Published in final edited form as:

ACS Chem Neurosci. 2018 January 17; 9(1): 80–84. doi:10.1021/acscchemneuro.7b00165.

Combined Proteomic-Molecular Epidemiology Approach to Identify Precision Targets in Brain Cancer

Ekaterina Mostovenko[†], Yanhong Liu^{‡,§}, E. Susan Amirian[‡], Spiridon Tsavachidis[‡], Georgina N. Armstrong[‡], Melissa L. Bondy^{‡,§}, and Carol L. Nilsson^{||}

[†]Department of Anatomy, Virginia Commonwealth University, 1217 E. Marshall St., Richmond, Virginia 23284 United States

[‡]Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas 77030, United States

[§]Department of Medicine, Baylor College of Medicine, Houston, Texas 77030, United States

^{||}Department of Clinical Sciences, Lund University, SE-221 84 Lund, Sweden

Abstract

Primary brain tumors are predominantly malignant gliomas. Grade IV astrocytomas (glioblastomas, GBM) are among the most deadly of all tumors; most patients will succumb to their disease within 2 years of diagnosis despite standard of care. The grim outlook for brain tumor patients indicates that novel precision therapeutic targets must be identified. Our hypothesis is that the cancer proteomes of glioma tumors may contain protein variants that are linked to the aggressive pathology of the disease. To this end, we devised a novel workflow that combined variant proteomics with molecular epidemiological mining of public cancer data sets to identify 10 previously unrecognized variants linked to the risk of death in low grade glioma or GBM. We hypothesize that a subset of the protein variants may be successfully developed in the future as novel targets for malignant gliomas.

Graphical Abstract

Protein Variants and GBM Risk

ORCID

Carol L. Nilsson: 0000-0002-2838-8751

Notes

The authors declare no competing financial interest.

Keywords

Glioblastoma; GBM; molecular epidemiology; proteomic; precision medicine; drug target

Malignant gliomas account for 70% of primary brain tumors, with an incidence rate of 5.27 per 100 000 persons per year.¹ Despite recent advances in treatment, including surgical resection followed by concurrent chemotherapy with radiation, the majority of these patients display progressive disease and subsequent death. The overall 5 and 10 year survival rates are 29% and 25%, respectively, in the United States (American Cancer Society; www.cancer.org). Glioma survival differs significantly by histology and age. The most common and aggressive form, glioblastoma (GBM, grade IV astrocytoma), has a median survival of approximately 12–17 months. Patients with anaplastic astrocytoma (grade III) survive 2–5 years; those with low-grade astrocytomas (LGG, grade II) survive 6–8 years.²

One hypothesis for tumor recurrence is the existence of glioma stem cells (GSCs) in areas of the brain that are surgically inaccessible or resistant to the standard-of-care (SOC) treatment that entails maximal safe tumor resection, radiation, and Temozolomide treatment.³ Recently, a comparative study of primary and recurrent tumor samples from 70 patients treated with radio- and chemotherapy showed a > 10-fold enrichment in tumor stem-like cells in the recurrent tumors.⁴ Those treatments also alter the phenotype of the GSCs, turning them into a more proliferative cell type. The underlying events that lead to the phenotypic change involve selective pressures upon the GSC populations, resulting in alterations in metabolic and signaling pathways that maintain cell proliferation and resistance to cytotoxic stimuli. Thus, the dismal clinical outlook of GBM is at least partly due to the ineffectiveness of SOC treatments on GSCs.

Our hypothesis is that protein variants linked to decreased survival times (mortality) may provide a new source of precision drug targets for the treatment of brain cancer. Our rationale for this approach is that structural differences in variant proteins can result in altered functions resulting in promotion of aggressive growth and recurrence of malignant tumors. Within this context, we have developed a new workflow to identify proteins linked to increased risk of death in brain cancer (Figure 1). Through a novel approach that combines variant proteomics with molecular epidemiology, we have identified 10 protein variants, not previously linked to brain cancer, that carry a statistically increased linked association to decreased survival in GBM or low grade glioma (LGG).

We previously published an approach to perform large-scale identification of variant proteins in glioma stem cells.⁵ Briefly, we performed whole proteome and transcriptome characterization in 36 low passage GSC lines isolated from patient tumors. The proteomic data was searched against a custom database that included protein variant features, derived from all chromosomes. The existence of 126 expressed protein variants was validated by use of transcript matching and mass spectrometric parallel reaction monitoring of variant peptides.

