Published in final edited form as:
Clin Cancer Res. 2012 June 1; 18(11): 3154–3162. doi:10.1158/1078-0432.CCR-11-2778.

SSBP2 variants are associated with survival in glioblastoma patients

Yuanyuan Xiao[1], **Paul A. Decker**[2], **Terri Rice**[3], **Lucie S. McCoy**[3], **Ivan Smirnov**[3], **Joseph S. Patoka**[3], **Helen M. Hansen**[3], **Joe L. Wiemels**[1,4], **Tarik Tihan**[5], **Michael D. Prados**[3], **Susan M. Chang**[3], **Mitchel S. Berger**[3], **Matthew L. Kosel**[2], **Brooke L. Fridley**[2], **Daniel H. Lachance**[6,7], **Brian Patrick O'Neill**[6], **Jan C. Buckner**[8], **Reid C. Thompson**[9], **L. B. Nabors**[10], **Jeffrey J. Olson**[11], **Steve Brem**[12], **Melissa H. Madden**[13], **James E. Browning**[13], **John K. Wiencke**[3,4,14], **Kathleen M. Egan**[13,14], **Robert B. Jenkins**[7,14], and **Margaret R. Wrensch**[3,4,14]

[1]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California 94143, USA

[2]Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, MN 55905, USA

[3]Department of Neurological Surgery, University of California, San Francisco, San Francisco, California 94143, USA

[4]Institute of Human Genetics, University of California, San Francisco, San Francisco, California 94143, USA

[5]Department of Pathology, University of California, San Francisco, San Francisco, California 94143, USA

[6]Department of Neurology, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA

[7]Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA

[8]Division of Medical Oncology, Mayo Clinic College of Medicine, Rochester, Minnesota

[9]Department of Neurological Surgery, Vanderbilt University Medical Center, Nashville, TN, 37232 USA

[10]Neuro-oncology Program, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Address Correspondence to: Yuanyuan Xiao, PhD, Department of Epidemiology and Biostatistics, University of California, San Francisco, 185 Berry St, Suite 5700, San Francisco, CA 94107, (yuanyuan.xiao@ucsf.edu).
[14]These authors represent the senior authors.

Conflict of Interest Disclosure:
Louis Burt Nabors has an uncompensated position at Merck KgaA.

Contributions of data:
The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000036C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #1U58 DP000807-01 awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the author(s) and endorsement by 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 is not intended nor should be inferred.
The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at "http://cancergenome.nih.gov".

[11]Department of Neurosurgery, Emory School of Medicine, Atlanta, GA 30322, USA

[12]Department of Neurosurgery, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33612, USA

[13]Department of Cancer Epidemiology & Genetics, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33612, USA

Abstract

Purpose—Glioblastoma is a devastating, incurable disease with few known prognostic factors. Here we present the *first genome-wide survival and validation study* for glioblastoma.

Methods—Cox regressions for survival with 314,635 inherited autosomal single nucleotide polymorphisms (SNPs) among 315 San Francisco Adult Glioma Study patients for discovery and three independent validation datasets (87 Mayo Clinic, 232 GliomaSE and 115 The Cancer Genome Atlas patients) were used to identify SNPs associated with overall survival for Caucasian glioblastoma patients treated with the current standard of care, resection, radiation and temozolomide (total n=749). Tumor expression of the gene that contained the identified prognostic SNP was examined in three separate datasets (total n=619). Genotype imputation was used to estimate hazard ratios (HRs) for SNPs that had not been directly genotyped.

Results—From the discovery and validation analyses, we identified a variant in *SSBP2* (single-stranded DNA-binding protein 2) on 5q14.1 associated with overall survival in combined analyses (hazard ratio (HR) = 1.64; $P = 1.3X10^{-6}$). Expression of *SSBP2* in tumors from three independent datasets also was significantly related to patient survival ($P = 5.3 X 10^{-4}$). Using genotype imputation, the *SSBP2* SNP rs17296479 had the strongest statistically significant genome-wide association with poorer overall patient survival (HR = 1.79; 95% CI: 1.45–2.22; $P = 1.0 X 10^{-7}$).

Conclusion—The minor allele of *SSBP2* SNP rs17296479 and the increased tumor expression of *SSBP2* were statistically significantly associated with poorer overall survival among glioblastoma patients. With further confirmation, previously unrecognized inherited variations influencing survival may warrant inclusion in clinical trials to improve randomization. Unaccounted for genetic influence on survival could produce unwanted bias in such studies.

Keywords

glioma; glioblastoma; GWAS; survival; epidemiology; *SSBP2*

Glioblastoma is a rapidly fatal form of primary brain cancer with few known prognostic factors. Major challenges of achieving complete patient follow-up, treatment heterogeneity and changing patterns of patient care over time have limited the feasibility of genome-wide cancer survival discovery with very few such studies published for any cancer site (1) and none thus far for glioblastoma. Moreover, candidate gene studies for glioblastoma survival have provided equivocal results (2–9) possibly due to the factors above or to inadequate gene coverage. In order to minimize these challenges, we focused this first genome-wide discovery and validation study for glioblastoma patient survival on carefully selected glioblastoma patient groups with follow-up and initial treatment with current standard of care.

METHODS

Study Subjects

Informed consent was obtained from each subject. The subject recruitment and studies were conducted after approval was obtained from the investigational review boards at each

participating site in accordance with assurances filed with and approved by the U.S. Department of Health and Human Services. (10, 11)

Discovery Study—Details of subject ascertainment for the San Francisco Adult Glioma Study (AGS) have been previously described. (10, 12, 13) The 315 glioblastoma patients in the present study are the subset who had received current standard-of-care treatment (resection, radiation and temozolomide) of the 525 glioblastoma patients whose results were used in the genome-wide association study reported by Wrensch et al. (10) after stringent sample quality control filtering. Among these patients, tumor characteristics (*IDH1* (n=173) and *TP53* (n=151) mutation status and EGFR copy number (n=173)) were available from ongoing studies. (14–16)

Validation Study—The Mayo Clinic study included 87 glioblastoma patients newly diagnosed between 2005 and 2008. Most cases were identified within 24 hours of diagnosis; some were initially diagnosed elsewhere and later had their diagnosis verified at the Mayo Clinic. Pathologic diagnosis was confirmed by review of the primary surgical material for all cases by two Mayo Clinic neuropathologists based on surgically resected material.

