ-only Cox model:
477 HR=3.77, 95% CI: 1.48-9.58, p=0.0054 / non-TMZ-only Cox model: HR=0.38, 95% CI: 0.11-1.39,
478 p=0.14). Individuals treated with temozolomide with the G polymorphism (n=19) had a median
479 survival time of 11.3 years, compared to 20.4 years for the 64 treated individuals with the T/T
480 genotype. Individuals without temozolomide as first course of treatment had median survival times
481 of 20.1 and 16.5 years, for carriers and non-carriers of the G allele, respectively.
482
483 *Grades 2/3 IDH mutant 1p/19q non-codeleted gliomas:*
484 Median survival time for the 119 individuals treated with temozolomide was 10.2 years,
485 compared to the 11.0 years for the 112 without first-course temozolomide treatment. Two SNPs
486 in the *LIG4* gene (rs3093772-C, rs3093774-T) had suggestively important TMZ interactive effects
487 on survival (**Supplemental Table 2**), however neither passed stage 2 thresholding.
488
489 *Grade 4 IDH mutant gliomas:*
490 This study included only 66 individuals with a grade 4 *IDH* mutant glioma and known
491 presence/absence of temozolomide as first-course of treatment. The 48 with temozolomide

492 treatment had a median survival of 42.7 months, compared to the 20.6 months for those 18
493 individuals without temozolomide. We did not have the sample size necessary to detect
494 differences, particularly in the stage 2 analysis. Full results are however made available in
495 **Supplemental Table 2**.
496
497 *IDH wildtype gliomas*:
498 A total of 913 individuals with *IDH* wildtype glioma and known presence/absence of
499 temozolomide during first-course treatment were included in this analysis. The 685 with known
500 presence of temozolomide had a median survival of 17.8 months, and the 228 with known
501 absence had a median survival of 15.1 months. A total of 26 SNPs passed stage 1 thresholding
502 with interaction p-values < 0.005. Of those 26, only rs1052133-G, a missense variant of *OGG1*
503 (Table 4), passed stage 2 thresholds, and is suggestive of altered survival outcomes for
504 temozolomide treated *IDH* wildtype gliomas (interaction p-value = 0.0012, TMZ KM p-value =
505 0.035, non-TMZ KM p-value = 0.29, TMZ-only Cox model: HR=0.78, 95% CI: 0.67-0.92, p=0.0025
506 / non-TMZ-only Cox model: HR=1.15, 95% CI: 0.85-1.56, p=0.35). For the 685 with known
507 temozolomide treatment, the median overall survival time for carriers of the G polymorphism was
508 18.4 months (n=277), compared to the 16.9 months for C/C genotypes (n=408). For the 228 with
509 known absence, carriers of the G polymorphism had a median survival time of 14.4 months
510 (n=102), compared to 16.1 for those with the C/C genotype (n=126).
511
512 *Grades 2/3 IDH wildtype gliomas:*
513 The 88 individuals with primary grade 2/3 IDH wildtype glioma who received temozolomide
514 during the initial course of treatment had a median survival time of 23.8 months after primary
515 diagnosis. The 61 who did not receive the drug during first course had a median survival of 22.9
516 months. 11 SNPs across two genes (*MSH3* and *LIG4*) reached p<0.005 in stage 1 interaction
517 analysis. Of those, nine SNPs (in a single LD block) in *MSH3* passed stage 2 filters (**Table 4**).

518 The association in *MSH3* was captured by the rs34799825-T polymorphism, an intron
519 variant, where the T allele was associated with drastically improved survival outcomes in the
520 temozolomide group (interaction p-value = 0.00031, TMZ KM p-value = 0.012, non-TMZ KM p-
521 value = 0.18, TMZ-only Cox model: HR=0.52, 95% CI: 0.27-1.01, p=0.053 / non-TMZ-only Cox
522 model: HR=5.19, 95% CI: 2.05-13.12, p=0.00051). For individuals with temozolomide, those with
523 the C/T or T/T genotype (n=20) had a median survival of 35.6 months (95% CI: 25.9-NA),
524 compared to 19.9 months (95% CI: 16.7-75.2) for those with the C/C genotype (n=68). For
525 individuals who did not receive temozolomide in first course of treatment, those with the C/T or
526 T/T genotypes (n=15) had a median survival time of 17.6 months (95% CI: 9.4-173.6), compared
527 to the C/C genotype (n=46) of 23.8 months (95% CI: 16.6-74.16).
528
529 Grade 4 IDH wildtype gliomas:
530 The 597 included individuals with grade 4 *IDH* wildtype and known temozolomide
531 treatment had a median survival of 17.1 months, while the 167 with confirmed absence of
532 temozolomide as first course treatment had a median 13.7 months survival. In a multivariable Cox
533 PH model, temozolomide treatment was associated with altered *IDH* wildtype GBM survival
534 (HR=0.8, 0.66-0.98, p=0.031).
535 Our stage 1 analysis nominated eight SNPs across two genes with p-values less than
536 0.005. Of those eight, rs149630102-T in *MSH2* passed stage 2 thresholding (**Table 4**,
537 **Supplementary File 2**). An intron variant, rs149630102-T was associated with temozolomide-
538 specific survival alterations (interaction p-value = 0.0017, TMZ KM p-value = 0.00015, non-TMZ
539 KM p-value = 0.078, TMZ-only Cox model: HR=1.74, 95% CI: 1.27-2.39, p=0.00052 / non-TMZ-
540 only Cox model: HR=0.41, 95% CI: 0.23-0.75, p=0.0038). For the temozolomide group, carriers
541 of the T allele (n=45) had a median survival of 13.2 months (95% CI: 12.3-16.7), compared to
542 17.6 months (95% CI: 16.5-18.5) for the 552 individuals with the A/A genotype. For the non-

543 temozolomide group, those with the T allele (n=16) had a median survival of 15.8 months (95%
544 CI: 9.6-46.9) compared to 13.7 months (95% 11.3-15.5) for those with the A/A genotype (n=151).
545
546 *Quantitative Trait Loci analysis:*
547 Using the Genotype-Tissue Expression (GTEx) project (Analysis release v8) and paired
548 genotype/RNA-seq data from the TCGA, we investigated quantitative trait loci associations for
549 each of the six suggestive SNPs from **Table 4** within the brain/tumor. Within the cortex,
550 rs1540354-T was an eQTL for RP11-640L9.2 (**Figure 5B**). Within the cerebellum, the same SNP,
551 rs1540354-T, was a splicing QTL for *LRRFIP2*. The rs71636247-G variant was a significant eQTL
552 for *MSH3* in the cerebellum, with the G allele associated with increased expression (**Figure 5D**).
553 Similarly, rs34799825-T was also a significant eQTL for *MSH3* in the cerebellum. Lastly,
554 rs149630102-T was a significant eQTL for *EPCAM* on chromosome 2 within the hypothalamus.
555 No SNPs were significant eQTLs for their associated genes within TCGA subtype-specific tumor
556 samples.
557
558 **Discussion**:
559
560 We investigated the association of 1393 putatively independent SNPs within 22 DNA
561 repair genes with overall survival outcomes of 2078 adults with glioma in a meta-analysis of three
562 datasets with molecular subtype specificity. We further investigated if these germline
563 polymorphisms modified survival specifically for patients treated with temozolomide. We observed
564 statistically significant polygenic associations with both overall survival, and temozolomide-
565 specific outcomes, broadly amongst *IDH* mutant gliomas.
566 Grade 2/3 *IDH* mutant gliomas are comprised of both astrocytoma and oligodendroglioma
567 tumors, both which have generally favorable prognoses in comparison with other glioma
568 subgroups. In this mixed group, we observed 4 polymorphisms (in *MLH1*, *MSH4*, *MSH3*, and

