Title

Using germline variants to estimate glioma and subtype risks

Journal

Neuro-Oncology, 21(4)

ISSN

1522-8517

Authors

Eckel-Passow, Jeanette E

Decker, Paul A

Kosel, Matt L

et al.

Publication Date

2019-03-18

DOI

10.1093/neuonc/noz009

Peer reviewed

Neuro-Oncology

21(4), 451–461, 2019 | doi:10.1093/neuonc/noz009 | Advance Access date 8 January 2019

Using germline variants to estimate glioma and subtype risks

Jeanette E. Eckel-Passow, Paul A. Decker, Matt L. Kosel, Thomas M. Kollmeyer, Annette M. Molinaro, Terri Rice, Alissa A. Caron, Kristen L. Drucker, Corinne E. Praska, Melike Pekmezci, Helen M. Hansen, Lucie S. McCoy, Paige M. Bracci, Bradley J. Erickson, Claudia F. Lucchinetti, Joseph L. Wiemels, John K. Wiencke, Melissa L. Bondy, Beatrice Melin, Terry C. Burns, Caterina Giannini, Daniel H. Lachance, Margaret R. Wrensch,[#] and Robert B. Jenkins[#]

Division of Biomedical Statistics & Informatics, Mayo Clinic, Rochester, Minnesota, USA (J.E.E., P.A.D., M.L.K.); Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA (T.M.K., A.A.C., K.L.D., C.E.P., C.G., D.H.L., R.B.J.); Department of Neurological Surgery, University of California San Francisco (UCSF), San Francisco, California, USA (A.M.M., T.R., H.M.H., L.S.M., J.K.W., M.R.W.); Department of Epidemiology and Biostatistics, UCSF, San Francisco, California, USA (A.M.M., P.M.B., J.K.W., M.R.W.); Department of Pathology, UCSF, San Francisco, California, USA (M.P.); Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA (B.J.E.); Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA (C.F.L., D.H.L.); Institute of Human Genetics, UCSF, San Francisco, California, USA (J.K.W., M.R.W.); Department of Medicine, Baylor College of Medicine, Houston, Texas, USA (M.L.B.); Department of Radiation Sciences, Faculty of Medicine, Umeå University, Umeå, Sweden (B.M.); Department of Neurologic Surgery, Mayo Clinic, Rochester, Minnesota, USA (T.C.B.); Center for Genetic Epidemiology, University of Southern California, Los Angeles, California, USA (J.L.W).

Corresponding Author: Robert Jenkins, M.D., Ph.D.; Mayo Clinic, 200 First Street SW, Rochester, MN 55905 (rjenkins@mayo.edu).

[#]These are co-senior authors.

Abstract

Background. Twenty-five single nucleotide polymorphisms (SNPs) are associated with adult diffuse glioma risk. We hypothesized that the inclusion of these 25 SNPs with age at diagnosis and sex could estimate risk of glioma as well as identify glioma subtypes.

Methods. Case-control design and multinomial logistic regression were used to develop models to estimate the risk of glioma development while accounting for histologic and molecular subtypes. Case-case design and logistic regression were used to develop models to predict isocitrate dehydrogenase (IDH) mutation status. A total of 1273 glioma cases and 443 controls from Mayo Clinic were used in the discovery set, and 852 glioma cases and 231 controls from UCSF were used in the validation set. All samples were genotyped using a custom Illumina OncoArray.

Results. Patients in the highest 5% of the risk score had more than a 14-fold increase in relative risk of developing an *IDH* mutant glioma. Large differences in lifetime absolute risk were observed at the extremes of the risk score percentile. For both *IDH* mutant 1p/19q non-codeleted glioma and *IDH* mutant 1p/19q codeleted glioma, the lifetime risk increased from almost null to 2.3% and almost null to 1.7%, respectively. The SNP-based model that predicted *IDH* mutation status had a validation concordance index of 0.85.

Conclusions. These results suggest that germline genotyping can provide new tools for the initial management of newly discovered brain lesions. Given the low lifetime risk of glioma, risk scores will not be useful for population screening; however, they may be useful in certain clinically defined high-risk groups.

Key Points

1. Using 25 glioma germline variants we developed a risk model to estimate glioma risk.
2. Using 25 germline variants we developed a model to distinguish IDH mutated versus wild-type glioma.

Importance of the Study

Genome-wide association studies identified variants in 25 regions that are associated with development of adult diffuse glioma. We show that these 25 germline variants can be used to develop a glioma subtype model that can be used to predict glioma subtype—for example, distinguishing less aggressive *IDH* mutated from more aggressive *IDH* wild-type glioma. Using the same 25 variants we also developed a glioma risk model to estimate relative and lifetime absolute risk. While the prevalence of glioma is too rare for population screening, the proposed risk model and subtype model could be used as another clinical biomarker to guide the clinical decision-making process.

Annually, glioma is diagnosed in approximately 20000 adults in the US.[1] Traditional diagnostic and prognostic features include age at diagnosis, sex, Karnofsky performance score, tumor histology, and tumor grade. However, determining the histologic type and grade can be challenging in adult gliomas. Recently it has become clear that adult gliomas can also be classified using various molecular genetic markers,[2–8] some of which are included in the 2016 World Health Organization (WHO) glioma classification guidelines.[9] In particular, the presence or absence of isocitrate dehydrogenase (*IDH*) mutation, chromosome arms 1p and 19q deletion (1p/19q codeletion), telomerase reverse transcriptase (*TERT*) promoter mutation, tumor protein 53 (TP53) immunoreactivity, and α-thalassemia/mental retardation syndrome X-linked (ATRX) immunoreactivity have been shown to be associated with patient outcome. Gliomas with *IDH* mutation and 1p/19q codeletion have the best prognosis, define tumors of oligodendroglial histology, and usually contain *TERT* promoter mutations. Gliomas with *IDH* mutation without 1p/19q codeletion have an intermediate prognosis and define tumors of astrocytic lineage; these gliomas usually have overexpression of TP53 and loss of ATRX expression. Gliomas without *IDH* mutation (ie, *IDH* wild-type) are most often primary glioblastomas (GBM), and these tumors have the poorest prognosis. Primary GBM often have *TERT* promoter mutations.[2,3,10,11]

Familial gliomas account for approximately 5% of glioma patients.[12–14] Thus, most cases of adult glioma are of unknown origin. Genome-wide association studies (GWAS) have identified germline single nucleotide polymorphisms (SNPs) in 25 regions that are associated with the development of adult diffuse glioma.[15–22] Some of these SNPs have been associated with risk of specific glioma molecular subtypes.[2,16,23] The strongest association is with the 8q24 SNP rs55705857, which confers an approximately 6.0-fold relative risk of *IDH* mutant gliomas.

