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 Guan3, Lianne Capistrano3, Allison Hsieh3, Veruna Kalsi3, Jaimie Sallee3, 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 with in 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=1.3x10^-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=8.9x10^-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=1.56x10^-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 logo 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 logo 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:
GG, VG, LC, AH, VK, JS, and SSF conceived of the study. GG and SSF wrote the main drafts of the manuscript. VG, LC, AG, VK, JS, GG, and SSF conducted literature searches. GG, GW, and SSF conducted statistical and computational analyses. GG, SSF, JWT, JLC, AMM, JLW, JEP, RBJ, LK, ERA, and MW advised on result interpretations. The primary data collection involved LM, HMH, AMM, JLW, TR, JEP, RBJ, JKW, and MW. All authors contributed to, reviewed, and approved the final manuscript.

**Introduction/Background:**

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

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

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

Many germline polymorphisms, particularly SNPs, in genes involved in DNA repair pathways, have been associated with altered genetic predispositions to cancers, such as glioma, colorectal, lung, breast, and others\(^{13-16}\). Since glioma patients are susceptible to TMZ-induced hypermutation, we hypothesized that polymorphisms in BER/MMR genes may be associated with altered glioma survival outcomes. To test this hypothesis, we conducted a large association study
of 1393 SNPs across 22 BER/MMR-associated genes and overall survival in 2078 adults with common glioma subtypes and further analyzed for TMZ-specific survival effects.

Methods:

Ethics:

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

Study Populations:

We analyzed three glioma datasets assembled based on genotyping platform and study population for a total sample size of 2078 cases with available genotyping, molecular subtyping (IDH mutation, 1p/19q codeletion status), tumor grade, treatment, and survival information. The first set, “AGS-Mayo”, included 875 cases from the University of California, San Francisco (UCSF) Adult Glioma Study (AGS) and Mayo Clinic who were genotyped on the Illumina OncoArray, as previously described. The second dataset, “AGS-i370”, included 518 cases from the UCSF Adult Glioma study genotyped on the Illumina HumanHap370duo panel. The third dataset included 685 glioma cases from The Cancer Genome Atlas (TCGA), with available molecular data genotyped on the Affymetrix 6.0 array. Subtype information for the TCGA dataset (IDH mutation, 1p/19q codeletion status, TERT mutation, primary diagnosis, grade) was downloaded from
Ceccarelli et al. 2016\textsuperscript{21}. These datasets data quality control, imputation, and sample selection details have been previously described\textsuperscript{22}.

For all analyses, samples were separated into major subtypes using prognostically significant molecular markers (presence/absence of somatic \textit{IDH} mutation and/or 1p/19q chromosomal arm codeletion) and primary tumor grades\textsuperscript{3}. In total, eight groups of glioma cases were separately analyzed in the present study, most coarsely separated by \textit{IDH} mutation, and then further stratified by 1p/19q codeletion status and tumor grade (full description available in Figure 1). Individuals with unknown grade or \textit{IDH} status were excluded from all analyses.

Recorded temozolomide usage (yes/no) during first round of treatment was available for all glioma cases from the UCSF AGS and Mayo Clinic datasets. Glioma cases included from the TCGA were generally treatment naïve.

\textit{Study Genes:}

We conducted a literature review on DNA repair and temozolomide, specifically on BER and MMR pathways, to identify genes known or suspected to play a role in these mechanisms. We searched PubMed, Web of Science, Scopus, and Embase until April 2, 2023. Search keywords included: “base-excision repair,” “germline mutation,” “hypermutation,” “glioma,” “mismatch base repair,” “single nucleotide polymorphism,” “temozolomide,” “TMZ mechanism of action,” and their synonyms. Candidate genes were included regardless of any reported relationship with TMZ. This search identified the following 22 candidate genes: \textit{MSH2}, \textit{MSH3}, \textit{MSH4}, \textit{MSH5}, \textit{MSH6}, \textit{MLH1}, \textit{MLH3}, \textit{PMS1}, \textit{PMS2}, \textit{XRCC1}, \textit{XRCC3}, \textit{ERCC1}, \textit{ERCC2}, \textit{LIG1}, \textit{LIG3}, \textit{LIG4}, \textit{APEX1}, \textit{ADPRT}, \textit{MBD4}, \textit{POLB}, \textit{OGG1}, and \textit{MUTYH}.

\textit{SNP selection:}

This study analyzed a total of 1393 SNPs with minor allele frequency (MAF) >0.01 (amongst European populations CEU, TSI, FIN, GBR, and IBS from the 1000 Genomes Project\textsuperscript{23})
which were available on all three arrays within the regions 1kb upstream or downstream of any of the 22 candidate DNA repair genes. To properly adjust for multiple testing, the quantity of putatively independent SNP signals was estimated using a linkage disequilibrium (LD) $r^2 >= 0.10$ (within European populations) filter for SNPs within each chromosome, resulting in 259 effectively independent SNPs across the 22 genes. This study therefore used a Bonferroni-adjusted p-value threshold of $0.05/259 = 0.00019$ to determine strict statistical significance for all survival analyses. A less stringent threshold of $p<0.005$ was used to highlight potentially important associations.