The role of protein-expressed genetic variants in relation to glioma prognosis is not well-studied. One recent report demonstrated that new potential therapeutic targets could be

identified successfully in GSCs through a combination of gene expression analysis, rtPCR, and immunoblotting of clinical samples.⁶ In this report, we applied survival analysis to investigate whether genetic variations, as expressed non-synonymous SNPs in our proteomic data, are associated with GBM and LGG survival in TCGA and M. D. Anderson Cancer Center (MDACC) data sets. Because GBM and LGG differ substantially in prognosis, survival analyses were performed on each subgroup separately.

RESULTS AND DISCUSSION

Patient Characteristics

The TCGA study included 380 GBM patients and 426 LGG patients, whereas the MDACC study consisted of 662 GBM (grade IV) and 562 LGG (264 grade-II and 298 grade-III) patients. Among the characteristics listed in Table 1, age showed consistent significance in association with glioma (both GBM and LGG) overall survival in the two studies. Moreover, radiation therapy and chemotherapy showed association with GBM patient data in the TCGA study, and both GBM and LGG overall survival in the MDACC study.

A comparison of the characteristics of the MDACC and TCGA studies revealed large discrepancies in age at diagnosis and survival time. For TCGA patients, the median survival time (MST) was 1.2 years for GBM and 7.9 years for LGG patients; whereas for the MDACC patients, the MST was 1.4 years for GBM (0.2 year longer than TCGA) and 5.6 years for LGG patients (2.3 years shorter than TCGA). The mean age was 60 for GBM and 42 for LGG patients in the TCGA Study; whereas for the MDACC patients mean age was 53 for GBM and 40 for LGG patients.

Associations between SNPs and Overall Survival

Of the 57 nsSNPs evaluated, back-translated from identified SAV peptides in the proteomic data, four showed a statistically significant correlation with GBM overall survival, and six were correlated with LGG overall survival according to the log-rank test and Cox regression analysis (Table 2 and Figure 2). For GBM survival, the strongest significance was associated with *PREX1* Ser1559Thr and *DNAAF5* Val632Ala. For LGG survival, the strongest associations were attained at *NUP210* Arg786Gln and *PCMT1* Val143Ile.

PREX1 (phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 protein, Figure 2A) has never been associated with glioma, but it does have a previous association of in breast cancer. Ebi et al. established that PI3K regulates MEK/ERK signaling in breast cancer via PREX1, a Rac-GEF.⁷ The finding that PREX1 Ser1559Thr has an association to decreased survival time in GBM is novel and warrants further studies. *DNAAF5* (dynein axonemal assembly factor 5) is a motility protein that has no previously known link to cancer. Heterozygosity for *DNAAF5* Val632Ala is associated with shorter survival time in GBM ($P = 0.016$, Figure 2B).

In the case of *NUP210* (nuclear pore membrane glycoprotein 210), data was available for individuals, both heterozygous and homozygous (rs2280084), in the data sets. The protein variant was highly significant in this study (P -value 0.006). The nuclear pore complex is a very large structure that bridges the nuclear envelope, forming a structure that regulates the

exchange of between the nucleus and the cytoplasm.⁸ NUP210 is a major component of the nuclear pore complex. An examination of Figure 2E shows an increased risk of death in GBM patients heterozygous for NUP210 Arg786Gln as well as a further increase when the patient is homozygous for the variant. This particular NUP variant was detected at the protein level in a variant proteomic screen of breast cancer cells.⁹ Expression of NUP210 has been detected in two human brain cancer cell lines, U251 and YKG-1.¹⁰

Protein-L-isoaspartate (D-aspartate) *O*-methyltransferase (PCMT1) Val143Ile was found to be associated with shorter survival times in the MDACC study ($P = 0.016$). This enzyme plays a role in protein repair by recognizing and converting D-aspartyl and L-isoaspartyl residues resulting from spontaneous deamidation back to the normal L-aspartyl form. It was found to be positive in brain cancer in a study of alternatively spliced mRNA;¹¹ however, the variant Val143Ile has not been previously linked to cancer.

In summary, our approach to combine proteomic and genetic epidemiological data in large patient cohorts revealed 10 protein variants with significant correlations to patient hazard risks in GBM and LGG. However, there are limitations in our study. First, some of our findings reach statistical significant level after multiple testing. Second, the population stratification differs between the TCGA and MDACC studies. Finally, one must recognize the genetic heterogeneity in a patient population that contains individuals with different treatments, due to the heterogeneity of histopathological and clinical presentation of the disease.