We hypothesized that we could use germline SNPs, along with age at diagnosis and sex, to estimate glioma risk and histologic and molecular subtype. We examined all 25 known glioma SNPs[2] and generated scores to estimate relative and lifetime absolute risk of glioma as well as risk of specific subtypes.

Methods

Subjects

Mayo Clinic case-control study

The Mayo Clinic glioma case-control study has been described previously.[2,17,22,24] This study was approved by the Mayo Clinic Office for Human Research Protection, and informed written consent was obtained from all participants. Cases were identified at diagnosis (at Mayo Clinic) or at the time of pathologic confirmation (diagnosed elsewhere and treated at Mayo Clinic); patients were at least 18 years of age and had a surgical resection or biopsy between 1973 and 2014. Patient clinical data were extracted from the electronic medical record. Controls were recruited through the Mayo Clinic Biobank, an institutional biorepository of subjects recruited from April 2009 through December 31, 2015. Participants provided consent to participate in future studies approved by the Biobank Access Committee. Controls were at least 18 years old and had no history of a previous brain tumor. The Biobank is supported by the Mayo Clinic Center for Individualized Medicine. Consenting participants provided blood, buccal, and/or saliva specimens and information during in-person or telephone interviews. A total of 1273 cases and 443 controls were evaluated.

UCSF Adult Glioma Study (AGS) case-control study

The UCSF case-control study includes participants of the San Francisco Bay Area Adult Glioma Study (AGS). This study was approved by the UCSF Committee on Human Research, and informed written consent was obtained from all participants. Details of subject recruitment for AGS have been reported previously.[2,12,15,17,22,25,26] Cases were adults (>18 y of age) with newly diagnosed, histologically confirmed grade II, III, or IV glioma. Population-based cases diagnosed between 1991 and 2009 and residing in the 6 San Francisco Bay Area counties were ascertained using the Cancer Prevention Institute of California's early case ascertainment system. Clinic-based cases diagnosed between 2002 and 2012 were recruited from the UCSF Neuro-oncology Clinic, regardless of place of residence. Between 2010 and 2012, controls were recruited from the UCSF general medicine phlebotomy clinic. Consenting participants provided blood, buccal, and/or saliva specimens and information during in-person or telephone interviews. A total of 852 cases and 231 controls were evaluated.

Genotyping

All Mayo Clinic and UCSF cases and controls were genotyped on the same custom Illumina OncoArray.[17] To note, GWAS results for 358 of 1273 (28%) Mayo cases, all 443 Mayo controls, 277 of 852 (33%) UCSF cases, and 229 of 231 (99%) UCSF controls were also reported previously.[17] Herein, we evaluated the previously confirmed 25 glioma risk SNPs.[15–22] Of these 25 SNPs, 10 were directly genotyped, whereas 15 were imputed with high quality ($R^2 > 0.93$; Supplementary Table 1).

Statistical Methods

Association of 25 known glioma risk SNPs with molecular subtypes

Standard SNP quality-control metrics were evaluated. Mayo Clinic and UCSF SNP data were each phased and imputed using the Michigan Imputation Server with the Haplotype Reference Consortium (HRC release 1) as the reference population. To account for glioma subtypes, an additive multinominal logistic regression model was used for each of the 25 SNPs to assess the association between each SNP and disease status:[27]

$$\ln \frac{P(Y_i = k)}{P(Y_i = control)} = \beta X + \varepsilon.$$

Y_i denotes the disease status of subject *i*, where control denotes the reference outcome and *k* denotes the 5 molecular subtypes of glioma based on *TERT* promoter mutation, *IDH* mutation, and 1p/19q codeletion: triple-negative (*IDH* wild-type, *TERT* wild-type, and 1p/19q non-codeleted), *TERT* mutation only, *IDH* mutation only, *TERT* and *IDH* mutation, and triple-positive (*IDH* mutant, *TERT* mutant, and 1p/19q codeleted).[2] The matrix **X** represents predictor variables (SNP, age, sex, and site), β is a vector of estimated coefficients, and ε is a vector of error terms. Genotype was coded as 0, 1, or 2 copies of the alternate allele for genotyped SNPs, whereas dosage was analyzed for imputed SNPs. All models adjusted for age (continuous), sex, and site (Mayo Clinic and UCSF). The overall F-statistic for the SNP main effect tests whether any of the molecular subtypes have an odds ratio significantly different than one. If the overall F-statistic was significant ($P < 0.002$; corrected for testing 25 SNPs), then contrast statements were created to determine which molecular subtypes had odds ratios that were significantly different than one.

Glioma risk models (case-control design)

Additive multinominal logistic regression models were used to develop 2 glioma risk models: (i) where subtypes were classified as GBM (grade IV) or non-GBM (grades II–III), and (ii) where subtypes were classified molecularly as *IDH* wild-type, *IDH* mutant 1p/19q non-codeleted, or *IDH* mutant 1p/19q codeleted. All risk models contained additive effects for age (continuous), sex, and the 25 known glioma risk SNPs; all variables were retained in the models. We utilized a 2-stage (discovery and validation) design[28]; risk models were built using Mayo Clinic glioma cases and controls and validated using UCSF cases and controls. Multinomial logistic regression was used to estimate odds ratios for having a particular glioma subtype by percentile of risk score. Risk score percentile categories were determined from the Mayo Clinic controls, and the middle category (45–55%) was used as the reference category in the multinomial logistic models. Lifetime absolute risk of developing specific subtypes of glioma at different risk score percentile categories was estimated by multiplying the absolute risk in the general population by the relative risk for each percentile category. This approach is appropriate, since the absolute risk of developing an adult diffuse glioma is low.[14,29]

Glioma subtype models (case-case design)

