HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2017 August 16.

Published in final edited form as:

Nat Genet. 2017 May ; 49(5): 789–794. doi:10.1038/ng.3823.

Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors

A full list of authors and affiliations appears at the end of the article.

Abstract

Genome-wide association studies (GWAS) have transformed our understanding of glioma susceptibility, but individual studies have had limited power to identify risk loci. We performed a meta-analysis of existing GWAS and two new GWAS, which totaled 12,496 cases and 18,190 controls. We identified five new loci for glioblastoma (GBM) at 1p31.3 (rs12752552; $P = 2.04 \times 10^{-9}$, odds ratio (OR) = 1.22), 11q14.1 (rs11233250; $P = 9.95 \times 10^{-10}$, OR = 1.24), 16p13.3 (rs2562152; $P = 1.93 \times 10^{-8}$, OR = 1.21), 16q12.1 (rs10852606; $P = 1.29 \times 10^{-11}$, OR = 1.18) and 22q13.1 (rs2235573; $P = 1.76 \times 10^{-10}$, OR = 1.15), as well as eight loci for non-GBM tumors at 1q32.1 (rs4252707; $P = 3.34 \times 10^{-9}$, OR = 1.19), 1q44 (rs12076373; $P = 2.63 \times 10^{-10}$, OR = 1.23), 2q33.3 (rs7572263; $P = 2.18 \times 10^{-10}$, OR = 1.20), 3p14.1 (rs11706832; $P = 7.66 \times 10^{-9}$, OR = 1.15), 10q24.33 (rs11598018; $P = 3.39 \times 10^{-8}$, OR = 1.14), 11q21 (rs7107785; $P = 3.87 \times 10^{-10}$, OR = 1.16), 14q12 (rs10131032; $P = 5.07 \times 10^{-11}$, OR = 1.33) and 16p13.3 (rs3751667; $P = 2.61 \times 10^{-9}$, OR = 1.18). These data substantiate that genetic susceptibility to GBM and non-GBM tumors are highly distinct, which likely reflects different etiology.

Correspondence should be addressed to B.S.M. (beatrice.melin@umu.se), R.S.H. (richard.houlston@icr.ac.uk) or M.L.B. (mbondy@bcm.edu).

[25]A full list of members and affiliations appears in the Acknowledgments

[41]These authors contributed equally to this work

[42]These authors jointly directed this work

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

M.L.B., B.S.M., R.S.H. and J.S.B.-S. managed the project; R.S.H., M.L.B., B.S.M., J.S.B.-S., R.B.J., Q.T.O., B.K. and M.R.W. drafted the manuscript; Q.T.O., K.L., B.K., J.E.E.-P. and P.A.D. performed statistical analyses; Y.C., K.L., Y.L. and B.K. performed bioinformatics analyses; B.S.M., J.S.B.-S., M.R.W., J.K.W., C.J., D.I., R.K.L., G.A., P.A.D., U.A., T.R., H.H., L.M., M.L.K., H.S., J.L.B., F.D., D.L, C.I.A, C.L., R.T.M., J. Shildkraut, F.A.-O., S. Sadetski, M. Scheurer, S. Shete, E.B.C., S.H.O., R.B.J., R.S.H. and M.L.B. developed the GICC protocol and performed sample acquisition; and P.R., S.C., M. Linet, Z.W. and M.Y. provided the National Cancer Institute (NCI) data. In the UK, P.B., A.S., M.J.S., S.J.F. and R.S.H. developed patient recruitment, performed sample acquisition and performed sample collection of cases; P.B. oversaw DNA isolation and storage, and performed case and control ascertainment, and supervision of DNA extractions. In Germany, M. Simon, M.M.N., H.-E.W., S. Schreiber and J. Schramm developed patient recruitment, and oversaw performed blood sample collection; M. Simon oversaw DNA isolation and storage and performed case and control ascertainment, and supervision of DNA extractions; and S. Herms, S. Heilmann and K.G. performed experimental work. In France, M. Sanson and J.-Y.D. developed patient recruitment; M. Labussière, A.-L.D.S., P.G., K.M., A.I. and K.H.-X. performed patient ascertainment. M. Lathrop performed laboratory management and oversaw genotyping of the French samples. All authors contributed to the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Glioma accounts for around 27% of all primary brain tumors and is responsible for approximately 13,000 cancer-related deaths in the United States each year[1, 2]. Gliomas can be broadly classified into GBM and lower-grade non-GBM tumors[3]. Gliomas typically have a poor prognosis irrespective of medical care, with the most common form, GBM, having a five-year survival rate of only 5% (ref. 4).

So far, no environmental exposures have been robustly linked to the risk of developing glioma, except for moderate to high doses of ionizing radiation, which accounts for a small proportion of cases[5]. Evidence for an inherited predisposition to glioma is provided by a number of rare inherited cancer syndromes, such as Turcot's and Li–Fraumeni syndromes, as well as neurofibromatosis. Even collec-tively, however, these account for little of the twofold familial risk of glioma[6]. Our understanding of the heritability of glioma has been transformed by recent GWAS, which have identified single-nucleotide polymorphisms (SNPs) at 13 loci influencing risk[7–14].