569 *MUTYH*) which, when combined, significantly stratified survival trajectories of individuals within
570 this tumor group. Cases with none of the four risk SNPs had a median survival of nine years
571 longer than those with any of the four. For three of the four polymorphisms, we presented possible
572 functional alterations within the brain. The rs4949677-A polymorphism was associated with
573 elevated *MSH4* expression both in normal brain and tumor tissues. The rs146901049-A
574 polymorphism, while located within the *MSH3* gene, was associated with a decreased expression
575 of *DHFR* within the brain. *DHFR*, highly expressed in gliomas, is a crucial metabolic enzyme in
576 the folic acid signaling pathway, and has been hypothesized to contribute to glioma cell
577 proliferation and to decreased chemosensitivity to temozolomide, and has previously been
578 nominated as a target for therapeutics[27,28]. The results presented here, however, provide
579 evidence that a germline decrease in *DHFR* expression (in normal brain) is detrimental to glioma
580 survival. We found that carriers of the G polymorphism at rs3219489 in *MUTYH* exhibited
581 improved survival compared to CC carriers amongst adults with grade 2/3 *IDH* mutant gliomas.
582 Located at chromosome locus 1p34.3—1p32.1, the *MUTYH* gene encodes for a BER enzyme,
583 MYH glycosylase, responsible for the repair of oxidized guanine incorrectly paired with adenine,
584 one of the most pervasive forms of oxidative damage, by excising the undamaged adenine
585 base[29,30]. Of all tissues available in the GTEx database, *MUTYH* is most highly expressed in the
586 brain (cerebellum). The *MUTYH* rs3219489 C to G polymorphism results in an amino acid change
587 from glutamine to histidine, which evidence from GTEx QTL analysis suggests results in
588 alternative splicing of the *MUTYH* protein in the brain, a likely functional alteration to the BER
589 mechanism and a potential mediator of glioma survival. The rs3219489-G allele has been
590 previously associated with decreased survival amongst lung cancer patients who received
591 platinum-based chemotherapy[31] and increased risk of recurrence in oral squamous cell carcinoma
592 patients receiving chemoradiotherapy[32]. Further germline alterations in *MUTYH* have been most
593 strongly linked with an elevated risk of colorectal cancer via inheritable colon polyposis, primarily
594 amongst individuals with European ancestry[33], and others have been extensively described in

595 terms of their ability to promote somatic mutations, including C>A mutations[34]. Further functional
596 work is needed to understand the role of rs3219489 in promoting somatic mutations and the
597 effects of this potential sQTL in glioma.

598 We identified four germline DNA repair polymorphisms (in *MSH4*, *ERCC1*, and *ERCC2*)
599 which were each associated with increased survival times of individuals with low grade 1p/19q
600 non-codeleted glioma. We demonstrated that the presence of any of these four was significantly
601 associated a nearly double increase in median survival times compared to those with none of the
602 polymorphisms. Based on eQTL analyses, the most common of the four polymorphisms,
603 rs3916788-C allele is shown to downregulate the mRNA expression of *ERCC2* in a variety of
604 tissues, and has been previously implicated with improved overall survival in a study of individuals
605 with esophageal squamous cell cancer[35]. Further, three of the associated polymorphisms are
606 located on the q arm of chromosome 19 (specifically, 19q13.32, an area near a putative glioma
607 suppressor region[36]), a copy of which is commonly lost in glioma tumors. We suspect that
608 disruption of these DNA repair pathways maybe related to the favorable prognosis of 1p/19q co-
609 deleted gliomas, however functional studies are required.

610 The role of temozolomide in treating glioblastomas is well-understood, with a clear
611 improvement in patient outcomes when combined with other standard-of-care treatments. Our
612 study found suggestive evidence of a single DNA repair germline risk factor (rs149630102-T in
613 *MSH2*) modifying this glioblastoma effect. Interestingly, the identified SNP was associated with
614 altered survival in opposite directions, depending on if the group received temozolomide. For the
615 group of grade 4 *IDH* wildtype gliomas without temozolomide, the T allele was suggestively
616 associated with improved survival, whereas the same allele was suggestive of decreased survival
617 for cases who received the drug. This could be related to the repair of TMZ-induced mutagenic
618 adducts and a threshold effect, providing protection in TMZ-naive subjects, and a detriment to
619 TMZ-treated subjects by repairing enough of the TMZ induced mutations to avoid cell death and
620 therefore promoting greater number of somatic mutations in a surviving clone. Using GTEx, we

621 found that rs149630102-T was associated with increased expression of the *EPCAM* gene in non-
622 cancer brain and was also suggestively associated with both *EPCAM* and *MSH2* expression in
623 TCGA glioblastoma tumor samples (not shown). A 2014 study demonstrated an association
624 between increased *EPCAM* expression and decreased grade 4 glioma survival time[37], which is
625 consistent with the line of evidence observed in this study for individuals who received
626 temozolomide. A 2015 study also presented in vivo evidence that decreased *MSH2* levels caused
627 a reduction in response to temozolomide in glioblastoma[38]. When comparing survival curves,
628 those without the rs149630102-T polymorphism and treated with temozolomide had the longest
629 median survival time of all groups, suggesting temozolomide may provide survival benefits only
630 to this group. To confirm this speculation, further work into the link between this polymorphism,
631 *EPCAM*, *MSH2*, other possible functional mechanisms, and the interaction with temozolomide is
632 both warranted and necessary.

633 As is well described by a growing body of literature, temozolomide-induced hypermutation at
634 recurrence is common amongst individuals with lower grade *IDH* mutant glioma, with only *MGMT*
635 promoter methylation level at primary diagnosis as a possible biomarker for hypermutation and
636 general sensitivity to the treatment[12]. While hypermutation and *MGMT* status are unknown in our
637 study participants, we identified two germline polymorphisms, in *MLH1* and *MSH3* genes, which
638 were associated with poor outcomes for individuals with *IDH* mutant glioma treated with
639 temozolomide, and no differential effect in the non-temozolomide group. Both SNPs were
640 significantly associated with altered expression of RNA transcripts in the brain, suggesting likely
641 functional alterations to the mismatch repair mechanisms during temozolomide-induced damage
642 repair of the tumor cells, possibly diminishing the efficacy of the mechanism. A reduction in *MLH1*
643 expression has been previously associated with temozolomide resistance in GBM cell lines[39],
644 consistent with the directional effect (non-significant) of rs1540354-T on *MLH1* expression. While
645 not conducted in glioma, previous cell-line analyses have demonstrated *MSH3* deficiency lead to
646 increased chemosensitivity and cell apoptosis, independent of the role of *MLH1*, in colon

647 carcinoma cells[40]. Our combined evidence of rs71636247-G having associations with both
648 decreased survival and increased *MSH3* expression in the brain remains consistent with this
649 previous observation. We show that the presence of either of these two polymorphisms resulted
650 in drastically decreased median survival times, even with rigorous significance thresholds. As
651 such, these two SNPs warrant further clinical investigation as our results generate the hypothesis
652 that temozolomide may not be an appropriate treatment for individuals harboring these germline
653 polymorphisms, and alternative treatments should be prioritized. Although not included here,
654 these results were replicated with only minor attenuation when excluding all grade 4 *IDH* mutant
655 glioma cases.