**Statistical Analysis:**

**Overall survival SNP analysis:**

We evaluated the relationship between germline genetic variants within DNA repair genes and overall survival outcomes of adults with subtype-specific gliomas. Single SNP survival analyses were conducted using SPAcox\(^{24}\) using Cox proportional hazards (PH) models with adjustments for age at diagnosis, sex, primary tumor grade, history of radiotherapy (binary), history of chemotherapy (binary), year of primary diagnosis, and the top 10 genetic ancestry principal components (PCs). No treatment or year of diagnosis information was available for the TCGA dataset. SNPs were modeled as continuous dosage values. Analyses were conducted separately for each of the three glioma datasets (AGS-Mayo, AGS-i370, TCGA) and results were combined in a fixed-effects meta-analysis using the software METAL\(^{25}\) to account for study-specific biases. SNP associations with $p<0.00019$ were considered statistically significant after multiple testing adjustments, and associations with $p<0.005$ were considered suggestively important. Within each subtype, SNPs where less than 10 cases carried any copy of the minor allele across the three studies were removed from consideration.

**SNP-Temozolomide interaction analysis:**

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

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

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

For each suggestively important SNP-TMZ interaction (\(p<0.005\) for \(\beta_3\)), we partitioned individuals into four groups according to temozolomide usage and presence of the minor allele, and performed unadjusted Kaplan-Meier (KM) survival analysis on these groups to further confirm if the candidate SNP was explicitly associated with altered survival outcomes only in those who received temozolomide, with no survival effect between the two non-temozolomide groups. Only SNPs with at least ten individuals in each group were included for KM analysis. SNPs with a KM \(p\)-value \(p<0.05\) in the temozolomide users’ group were retained. Amongst this set, SNPs with a KM \(p\)-value \(p<0.05\) in the non-temozolomide users’ group were discarded, as there was insufficient evidence to conclude that survival effects of the polymorphism were temozolomide-specific.

**DNA repair polymorphism burden analysis:**

DNA repair mutational burden scores (DRBS) were constructed to obtain estimates of the cumulative effect of polymorphisms on survival in each of the glioma subtypes. For each subtype with more than one suggestively important SNP association, highly correlated SNPs were filtered out (linkage disequilibrium, LD, \(r^2 > 0.5\), within 500kb), preferentially selecting those with the lowest \(p\)-value amongst suggestive variants \((p<0.005)\). For each individual in the subtype, a DRBS was
calculated as the integer count of effect polymorphisms amongst the set suggestive SNPs. For a
subtype with $k$ suggestive SNP associations, the score takes integer values in $\{0,1,...,k\}$. The
effect alleles were chosen such that the direction of effect was consistent across all SNPs. Cox
proportional hazard models adjusting for age, sex, grade, data source, and the top 10 genetic
PCs were used to determine the statistical significance of both the continuous DRBS and the
binary $>0$ versus $0$ DRBS on survival outcomes. Significant score associations were determined
using the above multiple testing adjusted $p$-value threshold of $p<0.00019$.

Results:

Sample Population:

Table 1 provides a descriptive summary of the 2078 included samples with available
genotyping/covariate information. Sample summaries for each analyzed subgroup can be found
in Supplementary Tables S1-S8.

DNA repair SNP associations with overall survival:

IDH mutant glioma:

The median survival time for the 813 included individuals with an $IDH$ mutant glioma
(grades 2, 3, or 4, and regardless of 1p/19q codeletion) was 11.3 years from primary diagnosis.
Of the 1393 SNPs analyzed for association with $IDH$ mutant survival, none reached strict
statistical significance ($p<0.00019$), however polymorphism in $MLH1$ had suggestively important
(meta $p<0.005$) associations with survival. The top tagging SNP of $MLH1$, based on $p$-value was
an intron variant, rs4647218-G (Hazard ratio, HR=0.81, 95% confidence interval: 0.61-0.95,$
$p=0.0035$). Carriers of two copies of the G allele had a median survival time of 14.3 years,
compared to 11.2 and 10.5 years for one or zero copies of G, respectively.
Grades 2/3 IDH mutant glioma:

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

To evaluate the effect of the co-occurrence of these polymorphisms, we constructed a DRBS, where individuals were grouped based on the quantity of effect polymorphisms (k=4 presented above) they harbor (0 to 4 sites with the ‘survival-decreasing’ polymorphism). We defined the four effect polymorphisms as: rs4647218-AA, rs4848677-AA, rs146901049-A, rs3219489-CC. In a multivariate cox model adjusting for age, sex, grade, data source, and the top 10 genetic PCs, the per-unit increase in the DRBS was significantly associated with decreased
survival outcomes (HR=1.52, 95% CI: 1.26-1.83, p=1.3x10^{-5}). Individuals carrying 3 of the reported polymorphisms (n=11) had a median survival time of 3.4 years, compared to 9.3 years for those carrying 2 polymorphisms (n=146), 12.8 years for those carrying only 1 polymorphism (n=374), and 20.4 years for those carrying none of the four (n=193) (Figure 2). In a similarly parameterized cox model, individuals having a polymorphism at any of the four sites described above (n=531) had decreased survival times (median 11.3 years) as compared to those carrying none of the four polymorphisms (HR=1.80, 95% CI: 1.26-2.59, p=0.0014). No substantial differences between WHO grade or age at diagnosis were observed between DRBS groups.