Two glioma subtype models were developed using logistic regression: predicting (i) GBM or non-GBM and (ii) *IDH* mutation status. We utilized a 2-stage design; Mayo Clinic glioma cases were used to develop the models, and UCSF glioma cases were used for validation. The subtype models contained additive effects for age (continuous), sex, and the 25 glioma risk SNPs; all variables were retained in the models. This full model was compared with a model that contained only additive effects for age and sex. Model discrimination was assessed using concordance index (c-index) and 95% confidence intervals (CIs). The c-index denotes the probability that a randomly selected patient who has an *IDH* mutation had a higher risk score than a patient who did not have an *IDH* mutation. The c-index is equal to the area under the receiver operating characteristic curve and ranges from 0.5 to 1. Model calibration was assessed by plotting observed versus predicted probabilities.[30]

Results

Association of 25 Known Glioma Risk SNPs with Molecular Subtypes

Using 1273 gliomas and 443 controls from Mayo Clinic and 852 gliomas and 231 controls from UCSF (Table 1), we evaluated the association of the 25 glioma risk SNPs with risk of the 5 molecular subtypes of glioma defined by *IDH* mutation, *TERT* promoter mutation, and 1p/19q codeletion.[2] We observed 3 categories of associations (Table 2, Supplementary Table 2). The first category consisted of the TP53 SNP, which was associated with all molecular subtypes except triple-negative glioma. The second category consisted of SNPs that were associated with gliomas that have an *IDH* mutation. The third category consisted of SNPs that were associated with *TERT* mutation only gliomas. *TERT* mutation only gliomas comprise largely primary GBM and *IDH* wild-type glioma.

Glioma Risk Models

Based on the association results described above, molecular subtypes were defined as *IDH* wild-type, *IDH* mutant 1p/19q non-codeleted, or *IDH* mutant 1p/19q codeleted. Using 402 Mayo Clinic glioma cases (165 *IDH* wild-type, 141 *IDH* mutant 1p/19q non-codeleted, 96 *IDH* mutant 1p/19q codeleted) and 443 Mayo Clinic controls (Table 1), coefficients from the multinomial logistic regression model were used to estimate risk scores associated with being *IDH* wild-type, *IDH* mutant 1p/19q non-codeleted, and *IDH* mutant 1p/19q codeleted (Fig. 1, Supplementary Table 3). The association of risk score by categories of glioma risk for each molecular subtype is provided in Fig. 2 and Table 3. Patients in the highest 5% of the *IDH* wild-type risk score have more than a 5-fold increased risk of developing an *IDH* wild-type glioma in comparison to patients with median risk scores. Patients in the highest 5% of the *IDH* mutant 1p/19q codeleted or *IDH* mutant 1p/19q non-codeleted risk score had more than a 14- and

Table 1 Patient and tumor characteristics for Mayo Clinic and UCSF glioma cases and controls

	Mayo Clinic		UCSF	
	Cases (N = 1273)	Controls (N = 443)	Cases (N = 852)	Controls (N = 231)
Age				
Median	48	56	51	54
Q1, Q3	36, 59	44, 66.5	40, 60	41, 64
Range	18–84	22–84	19–87	18–89
Sex				
Female	525 (41.2%)	193 (43.6%)	357 (41.9%)	110 (47.6%)
Male	748 (58.8%)	250 (56.4%)	495 (58.1%)	121 (52.4%)
Histology				
Astrocytoma	365 (28.7%)		178 (20.5%)	
Oligodendroglioma	195 (15.3%)		187 (21.6%)	
Oligoastrocytoma	232 (18.2%)		77 (8.9%)	
Glioblastoma	481 (37.8%)		410 (47.3%)	
Tumor Grade				
II	401 (31.5%)		273 (32%)	
III	391 (30.7%)		169 (19.8%)	
IV	481 (37.8%)		410 (48.1%)	
Major 2016 WHO Categories[9] /TCGA Molecular Subtypes[3]				
Missing	871		292	
IDH mutant 1p/19q codeleted	96 (23.9%)		92 (16.4%)	
IDH mutant 1p/19q non-codeleted	141 (35.1%)		133 (23.8%)	
IDH wild-type	165 (41%)		335 (59.8%)	
Eckel-Passow et al. Molecular Subtype[2]				
Missing	871		292	
Triple-negative	22 (5.5%)		65 (11.6%)	
TERT mutation only	143 (35.6%)		270 (48.2%)	
IDH mutation only	120 (29.9%)		117 (20.9%)	
TERT & *IDH* mutations	21 (5.2%)		16 (2.9%)	
Triple-positive	96 (23.9%)		92 (16.4%)	

TCGA = The Cancer Genome Atlas.

19-fold increased risk, respectively, of developing an *IDH* mutant glioma in comparison to patients with median risk scores. The molecular risk model was validated using UCSF glioma cases (335 *IDH* wild-type, 133 *IDH* mutant 1p/19q non-codeleted, 92 *IDH* mutant 1p/19q codeleted) and controls (Table 1). The association of risk score by categories of risk of glioma was similar to the Mayo Clinic series (Fig. 1, 2). Large differences in lifetime absolute risk of developing a particular molecular subtype of glioma was observed at the extremes of the risk score percentile categories (Table 3). The lifetime risk of developing an *IDH* wild-type glioma at the 5th and 95th percentiles of the risk score increased from 0.2% to 1.7%. For *IDH* mutant 1p/19q non-codeleted and *IDH* mutant 1p/19q codeleted gliomas, the lifetime risk increased from almost null to 2.3% and almost null to 1.7%, respectively.

Similar analyses were performed grouping gliomas as GBM versus non-GBM; the results are available in the Supplementary Materials.

