Imaging, Diagnosis, Prognosis

SSBP2 Variants Are Associated with Survival in Glioblastoma Patients

Yuanyuan Xiao1, Paul A. Decker5, Terri Rice2, Lucie S. McCoy2, Ivan Smirnov2, Joseph S. Patoka2, Helen M. Hansen2, Joe L. Wiemels1, Tarik Tihan3, Michael D. Prados2, Susan M. Chang5, Mitchel S. Berger2, Matthew L. Kosel6, Brooke L. Friddle5, Daniel H. Lachance7,8, Brian Patrick O’Neill7, Jan C. Buckner6, Reid C. Thompson9, Louis Burt Nabors10, Jeffrey J. Olson11, Steve Brem12, Melissa H. Madden12, James E. Browning12, John K. Wiencke2,4, Kathleen M. Egan12, Robert B. Jenkins8, and Margaret R. Wrensch2,4

Departments of 1Epidemiology and Biostatistics, 2Neurology and Biomedical Statistics and Informatics and 3Biomedical Sciences, University of California, San Francisco, San Francisco, California; Divisions of 4Neurological Surgery, 5Department of Neurosurgery, Emory School of Medicine, Atlanta, Georgia; Departments of 6Neurology and Pathology, 7Laboratory Medicine and Pathology, and 8Division of Cancer Epidemiology & Genetics, Mayo Clinic College of Medicine, Rochester, Minnesota; Departments of 9Neurology and 10Neurosurgical Oncology, Vanderbilt University Medical Center, Nashville, Tennessee; 11Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California; 12Department of Cancer Epidemiology & Genetics, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida; and 13Department of Neurosurgery, Hospital of the University of the Pennsylvania, Philadelphia, Pennsylvania

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

J.K. Wiencke, K.M. Egan, R.B. Jenkins, and M.R. Wrensch are senior authors.

Corresponding Author: Yuanyuan Xiao, Department of Epidemiology and Biostatistics, University of California, San Francisco 185 Berry St, Suite 5700, San Francisco, CA 94107. Phone: 415-514-8148; Fax: 415-514-8150; E-mail: yuanyuan.xiao@ucsf.edu

doi: 10.1158/1078-0432.CCR-11-2778

©2012 American Association for Cancer Research.

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.

Experimental Design: Cox regressions for survival with 314,635 inherited autosomal single-nucleotide polymorphisms (SNP) among 315 San Francisco Adult Glioma Study patients for discovery and three independent validation data sets [87 Mayo Clinic, 232 glioma patients recruited from several medical centers in Southeastern United States (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 data sets (total n = 619). Genotype imputation was used to estimate hazard ratios (HR) for SNPs that had not been directly genotyped.

Results: From the discovery and validation analyses, we identified a variant in single-stranded DNA-binding protein 2 (SSBP2) on 5q14.1 associated with overall survival in combined analyses (HR, 1.64; 1.3 × 10−6). Expression of SSBP2 in tumors from three independent data sets also was significantly related to patient survival (P = 5.3 × 10−10). 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 × 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. Clin Cancer Res; 18(11): 3154–62. ©2012 AACR.

Introduction

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. 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.
**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 single-stranded DNA-binding protein 2 (SSBP2) 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.

**Methods**

**Study subjects**

Informed consent was obtained from each subject. The subject recruitment and studies were conducted after approval was obtained from the Institutional 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 this 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 and colleagues (10) after stringent sample quality control filtering. Among these patients, tumor characteristics [\(\text{IDH1} (n = 173)\) and \(\text{TP53} (n = 151)\) mutation status and \(\text{EGFR}\) 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 2 Mayo Clinic neuropathologists based on surgically resected material.

The glioma patients recruited from several medical centers in Southeastern United States (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., nonrecurrent) 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 Cancer Genome Atlas (TCGA) data set was downloaded from 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 S1.

**Genotyping**

Genotyping for the AGS discovery subjects was conducted by deCODE Genetics using Illumina's HumanCNV370-duo BeadChip as previously described (10). After excluding single-nucleotide polymorphisms (SNP) with \(P < 10^{-5}\) for Hardy–Weinberg equilibrium in the AGS control samples (AGS participants that did not have glioma), or minor allele frequency less than 5%, or missing genotyping data more than 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 carried out with Illumina 610Quad SNP arrays as previously described (10). Genotyping for the GliomaSE study subjects was conducted with the Illumina Goldengate assay as previously described (11). Genotyping for the TCGA study subjects was conducted with Illumina 550 platform (17).