In conclusion, we have provided evidence to implicate 10 genetic polymorphisms as predictors of GBM and LGG survival. It is highly likely that the polymorphisms of these genes will be novel and potentially prognostic biomarkers for glioma survival and more importantly, potentially new therapeutic targets. With further confirmation, these previously unrecognized inherited variations influencing survival may warrant inclusion in clinical trials to improve randomization and validate new therapeutic approaches.

METHODS

Study Population and SNPs Data

Our study was based on two independent series of Caucasian glioma patients: (1) The Cancer Genome Atlas (TCGA) glioma study. TCGA study employed whole exome sequencing by use of the NimbleGen SeqCap EZ Human Exome Library v3.0. The genotype data and corresponding clinical information (age, gender, follow-up times, vital status) for TCGA data sets were downloaded from TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>, accessed in September 2015). Fifty-seven of the 70 SNPs have genotype data available from TCGA study. (2) The MDACC glioma study included 662 GBM and 560 low and middle grade gliomas.^{12,13} Treatment and survival data (dates of death or last contact) were collected retrospectively from medical record review for all patients. Dates of death were confirmed by querying the Social Security Death Index. The MDACC study were genotyped using the Human 610-Quad Bead Chips (Illumina, San Diego, CA). Twenty-one of the 70 SNPs had genotype data available from the Illumina 610 K chip to consider in the survival tests.

Statistical Methods

Survival time was defined as the time between the date of diagnosis and date of death for deceased patients or the last contact date for living patients. The overall survival time was estimated using Kaplan–Meier methods, and log-rank analysis was performed to compare survival curves between groups. Hazard ratios (HRs) and their corresponding 95% confidence intervals (CIs) were estimated using Cox regression, with adjustment for age, sex, extent of resection (not available in the TCGA study), chemotherapy, and radiation therapy.

Acknowledgments

Funding

National Institutes of Health grants (R01 CA119215 M.L.B, R01 CA070917 M.L.B, R01 CA139020 M.L.B, and K07 CA181480 to Y.L)

We thank the Cancer Genome Atlas consortium for making the SNP and clinical data available.

References

- Ohgaki H, Kleihues P. Population-Based Studies on Incidence, Survival Rates, and Genetic Alterations in Astrocytic and Oligodendroglial Gliomas. *J Neuropathol Exp Neurol*. 2005; 64:479–489. [PubMed: 15977639]
- Wen PY, Kesari S. Malignant Gliomas in Adults. *N Engl J Med*. 2008; 359:492–507. [PubMed: 18669428]
- Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, Marosi C, Vecht CJ, Mokhtari K, Wesseling P, Villa S, Eisenhauer E, Gorlia T, Weller M, Lacombe D, Cairncross JG, Mirimanoff RO. Effects of radiotherapy with concomitant and adjuvant Temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol*. 2009; 10:459–466. [PubMed: 19269895]
- Tamura K, Aoyagi M, Ando N, Ogishima T, Wakimoto H, Yamamoto M, Ohno K. Expansion of CD133-positive glioma cells in recurrent de novo glioblastomas after radiotherapy and chemotherapy. *J Neurosurg*. 2013; 119:1145–1155. [PubMed: 23991844]
- Lichti CF, Mostovenko E, Wadsworth PA, Lynch GC, Pettitt BM, Sulman EP, Wang Q, Lang FF, Rezeli M, Marko-Varga G, Végvári Á, Nilsson CL. Systematic Identification of Single Amino Acid Variants in Glioma Stem-Cell-Derived Chromosome 19 Proteins. *J Proteome Res*. 2015; 14:778–786. [PubMed: 25399873]
- Stangeland B, Mughal AA, Grieg Z, Sandberg CJ, Joel M, Nygård S, Meling T, Murrell W, Vik Mo EO, Langmoen IA. Combined expressional analysis, bioinformatics and targeted proteomics identify new potential therapeutic targets in glioblastoma stem cells. *Oncotarget*. 2015; 6:26192–26215. [PubMed: 26295306]
- Ebi H, Costa C, Faber AC, Nishtala M, Kotani H, Juric D, Della Pelle P, Song Y, Yano S, Mino-Kenudson M, Benes CH, Engelman JA. PI3K regulates MEK/ERK signaling in breast cancer via the Rac-GEF, P-Rex1. *Proc Natl Acad Sci U S A*. 2013; 110:21124–21129. [PubMed: 24327733]
- Lusk CP, King MC. The nucleus: keeping it together by keeping it apart. *Curr Opin Cell Biol*. 2017; 44:44–50. [PubMed: 28236735]
- Bunger MK, Cargile BJ, Sevinsky JR, Deyanova E, Yates NA, Hendrickson RC, Stephenson JL. Detection and Validation of Non-synonymous Coding SNPs from Orthogonal Analysis of Shotgun Proteomics Data. *J Proteome Res*. 2007; 6:2331–2340. [PubMed: 17488105]
- Fukushima T, Tezuka T, Shimomura T, Nakano S, Kataoka H. Silencing of Insulin-like Growth Factor-binding Protein-2 in Human Glioblastoma Cells Reduces Both Invasiveness and Expression