Glioma Subtype Models

The models described above estimated the relative risk and lifetime absolute risk of a patient developing an adult diffuse glioma. We hypothesized that once a glioma diagnosis is suspected, germline SNPs obtained from a simple blood test can also be used to determine the patient's subtype. Thus, we developed a model to predict *IDH* mutation status. Using Mayo Clinic glioma cases, the coefficients from a logistic model were used to estimate the probability of being *IDH* mutant (Fig. 3, Supplementary Table 4). The c-index associated with predicting *IDH* mutation status was 0.88 (95% CI: 0.84–0.91) (Supplementary Table 5). The model was well calibrated (Supplementary Fig. 1). To validate the model, model coefficients estimated from the Mayo Clinic cases were applied to the UCSF cases. The distribution of probabilities for the UCSF glioma cases was similar to the Mayo Clinic cases (Fig. 3). The validation

Table 2 Association of glioma risk SNPs with glioma molecular groups

Category	SNP	Cytoband	Gene(s)	Molecular Group[2]	OR	L95	U95	Molecular Group *P*-value	Overall F-statistic *P*-value
All gliomas	rs78378222	17p13.1	*TP53*	*TERT* mutation only	3.31	1.61	6.81	0.001177	**0.001371**
				TERT & *IDH* mutation	8.40	2.79	25.25	0.000151	
				IDH mutation only	3.92	1.69	9.11	0.001464	
				Triple-positive	3.17	1.33	7.51	0.008925	
IDH mutant gliomas	rs12076373	1q44	*AKT3*	*IDH* mutation only	0.64	0.45	0.91	0.011997	0.009662
				Triple-positive	0.64	0.44	0.92	0.015542	
	rs7572263	2q33.3	near *IDH1*	*IDH* mutation only	0.63	0.47	0.84	0.001997	0.004241
				Triple-positive	0.60	0.44	0.82	0.001218	
	rs11706832	3p14.1	*LRIG1*	*IDH* mutation only	1.39	1.10	1.76	0.005523	0.034659
				Triple-positive	1.34	1.06	1.70	0.013914	
	rs55705857	8q24.21	*CCDC26*	*TERT* & *IDH* mutation	4.61	2.31	9.19	0.000014	**2.18E-17**
				IDH mutation only	3.44	2.29	5.18	2.89E-09	
				Triple-positive	5.30	3.57	7.88	1.56E-16	
	rs7107785	11q21	*MAML2*	*IDH* mutation only	0.73	0.58	0.92	0.00882	0.007795
				Triple-positive	0.72	0.57	0.91	0.005793	
	rs648044	11q23.2	*ZBTB16*	*IDH* mutation only	0.70	0.55	0.88	0.002929	**0.000529**
				Triple-positive	0.64	0.51	0.82	0.000377	
	rs12803321	11q23.3	*PHLDB1*	*TERT* & *IDH* mutation	0.55	0.31	0.96	0.034161	0.020442
				IDH mutation only	0.71	0.55	0.92	0.008931	
	rs77633900	15q24.2	*ETFA*	*IDH* mutation only	1.59	1.08	2.34	0.018098	0.014347
				Triple-positive	1.97	1.36	2.85	0.000303	
TERT-mutant gliomas	rs634537	9p21.3	*CDKN2A, CDKN2B*	*TERT* mutation only	1.42	1.19	1.70	0.000132	0.003754
	rs11599775	10q25.2	*VTI1A*	*TERT* mutation only	0.78	0.64	0.94	0.010953	0.014432
	rs2297440	20q13.33	*RTEL1*	*TERT* mutation only	1.73	1.35	2.22	0.000012	**0.000969**
Other	rs10069690	5p15.33	*TERT*	Triple-negative	1.57	1.10	2.24	0.01211	**1.26E-08**
				TERT mutation only	1.94	1.59	2.38	1.00E-10	
				IDH mutation only	1.47	1.13	1.93	0.004383	
	rs75061358	7p11.2	near *EGFR*	*TERT* mutation only	1.80	1.30	2.49	0.000402	**0.000898**
				Triple-positive	1.72	1.15	2.57	0.008851	

Multinomial logistic regression was performed for each SNP separately; all models were adjusted for age, sex and site (Mayo, UCSF). SNPs with an overall multinomial F-statistic *P*-value < 0.05 are shown; *P*-value < 0.002 is in bold font. OR denotes odds ratio, L95 denotes lower 95% confidence interval, and U95 denotes upper 95% confidence interval. The results for all 25 SNPs and all molecular groups are provided in Supplementary Table 2.

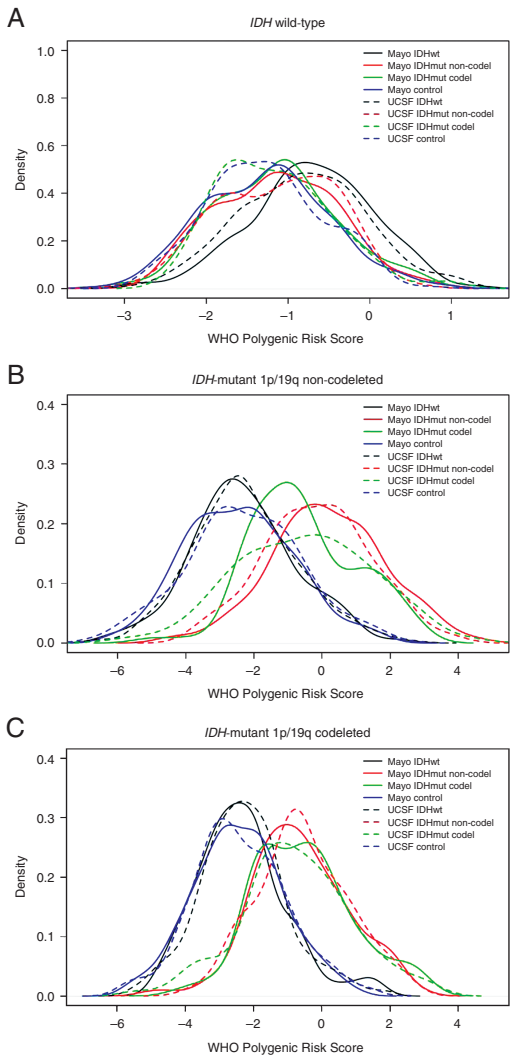

Fig. 1 Estimated glioma risk score from the multinomial logistic regression model that molecularly classified patients: estimated risk scores of being (A) *IDH* wild-type, (B) *IDH* mutant 1p/19q non-codeleted, and (C) *IDH* mutant 1p/19q codeleted. The model was developed using Mayo Clinic glioma cases and controls and validated using UCSF glioma cases and controls.

c-index associated with predicting *IDH* mutation status was 0.85 (95% CI: 0.82–0.88) in the UCSF cases (Supplementary Table 5). The model slightly overestimated the probability of being *IDH* mutant in the UCSF cases (Supplementary Fig. 1).

Similar analyses were performed predicting GBM versus non-GBM; the results are available in the Supplementary Materials.