Previous individual studies have had limited statistical power for the additional discovery of new glioma risk loci[15]. Therefore, to gain more comprehensive insight into glioma etiology, we performed a meta-analysis of previously published GWAS and two new GWAS, which allowed us to identify 13 new risk loci for glioma.

We analyzed GWAS SNP data that passed quality control for 12,496 cases (6,191 classified as GBM and 5,819 classified as non-GBM tumors) and 18,190 controls from eight studies with individuals of European ancestry, a new GWAS of 4,572 cases and 3,286 controls performed by the Glioma International Case Control Consortium (GICC) (Supplementary Table 1), a new GWAS of 1,591 cases and 804 controls from the University of California, San Francisco (UCSF)-Mayo, and six previously reported GWAS[9, 10, 13] totaling 6,405 cases and 14,100 controls (Supplementary Table 2). To increase genomic resolution, we imputed >10 million SNPs. Quantile–quantile (Q-Q) plots for SNPs with a minor allele frequency (MAF) >1% after imputation did not show evidence of substantive overdispersion (λ = 1.02–1.10, λ_{90} = 1.02–1.05; Supplementary Fig. 1). We derived joint ORs and 95% confidence intervals (CIs) under a fixed-effects model for each SNP with MAF >1% and associated per-allele principal component (PCA) corrected *P*-values for all glioma, GBM and non-GBM cases versus those for the controls (Fig. 1).

In the combined meta-analysis, among previously published glioma risk SNPs, those for all glioma at 17p13.1 (*TP53*), for GBM at 5p15.33 (*TERT*), 7p11.2 (*EGFR*), 9p21.3 (*CDKN2B–AS1*) and 20q13.33 (*RTEL1*), and for non-GBM tumors at 8q24.21 (*CCDC26*), 11q23.2, 11q23.3 (*PHLDB1*) and 15q24.2 (*ETFA*) showed even greater evidence for association (Supplementary Fig. 2 and Supplementary Table 3). SNPs at 10q25.2 and 12q12.1 for non-GBM tumors retained genome-wide significance (i.e., $P < 5.0 \times 10^{-8}$). Associations at the previously reported 3q26.2 (near *TERC*)[11] and 12q23.33 (*POLR3B*)[10] loci for GBM did not retain statistical significance (*P* values for the most associated SNPs are 2.68×10^{-5} and 1.60×10^{-5}, respectively; Supplementary Table 3).

In addition to previously reported loci, we identified genome-wide significant associations marking new risk loci (Table 1, Supplementary Fig. 3 and Supplementary Data 1) for GBM

at 1p31.3 (rs12752552; $P = 2.04 \times 10^{-9}$), 11q14.1 (rs11233250; $P = 9.95 \times 10^{-10}$), 16p13.3 (rs2562152; $P = 1.93 \times 10^{-8}$), 16q12.1 (rs10852606; $P = 1.29 \times 10^{-11}$) and 22q13.1 (rs2235573; $P = 1.76 \times 10^{-10}$) and for non-GBM tumors at 1q32.1 (rs4252707; $P = 3.34 \times 10^{-9}$), 1q44 (rs12076373; $P = 2.63 \times 10^{-10}$), 2q33.3 (rs7572263; $P = 2.18 \times 10^{-10}$), 3p14.1 (rs11706832; $P = 7.66 \times 10^{-9}$), 10q24.33 (rs11598018; $P = 3.39 \times 10^{-8}$), 11q21 (rs7107785; $P = 3.87 \times 10^{-10}$), 14q12 (rs10131032; $P = 5.07 \times 10^{-11}$) and 16p13.3 (rs3751667; $P = 2.61 \times 10^{-9}$). Conditional analysis confirmed the existence of two independent association signals at 7p11.2 (*EGFR*) as previously reported[7] but did not provide evidence for additional signals at any of the other established identified risk loci or at the 13 newly identified loci. Case-only analyses con-firmed the specificity of 11q14.1, 16p13.3 and 22q13.1 associations for GBM and of 1q44, 2q33.3, 3p14.1, 11q21 and 14q12 associations for non-GBM tumors (Fig. 2 and Supplementary Table 4). Collectively, our findings provide strong evidence for specific associations for the different glioma subtypes, consistent with their previously described distinctive molecular profiles, presumably resulting from different etiological pathways.

Across the new and known risk loci, we found a significant enrichment of overlap with enhancers in H9-Derived neuronal progenitor cells ($P = 8.2 \times 10^{-5}$; Supplementary Data 2). These observations support the assertion that the loci identified in the GWAS influence glioma risk through effects on neural *cis* regulatory networks and that they are strongly involved in transcriptional initiation and enhancement. To gain further insight into the biological basis for associations at the 13 new risk loci, we performed an expression quantitative trait loci (eQTL) analysis using RNA-seq data on ten regions of normal human brain from up to 103 individuals from the Genotype–Tissue Expression (GTEx) project[16] and blood eQTL data on 5,311 individuals from Westra *et al.*[17]. We used summary-level mendelian randomization (SMR)[18] analysis to test for a concordance between signals from GWAS and *cis* eQTL for genes within 1 Mb of the sentinel and correlated SNPs ($r^2 > 0.8$) at each locus (Supplementary Data 3) and derived b_{XY} statistics, which estimate the effect of gene expression on glioma risk. Additionally, for each of the risk SNPs at the 13 new loci (as well as the correlated variants), we examined published data[19, 20] and made use of the online resources HaploRegv4, RegulomeDB and SeattleSeq for evidence of functional effects (Supplementary Table 5).