The GliomaSE study included glioblastoma patients enrolled in a case-control study conducted at medical centers in the Southeast and diagnosed with a primary (e.g. non-recurrent) glioma between 2005 and 2010. (11) Patients were enrolled a median of 1 month following glioblastoma diagnosis (and a maximum of 4 months according to study protocol). The glioblastoma diagnosis was based on diagnostic pathology reports available for all patients in the study.

The TCGA dataset was downloaded from the Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/(17)). At the time of data retrieval from TCGA, alignment of sample identifiers yielded 181 glioblastoma patients with both genotype and clinical data, 115 of whom had resection, radiation and temozolomide treatment. The subject IDs of these 115 TCGA patients are listed in Supplementary Table 1.

Genotyping

Genotyping for the AGS discovery subjects was conducted by deCODE Genetics using Illumina's HumanCNV370-duo BeadChip as previously described (10). After excluding SNPs with $p < 10^{-5}$ for Hardy-Weinberg equilibrium in the AGS control samples (AGS participants that did not have glioma), or minor allele frequency < 5%, or missing genotyping data > 5% in the case groups, there were 314,635 autosomal SNPs to consider in the survival tests. Genotyping for the Mayo Clinic study subjects was performed using Illumina 610Quad SNP arrays as previously described (10). Genotyping for the GliomaSE study subjects was performed using the Illumina Goldengate assay as previously described. (10) Genotyping for the TCGA study subjects was performed using Illumina 550 platform. (17)

Statistical Analysis

Supplementary Figure 1 provides an overview of the three types of analyses conducted: 1. genome-wide constitutive discovery and validation of SNPs associated with glioblastoma patient overall survival, 2. functional validation of survival loci (association of gene expression in tumors with glioblastoma overall patient survival), and 3. fine mapping via genotype imputation.

Genome-wide Survival and Validation Analyses—Due to human subject IRB constraints, analyses on the raw genotype data were carried out separately at the AGS, Mayo

Clinic and GliomaSE sites (TCGA data were analyzed at the AGS site). Summary statistics were then submitted to the AGS site for combined analysis. For the AGS discovery study, we conducted Cox proportional hazards regression models to assess the association of each of the 314,635 SNPs with overall survival, adjusting for age (on a continuous scale) and sex. The SNP variable used in the model is coded as a continuous count of the number of minor alleles based on the additive genetic model. Per-allele hazard ratio (HR) and 95% confidence interval were obtained for each SNP. Statistical significance for each SNP was assessed with the Wald test. The same Cox proportional hazard models were used for all ensuing analyses of the validation datasets. The genomic inflation factor based on the genome-wide P values for the AGS discovery study was 1.04 indicating that systematic inflation of our survival association signals due to model misspecification, undetected genotyping error or hidden ancestry relationship was highly unlikely. The proportional hazards assumption for validated SNPs with a four-site combined $P \leq 10^{-5}$ was tested within each site using the Schoenfeld residuals, and SNPs with evidence for non-proportionality were removed from further consideration. Results for the non-proportionality test for rs7732320 are shown in Supplementary Table 2. Heterogeneity across the four studies for rs7732320 was assessed by Cochran's Q statistic. (18) As no significant heterogeneity across study sites was observed, a fixed effect model that used the inverse of the variance of the study-specific log(HR) estimates to give weights to the contribution of each study (19), was used to summarize results across studies. Specifically,

$$\widehat{\beta}_{Combined}=\frac{\sum_i \widehat{\beta}_i/\nu_i}{\sum_i /\nu_i}, \text{var}(\widehat{\beta}_{Combined})=\frac{1}{\sum_i 1/\nu_i},$$ where $\hat{\beta}_i$ and $\hat{\nu}_i$ are the log hazard ratio estimate and its variance for the i^{th} study respectively.

Functional Validation of Survival Loci—To examine associations of expression of the candidate gene, with survival, we assembled data from 619 primary glioblastoma samples from three published studies. (20–22) The Lee et al. (20) dataset described 218 glioblastoma expression samples including 132 samples from three previously published datasets as well as 86 new samples assembled into a single, unified dataset using Affymetrix U133A. The Murat et al. (21) dataset contains 75 glioblastoma expression samples using the Affymetrix U133. Normalized expression values using the standard RMA method for the Lee and Murat datasets were downloaded from the NCBI GEO database (GSE13041 and GSE7696). The TCGA dataset (22) has 326 primary glioblastoma expression samples using the Affymetrix U133A expression platform. Transcriptional class labels were obtained from the TCGA Advanced Working Group. (23) The updated labeling extends the original labeled set presented in Verhaak et al. (22) to previously unclassified samples. In total, we obtained 74 Proneurals, 45 Neurals, 93 Mesenchymals and 91 Classicals. For each of the three expression datasets, we carried out age and sex adjusted study specific survival analysis employing Cox models relating continuous gene expression data to patient survival, and then combined the study-specific HR estimates with a fixed effect model using the inverse variance approach. (19) Within the TCGA expression dataset, we also conducted expression subtype (Proneural, Neural, Classical and Mesenchymal) stratified survival analysis using a Cox model with the same specification. As treatment data were either missing or incomplete for these patients, we did not restrict the tumor gene expression analyses to patients with the current standard of care.

Fine Mapping via Imputation—Using MACH (24) and data from release 22 Phase II CEU HapMap data (MACH v 1.0.16) we imputed SNPs separately within each of the three studies with sufficient tagging SNPs (AGS, Mayo and TCGA). MACH implements a Markov Chain based algorithm to infer possible pairs of haplotypes for each individual's

genotypes (including untyped genotypes). We ran MACH with the default parameter values with the number of iterations of the Markov Chain set to 50 and the "greedy" option turned on. We then carried out study-specific Cox survival analysis using expected allele counts as the predictor for a total of 159 SNPs, whose variance ratios were larger than 0.5 for all three studies in order to exclude SNPs with poor quality imputed genotypes. Meta-analysis of the imputed data was performed in the same way as described above. To obtain survival signals independent of the most significant (imputed) SNP in the region, we included its expected counts in the Cox model as an additional covariate, along with the other covariates such as age and sex. All analyses were carried out using the R statistical package.