656 Resistance to therapeutics continues to be a major cause of treatment failure for individuals
657 with cancer. Anticipating how germline differences can alter response to temozolomide and other
658 possible chemotherapy options can help to influence the course of treatment clinicians take in the
659 care of patients, specifically if treatment should include temozolomide or an alternative drug. In
660 the treatment of *IDH* mutant 1p/19q codeleted oligodendrogliomas, there is controversy whether
661 either temozolomide or procarbazine, lomustine, and vincristine (PCV) in combination with
662 radiation therapy provides better patient outcomes, with the ongoing CODEL clinical trial
663 investigating the impact on progression-free survival for each drug[41]. Also, the recently completed
664 promising clinical trials of the drug, vorasidenib, an *IDH* inhibitor, presents another possible
665 treatment avenue for *IDH* mutant gliomas[42]. As therapeutic options progress, clinicians and
666 patients have more options in the treatment of *IDH* mutant gliomas, identifying predictors of
667 response to treatment, like the SNPs identified in this study, can help guide personalized
668 treatment courses.

669 Our findings should be considered with the following limitations. Our included meta-analysis
670 is from a set of historic cohorts. As such, we lack the full suite of somatic molecular markers to
671 adhere to strict WHO 2021 classification guidelines, and we lack knowledge of *MGMT*
672 methylation, a relatively recently identified and costly phenotype to measure. The cohorts also do

673 not include robust information on the recurrent tumor, including temozolomide-induced
674 hypermutation. Our results may not extend to non-European populations, as the included subjects
675 were chosen to have at least 70% genetically estimated European ancestry. Our study of
676 temozolomide was limited to a binary yes/no history of known drug usage during the first course
677 of treatment; we could not further investigate the role of timing and amounts of temozolomide
678 administered on survival outcomes. Ongoing studies which collect detailed treatment information
679 across the course of the disease will be key to determining if the effects presented in this study
680 are indeed isolated to just first-course treatment. Despite these limitations, this analysis is one of
681 the most extensive studies of glioma germline genetics and survival with available molecular
682 subtyping, long-term follow-up, and temozolomide utilization.

683 This study presents germline polygenic associations in DNA repair pathways which are
684 associated with significant impacts on overall survival and drug response in individuals with
685 specific glioma subtypes. We posit potential mechanisms using quantitative trait loci analysis for
686 a select set of SNPs. While further investigation into both a validation dataset and proper
687 functional studies are needed, this study lays the groundwork for the integration of germline SNPs
688 in determining accurate clinical management and the best therapeutic options in the treatment of
689 gliomas.

690

691 **References**:

692

693 1. Ostrom QT, Price M, Neff C, et al. CBTRUS Statistical Report: Primary Brain and Other
694 Central Nervous System Tumors Diagnosed in the United States in 2015–2019. *Neuro-*
695 *Oncol*. 2022;24(Supplement_5):v1-v95. doi:10.1093/neuonc/noac202

696 2. Weltgesundheitsorganisation. *WHO Classification of Tumours of the Central Nervous*
697 *System*. Revised 4th edition. (Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, eds.).
698 International Agency for Research on Cancer; 2016.

699 3. Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR. Genetic and molecular epidemiology of
700 adult diffuse glioma. *Nat Rev Neurol*. 2019;15(7):405-417. doi:10.1038/s41582-019-0220-2

701 4. Johnson DR, O'Neill BP. Glioblastoma survival in the United States before and during the
702 temozolomide era. *J Neurooncol*. 2012;107(2):359-364. doi:10.1007/s11060-011-0749-4

703 5. Nowell PC. The Clonal Evolution of Tumor Cell Populations: Acquired genetic lability
704 permits stepwise selection of variant sublines and underlies tumor progression. *Science*.
705 1976;194(4260):23-28. doi:10.1126/science.959840

706 6. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability — an evolving hallmark of
707 cancer. *Nat Rev Mol Cell Biol*. 2010;11(3):220-228. doi:10.1038/nrm2858

708 7. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell*.
709 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013

710 8. Singh N, Miner A, Hennis L, Mittal S. Mechanisms of temozolomide resistance in
711 glioblastoma - a comprehensive review. *Cancer Drug Resist*. Published online 2020.
712 doi:10.20517/cdr.2020.79

713 9. Johnson BE, Mazor T, Hong C, et al. Mutational Analysis Reveals the Origin and Therapy-
714 Driven Evolution of Recurrent Glioma. *Science*. 2014;343(6167):189-193.
715 doi:10.1126/science.1239947

716 10. Touat M, Li YY, Boynton AN, et al. Mechanisms and therapeutic implications of
717 hypermutation in gliomas. *Nature*. 2020;580(7804):517-523. doi:10.1038/s41586-020-2209-
718 9

719 11. Yu Y, Villanueva-Meyer J, Grimmer MR, et al. Temozolomide-induced hypermutation is
720 associated with distant recurrence and reduced survival after high-grade transformation of
721 low-grade *IDH* -mutant gliomas. *Neuro-Oncol*. 2021;23(11):1872-1884.
722 doi:10.1093/neuonc/noab081

723 12. Mathur R, Zhang Y, Grimmer MR, et al. MGMT promoter methylation level in newly
724 diagnosed low-grade glioma is a predictor of hypermutation at recurrence. *Neuro-Oncol*.
725 2020;22(11):1580-1590. doi:10.1093/neuonc/noaa059

726 13. Wang LE, Bondy ML, Shen H, et al. Polymorphisms of DNA Repair Genes and Risk of
727 Glioma. *Cancer Res*. 2004;64(16):5560-5563. doi:10.1158/0008-5472.CAN-03-2181

728 14. Lubbe SJ, Di Bernardo MC, Chandler IP, Houlston RS. Clinical Implications of the
729 Colorectal Cancer Risk Associated With *MUTYH* Mutation. *J Clin Oncol*. 2009;27(24):3975-
730 3980. doi:10.1200/JCO.2008.21.6853

731 15. Butkiewicz D, Rusin M, Enewold L, Shields PG, Chorazy M, Harris CC. Genetic
732 polymorphisms in DNA repair genes and risk of lung cancer. *Carcinogenesis*.
733 2001;22(4):593-597. doi:10.1093/carcin/22.4.593

734 16. Silva SN, Tomar M, Paulo C, et al. Breast cancer risk and common single nucleotide
735 polymorphisms in homologous recombination DNA repair pathway genes XRCC2, XRCC3,
736 NBS1 and RAD51. *Cancer Epidemiol*. 2010;34(1):85-92. doi:10.1016/j.canep.2009.11.002

737 17. Eckel-Passow JE, Drucker KL, Kollmeyer TM, et al. Adult diffuse glioma GWAS by
738 molecular subtype identifies variants in *D2HGDH* and *FAM20C*. *Neuro-Oncol*.
739 2020;22(11):1602-1613. doi:10.1093/neuonc/noaa117