Grades 2/3 IDH mutant 1p/19q codeleted glioma:

The median survival of the included 279 individuals with lower grade IDH mutant 1p/19q codeleted gliomas was 17.5 years from diagnosis. A single SNP in MSH2 had suggestive effects (p<0.005) on survival outcomes. The intron variant rs12615967-G in MSH2 was associated with poor prognosis (HR=2.04, 95% CI: 1.22-3.33, p=0.0032), where carriers of any copy of the G allele (n=69) had a median survival of 11.3 years, compared to 17.9 years for those with the common A/A genotype (n=210).

Grades 2/3 IDH mutant 1p/19q non-codeleted glioma:

The median survival of the 418 individuals with grade 2/3 IDH mutant gliomas without somatic 1p/19q codeletion was 10.3 years from diagnosis. SNPs in MSH4, ERCC2, and ERCC1 were suggestively associated (p<0.005) with overall survival. A set of SNPs in an LD block tagged by rs5745366-T in MSH4 were suggestively associated with improved survival (HR=0.33, 95% CI: 0.15-0.72, p=0.0010) where carriers of the C/T genotype (n=26) had a median survival of 13.6 years compared to 9.5 years for those with the common C/C genotype (n=392). Two SNPs in ERCC2, rs3916804-C and rs3916788-C (LD r^2=0.24) were associated with altered survival. The first, rs3916804-C, an intron variant was suggestively associated with improved survival
where individuals with the C/T or C/C genotype (n=98) lived a median 13.0 years post-surgery compared to 8.5 years for those with the T/T genotype (n=320). The second ERCC2 SNP, rs3916788-C was suggestively associated with improved survival times (HR=0.62, 95% CI: 0.47-0.83, p=0.0033), where carriers of the C/C genotype (n=66) had the longest survival, with 13.0 years post diagnosis, carriers of the C/A genotype (n=208) had a survival of 11.1 years, and carriers of the A/A genotype (n=144) had a median survival time of only 6.9 years. Lastly, a set of four SNPs in ERCC1, tagged by intron SNP rs144057986-C, were associated with improved survival outcomes (HR=0.58, 95% CI: 0.42-0.82, p=0.0045). Individuals carrying any copy of the C polymorphism (n=155) had a median survival of 13.3 years (95% CI: 11.1-NA), compared to only 8.0 years (95% CI: 6.6-10.5) for those with the T/T genotype (n=263).

To construct a DRBS, individuals were grouped based on the quantity of polymorphisms (k=4 presented above) they harbor (0 to 4 sites with ‘survival-increasing’ polymorphism). Individuals carrying 3 or 4 of the reported polymorphisms (n=40) had a median survival time of 14.5 years, compared to 11.5 years for those carrying 2 polymorphisms (n=147), 9.7 years for those carrying only 1 polymorphism (n=135), and 6.4 years for those carrying none of the four (n=96) (Figure 3). In a multivariate cox model adjusting for age, sex, grade, data source, and the top 10 genetic PCs, individuals having a polymorphism at any of the four sites described above (n=322) had significantly improved survival times as compared to those carrying none of the four polymorphisms (n=96) (HR=0.47, 95% CI: 0.33-0.69, p=8.07x10^-5). In a similarly parameterized cox model, the per-unit increase in the continuous DRBS was also significantly associated with improved survival outcomes (HR=0.67, 95% CI: 0.56-0.80, p=8.9x10^-6). No differences were observed between DRBS groups and age of diagnosis or WHO group.

Grade 4 IDH mutant glioma:
The median survival time for the 83 included individuals with grade 4 IDH mutated glioma tumors was 2.5 years. There were no significant or suggestively important SNP associations with survival in this group, likely due to small sample size.

IDH wildtype glioma:

The median survival for the 1265 included individuals with an IDH wildtype glioma was 16.1 months from time of primary diagnosis. None of the 1393 SNPs reached strict significance after multiple testing correction, however 3 SNPs in LD in MSH3 had p<0.0005 associations with IDH wildtype survival. The top polymorphism, rs6151863-A, an intron variant in MSH3 on chromosome 5, was associated with an increase in overall survival (HR=0.67, 95% CI: 0.50-0.89, p=0.00093), where individuals with any copy of the A allele (n=56) had a median survival time of 20.5 months compared to 15.7 for those with the common G/G genotype (n=1209).

Grade 2/3 IDH wildtype glioma:

A total of 222 individuals with grade 2/3 IDH wildtype gliomas were included in this analysis, with a median overall survival time of 22.4 months. A single SNP in MSH6, rs62139913-A, was suggestively associated with improved overall survival times (HR=0.26, 95% 0.09-0.76, p=0.0046). Carriers of any copy of the A allele (n=11) had a median survival of 6.5 years ((5% CI: 2.0-NA), compared to 1.8 years (95% CI: 1.6-2.2) for the 211 individuals with the G/G genotype.

Grade 4 IDH wildtype glioma:

The median survival for the 1043 included adults with grade 4 IDH wildtype gliomas was 15.4 months from initial diagnosis. There were no significant or suggestively important SNP associations with survival in this group, however the MSH3 association observed in IDH wildtype overall gliomas is only moderately attenuated when subsetting to grade 4, (rs6151863-A,
HR=0.73, 95% CI: 0.54-0.98, p=0.0064), where carriers of any copy of the A allele had a median survival of 19.1 months, compared to only 15.2 for those with the G/G genotype.