Discussion

Polygenic risk models have been reported in several cancers, including breast, ovarian, prostate, and chronic lymphocytic leukemia.[27,31–35] In glioma it has been shown that when GWAS analyses were performed by molecular subtype, SNPs with large and potentially clinically relevant effect sizes were identified.[15] Additionally, performing GWAS by molecular subtype may provide clues as to how gliomas develop. We evaluated the 25 known glioma risk variants and showed that the *TP53* germline variant is involved in the development of all gliomas. Variants in or near *AKT3*, *IDH1*, *LRIG1*, *CCDC26*, *MAML2*, *ZBTB16*, *PHLDB1*, and *ETFA* were associated with the development of *IDH* mutant glioma. And germline variants in or near *CDKN2A/B*, *VTI1A*, and *RTEL1* facilitate the development of *IDH* wild-type glioma. Similar associations by glioma subtype were recently reported, further validating the results.[23] Thus, we hypothesized that the inclusion of germline SNPs with age at diagnosis and sex might be useful for predicting risk of glioma and risk of specific glioma subtypes. Using 25 SNPs that have been shown to be associated with glioma risk, as well as age at diagnosis and sex, we developed models to estimate risk of glioma. Interestingly, in comparison to 5% of the controls, 42% and 38% of the Mayo *IDH* mutated 1p/19q non-codeleted glioma and *IDH* mutated 1p/19q codeleted glioma, respectively, had a risk score in the 95th–100th percentile of the risk score distribution. Thus, patients in the highest 5th percentile of risk score had more than a 14-fold increased risk of developing an *IDH* mutated glioma. This equates to an increased lifetime absolute risk from 0.12% in the general population to 2.3% (*IDH* mutated 1p/19q non-codeleted glioma) or 1.7% (*IDH* mutated 1p/19q codeleted glioma) for patients in the highest 5th percentile of risk score.

Molecular markers have been shown to be associated with prognosis in adult diffuse glioma and thus were recently incorporated into the 2016 WHO classification schema.[2–9] Currently, molecular characterization is typically determined from surgical specimens. Because information regarding patient prognosis, tumor aggressiveness, and treatment response can inform personalized treatment, recent efforts have focused on using images to classify gliomas into clinically relevant molecular groups prior to surgery.[36–40] Here, we evaluated the effectiveness of using the known 25 glioma risk SNPs to classify gliomas into clinically relevant groups. Specifically, we developed a model to predict *IDH* mutation status that had a validation c-index of 0.85.

In developing polygenic risk models it is important to determine how such models could be implemented in clinical practice to improve patient care. It was recently suggested that there are 3 applications of polygenic risk models: disease screening, therapeutic intervention, and life planning.[41] Because of the low absolute lifetime risk of glioma, population-level screening would result in numerous false positives and thus is not being suggested.[14] However, we hypothesize that risk models could help with characterizing suspicious brain lesions. Since characterization of suspected malignant brain tumors remains a challenge, even with improved imaging capabilities, a polygenic risk model could provide a quantitative measure of the likelihood of glioma that may help with interpretation of an MRI. Potential examples where these risk scores

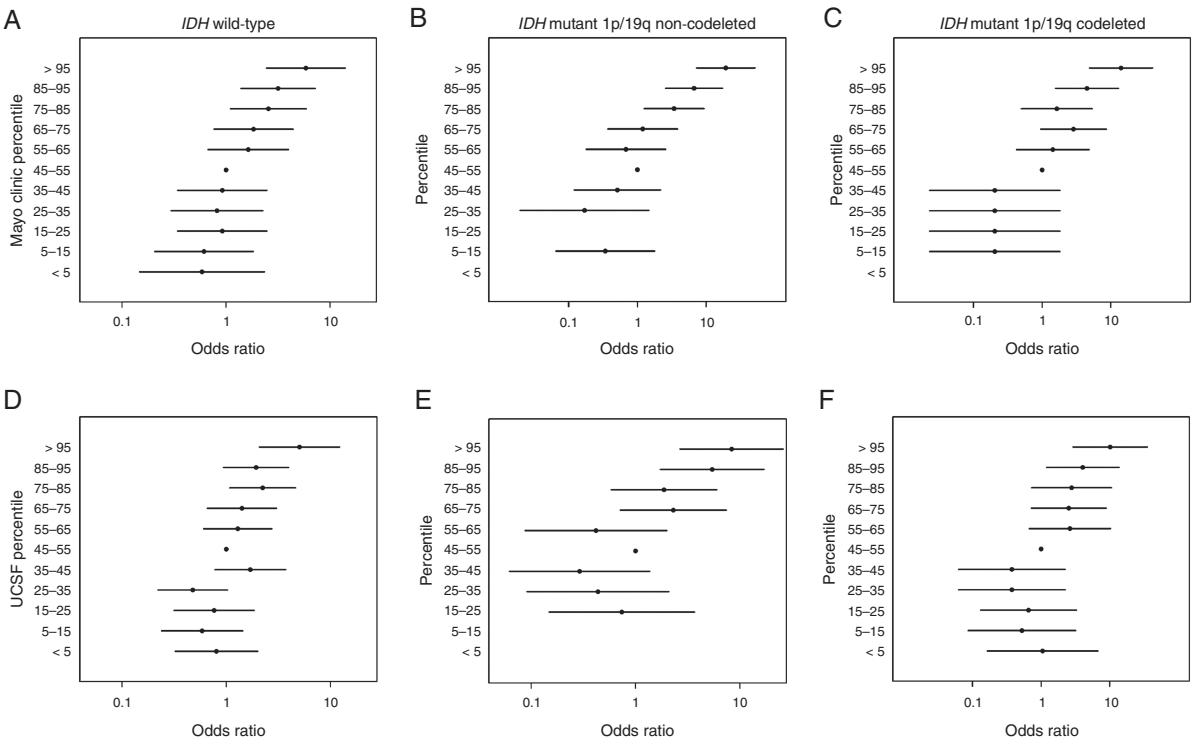

Fig. 2 Association between glioma risk score and relative risk of a specific glioma subtype, estimated from Mayo Clinic glioma cases and controls, for: (A) *IDH* wild-type, (B) *IDH* mutant 1p/19q non-codeleted, and (C) *IDH* mutant 1p/19q codeleted. Associations were validated using UCSF cases and controls for (D) *IDH* wild-type, (E) *IDH* mutant 1p/19q non-codeleted, and (F) *IDH* mutant 1p/19q codeleted. Odds ratios were calculated for percentiles of risk score relative to the middle category (45–55%) of risk scores.