740 18. Melin BS, Barnholtz-Sloan JS, Wrensch MR, et al. Genome-wide association study of
741 glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and
742 non-glioblastoma tumors. *Nat Genet*. 2017;49(5):789-794. doi:10.1038/ng.3823

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

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

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

752 22. Guerra G, Kachuri L, Wendt G, et al. The immunogenetics of viral antigen response is
753 associated with subtype-specific glioma risk and survival. *Am J Hum Genet*.
754 2022;109(6):1105-1116. doi:10.1016/j.ajhg.2022.04.011

755 23. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, et al. A global
756 reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
757 doi:10.1038/nature15393

758 24. Bi W, Fritsche LG, Mukherjee B, Kim S, Lee S. A Fast and Accurate Method for Genome-
759 Wide Time-to-Event Data Analysis and Its Application to UK Biobank. *Am J Hum Genet*.
760 2020;107(2):222-233. doi:10.1016/j.ajhg.2020.06.003

761 25. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide
762 association scans. *Bioinforma Oxf Engl*. 2010;26(17):2190-2191.
763 doi:10.1093/bioinformatics/btq340

764 26. Rizvi AA, Karaesmen E, Morgan M, et al. gwasurvivr: an R package for genome-wide
765 survival analysis. Schwartz R, ed. *Bioinformatics*. 2019;35(11):1968-1970.
766 doi:10.1093/bioinformatics/bty920

767 27. Zhao M, Tan B, Dai X, et al. DHFR/TYMS are positive regulators of glioma cell growth and
768 modulate chemo-sensitivity to temozolomide. *Eur J Pharmacol*. 2019;863:172665.
769 doi:10.1016/j.ejphar.2019.172665

770 28. Fawal MA, Jungas T, Davy A. Inhibition of DHFR targets the self-renewing potential of brain
771 tumor initiating cells. *Cancer Lett*. 2021;503:129-137. doi:10.1016/j.canlet.2021.01.026

772 29. Shibutani S, Takeshita M, Grollman AP. Insertion of specific bases during DNA synthesis
773 past the oxidation-damaged base 8-oxodG. *Nature*. 1991;349(6308):431-434.
774 doi:10.1038/349431a0

775 30. Michaels ML, Miller JH. The GO system protects organisms from the mutagenic effect of the
776 spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J Bacteriol*.
777 1992;174(20):6321-6325. doi:10.1128/jb.174.20.6321-6325.1992

778 31. Singh A, Singh N, Behera D, Sharma S. Genetic Investigation of Polymorphic OGG1 and
779 MUTYH Genes Towards Increased Susceptibility in Lung Adenocarcinoma and its Impact

780 on Overall Survival of Lung Cancer Patients Treated with Platinum Based Chemotherapy.
781 *Pathol Oncol Res*. 2019;25(4):1327-1340. doi:10.1007/s12253-017-0372-6

782 32. Senghore T, Chien HT, Wang WC, et al. Predictive value of genetic variants XRCC1
783 rs1799782, APEX1 rs1760944, and MUTYH rs3219489 for adjuvant concurrent
784 chemoradiotherapy outcomes in oral squamous cell carcinoma patients.
785 *Pharmacogenomics J*. 2020;20(6):813-822. doi:10.1038/s41397-020-0170-5

786 33. Nielsen M, Morreau H, Vasen HFA, Hes FJ. MUTYH-associated polyposis (MAP). *Crit Rev*
787 *Oncol Hematol*. 2011;79(1):1-16. doi:10.1016/j.critrevonc.2010.05.011

788 34. Robinson PS, Thomas LE, Abascal F, et al. Inherited MUTYH mutations cause elevated
789 somatic mutation rates and distinctive mutational signatures in normal human cells. *Nat*
790 *Commun*. 2022;13(1):3949. doi:10.1038/s41467-022-31341-0

791 35. Zhang R, Zhou F, Cheng L, et al. Genetic variants in nucleotide excision repair pathway
792 predict survival of esophageal squamous cell cancer patients receiving platinum-based
793 chemotherapy. *Mol Carcinog*. 2018;57(11):1553-1565. doi:10.1002/mc.22877

794 36. Smith JS, Tachibana I, Pohl U, et al. A Transcript Map of the Chromosome 19q-Arm Glioma
795 Tumor Suppressor Region. *Genomics*. 2000;64(1):44-50. doi:10.1006/geno.1999.6101

796 37. Chen X, Ma WY, Xu SC, et al. The overexpression of Epithelial cell adhesion molecule
797 (EpCAM) in glioma. *J Neurooncol*. 2014;119(1):39-47. doi:10.1007/s11060-014-1459-5

798 38. McFaline-Figueroa JL, Braun CJ, Stanciu M, et al. Minor Changes in Expression of the
799 Mismatch Repair Protein MSH2 Exert a Major Impact on Glioblastoma Response to
800 Temozolomide. *Cancer Res*. 2015;75(15):3127-3138. doi:10.1158/0008-5472.CAN-14-3616

801 39. Shinsato Y, Furukawa T, Yunoue S, et al. Reduction of MLH1 and PMS2 confers
802 temozolomide resistance and is associated with recurrence of glioblastoma. *Oncotarget*.
803 2013;4(12):2261-2270. doi:10.18632/oncotarget.1302

804 40. Park JM, Huang S, Tougeron D, Sinicrope FA. MSH3 mismatch repair protein regulates
805 sensitivity to cytotoxic drugs and a histone deacetylase inhibitor in human colon carcinoma
806 cells. *PloS One*. 2013;8(5):e65369. doi:10.1371/journal.pone.0065369

807 41. Jaeckle KA, Ballman KV, Van Den Bent M, et al. CODEL: phase III study of RT, RT + TMZ,
808 or TMZ for newly diagnosed 1p/19q codeleted oligodendroglioma. Analysis from the initial
809 study design. *Neuro-Oncol*. 2021;23(3):457-467. doi:10.1093/neuonc/noaa168

810 42. Mellinghoff IK, van den Bent MJ, Blumenthal DT, et al. Vorasidenib in IDH1- or IDH2-Mutant
811 Low-Grade Glioma. *N Engl J Med*. Published online June 4, 2023:NEJMoa2304194.
812 doi:10.1056/NEJMoa2304194