Quantitative Trait Loci analysis:

Using the Genotype-Tissue Expression (GTEx) project (Analysis release v8) and paired genotype/RNA-seq data from the TCGA, we investigated expression quantitative trait loci (eQTL) associations for each SNP from Table 2. Within GTEx, the rs4949677-A polymorphism was associated with increased MSH4 expression in the cerebellum (p=2.53e-11) amongst many other tissues (Figure 4A). We observed that the eQTL effect remained within tumor samples of 323 grade 2/3 IDH mutant glioma cases using paired tumor RNAseq and germline genotype data (Figure 4B). The rs146901049-A polymorphism, while located within the MSH3 gene, was associated with increased DHFR expression in the cortex (p=3.37e-5), an adjacent gene which shares a promoter region (Figure 4C). The association was directionally consistent, but not significant (p=0.36), amongst tumor samples of grade 2/3 IDH mutant glioma cases from the TCGA (not reported). The rs3219489-G SNP was a strong splicing QTL for MUTYH within the brain (p=1.51e-12), where the G allele was associated with significantly decreased intron excision ratios (Figure 4D). Bulk tissue gene expression for MUTYH across 54 tissue types revealed that the gene is most highly expressed in the brain (cerebellum) (Figure 4E). Amongst the remaining SNPs in Table 2, there were no significant QTL associations within the brain within GTEx, or within associated tumor types in TCGA.

SNP-TMZ Interaction Analysis:

A total of 1393 glioma cases (480 IDH mutant, 913 IDH wildtype) from the UCSF AGS and Mayo Clinic datasets with known temozolomide exposure during first round of treatment were included. Table 3 provides a demographic breakdown of each group separated by temozolomide usage.
In this section, we applied a two-stage analysis: For stage 1, we performed a Cox PH survival analysis on each of the 1393 SNPs with an included TMZ x SNP interaction term (significance threshold p<0.005). For stage 2, for all SNPs with significant TMZ-survival interaction, we calculated Kaplan Meier survival curves for the SNP of interest, separately for those with TMZ treatment and those without, to filter for TMZ-specific survival associations, using the log-rank p-values. All results which passed the 2-stage analysis are summarized in Table 4, along with temozolomide and non-temozolomide specific Cox model associations. The full results set for all considered glioma subtypes are available in Supplementary File 2.

**IDH mutant gliomas:**

The median overall survival time for the 263 included IDH mutant glioma cases with record of temozolomide during first course of treatment was 10.9 years (95% CI: 9.2-NA), compared to 13.1 years (95% CI: 11.3-17.5) for the 217 without temozolomide. 9 SNPs across four genes reached the threshold of p<0.005 for an interactive survival effect with temozolomide. Of those 11, two SNPs had significant (KM log rank p<0.05) associations with survival outcomes within case treated with temozolomide, and no association (p>0.05) with survival for those without temozolomide (Table 4).

The first, rs1540354-T, an intron variant in MLH1, was associated with decreased overall survival only for cases receiving temozolomide (interaction p-value = 0.0044, TMZ KM p-value = 0.00047, non-TMZ KM p-value = 0.71, TMZ-only Cox model: HR=1.70, 95% CI: 1.17-2.47, p=0.0058 / non-TMZ-only Cox model: HR=0.88, 95% CI:0.55-1.42, p=0.60). Among the 263 temozolomide treated individuals, those with T/T or A/T genotypes (n=84) had an overall survival 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 the A/A genotype. The untreated group had indistinguishable survival times: carriers of the T polymorphism (n=63) lived a median 13.8 years after diagnosis, compared to 13.1 years for those with the A/A genotype (n=153) (Figure 5A).
The second variant, rs71636247-G, an intron in MSH3 was also associated with decreased survival amongst cases treated with temozolomide (interaction p-value =0.0045, TMZ KM p-value = 0.025, non-TMZ KM p-value = 0.73, TMZ-only Cox model: HR=1.93, 95% CI: 1.14-3.27, p=0.015 / non-TMZ-only Cox model: HR=0.74, 95% CI: 0.40-1.38, p=0.35). The median survival for individuals treated with temozolomide carrying any copy of the G allele (n=28) was 6.2 years, compared to 12.2 years for the 235 with the A/A genotype. For those without temozolomide treatment, the median survival times were 12.2 and 13.2 years for those with (n=24) and without (n=193) any copy of the G allele (Figure 5C).

In constructing a DRBS using the two SNPs, individuals were grouped based on the quantity of polymorphisms (k=2 presented above) present in their germline DNA (0 to 2 sites with polymorphism). Amongst individuals with temozolomide, in a multivariate cox model adjusting for age, sex, grade, and the top 10 genetic PCs, a per unit increase in the DRBS was associated with decreased survival probability (HR=2.23, 95% CI: 1.55-3.20, p=1.56x10⁻⁵). Those with both polymorphisms (n=7) had a median survival time of 4.3 years, those with only one (n=98) survived a median 7.0 years after diagnosis, and of the n=158 with neither polymorphism, there have not been enough events to determine a median time (NA, 95% CI: 11.3-NA). For those not treated with temozolomide during first course of treatment, in an identically parameterized cox model, the per-unit increase in the DRBS was not associated with survival outcomes (HR=0.90, 95% CI: 0.63-1.28, p=0.56). Median survival times were NA, 11.8, and 13.6 years for those with DRBS scores of two, one, and zero, respectively (Figure 5E).