might be useful in clinical settings are assisting in differentiating contrast enhancing lesions. For example, differentiating high-grade glioma versus lymphoma versus demyelination, and indeterminate non-enhancing lesions for which glioma is necessarily in the differential diagnosis. The clinical findings and radiological appearance of central nervous system (CNS) lymphoma can be indistinguishable from high-grade glioma and current research is aimed at improving diagnostic accuracy to differentiate these tumors.[42,43] Similarly, tumefactive demyelinating lesions can sometimes appear very similar to high-grade glioma or CNS lymphoma.[44–47] Misdiagnosing a tumefactive demyelinating lesion as a brain tumor could result in the inappropriate use of radiation therapy, resulting in significant consequences.[45] Lucchinetti et al[44] analyzed 168 patients with biopsy-confirmed tumefactive demyelinating disease and reported that 31% were initially misdiagnosed and determined to not have tumefactive demyelinating disease; astrocytoma was the misdiagnosis in 39% of these cases. Thus, if appropriately clinically validated, the glioma risk model—which requires only a simple and inexpensive blood test—might be implemented as an ancillary measure to help define a difficult diagnosis.

While we hypothesize that the glioma risk model could be used to help interpret current disease screening modalities, such as MRI, we hypothesize that the glioma subtype model could be used for therapeutic intervention.[41] That is, to determine tumor aggressiveness (eg, *IDH* mutation status) prior to surgery in order to inform personalized treatment. Recent efforts have focused on using images to classify gliomas into clinically relevant molecular groups prior to surgery[36–40]; the glioma subtype model is a simple and inexpensive blood test that could also be utilized. Before utilizing polygenic models for therapeutic intervention, future work would need to evaluate the predictive accuracy of these models both along, as well as in combination, with radiology-based models.

There are some limitations with this study. Small numbers of subjects in some of the molecular subtypes may have limited the ability to detect associations with certain SNPs. Furthermore, for the reasons described below, the risk models discussed herein all require additional external validation, particularly within clinically or radiographically defined groups. Because there are limited GWAS data available on patients who also have tumor molecular data, some of the patients analyzed were included in previous glioma GWAS, as discussed in the Methods section: 28% of the Mayo cases and 33% of the UCSF cases were also analyzed previously,[17] which may increase the associations over what might be observed

Table 3 Relative risk (RR) and lifetime absolute risk of developing an *IDH* wild-type (*IDH*wt), *IDH* mutated 1p/19q non-codeleted (*IDH*mt noncodel), or *IDH* mutated 1p/19q codeleted (*IDH*mt codel) glioma at different risk score percentile categories

Risk Score % Category	Number Controls	*IDH*wt					*IDH*mt Noncodel					*IDH*mt Codel				
		Number *IDH*wt	Relative Risk (RR)	RR L95	RR U95	Absolute Risk (%)	Number *IDH*mt noncodel	Relative Risk (RR)	RR L95	RR U95	Absolute Risk (%)	Number *IDH*mt codel	Relative Risk (RR)	RR L95	RR U95	Absolute Risk (%)
<5	23	3	0.587	0.147	2.344	0.170	0	4.34E-08	0.000	Inf	5.21E-09	0	2.88E-07	0.000	Inf	3.46E-08
5–15	44	6	0.614	0.205	1.833	0.178	2	0.341	0.065	1.781	0.041	1	0.205	0.023	1.822	0.025
15–25	44	9	0.920	0.341	2.482	0.267	0	3.05E-08	0.000	Inf	3.66E-09	1	0.205	0.023	1.822	0.025
25–35	44	8	0.818	0.296	2.265	0.237	1	0.170	0.020	1.474	0.020	1	0.205	0.023	1.822	0.025
35–45	44	9	0.920	0.341	2.482	0.267	3	0.511	0.120	2.173	0.061	1	0.205	0.023	1.822	0.025
45–55	45	10	Reference	-	-	0.29	6	Reference	-	-	0.12	5	Reference	-	-	0.12
55–65	44	16	1.636	0.670	3.996	0.475	4	0.682	0.180	2.582	0.082	7	1.432	0.422	4.853	0.172
65–75	44	18	1.841	0.765	4.428	0.534	7	1.193	0.371	3.833	0.143	14	2.864	0.951	8.624	0.344
75–85	44	25	2.557	1.101	5.940	0.741	20	3.409	1.251	9.290	0.409	8	1.636	0.497	5.390	0.196
85–95	44	31	3.170	1.389	7.235	0.919	39	6.648	2.559	17.270	0.798	22	4.500	1.565	12.940	0.540
>95	23	30	5.870	2.448	14.072	1.702	59	19.239	7.230	51.192	2.309	36	14.087	4.872	40.733	1.690

Risk score percentile categories were determined from the Mayo Clinic controls, and the middle decile (45–55%) was used as the reference category in the logistic models to estimate RR (odds ratios). Lifetime absolute risk of specific molecular subtypes of glioma at different risk score percentile categories was estimated by multiplying the absolute risk in the general population by the RR for each percentile category. RR L95 and RR U95 denote the lower and upper 95% confidence interval for the corresponding RR.

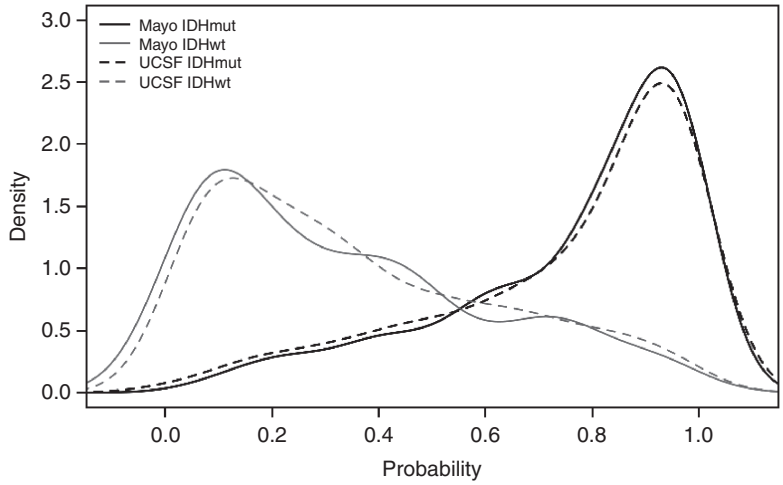

Fig. 3 Estimated probability of being *IDH* mutant from a logistic regression model that predicted *IDH* mutation status. Model was developed using Mayo Clinic glioma cases and validated using UCSF glioma cases.