At 16q12.1, the GBM association signal was significantly associated with *HEATR3* expression in nine of ten regions of the brain (PSMR = 3.38×10^{-6} to 6.55×10^{-10}; b_{XY} = 0.14–0.24; Supplementary Fig. 4 and Supplementary Data 3). The risk allele ‘C’ of rs10852606 that was associated with reduced *HEATR3* expression was consistent with differential expression of *HEATR3* being the functional basis of the 16q12.1 association. The observation that variation at 16q12.1 is associated with risk of testicular[21] (rs8046148; pairwise r^2 and D' with rs10852606 of 0.67 and 1.0, respectively) and esophageal[22] (rs4785204; pairwise r^2 and D' with rs10852606 of 0.16 and 1.0, respectively) cancer suggests that the locus has pleiotropic effects on tumor risk, which are compatible with generic effects as shown by the observation of a *HEATR3* eQTL signal in blood (*PSMR* = 5.84×10^{-11}; b_{XY} = 0.30).

Similarly, significant associations between gene expression and glioma risk were observed at the GBM loci 1p31.3 (*JAK1*, brain cortex and cerebellar hemisphere), 16p13.3 (*POLR3K*, whole blood) and 22q13.1 (*CTA-228A9.3*, brain cerebellum; *PICK1*, brain hippocampus) (Supplementary Fig. 4 and Supplementary Data 3). The non-GBM association at 1q32.1 marked by rs4252707 (Supplementary Fig. 3) maps to intron 8 of the gene encoding MDM4, a p53-binding protein. The SNP rs4252707 is in strong linkage disequilibrium (LD) with rs12031912 and rs12028476 ($r^2 = 0.92$), both of which map to the *MDM4* promoter. Although no significant eQTL was shown in any brain tissue, an association with *MDM4* was seen in blood (PSMR = 4.74×10^{-6}; $b_{XY} = 0.31$; Supplementary Fig. 4 and Supplementary Data 3). Overexpression of *MDM4* is a feature in glioma tumors containing wild-type *TP53* and no amplification of the *MDM2* gene, consistent with *MDM4* amplification being a mechanism by which the p53-dependent growth control is inactivated[23].

The 1q44 association with non-GBM that is marked by rs12076373 maps to intron 8 of *AKT3*, whose encoded product is one of the major downstream effectors of phosphatidylinositol 3-kinase (PI3K) and is highly expressed during active neurogenesis, with haploinsufficiency causing postnatal microcephaly and agenesis of the corpus callosum[24]. Notably, *AKT3* is hyper-expressed in glioma, thus having a role in tumor viability by activating DNA repair[25]. Although rs12076373 does not map to a regulatory element, the correlated SNPs rs12124113 ($r^2 = 0.94$) and rs59953491 ($r^2 = 0.90$) locate within an enhancer element in brain cells and tissues, including H9-derived neuronal progenitor cultured cells, cortex-derived primary cultured neurospheres and NH-A astrocytes.

The 3p14.1 association with non-GBM that is marked by rs11706832 localizes to intron 2 of *LRIG1*. Although we did not identify an eQTL in this gene, *LRIG1* is highly expressed in the brain and is a pan-negative regulator of the epidermal growth factor receptor (EGFR) signaling pathway, which inhibits hypoxia-induced vasculogenic mimicry via EGFR–PI3K–AKT pathway suppression and epithelial-to-mesenchymal transition[26]. Reduced *LRIG1* expression is linked to tumor aggressiveness, temozolomide resistance and radio-resistance[27, 28]. We have previously shown an association for glioma at *EGFR* (7p11.2)[7], which is well established to be pivotal in both the initiation of primary GBM and the progression of lower-grade glioma to grade IV. Although speculative, our new findings now suggest a more extensive pathway involving variation at *LRIG1* and *AKT3*.

Of particular interest is rs7572263, which maps to 2q33.3, localizes within intron 3 of *C2orf80* and is 50 kb telomeric to *IDH1*. Mutation of *IDH1* is a driver for gliomagenesis[29, 30] and is responsible for the CpG island methylator (G-CIMP) phenotype[31, 32]. Mutations in *IDH1* predominate in non-GBM glioma[33, 34]; therefore, the association at 2q33.3 is plausible as a basis for susceptibility to non-GBM glioma. In the absence of convincing eQTL or other functional support, this does not preclude *C2orf80* or another gene mapping to the region of LD as being the functional basis for the 2q33.3 association.