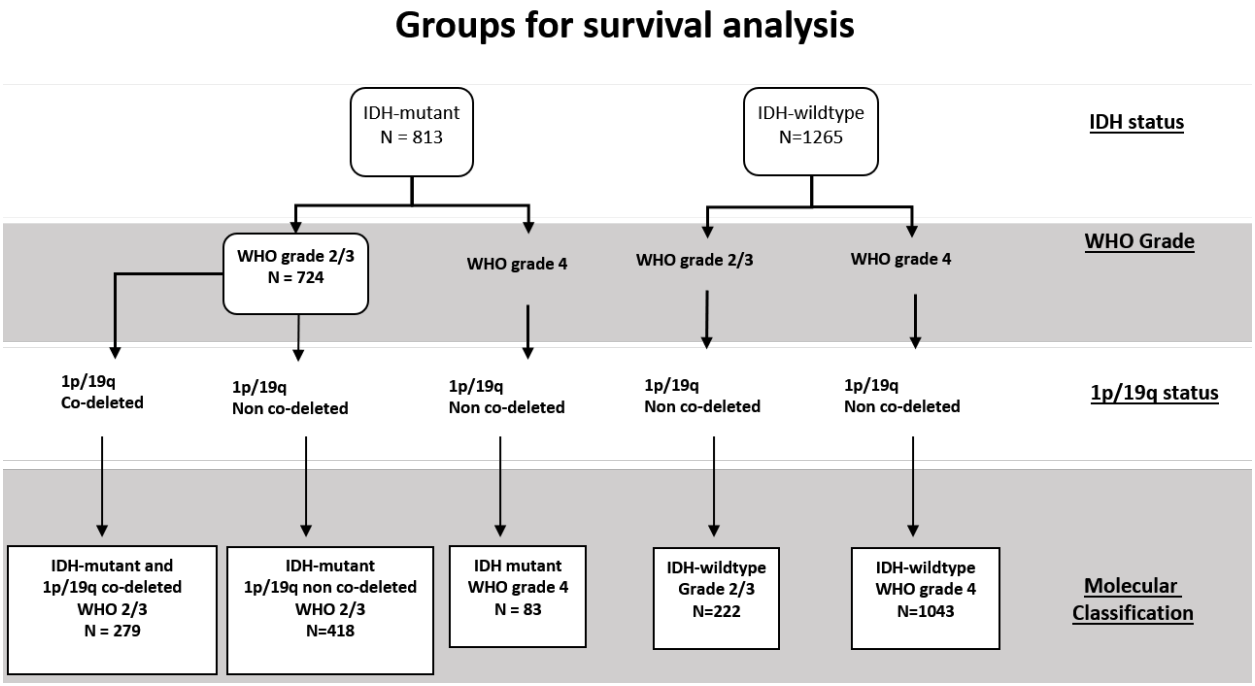
813

814

815 Figures and Tables for '**The impact of germline variants in DNA repair pathways on**

816 **survival and temozolomide toxicity in adults with glioma'**

817



818

819 ***Figure 1: Included glioma subgroups based on prognostically significant molecular***
820 ***markers and tumor grade.*** *Tumor types were coarsely split first by IDH mutation status, and*
821 *then by further by tumor grade at primary diagnosis. Grade 2/3 IDH mutant gliomas were further*
822 *stratified by the 1p/19q chromosomal arm codeletion phenotype. Boxes with included sample*
823 *size are used to highlight the eight subgroups analyzed in this study.*

824

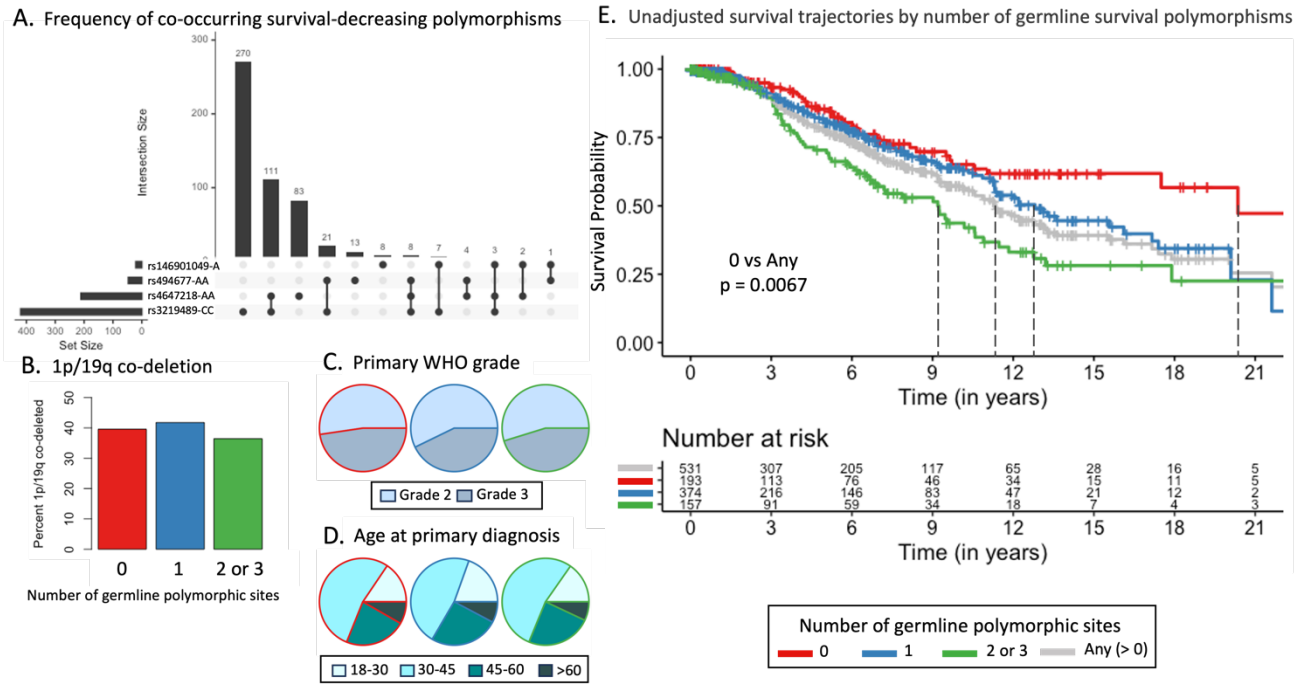


Figure 2: Co-occurrence of four germline DNA repair polymorphisms in MLH1, MSH4, MSH3, and MUTYH, and the associated survival trajectories of individuals with grade 2/3 IDH mutant gliomas. *Individuals were grouped based on the presence of four suggestively important survival SNPs: rs1046901049-A (MSH3), rs494677-AA (MSH4), rs4647218-AA (MLH1), and rs3219489-CC (MUTYH). The effect polymorphisms were chosen such that each had a survival decreasing effect. Possible groups consist of individuals with zero to four of the highlighted polymorphisms. Individuals with 2 or more were grouped due to small group sizes. A) Upset plot highlighting the common co-occurrences of the four suggestive survival SNPs. B) The percentage of cases with 1p/19q codeletion phenotype, by group. C) Pie charts highlighting the distribution of tumor grades within each of the four groups. Border color indicates the group. D) Pie charts highlighting the age distribution of individuals within each group. Border color indicates the group. D) Unadjusted percent survival distributions for adults with Grade 2/3 IDH mutant gliomas, stratified by the count of highlighted DNA repair polymorphisms, demonstrates the associated survival disadvantage of present polymorphisms. A curve highlighting individuals with at least one of the polymorphisms is also included (grey). Dashed vertical lines represent median survival times, measured in years since diagnosis. Cases were collected from the UCSF Adult Glioma Study, the Mayo Clinic, and The Cancer Genome Atlas.*