of Progression-associated Gene CD24. *J Biol Chem.* 2007; 282:18634–18644. [PubMed: 17475624]

11. Wang Z, Lo HS, Yang H, Gere S, Hu Y, Buetow KH, Lee MP. Computational Analysis and Experimental Validation of Tumor-associated Alternative RNA Splicing in Human Cancer. *Cancer Res.* 2003; 63:655–657. [PubMed: 12566310]
12. Liu Y, Shete S, Etzel CJ, Scheurer M, Alexiou G, Armstrong G, Tsavachidis S, Liang FW, Gilbert M, Aldape K, Armstrong T, Houlston R, Hosking F, Robertson L, Xiao Y, Wiencke J, Wensch M, Andersson U, Melin BS, Bondy M. Polymorphisms of LIG4, BTBD2, HMGA2, and RTEL1 Genes Involved in the Double-Strand Break Repair Pathway Predict Glioblastoma Survival. *J Clin Oncol.* 2010; 28:2467–2474. [PubMed: 20368557]
13. Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, Simon M, Marie Y, Boisselier B, Delattre JY, Hoang-Xuan K, Hallani SE, Idbaih A, Zelenika D, Andersson U, Henriksson R, Bergenheim AT, Feychting M, Lonn S, Ahlbom A, Schramm J, Linnebank M, Hemminki K, Kumar R, Hepworth SJ, Price A, Armstrong G, Liu Y, Gu X, Yu R, Lau C, Schoemaker M, Muir K, Swerdlow A, Lathrop M, Bondy M, Houlston RS. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet.* 2009; 41:899–904. [PubMed: 19578367]

Figure 1.
Workflow to define protein variants as potential precision targets in brain cancer.

Figure 2.

Kaplan–Meier survival plots for the most significant protein variants derived from nsSNPs listed in Table 2. Panels (A)–(D) show GBM survival: (A) *PREX1*, rs3936192; (B) *DNAAF5*, rs4720951; (C) *SDHA*, rs6962; (D) *MTHFD1*, rs2236225. Panels (E)–(H) are linked to LGG survival: (E) *NUP210*, rs2280084; (F) *NSFL1C*, rs9575; (G) *TOR1AIP1*, rs1281378; (H) *HEPACAM*, rs10790715. The *Y*-axis is survival probability, and the *X*-axis is survival time. Thus, each line represents survival over time for the individuals with that genotype. If one line is higher than the other across the graph, then individuals with that genotype had better survival over time compared to those with the other genotype. Genotype AA is the wild-type homozygous; AB is for heterozygous; BB is for variant homozygous.

Table 1
Demographics and Treatment Information for the Study Population for the Two Different Data Sets^a

variable	GBM				LGG			
	no. patients (death) ^b	MST (y)	log-rank <i>P</i>	Cox regression HR (95% CI) ^c	no. patients (death) ^b	MST (y)	log-rank <i>P</i>	Cox regression HR (95% CI) ^c
TCGA Study	<i>N</i> = 380				<i>N</i> = 426			
<i>N</i> , survival	380 (252)	1.15			426 (65)	7.88		
mean (SD)	1.2 (1.3)				3.4 (3.1)			
range	0.01–10.6				0.6–11.2			
age diagnosis (y)			<0.0001			<0.0001		
mean (SD)	59.8 ± 13.6				43 ± 13.3			
range	21–89				14–87			
<48	68 (36)	1.93		reference	283 (31)	8.77		reference
48	312 (216)	1.06		2.23 (1.55–3.20)	143 (34)	4.83		3.89 (2.33–6.49)
sex			0.56			0.75		
male	240 (162)	1.08		reference	235 (33)	8.77		reference
female	140 (90)	1.30		0.77 (0.59–1.01)	191 (32)	7.88		1.08 (0.67–1.77)
radiation therapy			<0.0001			0.1		
yes	283 (170)	1.23		reference	232 (43)	6.26		reference
no	97 (82)	0.24		2.81 (2.15–3.66)	194 (22)	9.78		0.65 (0.39–1.09)
chemotherapy			<0.0001			0.24		
yes	292 (184)	1.15		reference	200 (38)	6.26		reference
no	88 (68)	0.33		1.89 (1.43–2.51)	226 (27)	9.78		0.75 (0.45–1.22)
MDACC Study								
<i>N</i> , survival	662 (554)	1.4			562 (283)	5.6		
mean (SD)	2.2 ± 2.4				6.3 ± 4.5			
range	0.01–17.3				0.03–24.5			
age diagnosis (y)			<0.0001			<0.0001		
mean (SD)	52.5 ± 11.7				40.2 ± 12.1			
range	18–86				18–80			
<48	196 (143)	1.8		reference	411 (184)	6.9		reference
48	466 (411)	1.3		1.85 (1.53–2.24)	151 (99)	2.4		2.26 (1.77–2.89)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