RESULTS

Patient characteristics (age, sex and median survival) for the four datasets (AGS, the Mayo Clinic, GliomaSE and TCGA) are described in Table 1. The majority of the observed survival Cox regression *P* values for 314,635 SNPs from the AGS discovery dataset conformed to the identity line in the Q-Q plot, whereas 90 SNPs showed significant deviation from expectation at $P = 10^{-4}$ (Supplementary Figure 2). We submitted these 90 SNPs for validation in Mayo Clinic patients of which 78 passed quality control. Ten of these SNPs had $P < 10^{-5}$ in the combined analysis using a fixed-effect model. (25) Examination of these 10 SNPs in two additional patient groups, GliomaSE and TCGA patients, yielded one SNP, rs7732320, that had discovery and validation combined $P < 10^{-5}$ for survival and had proportionality of hazards in all four datasets (Table 2 and Supplementary Table 3). The associations of this SNP with patient survival were in the same direction across the studies and had a combined validation P = 0.008 and a combined discovery-validation P = $1.3X10^{-6}$. There was no evidence of heterogeneity of the HR estimates across the four studies (Table 2). Effect modification by age at diagnosis for rs7732320 was evaluated in the AGS discovery data by the significance of the interaction term between age at diagnosis and the SNP; no statistical significant interaction was detected. In the AGS discovery data, the median survival time for the three groups of patients with 0, 1, and 2 adverse alleles of rs7732320 were 17.8, 13.4 and 10.6 months respectively.

Rs7732320 is located in the intronic region of *SSBP2;* we therefore investigated whether patient survival was associated with the transcript levels of *SSBP2* among 619 patients in three publically available glioblastoma gene expression datasets (Lee et al., (20) Murat et al., (21) and TCGA (22);see Methods and Supplementary Figure 1). We observed a strong and significant association of *SSBP2* expression with poorer overall survival (HR = 1.22; 95% CI: 1.09 – 1.36; P = 5.3×10^{-4}) and the association was consistent across the three expression datasets (Table 3). No effect modification by age at diagnosis was found for the association of *SSBP2* tumor expression with survival in any of the three expression datasets. Additionally, among TCGA glioblastoma patients, the HR for patient survival associated with tumor *SSBP2* expression was highest and statistically significant only among patients with the previously described (22) proneural signature (HR = 1.44; 95%CI: 1.10 – 1.89; P = 0.007) (Table 3). Consistent with this finding, we found that proneural glioblastoma patients expressed the lowest amount of *SSBP2* compared to the other subtypes (Wilcoxon P = $2.16X10^{-12}$; Figure 1A). Intriguingly, even though the overall survival for patients of the proneural subtype was not significantly different from the other gene expression subtypes (log rank P = 0.21; Figure 1B), significant survival differences were observed for the proneural *SSBP2*-negative patients (Figure 1C), arbitrarily defined as the subset of patients with lower than 25 percentile of *SSBP2* expression in the proneural group. We observed significantly better survival for proneural *SSBP2*-negative patients (median survival time: 28.8 months) than proneural *SSBP2*-positive patients (median survival of time: 12.4 months) and all other non-proneural glioblastoma patients (median survival time: 13.8 months). Proneural *SSBP2*-negative status remained a significant prognostic factor for longer survival

(Cox P = $9.7X10^{-3}$) in Cox multivariate analysis after adjusting for patient age at diagnosis and sex.

The proneural expression subtype has recently been linked to a subset of tumors exhibiting a glioma-CpG island methylator phenotype (G-CIMP)(26). To understand the relationship between *SSBP2* and the G-CIMP signature, we compared the *SSBP2* genotype and tumor expression in the set of TCGA glioblastoma samples with available G-CIMP status. Of the 241 TCGA samples with concomitant tumor expression and G-CIMP information, 24 were G-CIMP positive and they expressed a much lower level of *SSBP2* than the 217 G-CIMP negative tumors (Wilcoxon P = $3.54X10^{-4}$). Of the 151 TCGA samples with attendant *SSBP2* genotype and G-CIMP information, 2 out of the 16 (12.5%) G-CIMP positive glioblastoma patients belonged to the group with at least one copy of the adverse allele T, in contrast to a much higher proportion (28.4%, 38 out of 135) in the G-CIMP negative glioblastomas. Because of small sample sizes, validating the relationship between SSBP2 genotype, expression and G-CIMP status will require further studies.

Interestingly, *IDH1* mutation status was not found to be associated with the *SSBP2* genotype in either of the AGS and TCGA datasets, nor was it linked to *SSBP2* tumor expression in the TCGA dataset. For *TP53* mutation, we detected an increased frequency of the risk allele T of *SSBP2* in *TP53* mutated glioblastoma patients (OR = 2.35; 95% CI = 1.06–5.19; P = 0.03) in the TCGA dataset. However, this association was not found in the AGS dataset. Next, in order to perform a multivariate analysis incorporating both patient genotypes and tumor markers that are related to survival in glioblastoma patients, we used the AGS dataset, for whom 143 of the 315 patients with standard-of-care treatment had data on *TP53* and *IDH1* mutation status, and *EFGR* amplification. Unfortunately only 35 of the 115 TCGA patients with standard-of-care treatment had both *IDH1* and *TP53* mutation data. In a Cox multivariate regression including age, sex, *IDH1* mutation status, *EGFR* copy number, *TP53* mutation and *SSBP2* rs7732320 genotype, *SSBP2* genotype remained an independent predictor of poorer survival (HR=1.99; 95%CI: 1.32–3.00, P=0.001, n=143)

Taken together, the findings above present a consistent connection by showing that both the adverse *SSBP2* inherited variant and increased *SSBP2* expression in tumors are associated with shorter survival time in glioblastoma patients and that the relationship is most evident among patients with the proneural expression signature. A test for the statistical interaction between the *SSBP2* SNP rs773232 and its tumor expression was performed in the TCGA dataset by inclusion of the cross-product term in the Cox model and assessed by use of the likelihood ratio test. No statistical significant interaction effect (P=0.66) was observed.