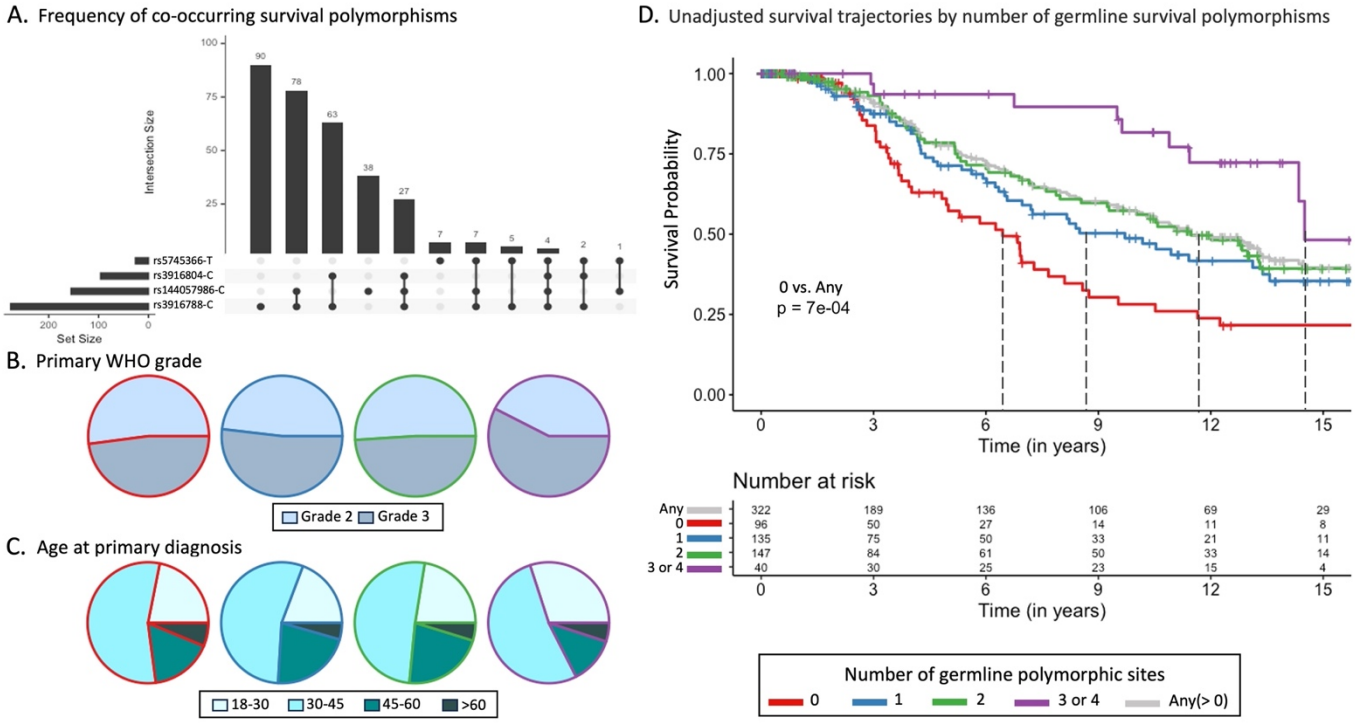


Figure 2: Co-occurrence of four germline DNA repair polymorphisms in MSH4, ERCC2, and ERCC1, and the associated survival trajectories of individuals with grade 2/3 IDH mutant 1p/19q non-codeleted gliomas. *Individuals were grouped based on the presence of four suggestively important survival SNPs: rs5745366-T (MSH4), rs3918604-C (ERCC2), rs144057986-C (ERCC1), rs3916788-C (ERCC2). Groups consist of individuals with zero to four of the highlighted polymorphisms. Individuals with three or four were grouped due to small group sizes. A) Upset plot highlighting the common co-occurrences of the four suggestive survival SNPs. B) Pie charts highlighting the distribution of tumor grades within each of the four groups. Border color indicates the group. C) Pie charts highlighting the age distribution of individuals within each group. Border color indicates the group. D) Unadjusted percent survival distributions for adults with Grade 2/3 IDH mutant 1p/19q non-codeleted gliomas, stratified by the count of highlighted DNA repair polymorphisms, demonstrates the associated survival benefits of accumulated polymorphisms. A curve highlighting individuals with at least one of the polymorphisms is also included (grey). Dashed vertical lines represent median survival times, measured in years since diagnosis. Cases were collected from the UCSF Adult Glioma Study, the Mayo Clinic, and The Cancer Genome Atlas.*