The maintenance of telomeres is central to cell immortalization, and it has a central role in gliomagenesis[35]. We have previously shown that the risk of GBM is strongly linked to genetic variation in the telomere-related genes *TERT* (5p15.33) and *RTEL1* (20q13.33), and possibly also *TERC* (3q26.2)[8, 9, 11]. The 10q24.33 association with non-GBM that is marked by rs11598018 lies intronic to *OBFC1*, which functions in a telomere-associated complex that protects telomeres independently of POT1 (ref. 36). The CST complex, whose components are encoded by *OBFC1, CTC1*, and *TEN1*, competes with shelterin for telomeric-DNA-inhibiting telomerase-based telomere extension[37]. The significant association between the risk of non-GBM tumors and *OBFC1* variation is particularly of note in light of our recent exome-sequencing report demonstrating that rare germline loss-of-function mutations in genes that encode components of the shelterin complex are a cause of familial oligodendroglioma[38]. The glioma risk alleles at *TERT, TERC* and *OBFC1* are associated with increased leukocyte telomere length, thereby supporting a relationship between genotype and biology (Supplementary Table 6)[35, 39, 40]. However, the *RTEL1* locus is not consistent with such a postulate, and recent data that have not shown a relationship between mutations in the *TERT* promoter and telomere length in glioma[41] raise the possibility of a role for extratelomeric effects.

The deregulation of pathways involved in telomere length and EGFR signaling are thus consistent with glioma risk being governed by pathways that are important in the longevity of glial cells, and they substantiate early observations that genetic susceptibility to GBM and non-GBM tumors is highly distinct, presumably reflecting different etiologies between GBM and non-GBM tumors (Fig. 2).

The other associations we identified mark genes with varying degrees of plausibility for having a role in glioma oncogenesis. The GBM association at 16p13.33 marked by rs2562152 localizes 3 kb telomeric to *MPG*, which encodes a *N*-methylpurine DNA glycosylase whose expression is linked to temozolomide resistance in glioma[42]. Although attractive as a candidate, the only genes for which there was found to be a significant association between expression and glioma risk were *POLR3K* and *C16ORF33* in blood (Supplementary Fig. 4 and Supplementary Data 3). At 1p31.3, only *JAK1* provided convincing evidence for a significant eQTL with glioma risk SNPs in brain tissue. The strongest association was shown in the cortex ($PSMR = 1.61 \times 10^{-6}$; $b_{XY} = 0.22$; Supplementary Fig. 4 and Supplementary Data 3), with the risk allele ‘T’ of rs12752552 showing increased *JAK1* expression. The *cis*-eQTL signal for *JAK1* in the cortex maps from 65.3 Mb to 65.35 Mb and shows a consistent direction of effect with the glioma-associated SNPs. JAK1–STAT6 signaling is increasingly being recognized to be relevant in glioma progression[43]. Hence, although *JAK1* remains an attractive candidate mechanistic basis for the glioma association at 1p31.3, we cannot exclude the possibility that the cluster of SNPs between 65.3 Mb and 65.35 Mb contains the true causal variant. In the absence of functional data, potential target genes for associations at 11q14.1 (GBM), 16p13.3 (non-GBM), 11q21 (non-GBM) and 14q12 (non-GBM) remain to be elucidated.

In conclusion, we have performed the largest glioma GWAS to date and have identified 13 new glioma risk loci, thereby providing further evidence for a polygenic basis of genetic susceptibility to glioma. Histological classification of glioma is, in part, being superseded by

molecular profiling[34, 44]; hence, it is important to understand the biology behind these risk variants in the context of molecularly defined glioma subtypes. Currently identified risk SNPs for glioma account for, at best, ~27% and ~37% of the familial risk of GBM and non-GBM tumors, respectively (Supplementary Table 7). Therefore, further GWAS-based analyses in concert with functional analyses should lead to additional insights into the biology and etiological basis of the different glioma histologies. Notably, such information can inform gene discovery initiatives and thus have a measurable effect on the successful development of new therapeutic agents.

ONLINE METHODS

Ethics

Collection of patient samples and associated clinico-pathological information was undertaken with written informed consent and relevant ethical review board approval at the respective study centers in accordance with the tenets of the Declaration of Helsinki. Specifically informed consent and ethical board approval was obtained from the South-East Multicentre Research Ethics Committee (MREC) (UK), the Scottish MREC (UK), the APHP ethical committee-CPP (Comité de Protection des Personnes) (France), the Ethics Commission of the Medical Faculty of the University of Bonn (Germany), the University of Texas MD Anderson Cancer Institutional Review Board (USA), the Mayo Clinic Office for Human Research Protection (USA), the UCSF Committee on Human Research (USA), the University Hospitals of Cleveland Institutional Review Board (USA) and the Cleveland Clinic Institutional Review Board (board for the Case Comprehensive Cancer Center) (USA). The diagnosis of glioma (ICDO-3 codes 9380-9480 or equivalent) was established through histology in all cases in accordance with World Health Organization guidelines. Every effort was made to classify tumors as GBM or non-GBM.