To further localize the association with survival in the 5q14.1 region around rs7732320, we imputed non-genotyped SNPs in the entire genomic locus of *SSBP2* with a 100kb extension at its 3′ end from 80,680,000 – 80,980,000 on chromosome 5. The Hapmap II CEU dataset (27) contained 217 SNPs in this region (the AGS dataset had 31 SNPs). Out of the 186 (217 minus 31) imputed SNPs, 159 had good imputation quality for AGS, Mayo, and TCGA. Meta-analysis using a fixed effect model to combine study-specific HR estimates from age-gender adjusted Cox models shows a genome-wide statistically significant association of patient survival with the SNP rs17296479 (P = $1.0 X 10^{-7}$; see Figure 2 and Supplementary Table 4), which is located ~8kb centromeric of rs7732320. Two SNPs, rs12187089 and rs11738172, located between these two markers, also displayed strong associations with patient survival, with P = $1.2 X 10^{-7}$ and $2.3 X 10^{-7}$ respectively. These four SNPs are highly linked with each other ($r^2 > 0.8$). The smallest combined nominal P value from multivariate Cox models of patient survival with the remaining SNPs adjusting for rs17296479 genotype was 0.061, suggesting that there were no residual independent survival signals remaining.

DISCUSSION

Major strengths of this study include: 1. a large group of glioblastoma patients in the discovery study (AGS) with initial standard of care treatment of resection, radiation, and temozolomide; 2. three independent validation studies restricted to patients also treated with standard of care; 3. direct functional analysis of tumor gene expression at discovered loci at different levels; and 4. imputation to localize the SNPs most strongly associated with patient survival. Limitations of this study include the lack of detailed temozolomide dosing or timing information, and the fact that subsequent treatments at patient relapse are not included as part of the analysis. Another limitation is that tumor expression data was not available for most of the patients for whom constitutive genotyping data was available, but TCGA data did provide one group of patients with both tumor expression and constitutive genotyping. Recently, Colman et al. (28) found an approximately 3-fold hazard ratio for overall survival for glioblastoma associated with a 9-gene tumor expression signature among patients treated with temozolomide. In our analysis, we have identified a distinct subset of proneural patients with low *SSBP2* expression with a median survival time that was more than twice as long as the other glioblastoma patients. In addition, the *SSBP2* risk allele conferred a 1.64 fold increase in rate of death. As our survival analyses are done using a single SNP covariate, the inclusion of additional SNPs in combinations with tumor makers may lead to improved prognostic ability. Such an undertaking is an important future direction for research.

Despite assembling the largest sample size yet available of standard of care treated primary glioblastoma patients with genome-wide SNPs and survival data, our study is still exploratory with relatively small sample size compared to case-control genome-wide studies. The observed associations between the *SSBP2* SNP and glioblastoma patient overall survival did not reach nominal genome-wide significance in the discovery study. However, genotype imputation identified an untyped SNP (rs17296479) in *SSBP2* achieving genome-wide significance (Bonferroni corrected $p=1.0*10^{-7}*314,635 = 0.03$). Nevertheless, preventing false positive discoveries is a pertinent issue in such a large-scale study involving so many statistical tests. Consequently, we sought additional functional validation of the discovered loci by assessing their tumor gene expression association with survival. We believe these additional exercises improved our chances of deriving results that can be replicated in future studies as well as inform future functional studies.

We report here persuasive evidence for the genotypic and transcriptional association of the *SSBP2* locus with patient survival. However, establishing the nature of the regulatory relationship between the two awaits further in-depth experimental investigation. It is also possible that the variant is associated the natural history of the disease; leading to differences in time of diagnosis for carriers versus non-carriers. As yet, the variant has not been associated with glioma risk.

Using imputation for fine mapping, we identified four linked SNPs (rs17296479, rs12187089, rs11738172 and rs7732320), spanning ~12 kb at the 3′ end of *SSBP2,* that are strongly associated with patient survival. Although all four SNPs are non-coding, their immediate proximity to the gene and the ample evidence for epigenetic modifications within the region supports a possible role in transcriptional regulation of *SSBP2*. First, the histone methylation marker H3K4Me1 for enhancer elements has a broad peak encompassing three of the four variants (See Supplementary Figure 3). Second, there are three un-annotated human transcripts (AK024171, AK054959 and CR608789) located in the same region, just downstream of *SSBP2*, suggestive of a transcriptionally active genomic interval. Last and most importantly, the direct functional evidence relating the variant rs7732320 to *SSBP2* expression in glioblastomas and the unequivocal associations of patient survival with *SSBP2*

inherited variants and *SSBP2* expression levels in tumors point to a *cis* effect of the variant(s) with the disruption of the transcriptional control of *SSBP2* as the likely functional mechanism. The genotyped and imputed variants could either tag the principal association with survival attributable to this 5q14.1 locus or they themselves could be the principal culprits. Comprehensive resequencing efforts and further functional analysis will be required to unambiguously identify the causal variants.

As further evidence of the biologic plausibility of these findings, *SSBP2* has been reported to be involved in the maintenance of genome stability (29) and has been implicated in transcriptional signatures in several cancers including leukemia, (30) pancreatic cancer, (31) oligodendroglial tumors, (32) and esophageal squamous cell carcinoma. (29) A direct confirmation of the link between *SSBP2* and survival in brain cancer is further proffered by Shaw et al., (32) in which the expression of *SSBP2* was shown to be associated with response to chemotherapy in patients with oligodendroglial tumors. Evidence that the genotypic association of *SSBP2* with patient survival appears to be independent of tumor *IDH1* mutation status and strongest among patients with a proneural/G-CIMP expression signature suggests *SSBP2* may contribute to glioblastoma pathogenesis.