Grades 2/3 IDH mutant gliomas:

Amongst the grade 2/3 IDH mutant cohort of 410 cases, 213 had known temozolomide treatment, and 197 had confirmed absence of temozolomide treatment. The unadjusted median survival times were 13.4 years for those with treatment (95% CI: 10.9-17.3) and 15.2 years for those without (95% CI: 12.8-18.2). 2 SNPs in XRCC1 reached significance (p<0.005) in stage 1.
Neither passed stage 2 filters in separate KM analyses for cases with and without TMZ treatment (Supplemental Table 2).

Grades 2/3 IDH mutant 1p/19q codeleted gliomas:

The median survival time for grade 2/3 IDH mutant gliomas was further 1p/19q codeletion was 17.5 years for those treated with temozolomide (n=83), and 17.9 for those without the treatment (n=69). Nine SNPs across four genes had a suggestively important TMZ-SNP interaction term (Supplementary File 2). Of those, four SNPs in *MSH2*, tagged by lead variant rs6739909-G (Table 4), passed stage 2 Kaplan-Meier filters. Polymorphism rs6739909-G was associated with decreased survival amongst individuals treated with temozolomide (interaction p-value = 0.0045, TMZ KM p-value = 0.030, non-TMZ KM p-value = 0.13, TMZ-only Cox model: 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, p=0.14). Individuals treated with temozolomide with the G polymorphism (n=19) had a median survival time of 11.3 years, compared to 20.4 years for the 64 treated individuals with the T/T genotype. Individuals without temozolomide as first course of treatment had median survival times of 20.1 and 16.5 years, for carriers and non-carriers of the G allele, respectively.

Grades 2/3 IDH mutant 1p/19q non-codeleted gliomas:

Median survival time for the 119 individuals treated with temozolomide was 10.2 years, compared to the 11.0 years for the 112 without first-course temozolomide treatment. Two SNPs in the *LIG4* gene (rs3093772-C, rs3093774-T) had suggestively important TMZ interactive effects on survival (Supplemental Table 2), however neither passed stage 2 thresholding.

Grade 4 IDH mutant gliomas:

This study included only 66 individuals with a grade 4 *IDH* mutant glioma and known presence/absence of temozolomide as first-course of treatment. The 48 with temozolomide
treatment had a median survival of 42.7 months, compared to the 20.6 months for those 18 individuals without temozolomide. We did not have the sample size necessary to detect differences, particularly in the stage 2 analysis. Full results are however made available in Supplemental Table 2.

**IDH wildtype gliomas:**

A total of 913 individuals with IDH wildtype glioma and known presence/absence of temozolomide during first-course treatment were included in this analysis. The 685 with known presence of temozolomide had a median survival of 17.8 months, and the 228 with known absence had a median survival of 15.1 months. A total of 26 SNPs passed stage 1 thresholding with interaction p-values < 0.005. Of those 26, only rs1052133-G, a missense variant of OGG1 (Table 4), passed stage 2 thresholds, and is suggestive of altered survival outcomes for temozolomide treated IDH wildtype gliomas (interaction p-value = 0.0012, TMZ KM p-value = 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 / non-TMZ-only Cox model: HR=1.15, 95% CI: 0.85-1.56, p=0.35). For the 685 with known temozolomide treatment, the median overall survival time for carriers of the G polymorphism was 18.4 months (n=277), compared to the 16.9 months for C/C genotypes (n=408). For the 228 with known absence, carriers of the G polymorphism had a median survival time of 14.4 months (n=102), compared to 16.1 for those with the C/C genotype (n=126).

**Grades 2/3 IDH wildtype gliomas:**

The 88 individuals with primary grade 2/3 IDH wildtype glioma who received temozolomide during the initial course of treatment had a median survival time of 23.8 months after primary diagnosis. The 61 who did not receive the drug during first course had a median survival of 22.9 months. 11 SNPs across two genes (MSH3 and LIG4) reached p<0.005 in stage 1 interaction analysis. Of those, nine SNPs (in a single LD block) in MSH3 passed stage 2 filters (Table 4).
The association in MSH3 was captured by the rs34799825-T polymorphism, an intron variant, where the T allele was associated with drastically improved survival outcomes in the temozolomide group (interaction p-value = 0.00031, TMZ KM p-value = 0.012, non-TMZ KM p-value = 0.18, TMZ-only Cox model: HR=0.52, 95% CI: 0.27-1.01, p=0.053 / non-TMZ-only Cox model: HR=5.19, 95% CI: 2.05-13.12, p=0.00051). For individuals with temozolomide, those with the C/T or T/T genotype (n=20) had a median survival of 35.6 months (95% CI: 25.9-NA), compared to 19.9 months (95% CI: 16.7-75.2) for those with the C/C genotype (n=68). For individuals who did not receive temozolomide in first course of treatment, those with the C/T or T/T genotypes (n=15) had a median survival time of 17.6 months (95% CI: 9.4-173.6), compared to the C/C genotype (n=46) of 23.8 months (95% CI: 16.6-74.16).