GWAS data sets

GICC, UK, French, German, MDA, SFAGS and GliomaScan—Studies participating in GICC are described in Amirian *et al.*[46] and in Supplementary Table 1. Briefly, they comprise 5,189 glioma cases and 3,827 controls that were ascertained through centers in the USA, Denmark, Sweden and the UK. Cases had newly diagnosed glioma, and controls had no personal history of central nervous system tumor at the time of ascertainment. Detailed information regarding recruitment protocol is given in Amirian *et al.*[46]. Cases and controls were genotyped using the Illumina Oncoarray according to the manufacturer's recommendations (Illumina Inc.). Individuals with a call rate <99%, as well as all individuals evaluated to be of non-European ancestry (<80% estimated European ancestry using the FastPop[47] procedure developed by the GAMEON consortium with HapMap version 2 CEU, JPT/CHB and YRI populations as a reference; Supplementary Fig. 5), were excluded. For pairs of apparent first-degree relatives, we removed the control from a case–control pair; otherwise, we excluded the individual with the lower call rate. SNPs with a call rate <95% were excluded as were those with a MAF <0.01 or those displaying significant deviation from the Hardy–Weinberg equilibrium (HWE) (i.e., $P < 10^{-5}$). After performing these quality-control measures, there were 4,572 cases and 3,286 controls remaining for downstream analyses.

The UK, French, German, MDA, SFAGS and GliomaScan GWAS of non-overlapping case–control series of Northern European ancestry have been the subject of previous studies. Briefly, the UK GWAS[7, 8, 10] was based on 636 cases (401 males; mean age 46 years) who were ascertained through the INTERPHONE study[48]. Individuals from the 1958 Birth Cohort (n = 2,930) served as a source of controls. The French GWAS[7, 10] comprised 1,495 patients with glioma who were ascertained through the Service de Neurologie Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris. The controls (n = 1,213) were ascertained from the SU.VI.MAX (Supplementation en Vitamines et Mineraux Antio Xydants) study of 12,735 healthy subjects (women aged 35–60 years; men aged 45–60 years)[49]. The German GWAS[10] comprised 880 patients who had undergone surgery for a glioma at the Department of Neurosurgery, University of Bonn Medical Center, between 1996 and 2008. Control subjects were taken from three population studies: KORA (Co-operative Health Research in the Region of Augsburg; n = 488)[50]; POPGEN (Population Genetic Cohort; n = 678)[51] and the Heinz Nixdorf Recall study (n = 380)[52]. Standard quality-control measures were applied to the UK, French and German GWAS and have previously been reported. The MDA GWAS[8] was based on 1,281 cases (786 males; mean age 47 years) who were ascertained through the MD Anderson Cancer Center, Texas, between 1990 and 2008. Individuals from the Cancer Genetic Markers of Susceptibility (CGEMS, n = 2,245) studies served as controls[53, 54]. Quality-control measures were applied as per the primary GWAS. The UCSF adult glioma case–control study (SFAGS–GWAS) included participants of the San Francisco Bay Area Adult Glioma Study (AGS). Details of subject recruitment for AGS have been reported previously[9, 12, 34, 55, 56]. Briefly, cases were adults (>18 years of age) with newly diagnosed, histologically confirmed glioma. Population-based cases who were diagnosed between 1991 and 2009 (series 1–4) and who were residing in the six San Francisco Bay area counties were ascertained using the Cancer Prevention Institute of California's early-case ascertainment system. Clinic-based cases who were diagnosed between 2002 and 2012 (series 3–5) were recruited from the UCSF Neuro-oncology Clinic, regardless of the place of residence. From 1991 to 2010, population-based controls from the same residential area as the population-based cases were identified using random digit-dialing and were frequency matched to population-based cases for age, gender and ethnicity. Between 2010 and 2012, all controls were selected from the UCSF general medicine phlebotomy clinic. Clinic-based controls were matched to clinic-based glioma cases for age, gender and ethnicity. Consenting participants provided blood, buccal and/or saliva specimens, and information, during in-person or telephone interviews. A total of 677 cases and 3,940 controls (including 3,347 Illumina iControlDB iControls) were used in the current analysis. For the GliomaScan GWAS[13], in addition to the published analysis, we excluded samples from the ATBC (Finnish study) and controls from NSHDS due to exhibiting outlying population ancestry after manual inspection of PCA plots. In total 1,653 cases and 2,725 controls were used in the current study.