With further confirmation, these previously unrecognized inherited variations influencing survival may warrant inclusion in clinical trials to improve randomization and validate new therapeutic approaches. The genes identified here by SNP tags may represent potential targets for developing new drug therapies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: Work at University of California, San Francisco was supported by the National Institutes of Health (grant numbers R01CA52689 P50CA097257), as well as the National Brain Tumor Foundation, the UCSF Lewis Chair in Brain Tumor Research and by donations from families and friends of John Berardi, Helen Glaser, Elvera Olsen, Raymond E. Cooper, and William Martinusen. Work at the Mayo Clinic was supported by the National Institutes of Health (grant numbers P50CA108961 and P30 CA15083) and the Bernie and Edith Waterman Foundation. Work at Moffitt Cancer Center for the GliomaSE study was supported the National Institutes of Health (CAR01116174) as well as institutional funding from the Moffitt Cancer Center, Tampa, FL and the Vanderbilt-Ingram Comprehensive Cancer Center, Nashville, TN.

The authors wish to acknowledge study participants, the clinicians and research staffs at participating medical centers, Kenneth Aldape, deCODE genetics, Drs. Bernd Scheithauer and Caterina Gianinni, the Mayo Clinic Comprehensive Cancer Center Biospecimens and Processing, Celia Sigua, Marek Wloch, and Ms. Anna Konidari.

References

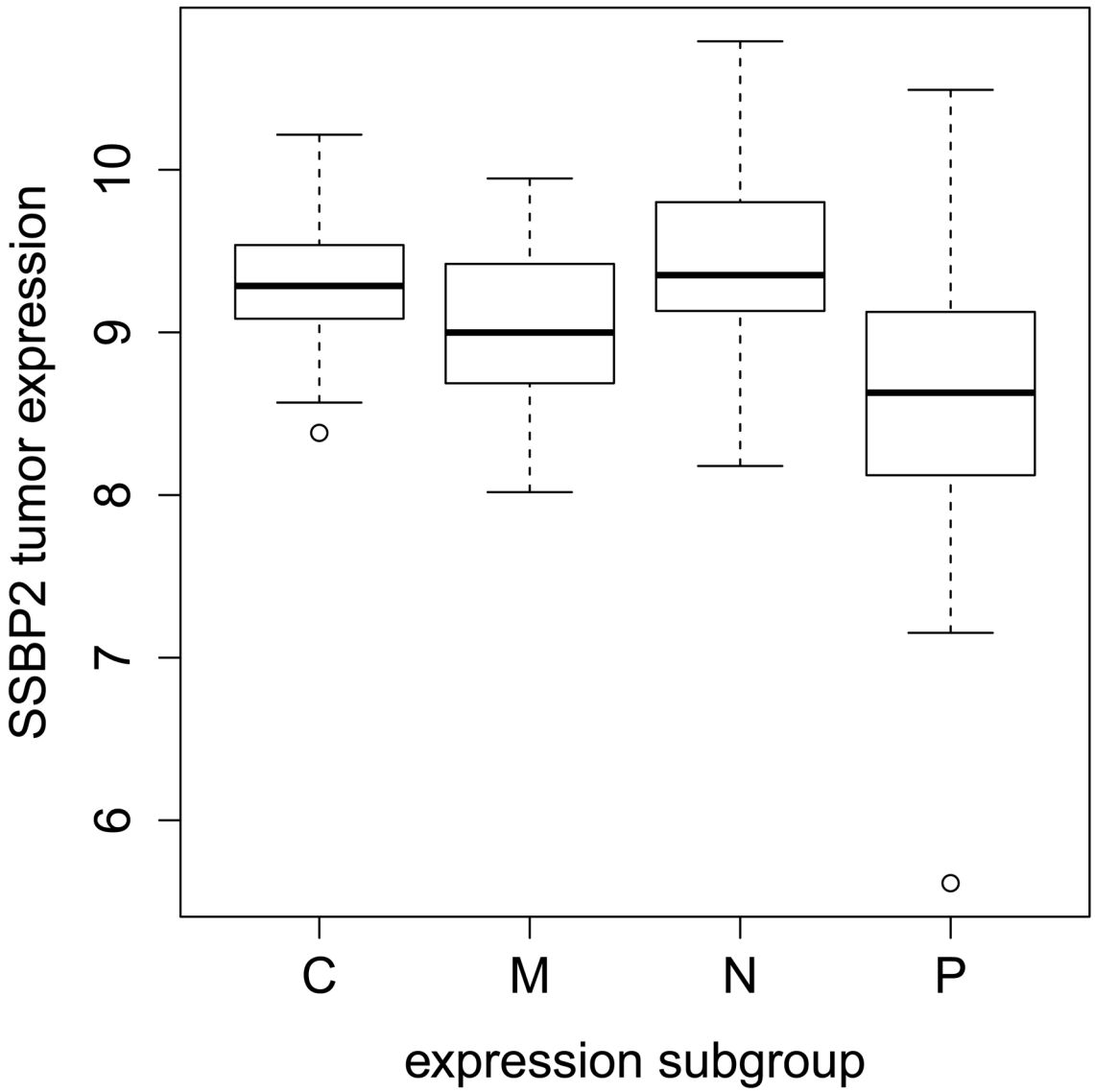
1. Wu X, Ye Y, Rosell R, Amos CI, Stewart DJ, Hildebrandt MA, et al. Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy. J Natl Cancer Inst. 2011; 103:817–25. [PubMed: 21483023]
2. Andersson U, Osterman P, Sjostrom S, Johansen C, Henriksson R, Brannstrom T, et al. MNS16A minisatellite genotypes in relation to risk of glioma and meningioma and to glioblastoma outcome. Int J Cancer. 2009; 125:968–72. [PubMed: 19405125]
3. Bhowmick DA, Zhuang Z, Wait SD, Weil RJ. A functional polymorphism in the EGF gene is found with increased frequency in glioblastoma multiforme patients and is associated with more aggressive disease. Cancer Res. 2004; 64:1220–3. [PubMed: 14973082]
4. Liu Y, Shete S, Etzel CJ, Scheurer M, Alexiou G, Armstrong G, et al. Polymorphisms of LIG4, BTBD2, HMGA2, and RTEL1 genes involved in the double-strand break repair pathway predict glioblastoma survival. J Clin Oncol. 2010; 28:2467–74. [PubMed: 20368557]