variable	GBM					LGG				
	no. patients (death) ^b	MST (y)	log-rank <i>P</i>	Cox regression HR (95% CI) ^c	no. patients (death) ^b	MST (y)	log-rank <i>P</i>	Cox regression HR (95% CI) ^c		
sex			0.18				0.12			
male	412 (351)	1.4		reference	349 (185)	5.2		reference		
female	250 (203)	1.5		0.89 (0.75–1.06)	213 (98)	6.2		0.82 (0.64–1.05)		
radiation therapy			0.004				0.003			
yes	601 (498)	1.5		reference	462 (242)	5.7		reference		
no	5 (5)	1.1		3.39 (1.40–8.22)	38 (7)	6.5		0.34 (0.16–0.72)		
chemotherapy			0.0001				0.0001			
yes	554 (451)	1.6		reference	434 (235)	5.3		reference		
no	14 (14)	0.5		4.55 (2.66–7.78)	44 (5)	7.4		0.21 (0.09–0.50)		
extent of resection			<0.0001				0.0001			
biopsy	121 (108)	1.1		reference	224 (132)	4.4		reference		
subtotal	215 (189)	1.3		0.78 (0.62–0.99)	163 (82)	6.2		0.81 (0.61–1.07)		
gross	193 (150)	1.8		0.49 (0.39–0.63)	108 (34)	7.6		0.44 (0.30–0.64)		

^a Abbreviations: HR, hazard ratio; MST, median survival time (year).

^b Numbers do not add up to the column totals due to missing values.

^c Adjust for age, sex, chemo, and radiation therapy for the TCGA Study; and adjusted for diagnosis age, sex, chemo, and radiation therapy and extent of resection for the MDACC Study.

Table 2

Ten Significant Protein Variants Identified in TCGA and MDACC Studies^a

gene	variant	SNP ID	study data set	HR (95% CI) ^b	P
GBM					
<i>PRRX1</i>	Ser1559Thr	rs3936192	TCGA Study	0.62 (0.44–0.86)	0.005
<i>DNAAF5</i>	Val632Ala	rs4720951	TCGA Study	1.4 (1.06–1.85)	0.016
<i>SDHA</i>	Val657Ile	rs6962	TCGA Study	1.56 (1.04–2.33)	0.031
<i>MTHFD1</i>	Arg653Gln	rs2236225	TCGA Study	1.14 (1.01–1.3)	0.036
LGG					
<i>NUP210</i>	Arg786Gln	rs2280084	TCGA Study	0.49 (0.30–0.82)	0.006
<i>PCMT1</i>	Val143Ile	rs4816	MDACC Study	0.79 (0.65–0.96)	0.016
<i>TOR1AIP1</i>	Met146Thr	rs1281378	TCGA Study	0.51 (0.28–0.93)	0.028
<i>HEPACAM</i>	Met218Val	rs10790715	TCGA Study	8.87 (1.23–64.2)	0.030
<i>P2RX7</i>	Gln460Arg	rs2230912	MDACC Study	0.79 (0.63–0.98)	0.035
<i>NSFL1C</i>	Asp290Asn	rs9575	TCGA Study	0.59 (0.35–1)	0.050

^a Abbreviations: Hazard ratio (HR), confidence interval (CI), and P-value (P).

^b Adjust for age, sex, chemo, and radiation therapy for the TCGA Study; and adjusted for diagnosis age, sex, chemo, and radiation therapy and extent of resection for the MDACC Study.