**Statistical analysis**

Supplementary Fig. S1 provides an overview of the 3 types of analyses conducted: (i) genome-wide constitutive discovery and validation of SNPs associated with glioblastoma patient overall survival, (ii) functional validation of survival loci (association of gene expression in tumors with glioblastoma overall patient survival), and (iii) 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 HR and 95% CI 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 data sets. The genomic inflation factor was calculated for each data set. The SNP with the highest genomic inflation factor was removed (17).
assumption for validated SNPs with a 4-site combined \( P \leq 10^{-3} \) was tested within each site with the Schoenfeld residuals, and SNPs with evidence for nonproportionality were removed from further consideration. Results for the nonproportionality test for rs7732320 are shown in Supplementary Table S2. Heterogeneity across the 4 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 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

\[
\hat{\beta}_{\text{Combined}} = \frac{\sum \hat{\beta}_i / v_i}{\sum 1 / v_i}, \quad \text{var}(\hat{\beta}_{\text{Combined}}) = \frac{1}{\sum 1 / v_i}.
\]

where \( \hat{\beta}_i \) and \( v_i \) are the log HR 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 3 published studies (20–22). The Lee and colleagues (20) data set described 218 glioblastoma expression samples including 132 samples from 3 previously published data sets as well as 86 new samples assembled into a single, unified data set with Affymetrix U133A. The Murat and colleagues (21) data set contains 75 glioblastoma expression samples using the Affymetrix U133. Normalized expression values using the standard RMA method for the Lee and Murat data sets were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus database (GSE13041 and GSE7696). The TCGA data set (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 and colleagues (22) to previously unclassified samples. In total, we obtained 74 proneurals, 45 neurals, 93 mesenchymals, and 91 classicals. For each of the 3 expression data sets, 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 data set, 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 CELI HapMap data (MACH version 1.0.16), we imputed SNPs separately within each of the 3 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 with expected allele counts as the predictor for a total of 159 SNPs, whose variance ratios were larger than 0.5 for all 3 studies to exclude SNPs with poor quality imputed genotypes. Meta-analysis of the imputed data was carried out 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 conducted by the R statistical package.

**Results**

Patient characteristics (age, sex, and median survival) for the 4 data sets (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 data set conformed to the identity line in the Q-Q plot, whereas 90 SNPs showed significant deviation from expectation at \( P = 10^{-6} \) (Supplementary Fig. S2). 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 2 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 4 data sets (Table 2 and Supplementary Table S3). 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 \( P = 1.3 \times 10^{-6} \). There was no evidence of heterogeneity of the HR estimates across the 4 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 3 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 3 publically available glioblastoma gene expression data sets [Lee and colleagues (20), Murat and colleagues (21), and TCGA (22); see Methods and Supplementary Fig. S1]. We observed a strong and significant association of SSBP2 expression with poorer overall survival (HR, 1.22; 95% CI, 1.09–1.36; \( P = 5.3 \times 10^{-4} \)), and the association was consistent across the 3 expression data sets (Table 3). No effect modification by age at diagnosis was found for the association of SSBP2 tumor expression with survival in any
SSBP2 Variants Are Associated with Survival in GBM Patients

Table 1. Characteristics of glioblastoma patients used in discovery (University of California, San Francisco, 1997–2008) and validation sets (Mayo Clinic, GliomaSE, and TCGA)

<table>
<thead>
<tr>
<th></th>
<th>Discovery: AGS</th>
<th>Validation I: Mayo</th>
<th>Validation II: GliomaSE</th>
<th>Validation III: TCGA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (events/total)</td>
<td>Median survival (mo)</td>
<td>N (events/total)</td>
<td>Median survival (mo)</td>
</tr>
<tr>
<td>Total</td>
<td>270/315</td>
<td>17.1</td>
<td>64/87</td>
<td>16.3</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (interquartile range)</td>
<td>55 (17.3)</td>
<td>54 (15.0)</td>
<td>59 (17.4)</td>
<td>57 (18.0)</td>
</tr>
<tr>
<td>HR (95% CI), P*</td>
<td>1.03 (1.02–1.04), 7.3E-09</td>
<td>1.03 (1.01–1.08), 4.3E-03</td>
<td>1.02 (1.01–1.04), 3.0E-04</td>
<td>1.03 (1.01–1.05), 5.7E-04</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>92/101</td>
<td>15.4</td>
<td>22/35</td>
<td>16.3</td>
</tr>
<tr>
<td>Male</td>
<td>178/214</td>
<td>17.2</td>
<td>42/52</td>
<td>16.8</td>
</tr>
<tr>
<td>HR (95% CI), P*</td>
<td>0.82 (0.64–1.05), 0.12</td>
<td>1.10 (0.66–1.85), 0.72</td>
<td>0.98 (0.73–1.33), 0.92</td>
<td>1.07 (0.68–1.68), 0.77</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>315 (100%)</td>
<td>87 (100%)</td>
<td>232 (100%)</td>
<td>115 (100%)</td>
</tr>
</tbody>
</table>

Abbreviations: GliomaSE, glioma patients recruited from several medical centers in Southeastern United States; TCGA, The Cancer Genome Atlas.