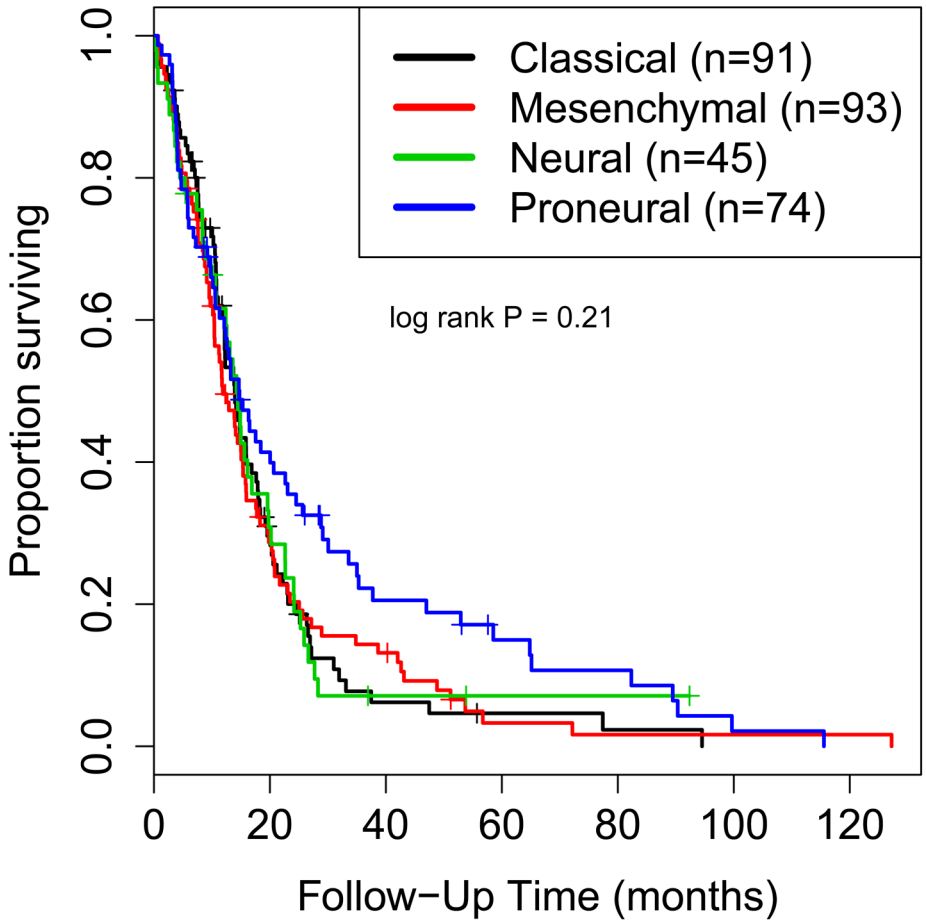
5. Okcu MF, Selvan M, Wang LE, Stout L, Erana R, Airewele G, et al. Glutathione S-transferase polymorphisms and survival in primary malignant glioma. Clin Cancer Res. 2004; 10:2618–25. [PubMed: 15102663]
6. Scheurer ME, Amirian E, Cao Y, Gilbert MR, Aldape KD, Kornguth DG, et al. Polymorphisms in the interleukin-4 receptor gene are associated with better survival in patients with glioblastoma. Clin Cancer Res. 2008; 14:6640–6. [PubMed: 18927306]
7. Sjostrom S, Andersson U, Liu Y, Brannstrom T, Broholm H, Johansen C, et al. Genetic variations in EGF and EGFR and glioblastoma outcome. Neuro Oncol. 2010; 12:815–21. [PubMed: 20197289]
8. Sjostrom S, Wibom C, Andersson U, Brannstrom T, Broholm H, Johansen C, et al. Genetic variations in VEGF and VEGFR2 and glioblastoma outcome. J Neurooncol. 2010
9. Wang L, Wei Q, Wang LE, Aldape KD, Cao Y, Okcu MF, et al. Survival prediction in patients with glioblastoma multiforme by human telomerase genetic variation. J Clin Oncol. 2006; 24:1627–32. [PubMed: 16575014]
10. Wrensch M, Jenkins RB, Chang JS, Yeh RF, Xiao Y, Decker PA, et al. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. Nat Genet. 2009; 41:905–8. [PubMed: 19578366]
11. Egan KM, Thompson RC, Nabors LB, Olson JJ, Brat DJ, Larocca RV, et al. Cancer susceptibility variants and the risk of adult glioma in a US case-control study. J Neurooncol. 2011
12. Felini MJ, Olshan AF, Schroeder JC, Carozza SE, Miike R, Rice T, et al. Reproductive factors and hormone use and risk of adult gliomas. Cancer Causes Control. 2009; 20:87–96. [PubMed: 18766447]
13. Wrensch M, McMillan A, Wiencke J, Wiemels J, Kelsey K, Patoka J, et al. Nonsynonymous coding single-nucleotide polymorphisms spanning the genome in relation to glioblastoma survival and age at diagnosis. Clin Cancer Res. 2007; 13:197–205. [PubMed: 17200355]
14. Christensen BC, Smith AA, Zheng S, Koestler DC, Houseman EA, Marsit CJ, et al. DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma. J Natl Cancer Inst. 2011; 103:143–53. [PubMed: 21163902]
15. Wiencke JK, Aldape K, McMillan A, Wiemels J, Moghadassi M, Miike R, et al. Molecular features of adult glioma associated with patient race/ethnicity, age, and a polymorphism in O6-methylguanine-DNA-methyltransferase. Cancer Epidemiol Biomarkers Prev. 2005; 14:1774–83. [PubMed: 16030116]
16. Wrensch M, Wiencke JK, Wiemels J, Miike R, Patoka J, Moghadassi M, et al. Serum IgE, tumor epidermal growth factor receptor expression, and inherited polymorphisms associated with glioma survival. Cancer Res. 2006; 66:4531–41. [PubMed: 16618782]
17. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–8. [PubMed: 18772890]
18. Cochran, WG. Biometrics. Washington, DC, USA: International Biometric Society; 1954. The Combination of Estimates from Different Experiments; p. 101-29.
19. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med. 1999; 18:321–59. [PubMed: 10070677]
20. Lee Y, Scheck AC, Cloughesy TF, Lai A, Dong J, Farooqi HK, et al. Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age. BMC Med Genomics. 2008; 1:52. [PubMed: 18940004]
21. Murat A, Migliavacca E, Gorlia T, Lambiv WL, Shay T, Hamou MF, et al. Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. J Clin Oncol. 2008; 26:3015–24. [PubMed: 18565887]
22. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010; 17:98–110. [PubMed: 20129251]
23. Huse JT, Phillips HS, Brennan CW. Molecular subclassification of diffuse gliomas: seeing order in the chaos. Glia. 2011; 59:1190–9. [PubMed: 21446051]

24. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–34. [PubMed: 21058334]
25. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med. 1998; 17:2815–34. [PubMed: 9921604]
26. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell. 2010; 17:510–22. [PubMed: 20399149]
27. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007; 449:913–8. [PubMed: 17943131]
28. Colman H, Zhang L, Sulman EP, McDonald JM, Shooshtari NL, Rivera A, et al. A multigene predictor of outcome in glioblastoma. Neuro Oncol. 2010; 12:49–57. [PubMed: 20150367]
29. Huang Y, Chang X, Lee J, Cho YG, Zhong X, Park IS, et al. Cigarette smoke induces promoter methylation of single-stranded DNA-binding protein 2 in human esophageal squamous cell carcinoma. Int J Cancer. 2011; 128:2261–73. [PubMed: 20658532]
30. Liang H, Samanta S, Nagarajan L. SSBP2, a candidate tumor suppressor gene, induces growth arrest and differentiation of myeloid leukemia cells. Oncogene. 2005; 24:2625–34. [PubMed: 15782145]
31. Baine MJ, Chakraborty S, Smith LM, Mallya K, Sasson AR, Brand RE, et al. Transcriptional profiling of peripheral blood mononuclear cells in pancreatic cancer patients identifies novel genes with potential diagnostic utility. PLoS One. 2011; 6:e17014. [PubMed: 21347333]
32. Shaw EJ, Haylock B, Husband D, du Plessis D, Sibson DR, Warnke PC, et al. Gene expression in oligodendroglial tumors. Anal Cell Pathol (Amst). 2010; 33:81–94. [PubMed: 20966545]

Statement of Translational Relevance

Glioblastoma is the most fatal form of primary brain cancer and only a few prognostic factors, age, initial Karnofsky performance status and some treatments, are known. Reliable genetic prognostic markers are still not well established. We present the first genome-wide survival and validation study for glioblastoma patients treated with the current standard of care, resection, radiation and temozolomide. Using Cox regressions for genome-wide survival analysis, followed by functional validation in tumor expression and genotype imputation, we identified a variant in *SSBP2* (single-stranded DNA-binding protein 2) and the tumor expression of *SSBP2* to be significantly associated with patient survival. Identification and characterization of the role of genetic variation in predicting glioblastoma patient survival may help optimize clinical trial study design and individualize patient treatment plans.





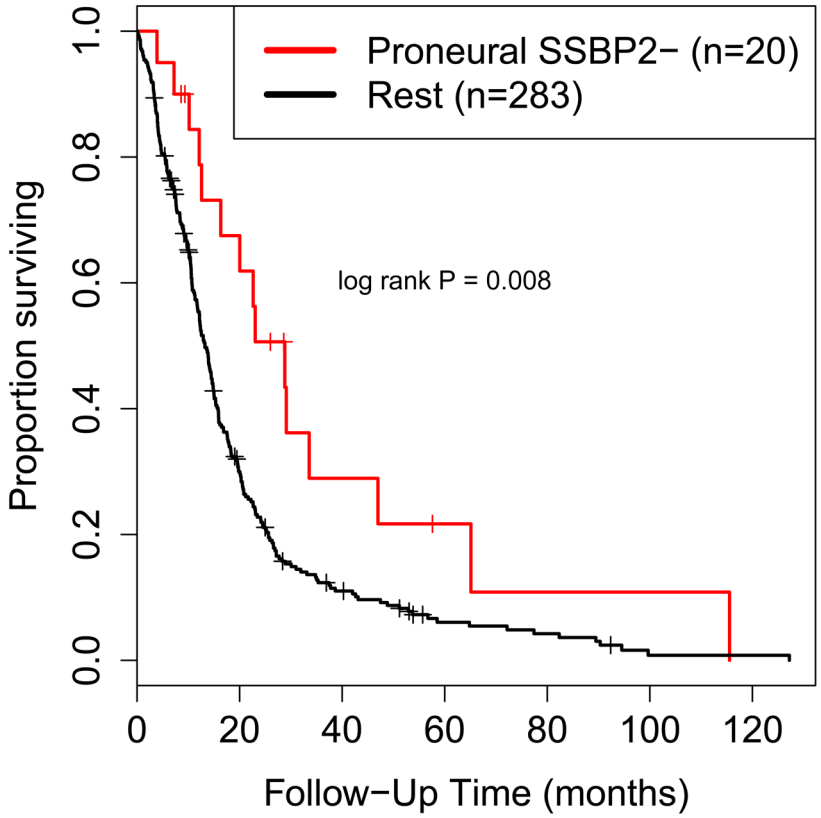


Figure 1.

(A) Boxplots of *SSBP2* tumor expression by previously assigned TCGA expression groups in 303 glioblastomas: C, classical; M, mesenchymal; N, neural; and P, proneural. (B) Kaplan-Meier survival curves for the four TCGA expression groups. (C) Kaplan-Meier survival curves based on *SSBP2* expression and TCGA expression groups. The "Proneural *SSBP2*-" group is designated as the subset of 20 patients with lower than 25 percentile expression of *SSBP2* expression in the proneural group.