in a completely independent training set. Additionally, because 15 of the 25 SNPs were imputed using data from a custom Illumina OncoArray (Supplementary Table 1), a custom clinical assay that directly genotypes all 25 SNPs will be needed and is currently in development. We acknowledge that epistasis is important, but significant SNP-SNP interactions have yet to be identified and thus were not interrogated in the risk models. While we did not include these interactions in our models, future work should include analyzing large cohorts that are adequately powered to evaluate interactions in predicting glioma risk.[48] There are likely additional variables that should be considered in the risk models such as Karnofsky performance score, history of seizure, family history of brain cancer, etc. However, these variables are often difficult to capture accurately. For example, while family history could be helpful, patients often have a difficult time differentiating gliomas from brain metastases or other primary brain tumors.

The discovery of germline risk SNPs for glioma has altered our concepts of how these tumors arise and opened new avenues for etiologic research; however, they have not yet altered neuro-oncology practice. Using 25 SNPs, patient age, and sex, we developed risk models to estimate relative and lifetime absolute risk and subtype models to predict glioma subtypes. We propose that these models could be useful for disease screening, therapeutic intervention, and life planning. This could impact neurologic, neurosurgical, and neuro-oncologic patient management, potentially influencing optimal long-term outcomes for diffuse adult glioma patients.

Supplementary Material

Supplementary data are available at *Neuro-Oncology* online.

Keywords

classification | genotype | glioblastoma | glioma | polygenic

Funding

Work at Mayo Clinic was supported by the National Institutes of Health (NIH; P50CA108961, P30 CA15083, RC1NS068222Z), Bernie and Edith Waterman Foundation, and Ting Tsung and Wei Fong Chao Family Foundation. Work at University of California, San Francisco, was supported by NIH (grant numbers R01CA52689, P50CA097257, R01CA126831, R01CA139020, R01CA163687, and R25CA112355), as well as the Loglio Collective, the National Brain Tumor Foundation, the Stanley D. Lewis and Virginia S. Lewis Endowed Chair in Brain Tumor Research, the Robert Magnin Newman Endowed Chair in Neuro-oncology, and by donations from families and friends of John Berardi, Helen Glaser, Elvera Olsen, Raymond E. Cooper, and William Martinusen.

Acknowledgments

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 HHSN261201000140C 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 # U58DP003862-01 awarded to the California Department of Public Health. This

publication was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 RR024131. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. 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 authors acknowledge study participants, clinicians and research staff at participating medical centers, Katherine Cornelius, Mayo Clinic Comprehensive Cancer Center Biospecimens and Processing and Genotyping Shared Resources, UCSF Diller Cancer Center Genomics Core and The Gliogene Consortium.

Conflict of interest statement. There are no conflicts of interest to disclose.

Authorship statement. Experimental design: JEEP, AMM, DHL, MW, RBJ. Implementation: TMK, TR, AC, KLD, CP, MP, HMH, LSM, PMB, JW, JKW, MLB, BM, TCB, CG. Analysis and interpretation of the data: JEEP, PAD, MLK, AMM, BJE, CL, DHL, MW, RBJ. All authors were involved in the writing of the manuscript and have read and approved the final version.

References

1. Ostrom QT, Gittleman H, Fulop J, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008–2012. *Neuro Oncol.* 2015;17(Suppl 4):iv1–iv62.
2. Eckel-Passow JE, Lachance DH, Molinaro AM, et al. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N Engl J Med.* 2015;372(26):2499–2508.
3. Brat DJ, Verhaak RG, Aldape KD, et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med.* 2015;372(26):2481–2498.
4. Ceccarelli M, Barthel FP, Malta TM, et al; TCGA Research Network. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell.* 2016;164(3):550–563.
5. Killela PJ, Pirozzi CJ, Healy P, et al. Mutations in IDH1, IDH2, and in the TERT promoter define clinically distinct subgroups of adult malignant gliomas. *Oncotarget.* 2014;5(6):1515–1525.
6. Labussière M, Boisselier B, Mokhtari K, et al. Combined analysis of TERT, EGFR, and IDH status defines distinct prognostic glioblastoma classes. *Neurology.* 2014;83(13):1200–1206.
7. Labussière M, Di Stefano AL, Gleize V, et al. TERT promoter mutations in gliomas, genetic associations and clinico-pathological correlations. *Br J Cancer.* 2014;111(10):2024–2032.
8. Leeper HE, Caron AA, Decker PA, Jenkins RB, Lachance DH, Giannini C. IDH mutation, 1p19q codeletion and ATRX loss in WHO grade II gliomas. *Oncotarget.* 2015;6(30):30295–30305.
9. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 2016;131(6):803–820.
10. Brennan CW, Verhaak RG, McKenna A, et al; TCGA Research Network. The somatic genomic landscape of glioblastoma. *Cell.* 2013;155(2):462–477.
11. Pekmezci M, Rice T, Molinaro AM, et al. Adult infiltrating gliomas with WHO 2016 integrated diagnosis: additional prognostic roles of ATRX and TERT. *Acta Neuropathol.* 2017;133(6):1001–1016.
12. Wrensch M, Lee M, Miike R, et al. Familial and personal medical history of cancer and nervous system conditions among adults with glioma and controls. *Am J Epidemiol.* 1997;145(7):581–593.
13. Malmer B, Grönberg H, Bergenheim AT, Lenner P, Henriksson R. Familial aggregation of astrocytoma in northern Sweden: an epidemiological cohort study. *Int J Cancer.* 1999;81(3):366–370.
14. Rice T, Lachance DH, Molinaro AM, et al. Understanding inherited genetic risk of adult glioma - a review. *Neurooncol Pract.* 2016;3(1):10–16.
15. Jenkins RB, Xiao Y, Sicotte H, et al. A low-frequency variant at 8q24.21 is strongly associated with risk of oligodendroglial tumors and astrocytomas with IDH1 or IDH2 mutation. *Nat Genet.* 2012;44(10):1122–1125.
16. Kinnersley B, Labussière M, Holroyd A, et al. Genome-wide association study identifies multiple susceptibility loci for glioma. *Nat Commun.* 2015;6:8559.
17. Melin BS, Barnholtz-Sloan JS, Wrensch MR, et al; GliomaScan Consortium. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat Genet.* 2017;49(5):789–794.
18. Rajaraman P, Melin BS, Wang Z, et al. Genome-wide association study of glioma and meta-analysis. *Hum Genet.* 2012;131(12):1877–1888.
19. Shete S, Hosking FJ, Robertson LB, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet.* 2009;41(8):899–904.
20. Stacey SN, Sulem P, Jonasdottir A, et al; Swedish Low-risk Colorectal Cancer Study Group. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet.* 2011;43(11):1098–1103.
21. Walsh KM, Codd V, Smirnov IV, et al; ENGAGE Consortium Telomere Group. Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk. *Nat Genet.* 2014;46(7):731–735.
22. Wrensch M, Jenkins RB, Chang JS, et al. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat Genet.* 2009;41(8):905–908.
23. Labreche K, Kinnersley B, Berzero G, et al. Diffuse gliomas classified by 1p/19q co-deletion, TERT promoter and IDH mutation status are associated with specific genetic risk loci. *Acta Neuropathol.* 2018;135(5):743–755.
24. Jenkins RB, Wrensch MR, Johnson D, et al. Distinct germ line polymorphisms underlie glioma morphologic heterogeneity. *Cancer Genet.* 2011;204(1):13–18.
25. Felini MJ, Olshan AF, Schroeder JC, et al. Reproductive factors and hormone use and risk of adult gliomas. *Cancer Causes Control.* 2009;20(1):87–96.
26. Wiemels JL, Wiencke JK, Sison JD, Miike R, McMillan A, Wrensch M. History of allergies among adults with glioma and controls. *Int J Cancer.* 2002;98(4):609–615.
27. Qian DC, Han Y, Byun J, et al. A novel pathway-based approach improves lung cancer risk prediction using germline genetic variations. *Cancer Epidemiol Biomarkers Prev.* 2016;25(8):1208–1215.