Grade 4 IDH wildtype gliomas:

The 597 included individuals with grade 4 IDH wildtype and known temozolomide treatment had a median survival of 17.1 months, while the 167 with confirmed absence of temozolomide as first course treatment had a median 13.7 months survival. In a multivariable Cox PH model, temozolomide treatment was associated with altered IDH wildtype GBM survival (HR=0.8, 0.66-0.98, p=0.031).

Our stage 1 analysis nominated eight SNPs across two genes with p-values less than 0.005. Of those eight, rs149630102-T in MSH2 passed stage 2 thresholding (Table 4, Supplementary File 2). An intron variant, rs149630102-T was associated with temozolomide-specific survival alterations (interaction p-value = 0.0017, TMZ KM p-value = 0.00015, non-TMZ KM p-value = 0.078, TMZ-only Cox model: HR=1.74, 95% CI: 1.27-2.39, p=0.00052 / non-TMZ-only Cox model: HR=0.41, 95% CI: 0.23-0.75, p=0.0038). For the temozolomide group, carriers of the T allele (n=45) had a median survival of 13.2 months (95% CI: 12.3-16.7), compared to 17.6 months (95% CI: 16.5-18.5) for the 552 individuals with the A/A genotype. For the non-
temozolomide group, those with the T allele (n=16) had a median survival of 15.8 months (95% CI: 9.6-46.9) compared to 13.7 months (95% 11.3-15.5) for those with the A/A genotype (n=151).

Quantitative Trait Loci analysis:

Using the Genotype-Tissue Expression (GTEx) project (Analysis release v8) and paired genotype/RNA-seq data from the TCGA, we investigated quantitative trait loci associations for each of the six suggestive SNPs from Table 4 within the brain/tumor. Within the cortex, rs1540354-T was an eQTL for RP11-640L9.2 (Figure 5B). Within the cerebellum, the same SNP, rs1540354-T, was a splicing QTL for LRRFIP2. The rs71636247-G variant was a significant eQTL for MSH3 in the cerebellum, with the G allele associated with increased expression (Figure 5D). Similarly, rs34799825-T was also a significant eQTL for MSH3 in the cerebellum. Lastly, rs149630102-T was a significant eQTL for EPCAM on chromosome 2 within the hypothalamus. No SNPs were significant eQTLs for their associated genes within TCGA subtype-specific tumor samples.

Discussion:

We investigated the association of 1393 putatively independent SNPs within 22 DNA repair genes with overall survival outcomes of 2078 adults with glioma in a meta-analysis of three datasets with molecular subtype specificity. We further investigated if these germline polymorphisms modified survival specifically for patients treated with temozolomide. We observed statistically significant polygenic associations with both overall survival, and temozolomide-specific outcomes, broadly amongst IDH mutant gliomas.

Grade 2/3 IDH mutant gliomas are comprised of both astrocytoma and oligodendroglioma tumors, both which have generally favorable prognoses in comparison with other glioma subgroups. In this mixed group, we observed 4 polymorphisms (in MLH1, MSH4, MSH3, and
MUTYH) which, when combined, significantly stratified survival trajectories of individuals within this tumor group. Cases with none of the four risk SNPs had a median survival of nine years longer than those with any of the four. For three of the four polymorphisms, we presented possible functional alterations within the brain. The rs4949677-A polymorphism was associated with elevated MSH4 expression both in normal brain and tumor tissues. The rs146901049-A polymorphism, while located within the MSH3 gene, was associated with a decreased expression of DHFR within the brain. DHFR, highly expressed in gliomas, is a crucial metabolic enzyme in the folic acid signaling pathway, and has been hypothesized to contribute to glioma cell proliferation and to decreased chemosensitivity to temozolomide, and has previously been nominated as a target for therapeutics\textsuperscript{27,28}. The results presented here, however, provide evidence that a germline decrease in DHFR expression (in normal brain) is detrimental to glioma survival. We found that carriers of the G polymorphism at rs3219489 in MUTYH exhibited improved survival compared to CC carriers amongst adults with grade 2/3 IDH mutant gliomas. Located at chromosome locus 1p34.3—1p32.1, the MUTYH gene encodes for a BER enzyme, MYH glycosylase, responsible for the repair of oxidized guanine incorrectly paired with adenine, one of the most pervasive forms of oxidative damage, by excising the undamaged adenine base\textsuperscript{29,30}. Of all tissues available in the GTEx database, MUTYH is most highly expressed in the brain (cerebellum). The MUTYH rs3219489 C to G polymorphism results in an amino acid change from glutamine to histidine, which evidence from GTEx QTL analysis suggests results in alternative splicing of the MUTYH protein in the brain, a likely functional alteration to the BER mechanism and a potential mediator of glioma survival. The rs3219489-G allele has been previously associated with decreased survival amongst lung cancer patients who received platinum-based chemotherapy\textsuperscript{31} and increased risk of recurrence in oral squamous cell carcinoma patients receiving chemoradiotherapy\textsuperscript{32}. Further germline alterations in MUTYH have been most strongly linked with an elevated risk of colorectal cancer via inheritable colon polyposis, primarily amongst individuals with European ancestry\textsuperscript{33}, and others have been extensively described in
terms of their ability to promote somatic mutations, including C>A mutations. Further functional work is needed to understand the role of rs3219489 in promoting somatic mutations and the effects of this potential sQTL in glioma.