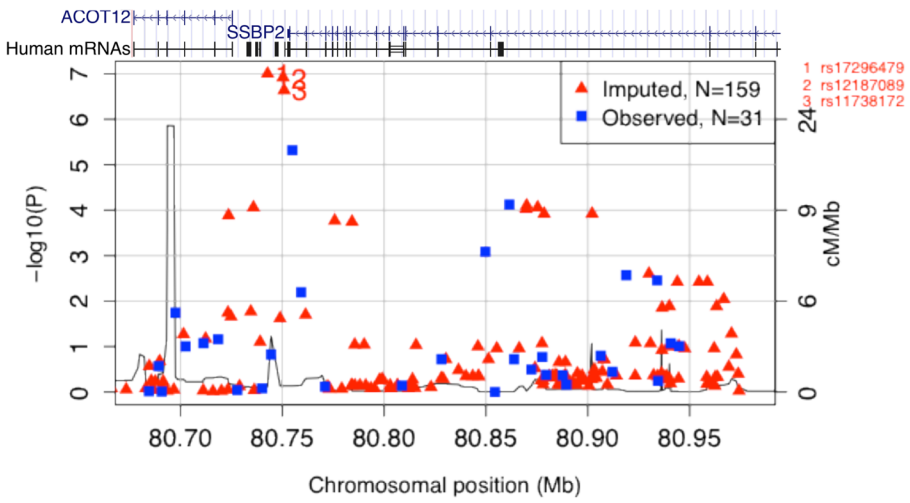


Figure 2.

Association of genetic variants near *SSBP2* with survival using data from uniformly treated glioblastoma patients. We used data from the San Francisco Adult Glioma Study, the Mayo Clinic and The Cancer Genome Atlas studies for imputation. Evidence for association at each SNP, measured as the combined -$\log_{10}$ P-value, is represented along the y-axis. The x-axis represents the placement of each SNP on chromosome 5 in genome build 36. Results for directly genotyped SNPs are colored in blue, and imputed SNPs in red. Association results are superimposed on a black line that summarizes the local recombination rate map. The upper panel indicates known RefSeq and mRNA coding sequences in the region.

Table 1

Patient characteristics of glioblastoma patients used in discovery (University of California, San Francisco, 1997–2008) and validation sets (Mayo Clinic, GliomaSE and TCGA).

	Discovery: AGS		Validation I: Mayo		Validation II: GliomaSE		Validation III: TCGA	
	N (Events/Total)	Median Survival (months)	N (Events/Total)	Median Survival (months)	N (Events/Total)	Median Survival (months)	N (Events/Total)	Median Survival (months)
Total	270/315	17.1	64/87	16.3	137/232	16.4	78/115	17.8
Age at diagnosis								
Median (interquartile range)	55 (17.3)		54 (15.0)		59 (17.4)		57 (18.0)	
HR (95% CI), P[a]	1.03 (1.02–1.04), 7.3E-09		1.03 (1.01–1.08), 4.3E-03		1.02 (1.01–1.04), 3.0E-04		1.03 (1.01–1.05), 5.7E-04	
Sex								
Female	92/101	15.4	22/35	16.3	55/85	16.4	34/48	20.4
Male	178/214	17.2	42/52	16.8	82/146	17.1	44/67	16.5
HR (95% CI), P[a]	0.82 (0.64–1.05), 0.12		1.10 (0.66–1.85), 0.72		0.98 (0.73–1.33), 0.92		1.07 (0.68–1.68), 0.77	
Race								
White	315 (100%)		87 (100%)		232 (100%)		115 (100%)	

Abbreviations: GliomaSE, glioma patients recruited from several medical centers in Southeastern United States; TCGA, The Cancer Genome Atlas; HR, hazard ratio; CI, confidence interval.

[a]P-values from log-additive Cox Proportional Hazards model adjusted for age at diagnosis (on a continuous scale) and sex.

Table 2

Association of rs7732320 genotype with survival for glioblastoma multiforme patients with initial standard of care (resection, radiation and temozolomide) treatment discovered in a genome-wide association study (University of California, San Francisco, 1997–2008, Adult Glioma Study (AGS[10])) and validated in three independent studies (Mayo Clinic[10], GliomaSE[11] and The Cancer Genome Atlas (TCGA[17])) based on combined P < 1E-5.

	Discovery AGS		Combined Validation (3 sites)		Heterogeneity Test (4 sites)		Combined Statistics (4 sites)	
SNP	**HR (95% CI)**	**P[a]**	**HR (95% CI)**	**P[b]**	**Q**	**P**	**HR (95% CI)**	**P[c]**
rs7732320 (*SSBP2, Chr 5, MA=T, MAF=0.11*)	1.80 (1.36–2.30)	3.07E-05	1.48 (1.11–1.99)	0.008	1.58	0.66	1.64 (1.34–2.00)	1.30E-06

Abbreviations: GliomaSE, glioma patients recruited from several medical centers in Southeastern United States; MA, minor allele; MAF, minor allele frequency;

[a] P-values from log-additive Cox Proportional Hazards model adjusted for age at diagnosis (on a continuous scale) and sex.

[b] P-values based on combining summary statistics from the three validation studies of Mayo, GliomaSE and TCGA using a fixed effect model with inverse variance weights.

[c] P-values based on combining summary statistics from all four study sites (AGS, Mayo, GliomaSE and TCGA) using a fixed effect model with inverse variance weights.

Table 3

Association of tumor gene expression and survival in glioblastoma multiforme cases using data from three different sources.

		Lee et al. (N=218)[20]		Murat et al. (N=75)[21]		TCGA (N=326)[22]				Combined (N=619)	
		HR (95% CI)[a]	P	HR (95% CI)[a]	P	HR (95% CI)[a]	P	Events/N	MST	HR (95% CI)	P
SSBP2	All subtypes	1.18 (1.01–1.38)	0.034	1.48 (0.88–2.51)	0.14	1.24 (1.05–1.47)	0.013			1.22 (1.09–1.36)	0.00053
	Proneural					1.44 (1.10–1.89)	0.007	65/74	14.7 (11.3–23.0)		
	Neural					1.19 (0.58–2.46)	0.63	40/45	14.3 (10.7–19.8)		
	Mesenchymal					1.27 (0.77–2.07)	0.35	86/93	11.9 (10.4–15.4)		
	Classical					1.25 (0.72–2107)	0.43	78/91	13.9 (12.1–17.6)		

Abbreviations: MST, median survival time (in months) (95% CI); TCGA, The Cancer Genome Atlas.

[a]Adjusted for age.