GWAS data from the seven studies were imputed to >10 million SNPs with IMPUTE2 (v2.3)[57] software using a merged reference panel consisting of data from the 1000 Genomes Project (phase 1 integrated release 3, March 2012)[58] and UK10K (ALSPAC, EGAS00001000090 and EGAD00001000195, and TwinsUK EGAS00001000108 and EGAS00001000194 studies). Genotypes were aligned to the positive strand in both

imputation and genotyping. Imputation was conducted separately for each study, and in each the data were pruned to a common set of SNPs between cases and controls before imputation. We set thresholds for imputation quality to retain potential risk variants with MAF > 0.01. Poorly imputed SNPs, defined by an information measure <0.40 with IMPUTE2, were excluded, as were SNPs exhibiting a significant deviation from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-8}$) in controls. Test of association between imputed SNPs and glioma was performed using SNPTEST (v2.5)[59] under an additive frequentist model. The adequacy of the case–control matching and the possibility of differential genotyping of cases and controls were formally evaluated using Q-Q plots of test statistics (Supplementary Fig. 1). Where appropriate, principal components, generated using common SNPs, were included in the analysis to limit the effects of cryptic population stratification that otherwise might cause inflation of test statistics. Principal components, based on genotyped SNPs, were generated for the GICC, GliomaScan, MDA-GWAS and SFAGS studies using PLINK[60]. Eigenvectors for the German GWAS were inferred using smartpca (part of EIGENSOFTv2.4)[61] by merging cases and controls with Phase II HapMap samples[10]. PCA plots for all studies are provided in Supplementary Figure 4.

UCSF-Mayo GWAS—The UCSF-Mayo study comprised Mayo cases ($n = 945$) and UCSF cases ($n = 574$) and Mayo Clinic Biobank control ($n = 806$) data. The Mayo Clinic case–control study has been described previously[9, 34, 62]. Briefly, adult cases (>18 years of age) were identified at diagnosis (diagnosed at Mayo Clinic) or at pathologic confirmation (diagnosed elsewhere and treated at Mayo Clinic), and the patients had a surgical resection or biopsy between 1973 and 2014. Consenting participants provided blood, buccal and/or saliva specimens, and information, during in-person or telephone interviews. This analysis used 574 non-overlapping cases from the UCSF Adult Glioma Study described above. Mayo Clinic and UCSF cases were genotyped using the Illumina Oncoarray. The Mayo Clinic Biobank controls comprised volunteers who donated biological specimens and provided risk factor data, access to clinical data obtained from the medical record and consent to participate in any study approved by the Access Committee. Recruitment for the Mayo Clinic Biobank took place from April 2009 through December 2015. Although participants could be unselected volunteers, the vast majority of participants were contacted as part of a pre-scheduled medical examination in the Department of Medicine, Divisions of Community Internal Medicine, Family Medicine and General Internal Medicine at Mayo Clinic sites in Rochester (Minnesota), Jacksonville (Florida), and the Mayo Clinic Health System sites in La Crosse and Onalaska (Wisconsin). All individuals were aged 18 years and older at the time of consent. Illumina Omni Express genotyping arrays were run on the 806 Mayo Clinic Biobank participants.

Quality-control analyses were performed on each cohort separately (Mayo cases, UCSF cases and Mayo Clinic Biobank controls). SNPs with call rates <95% were removed, followed by removal of subjects with call rates <95%. Concordance of replicate samples was assessed, and the sample with the higher call rate was retained. Subject's sex was verified using the sex check option in PLINK. Relationship checking was performed by estimating the proportion of alleles shared identical by descent (IBD) for all pairs of subjects in PLINK[60]. STRUCTURE[63] was used to assess population admixture with 1000 Genomes as

a reference. Subjects indicated to be non-Caucasian were excluded. Prior to imputation, SNPs were tested for HWE, and SNPs with HWE $P < 10^{-6}$ were removed. Mayo Clinic, UCSF and Mayo Clinic Biobank SNP data were each phased and imputed using the Michigan Imputation Server with the Haplotype Reference Consortium (release 1; http://www.haplotype-reference-consortium.org) as reference. Genotypes were forward-strand-aligned to the 1000 Genomes reference, and for ambiguous SNPs the Browning strand checking utility was used (http://faculty.washington.edu/sguy/beagle/strand_switching/strand_switching.html). PCA was used to correct for population stratification using SNPs common to cases and controls. The first three principal components were significantly ($P < 0.05$) associated with case–control status. An additive logistic regression model was used to assess the association between each SNP and disease status, with genotype being coded as 0, 1 or 2 copies of the minor allele, adjusted for age, sex and the first three principal components.

Meta-analysis and additional statistical analyses

Meta-analyses were performed using the fixed-effects inverse-variance method based on the β-esti-mates and standard errors from each study using META (v1.6)[64]. Cochran's Q-statistic was used to test for heterogeneity, and the I^2 statistic was used to quantify the proportion of the total variation due to heterogeneity[65], taking I^2 values >75 to indicate significant heterogeneity. Using the meta-analysis summary statistics and LD correlations from a reference panel of the 1000 Genomes Project combined with UK10K, we used GCTA[66, 67] to perform conditional association analysis. Association statistics were calculated for all SNPs, conditioning on the top SNP in each locus showing genome-wide significance. This was carried out in a step-wise fashion. We performed a case-only analysis to test for differences in SNP-risk-allele frequency between GBM and non-GBM tumors.