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

The role of temozolomide in treating glioblastomas is well-understood, with a clear improvement in patient outcomes when combined with other standard-of-care treatments. Our study found suggestive evidence of a single DNA repair germline risk factor (rs149630102-T in MSH2) modifying this glioblastoma effect. Interestingly, the identified SNP was associated with altered survival in opposite directions, depending on if the group received temozolomide. For the group of grade 4 IDH wildtype gliomas without temozolomide, the T allele was suggestively associated with improved survival, whereas the same allele was suggestive of decreased survival for cases who received the drug. This could be related to the repair of TMZ-induced mutagenic adducts and a threshold effect, providing protection in TMZ-naive subjects, and a detriment to TMZ-treated subjects by repairing enough of the TMZ induced mutations to avoid cell death and therefore promoting greater number of somatic mutations in a surviving clone. Using GTEx, we
found that rs149630102-T was associated with increased expression of the *EPCAM* gene in non-cancer brain and was also suggestively associated with both *EPCAM* and *MSH2* expression in TCGA glioblastoma tumor samples (not shown). A 2014 study demonstrated an association between increased *EPCAM* expression and decreased grade 4 glioma survival time\(^{37}\), which is consistent with the line of evidence observed in this study for individuals who received temozolomide. A 2015 study also presented in vivo evidence that decreased *MSH2* levels caused a reduction in response to temozolomide in glioblastoma\(^{38}\). When comparing survival curves, those without the rs149630102-T polymorphism and treated with temozolomide had the longest median survival time of all groups, suggesting temozolomide may provide survival benefits only to this group. To confirm this speculation, further work into the link between this polymorphism, *EPCAM*, *MSH2*, other possible functional mechanisms, and the interaction with temozolomide is both warranted and necessary.

As is well described by a growing body of literature, temozolomide-induced hypermutation at recurrence is common amongst individuals with lower grade IDH mutant glioma, with only *MGMT* promoter methylation level at primary diagnosis as a possible biomarker for hypermutation and general sensitivity to the treatment\(^{12}\). While hypermutation and *MGMT* status are unknown in our study participants, we identified two germline polymorphisms, in *MLH1* and *MSH3* genes, which were associated with poor outcomes for individuals with IDH mutant glioma treated with temozolomide, and no differential effect in the non-temozolomide group. Both SNPs were significantly associated with altered expression of RNA transcripts in the brain, suggesting likely functional alterations to the mismatch repair mechanisms during temozolomide-induced damage repair of the tumor cells, possibly diminishing the efficacy of the mechanism. A reduction in *MLH1* expression has been previously associated with temozolomide resistance in GBM cell lines\(^{30}\), consistent with the directional effect (non-significant) of rs1540354-T on *MLH1* expression. While not conducted in glioma, previous cell-line analyses have demonstrated *MSH3* deficiency lead to increased chemosensitivity and cell apoptosis, independent of the role of *MLH1*, in colon
carcinoma cells\textsuperscript{40}. Our combined evidence of rs71636247-G having associations with both
dehased survival and increased \textit{MSH3} expression in the brain remains consistent with this
previous observation. \textit{We show that the presence of either of these two polymorphisms resulted}
in drastically decreased median survival times, even with rigorous significance thresholds. As
such, these two SNPs warrant further clinical investigation as our results generate the hypothesis
that temozolomide may not be an appropriate treatment for individuals harboring these germline
polymorphisms, and alternative treatments should be prioritized. Although not included here,
these results were replicated with only minor attenuation when excluding all grade 4 \textit{IDH} mutant
glioma cases.

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

Our findings should be considered with the following limitations. Our included meta-analysis
is from a set of historic cohorts. As such, we lack the full suite of somatic molecular markers to
adhere to strict WHO 2021 classification guidelines, and we lack knowledge of \textit{MGMT}
methylation, a relatively recently identified and costly phenotype to measure. The cohorts also do
not include robust information on the recurrent tumor, including temozolomide-induced hypermutation. Our results may not extend to non-European populations, as the included subjects were chosen to have at least 70% genetically estimated European ancestry. Our study of temozolomide was limited to a binary yes/no history of known drug usage during the first course of treatment; we could not further investigate the role of timing and amounts of temozolomide administered on survival outcomes. Ongoing studies which collect detailed treatment information across the course of the disease will be key to determining if the effects presented in this study are indeed isolated to just first-course treatment. Despite these limitations, this analysis is one of the most extensive studies of glioma germline genetics and survival with available molecular subtyping, long-term follow-up, and temozolomide utilization.

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

References:


31. Singh A, Singh N, Behera D, Sharma S. Genetic Investigation of Polymorphic OGG1 and MUTYH Genes Towards Increased Susceptibility in Lung Adenocarcinoma and its Impact
on Overall Survival of Lung Cancer Patients Treated with Platinum Based Chemotherapy.