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

of the 3 expression data sets. In addition, 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 with the other subtypes (Wilcoxon P = 2.16 × 10^{-12}; Fig. 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; Fig. 1B), significant survival differences were observed for the proneural SSBP2-negative patients (Fig. 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 time, 12.4 months) and all other nonproneural glioblastoma patients (median survival time, 13.8 months). Proneural SSBP2-negative status remained a significant prognostic factor for longer survival (Cox P = 9.7 × 10^{-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

Table 2. Association of rs7732320 genotype with overall survival for glioblastoma multiforme patients with initial standard of care (resection, radiation, and temozolomide) treatment

<table>
<thead>
<tr>
<th>SNP</th>
<th>Discovery AGS</th>
<th>Combined validation (3 sites)</th>
<th>Heterogeneity test (4 sites)</th>
<th>Combined statistics (4 sites)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7732320 (SSBP2, Chr 5, MA = T, MAF = 0.11)</td>
<td>1.80 (1.36–2.30)</td>
<td>1.48 (1.11–1.99)</td>
<td>0.008</td>
<td>1.58</td>
</tr>
</tbody>
</table>

NOTE: SNP discovered in a genome-wide survival association study [University of California, San Francisco, 1997–2008, AGS (10)] and validated in 3 independent studies [Mayo Clinic (10), GliomaSE (11), and TCGA (17)] based on combined P < 1E-5.

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

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

*P values based on combining summary statistics from the 3 validation studies of Mayo, GliomaSE, and TCGA using a fixed effect model with inverse variance weights.

*P values based on combining summary statistics from all 4 study sites (AGS, Mayo, GliomaSE, and TCGA) using a fixed effect model with inverse variance weights.
methylator phenotype (G-CIMP; ref. 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 \( P = 3.54 \)

### Table 3. Association of gene expression and survival in glioblastoma multiforme cases with data from 3 different sources

<table>
<thead>
<tr>
<th>Source</th>
<th>N</th>
<th>HR (95% CI)^a</th>
<th>P</th>
<th>HR (95% CI)^a</th>
<th>P</th>
<th>Events/N</th>
<th>MST</th>
<th>HR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSBP2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All subtypes</td>
<td>218</td>
<td>1.18 (1.01–1.38)</td>
<td>0.034</td>
<td>1.48 (0.88–2.51)</td>
<td>0.14</td>
<td></td>
<td></td>
<td>1.24 (1.05–1.47)</td>
<td>0.013</td>
</tr>
<tr>
<td>Proneural</td>
<td></td>
<td>1.44 (1.10–1.89)</td>
<td>0.007</td>
<td>65/74</td>
<td></td>
<td></td>
<td></td>
<td>14.7 (11.3–23.0)</td>
<td></td>
</tr>
<tr>
<td>Neural</td>
<td></td>
<td>1.19 (0.58–2.46)</td>
<td>0.63</td>
<td>40/45</td>
<td></td>
<td></td>
<td></td>
<td>14.3 (10.7–19.8)</td>
<td></td>
</tr>
<tr>
<td>Mesenchymal</td>
<td></td>
<td>1.27 (0.77–2.07)</td>
<td>0.35</td>
<td>86/93</td>
<td></td>
<td></td>
<td></td>
<td>11.9 (10.4–15.4)</td>
<td></td>
</tr>
<tr>
<td>Classical</td>
<td></td>
<td>1.25 (0.72–2.17)</td>
<td>0.43</td>
<td>78/91</td>
<td></td>
<td></td>
<td></td>
<td>13.9 (12.1–17.6)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: MST, median survival time (in mo).
^aAdjusted for age.

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 4 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 versus the rest of the TCGA patients.
SSBP2 Variants Are Associated with Survival in GBM Patients

Figure 2. Association of genetic variants near SSBP2 with survival using data from uniformly treated glioblastoma patients. We used data from the San Francisco AGS, 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 marked with squares, and imputed SNPs with triangles. 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.