28. Molinaro AM, Wrensch MR, Jenkins RB, Eckel-Passow JE. Statistical considerations on prognostic models for glioma. *Neuro Oncol*. 2016;18(5):609–623.
29. Dupont WD, Plummer WD Jr. Understanding the relationship between relative and absolute risk. *Cancer*. 1996;77(11):2193–2199.
30. Harrell FE. *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2001.
31. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5): djv036.
32. Muranen TA, Mavaddat N, Khan S, et al. Polygenic risk score is associated with increased disease risk in 52 Finnish breast cancer families. *Breast Cancer Res Treat*. 2016;158(3):463–469.
33. Kuchenbaecker KB, McGuffog L, Barrowdale D, et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J Natl Cancer Inst*. 2017;109(7): djw032.
34. Lecarpentier J, Silvestri V, Kuchenbaecker KB, et al; EMBRACE; GEMO Study Collaborators; HEBON; KConFab Investigators. Prediction of breast and prostate cancer risks in male BRCA1 and BRCA2 mutation carriers using polygenic risk scores. *J Clin Oncol*. 2017;35(20):2240–2250.
35. Kleinstern G, Camp NJ, Goldin LR, et al. Association of polygenic risk score with the risk of chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis. *Blood*. 2018;131(23):2541–2551.
36. Jakola AS, Zhang YH, Skjulsvik AJ, et al. Quantitative texture analysis in the prediction of IDH status in low-grade gliomas. *Clin Neurol Neurosurg*. 2018;164:114–120.
37. Park YW, Han K, Ahn SS, et al. Prediction of IDH1-Mutation and 1p/19q-codeletion status using preoperative MR imaging phenotypes in lower grade gliomas. *AJNR Am J Neuroradiol*. 2018;39(1):37–42.
38. Jiang S, Zou T, Eberhart CG, et al. Predicting IDH mutation status in grade II gliomas using amide proton transfer-weighted (APTw) MRI. *Magn Reson Med*. 2017;78(3):1100–1109.
39. Korfiatis P, Kline TL, Lachance DH, Parney IF, Buckner JC, Erickson BJ. Residual deep convolutional neural network predicts MGMT Methylation status. *J Digit Imaging*. 2017;30(5):622–628.
40. Akkus Z, Ali I, Sedlář J, et al. Predicting deletion of chromosomal Arms 1p/19q in low-grade gliomas from MR images using machine intelligence. *J Digit Imaging*. 2017;30(4):469–476.
41. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. 2018;19(9):581–590.
42. Neska-Matuszewska M, Bladowska J, Sąsiadek M, Zimny A. Differentiation of glioblastoma multiforme, metastases and primary central nervous system lymphomas using multiparametric perfusion and diffusion MR imaging of a tumor core and a peritumoral zone-Searching for a practical approach. *PLoS One*. 2018;13(1):e0191341.
43. Lin X, Lee M, Buck O, et al. Diagnostic accuracy of T1-weighted dynamic contrast-enhanced-MRI and DWI-ADC for differentiation of glioblastoma and primary CNS lymphoma. *AJNR Am J Neuroradiol*. 2017;38(3):485–491.
44. Lucchinetti CF, Gavrilova RH, Metz I, et al. Clinical and radiographic spectrum of pathologically confirmed tumefactive multiple sclerosis. *Brain*. 2008;131(Pt 7):1759–1775.
45. Algahtani H, Shirah B, Alassiri A. Tumefactive demyelinating lesions: a comprehensive review. *Mult Scler Relat Disord*. 2017;14:72–79.
46. Abdoli M, Freedman MS. Neuro-oncology dilemma: tumour or tumefactive demyelinating lesion. *Mult Scler Relat Disord*. 2015;4(6):555–566.
47. Kim DS, Na DG, Kim KH, et al. Distinguishing tumefactive demyelinating lesions from glioma or central nervous system lymphoma: added value of unenhanced CT compared with conventional contrast-enhanced MR imaging. *Radiology*. 2009;251(2):467–475.
48. Crawford L, Zeng P, Mukherjee S, Zhou X. Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. *PLoS Genet*. 2017;13(7):e1006869.