ENCODE and chromatin state dynamics

Risk SNPs and their proxies (i.e., $r^2 > 0.8$ in the 1000 Genomes EUR reference panel) were annotated for putative functional effect using HaploReg (v4)[68], RegulomeDB[69] and SeattleSeq Annotation[70]. These servers make use of data from ENCODE, genomic evolutionary rate profiling (GERP) conservation metrics, combined annotation-dependent depletion (CADD) scores and PolyPhen scores. We searched for overlap of associated SNPs with enhancers defined by the FANTOM5 enhancer atlas[19], annotating by overlap with ubiquitous, permissive and robust enhancers, as well as enhancer–promoter correlations and enhancers specifically expressed in astrocytes, neuronal stem cells and brain tissue. Similarly, we searched for overlap with 'super-enhancer' regions, as defined by Hnisz *et al.*[20], restricting analysis to data from U87 GBM cells, astrocyte cells and brain tissue. We additionally made use of 15-state chromHMM data from H1- and H9-derived neuronal progenitor cells available from the Epigenome Roadmap Project[71]. Enhancer enrichment analysis was carried out using HaploReg (v4.0)[68]. Briefly, from a query list of variants, the overlap with enhancers in each of 107 cell types, as predicted from the Roadmap Epigenomics Project[71] chromatin-state segmentations, was calculated. A binomial test for enrichment was performed against a background set of all (i) 1000 Genomes variants with MAF > 0.05 and (ii) all unique GWAS loci in the European population. We applied a cutoff of $P < 3.94 \times 10^{-4}$ corresponding to a Bonferroni correction for 127 cell lines and tissues.

Expression quantitative trait loci (eQTL) analysis

To examine the relationship between SNP genotype and gene expression, we carried out summary-data-based mendelian randomization (SMR) analysis as per Zhu *et al.*[18] (at http://cnsgenomics.com/software/smr/index.html). We used publicly available brain tissue data from the GTEx[16] (http://www.gtexportal.org) v6p release. Briefly, GWAS summary statistics files were generated from the meta-analysis. Reference files were generated from merging 1000 Genomes phase 3 and UK10K (ALSPAC and TwinsUK) vcfs. Summary eQTL files for GTEx samples were generated from downloaded v6p "all_snpgene_pairs" files. Besd files were generated from these summary eQTL files using the –make-besd command. Additionally, we analyzed downloaded whole-blood eQTL data from Westra *et al.*[17]. Results from the SMR test for each of the 13 new glioma loci are reported in Supplementary Data 3. As previously advocated[18], only probes with at least one eQTL P value $<5.0 \times 10^{-8}$ were considered for SMR analysis. We set a threshold for the SMR test of $P_{SMR} < 1.06 \times 10^{-4}$ corresponding to a Bonferroni correction for 473 tests (473 probes with a top eQTL $P < 5.0 \times 10^{-8}$ across the 13 loci, 10 brain regions and Westra data set). For all genes passing this threshold, we generated plots of the eQTL and GWAS associations at the locus, as well as plots of GWAS and eQTL effect sizes (i.e., corresponding to input for the HEIDI heterogeneity test). HEIDI test P values <0.05 were taken to indicate significant heterogeneity. Respective SMR plots for significant eQTLs are shown in Supplementary Figure 4.

Additional statistical and bioinformatics analysis

Estimates of individual variance in risk associated with glioma risk SNPs was carried out using the method described in Pharoah *et al.*[72], assuming the familial risk of high-grade and low-grade glioma to be 1.76 and 1.54, respectively, from analysis of the Swedish series in Scheurer *et al.*[73]. Briefly, for a single allele (i) of frequency p, relative risk R and ln risk r, the variance (V_i) of the risk distribution due to that allele is given by:

$$V_i=(1-p)^2E^2+2p(1-p)(r-E)^2+p^2(2r-E)^2$$

Where E is the expected value of r given by:

$$E=2p(1-p)r+2p^2r$$

For multiple risk alleles, the distribution of risk in the population tends toward the normal with variance:

$$V=\sum V_i$$

The total genetic variance (V) for all susceptibility alleles has been estimated to be √1.77. Thus, the fraction of the genetic risk explained by a single allele is given by:

$$V_i/V$$

LD metrics were calculated in vcftools (v0.1.12b)[74] using UK10K data and plotted using visPIG[75]. LD blocks were defined on the basis of HapMap recombination rate (cM/Mb), as defined using the Oxford recombination hotspots and on the basis of distribution of confidence intervals defined by Gabriel *et al.*[76].