\[ P = 1.99 \times 10^{-4} \]. Of the 131 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 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 data sets, nor was it linked to SSBP2 tumor expression in the TCGA data set. 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, 1.26–4.49; \( P = 0.03 \)) in the TCGA data set. However, this association was not found in the AGS data set. Next, to conduct a multivariate analysis incorporating both patient genotypes and tumor markers that are related to survival in glioblastoma patients, we used the AGS data set, for whom 143 of the 315 patients with standard-of-care treatment had data on TP53 and IDH1 mutation status, and EGFR 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 rs7732320 and its tumor expression was carried out in the TCGA data set 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 nongenotyped SNPs in the entire genomic locus of SSBP2 with a 100 kb extension at its 3’ end from 80,680,000–80,980,000 on chromosome 5. The Hapmap II CEU data set (27) contained 217 SNPs in this region (the AGS data set 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 (\( P = 1.0 \times 10^{-7} \); see Fig. 2 and Supplementary Table S4), which is located approximately 8 kb centromeric of rs7732320. Two SNPs, rs12187089 and rs11738172, located between these 2 markers, also displayed strong associations with patient survival, with \( P = 1.2 \times 10^{-7} \) and \( 2.3 \times 10^{-7} \), respectively. These 4 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: (i) a large group of glioblastoma patients in the discovery study (AGS) with initial standard of care treatment of resection, radiation, and temozolomide; (ii) three independent validation studies...
restricted to patients also treated with standard of care; (iii) direct functional analysis of tumor gene expression at discovered loci at different levels; and (iv) 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 were not available for most of the patients for whom constitutive genotyping data were available, but TCGA data did provide one group of patients with both tumor expression and constitutive genotyping. Recently, Colman and colleagues (28) found an approximately 3-fold HR 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 with 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, genotypic imputation identified an untyped SNP (rs17296479) in SSBP2 achieving genome-wide significance (Bonferroni corrected \( P = 1.0 \times 10^{-7} \times 314,635 = 0.03 \)). 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 2 awaits further in-depth experimental investigation. It is also possible that the variant is associated with the natural history of the disease; leading to differences in time of diagnosis for carriers versus noncarriers. As yet, the variant has not been associated with glioma risk.

Using imputation for fine mapping, we identified 4 linked SNPs (rs17296479, rs12187089, rs11738172, and rs7732320), spanning approximately 12 kb at the 3′ end of SSBP2, that are strongly associated with patient survival. Although all 4 SNPs are noncoding, 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 3 of the 4 variants (See Supplementary Fig. S3). Second, there are 3 unannotated 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 and colleagues (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 seems 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.

Disclosure of Potential Conflicts of Interest

M.S. Berger: consultant/advisory board, JIVI Health Services and Pharmaco-Kinesis Corp. L.B. Nabors has an uncompensated position at Merck KGaA. The other authors disclosed no potential conflicts of interest.

Authors’ Contributions

Development of methodology: Y. Xiao, J.L. Wiemels, M.R. Wiensch

Xiao et al.


Study supervision: B.L. Friddy, R.C. Thompson, J.J. Olson, M.H. Madden, K.M. Egan, R.B. Jenkins, M.R. Wrensch

Acknowledgments

The authors thank study participants, the clinicians, and research staffs at participating medical centers, Kenneth Adlape, deCODE genetics, the late Dr. Bernd Scheithauer, Dr. Caterina Gianinni, the Mayo Clinic Comprehensive Cancer Center Biorepositories and Processing, Celia Siga, Marek Wloch, and Ms. Anna Konidari. The ideas and opinions expressed herein are those of the author(s) and endorsement by the 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 National Cancer Institute and National Human Genome Research Institute. Information about TCGA and the investigators and institutions that constitute the TCGA research network can be found at “http://cancergene.gov”.

References


Grant Support

The work at University of California, San Francisco was supported by the NIH (grant numbers RO1CA52689 and 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 Martinussen. The work at the Mayo Clinic was supported by the NIH (grant numbers P50CA108961 and P30 CA15083) and the Bermie and Edith Waterman Foundation. The work at Moffitt Cancer Center for the GliomaSE study was supported by the NIH (CA0011674) as well as institutional funding from the Moffitt Cancer Center, Tampa, FL, and the Vanderbilt Ingram Comprehensive Cancer Center, Nashville, TN. 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 103855; 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 Registrars, under agreement #11587 DP000807-1 awarded to the Public Health Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 27, 2011; revised February 23, 2012; accepted March 12, 2012; published OnlineFirst April 3, 2012.

SSBP2 Variants Are Associated with Survival in GBM Patients

www.aacrjournals.org Clin Cancer Res; 18(11) June 1, 2012 3161

3161
glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010;17:98–110.


SSBP2 Variants Are Associated with Survival in Glioblastoma Patients

Yuanyuan Xiao, Paul A. Decker, Terri Rice, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-11-2778

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2012/04/03/1078-0432.CCR-11-2778.DC1

Cited articles
This article cites 31 articles, 13 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/18/11/3154.full.html#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.