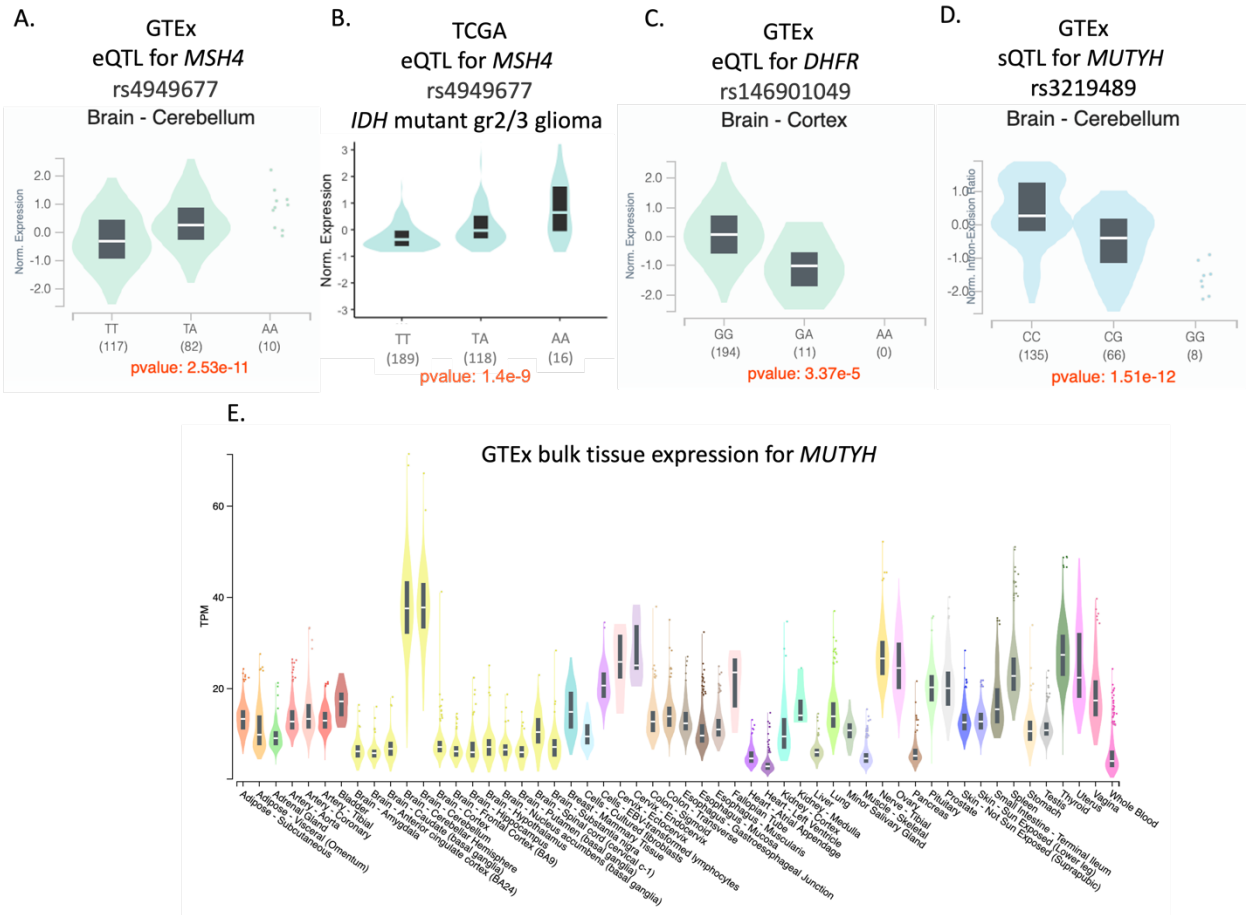


Figure 4: Statistically significant brain-specific QTL associations in cancer-free donors from the Genotype-Tissue Expression (GTEx) project and in tumor tissues from TCGA LGG/GBM projects amongst BER/MMR SNPs with suggestive effects on glioma survival. *We utilized data from GTEx and TCGA to explore functional changes associated with each of the 11 SNPs suggestively associated with altered glioma survival outcomes. Three SNPs had significant QTL associations within normal brain tissue, with one demonstrating the same QTL association within tumor samples. (A) The A allele of rs4949677 is associated with increased MSH4 mRNA expression within the brain. (B) The A allele of rs4949677 is associated with increased MSH4 mRNA expression within tumor tissue of IDH mutant grade 2/3 glioma samples. (C). The A allele of rs1496901049 is associated with decreased expression of DHFR within cancer-free brain cortex samples. (D). The G allele of rs3219489 is associated with a significant decrease in the intron-excision ratio of MUTYH mRNA within healthy brain, thus is a splicing QTL. (E**).** In complement with the result of (D), MUTYH is highly expressed in the cerebellum region of the brain, the highest across all 54 tissue types.*

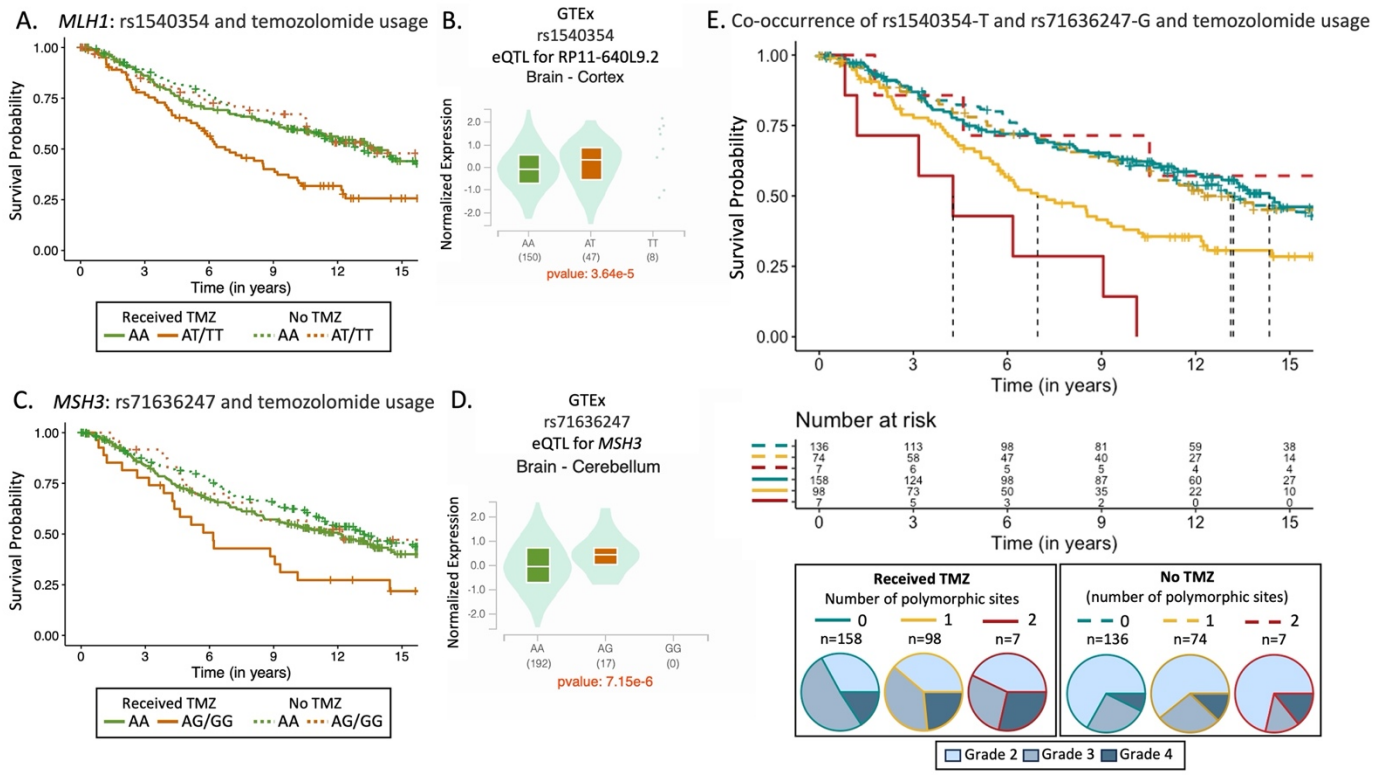


Figure 5: Germline polymorphisms in MLH1 and MSH3 are associated with negative survival outcomes amongst cases with IDH mutant glioma receiving temozolomide during first course of treatment. A) *Unadjusted percent survival distributions for adults with IDH mutant glioma tumors with or without temozolomide treatment further stratified by germline polymorphism rs1540354-T demonstrates the SNP is associated with negative survival outcomes only for cases who were treated with temozolomide.* ***B)*** *GTEx eQTL analysis from cancer-free brain tissues demonstrates rs1540354-T is associated with increased expression of RP11-640L9.2 in the cortex. C) Unadjusted percent survival distributions for adults with IDH mutant glioma tumors with or without temozolomide treatment further stratified by germline polymorphism rs71636247-G demonstrates the SNP is associated with negative survival outcomes only for cases who were treated with temozolomide.* ***D)*** *GTEx eQTL analysis from cancer-free brain tissues demonstrates rs71636247-G is associated with increased expression of MSH3 in the cerebellum.* ***E)*** *Unadjusted percent survival distributions for adults with IDH mutant glioma tumors with or without a history of temozolomide treatment, stratified by the co-occurrence of germline DNA repair SNPs rs1540354-T (MLH1) and rs71636247-G (MSH3) demonstrate that the combination of both SNPs is associated with the worst overall survival outcomes, only for individuals who received temozolomide. The SNPs have no association with survival in the non-temozolomide group. Vertical dashed lines represent median overall survival times, measured in years since primary diagnosis. Pie charts describe the distribution of tumor grades within each of the six groups. All cases were a part of either the UCSF Adult Glioma Study or the Mayo Clinic datasets.*

Table 1: Clinical and molecular summary of included adults with glioma across three datasets