Figures and Tables for ‘The impact of germline variants in DNA repair pathways on survival and temozolomide toxicity in adults with glioma’

Groups for survival analysis

Figure 1: Included glioma subgroups based on prognostically significant molecular markers and tumor grade. Tumor types were coarsely split first by IDH mutation status, and then by further by tumor grade at primary diagnosis. Grade 2/3 IDH mutant gliomas were further stratified by the 1p/19q chromosomal arm codeletion phenotype. Boxes with included sample size are used to highlight the eight subgroups analyzed in this study.
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 rs146901049 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

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

NA: not available information

*Group contains IDH wildtype and mutant. World Health Organization 2021 guidelines now classify IDH mutant Glioblastoma as Grade 4 Astrocytoma
Table 2: Multivariate cox proportional hazards analysis of genetic variants in mismatch repair and base excision repair genes with overall survival amongst major glioma molecular subtypes.

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

<table>
<thead>
<tr>
<th></th>
<th>IDH mutant</th>
<th>No TMZ</th>
<th>IDH wildtype</th>
<th>No TMZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>263</td>
<td>217</td>
<td>685</td>
<td>228</td>
</tr>
<tr>
<td>Age at diagnosis (Median + IQR)</td>
<td>40 (IQR: 16)&quot;</td>
<td>39 (IQR: 18)</td>
<td>58 (IQR: 16)</td>
<td>58 (IQR: 18)</td>
</tr>
<tr>
<td>Sex (% Male)</td>
<td>60.84%</td>
<td>57.60%</td>
<td>64.82%</td>
<td>58.77%</td>
</tr>
<tr>
<td>Self-report race (% White)</td>
<td>98.10%</td>
<td>98.16%</td>
<td>98.25%</td>
<td>100%</td>
</tr>
<tr>
<td>1p/19q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Codeleted</td>
<td>31.56%</td>
<td>31.80%</td>
<td>0.15%</td>
<td>1.75%</td>
</tr>
<tr>
<td>Non-codeleted</td>
<td>63.12%</td>
<td>59.91%</td>
<td>97.81%</td>
<td>93.42%</td>
</tr>
<tr>
<td>NA</td>
<td>5.32%</td>
<td>8.29%</td>
<td>2.04%</td>
<td>4.83%</td>
</tr>
<tr>
<td>TERT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wildtype</td>
<td>52.85%</td>
<td>53.00%</td>
<td>16.79%</td>
<td>17.98%</td>
</tr>
<tr>
<td>Mutant</td>
<td>36.88%</td>
<td>33.18%</td>
<td>77.23%</td>
<td>69.74%</td>
</tr>
<tr>
<td>NA</td>
<td>10.27%</td>
<td>13.82%</td>
<td>5.99%</td>
<td>12.28%</td>
</tr>
<tr>
<td>Grade (% of total)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>35.36%</td>
<td>64.98%</td>
<td>3.36%</td>
<td>13.16%</td>
</tr>
<tr>
<td>3</td>
<td>45.63%</td>
<td>25.81%</td>
<td>9.49%</td>
<td>13.60%</td>
</tr>
<tr>
<td>4</td>
<td>19.01%</td>
<td>9.22%</td>
<td>87.15%</td>
<td>73.25%</td>
</tr>
<tr>
<td>Histology (% of total)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Astrocytoma</td>
<td>36.88%</td>
<td>34.56%</td>
<td>11.24%</td>
<td>15.79%</td>
</tr>
<tr>
<td>Oligodendroglioma</td>
<td>27.00%</td>
<td>32.72%</td>
<td>0.73%</td>
<td>6.58%</td>
</tr>
<tr>
<td>Glioblastoma</td>
<td>19.01%</td>
<td>8.29%</td>
<td>87.15%</td>
<td>73.25%</td>
</tr>
<tr>
<td>Oligoastrocytoma</td>
<td>17.11%</td>
<td>24.42%</td>
<td>0.88%</td>
<td>4.39%</td>
</tr>
<tr>
<td>Diagnosis Year</td>
<td>2007</td>
<td>2003</td>
<td>2007</td>
<td>2002</td>
</tr>
<tr>
<td>% of cases diagnosed pre-1999</td>
<td>0.76%</td>
<td>29.95%</td>
<td>0%</td>
<td>39.47%</td>
</tr>
<tr>
<td>% of cases diagnosed pre-2005</td>
<td>34.60%</td>
<td>67.74%</td>
<td>34.60%</td>
<td>78.95%</td>
</tr>
<tr>
<td>% Censored (no known death event)</td>
<td>45.63%</td>
<td>43.78%</td>
<td>3.36%</td>
<td>6.58%</td>
</tr>
<tr>
<td>% Censored before 2 years</td>
<td>7.22%</td>
<td>4.61%</td>
<td>0.58%</td>
<td>0.88%</td>
</tr>
<tr>
<td>Radiation (% known RT)</td>
<td>72.24%</td>
<td>52.53%</td>
<td>97.52%</td>
<td>87.72%</td>
</tr>
</tbody>
</table>
Table 4: Association analysis of genetic variants in mismatch repair and base excision repair genes with temozolomide-specific survival amongst glioma molecular subtypes.

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

*: P-value of SNP x TMZ interaction term in Cox PH model
KM P: Unadjusted Kaplan-Meier log rank p-value