	UCSF AGS	*Mayo Clinic*	*TCGA*
Number of cases	1158	235	685
Age at diagnosis (Median + IQR)	52 (IQR: 20)	50 (IQR: 24)	52 (IQR: 26)
Sex (% Male)	61.92%	62.13%	59.27%
Self-report race (% White)	99.22%	94.89%	NA
IDH			
Wildtype	67.44%	56.17%	51.39%
Mutant	32.56%	43.83%	48.61%
1p/19q			
Codeleted	10.88%	13.19%	18.54%
Non-codeleted	86.44%	75.74%	80.15%
NA	2.68%	11.06%	1.31%
TERT			
Wildtype	29.10%	31.06%	20.44%
Mutant	60.88%	64.68%	18.83%
NA	10.02%	4.26%	60.73%
Grade			
2	20.47%	21.28%	26.28%
3	17.53%	29.36%	30.22%
4	62.00%	49.36%	43.50%
Histology			
Astrocytoma	19.86%	23.40%	20.58%
Oligodendroglioma	12.69%	6.38%	21.02%
Glioblastoma*	62.26%	47.23%	43.50%
Oligoastrocytoma	5.18%	22.98%	14.89%
Diagnosis Year			
Median (min, max, IQR)	2004 (1991, 2012, IQR: 7)	2008 (1991, 2013, IQR: 5)	NA
% of cases diagnosed pre-1999	12.52%	5.11%	NA
% of cases diagnosed pre-2005	52.42%	20.43%	NA
% Censored (no known death event)	16.67%	25.53%	59.71%
% Censored before 2 years	0.26%	13.62%	47.30%
Chemotherapy (% known Chemo)	78.50%	80.00%	NA
Temozolomide (% known TMZ)	69.78%	59.57%	NA
Radiation (% known RT)	83.77%	85.96%	NA

NA: not available information

*Group contains *IDH* wildtype and mutant. World Health Organization 2021 guidelines now classify IDH mutant Glioblastoma as Grade 4 Astrocytoma

907 **Table 2: Multivariate cox proportional hazards analysis of genetic variants in mismatch**
908 **repair and base excision repair genes with overall survival amongst major glioma**
909 **molecular subtypes.**

Gene	Chr.	SNP	Effect Allele	Allele Freq.	HR (95% CI)	*P*
***IDH* mutant glioma**						
MLH1	3	rs4647218	G	0.44	0.79 (0.68-0.93)	0.0034
***IDH* mutant grade 2/3 glioma**						
MLH1	3	rs4647218	G	0.44	0.74 (0.61-0.88)	0.0014
MSH4	1	rs4949677	A	0.26	1.33 (1.07-1.65)	0.0018
MSH3	5	rs146901049	A	0.017	3.28 (1.70-6.33)	0.0029
MUTYH	1	rs3219489	G	0.25	0.69 (0.54-0.88)	0.0034
***IDH* mutant grade 2/3 1p/19q codeleted glioma**						
MSH2	2	rs12615967	G	0.12	2.04 (1.22-3.33)	0.0032
***IDH* mutant grade 2/3 1p/19q non-codeleted glioma**						
MSH4	1	rs5745366	T	0.036	0.33 (0.15-0.72)	0.0010
ERCC2	19	rs3916804	C	0.16	0.44 (0.28-0.70)	0.0027
ERCC2	19	rs3916788	C	0.42	0.62 (0.47-0.83	0.0033
ERCC1	19	rs144057986	C	0.18	0.58 (0.42-0.82)	0.0045
***IDH* wildtype glioma**						
MSH3	5	rs6151863	A	0.026	0.67 (0.50-0.89)	0.0009
***IDH* wildtype grade 2/3 glioma**						
MSH6	2	rs62139913	A	0.027	0.26 (0.09-0.76)	0.0046
***IDH* wildtype grade 4 glioma**						
MSH3	5	rs6151863	A	0.026	0.73 (0.54-0.98)	0.0064

910
911

912 **Table 3: Clinical and molecular summary of included adults with glioma with known**

913 **usage/absence of temozolomide treatment.**

	IDH mutant		***IDH wildtype***	
	Received TMZ	**No TMZ**	**Received TMZ**	**No TMZ**
Number of cases	263	217	685	228
Age at diagnosis (Median + IQR)	40 (IQR: 16)"	39 (IQR: 18)	58 (IQR: 16)	58 (IQR: 18)
Sex (% Male)	60.84%	57.60%	64.82%	58.77%
Self-report race (% White)	98.10%	98.16%	98.25%	100%
1p/19q				
Codeleted	31.56%	31.80%	0.15%	1.75%
Non-codeleted	63.12%	59.91%	97.81%	93.42%
NA	5.32%	8.29%	2.04%	4.83%
TERT				
Wildtype	52.85%	53.00%	16.79%	17.98%
Mutant	36.88%	33.18%	77.23%	69.74%
NA	10.27%	13.82%	5.99%	12.28%
Grade (% of total)				
2	35.36%	64.98%	3.36%	13.16%
3	45.63%	25.81%	9.49%	13.60%
4	19.01%	9.22%	87.15%	73.25%
Histology (% of total)				
Astrocytoma	36.88%	34.56%	11.24%	15.79%
Oligodendroglioma	27.00%	32.72%	0.73%	6.58%
Glioblastoma	19.01%	8.29%	87.15%	73.25%
Oligoastrocytoma	17.11%	24.42%	0.88%	4.39%
Diagnosis Year	2007	2003	2007	2002
Median (min, max, IQR)	(1998, 2013, IQR: 5)	(1991, 2012, IQR: 10)	(2001, 2013, IQR: 5)	(1991, 2013, IQR: 6)
% of cases diagnosed pre-1999	0.76%	29.95%	0%	39.47%
% of cases diagnosed pre-2005	34.60%	67.74%	34.60%	78.95%
% Censored (no known death event)	45.63%	43.78%	3.36%	6.58%
% Censored before 2 years	7.22%	4.61%	0.58%	0.88%
Radiation (% known RT)	72.24%	52.53%	97.52%	87.72%

914

915

916 **Table 4: Association analysis of genetic variants in mismatch repair and base excision**

917 **repair genes with temozolomide-specific survival amongst glioma molecular subtypes.**

Gene	Chr.	SNP	Effect Allele	Allele Freq.	*P**	TMZ cox PH model			Non-TMZ cox PH model		
						KM *P*	HR (95% CI)	*P*	KM *P*	HR (95% CI)	*P*
***IDH* mutant glioma**											
MLH1	3	rs1540354	T	0.18	0.0026	0.00047	1.70 (1.17-2.47)	0.0058	0.71	0.88 (0.55-1.42)	0.60
MSH3	5	rs71636247	G	0.043	0.0045	0.025	1.93 (1.14-3.27)	0.015	0.73	0.74 (0.40-1.38)	0.35
***IDH* mutant grade 2/3 1p/19q codeleted glioma**											
MSH2	2	rs6739909	G	0.11	0.0045	0.030	3.77 (1.48-9.58)	0.0054	0.13	0.38 (0.11-1.39)	0.14
***IDH* wildtype glioma**											
OGG1	1	rs1052133	G	0.23	0.0012	0.034	0.78 (0.67-0.92)	0.0025	0.29	1.15 (0.85-1.56)	0.35
***IDH* wildtype grade 2/3 glioma**											
MSH3	5	rs34799825	T	0.13	0.0003	0.012	0.52 (0.27-1.01)	0.053	0.18	5.19 (2.05-13.12)	0.0005
***IDH* wildtype grade 4 glioma**											
MSH2	2	rs149630102	T	0.040	0.0017	0.00015	1.74 (1.27-2.39)	0.00052	0.078	0.41 (0.23-0.75)	0.0038

918 *: P-value of SNP x TMZ interaction term in Cox PH model

919 KM *P*: Unadjusted Kaplan-Meier log rank p-value

920

921