Data availability

Genotype data from the GICC GWAS are available from the database of Genotypes and Phenotypes (dbGaP) under accession phs001319.v1.p1. Additionally, genotypes from the GliomaScan GWAS can be accessed through dbGaP accession phs000652.v1.p1. Data from the other studies are available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Beatrice S Melin[1,41], Jill S Barnholtz-Sloan[2,41], Margaret R Wrensch[3,4,41], Christoffer Johansen[5,41], Dora Il'yasova[6,7,8,41], Ben Kinnersley[9,41], Quinn T Ostrom[2], Karim Labreche[9,10], Yanwen Chen[2], Georgina Armstrong[11], Yanhong Liu[11], Jeanette E Eckel-Passow[12], Paul A Decker[12], Marianne Labussière[10], Ahmed Idbaih[10,13], Khe Hoang-Xuan[10,13], Anna-Luisa Di Stefano[10,13], Karima Mokhtari[10,13], Jean-Yves Delattre[10,13], Peter Broderick[9], Pilar Galan[14], Konstantinos Gousias[15], Johannes Schramm[15], Minouk J Schoemaker[9], Sarah J Fleming[16], Stefan Herms[16], Stefanie Heilmann[17], Markus M Nöthen[17], Heinz-Erich Wichmann[18,19,20], Stefan Schreiber[21], Anthony Swerdlow[9,22], Mark Lathrop[23], Matthias Simon[15], Marc Sanson[10,13], Ulrika Andersson[1], Preetha Rajaraman[24], Stephen Chanock[24], Martha Linet[24], Zhaoming Wang[24], Meredith Yeager[24], GliomaScan Consortium[25], John K Wiencke[3,4], Helen Hansen[3], Lucie McCoy[3], Terri Rice[3], Matthew L Kosel[12], Hugues Sicotte[12], Christopher I Amos[26], Jonine L Bernstein[27], Faith Davis[28], Dan Lachance[29], Ching Lau[30], Ryan T Merrell[31], Joellen Shildkraut[7,8], Francis Ali-Osman[7,32], Siegal Sadetzki[33,34], Michael Scheurer[30], Sanjay Shete[35], Rose K Lai[36,42], Elizabeth B Claus[37,38,42], Sara H Olson[27,42], Robert B Jenkins[39,42], Richard S Houlston[9,40,42], and Melissa L Bondy[11,42]

Affiliations

[1]Department of Radiation Sciences, Umeå University, Umeå, Sweden

[2]Case Comprehensive Cancer Center, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

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

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

[5]Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark and Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

[6]Department of Epidemiology and Biostatistics, School of Public Health, Georgia State University, Atlanta, Georgia, USA

[7]Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina, USA

[8]Cancer Control and Prevention Program, Department of Community and Family Medicine, Duke University Medical Center, Durham, North Carolina, USA

[9]Division of Genetics and Epidemiology, Institute of Cancer Research, London, UK

[10]Sorbonne Universités UPMC Univ Paris 06, INSERM CNRS, U1127, UMR 7225, ICM, Paris, France

[11]Department of Medicine, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, USA

[12]Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota, USA

[13]AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Service de neurologie 2-Mazarin, Paris, France

[14]Université Paris 13 Sorbonne Paris Cité, INSERM U557, INRA U1125, CNAM, Paris, France

[15]Department of Neurosurgery, University of Bonn Medical Center, Bonn, Germany

[16]Centre for Epidemiology and Biostatistics, Faculty of Medicine and Health, University of Leeds, Leeds, UK

[17]Institute of Human Genetics, University of Bonn, Bonn, Germany

[18]Helmholtz Center Munich, Institute of Epidemiology I, Munich, Germany

[19]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany

[20]Institute of Medical Statistics and Epidemiology, Technical University Munich, Munich, Germany

[21]1st Medical Department, University Clinic Schleswig–Holstein, Campus Kiel, Kiel, Germany

[22]Division of Breast Cancer Research, Institute of Cancer Research, London, UK

[23]Génome Québec, Department of Human Genetics, McGill University, Montreal, Quebec, Canada

[24]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

[26]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA

[27]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA

[28]School of Public Health, University of Alberta, Edmonton, Alberta, Canada

[29]Department of Neurology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, Minnesota, USA

[30]Department of Pediatrics, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, USA

[31]Department of Neurology, NorthShore University HealthSystem, Evanston, Illinois, USA

[32]Department of Surgery, Duke University Medical Center, Durham, North Carolina, USA

[33]Cancer and Radiation Epidemiology Unit, Gertner Institute, Chaim Sheba Medical Center, Tel Hashomer, Israel

[34]Department of Epidemiology and Preventive Medicine, School of Public Health, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

[35]Department of Biostatistics, University of Texas Maryland Anderson Cancer Center, Houston, Texas, USA

[36]Departments of Neurology and Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

[37]School of Public Health, Yale University, New Haven, Connecticut, USA

[38]Department of Neurosurgery, Brigham and Women's Hospital, Boston, Massachusetts, USA

[39]Department of Laboratory Medicine and Pathology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, Minnesota, USA

[40]Division of Molecular Pathology, Institute of Cancer Research, London, UK

Acknowledgments

We are grateful to all of the patients and individuals for their participation, and we would also like to thank the clinicians and other hospital staff members, cancer registries and the study staff members in the respective centers who contributed to the blood sample and data collection.

The GICC was supported by grants from the US National Institutes of Health (NIH) (R01CA139020 (M.L.B. and B.S.M.), R01CA52689 (M.R.W.), R01CA52689 (M.L.B.) and P30CA125123 (M. Scheurer). Additional support was provided by the McNair Medical Institute (M. Scheurer) and the Population Sciences Biorepository at Baylor College of Medicine (M. Scheurer).

In Sweden, work was additionally supported by Acta Oncologica through the Royal Swedish Academy of Science (B.S.M.'s salary) and by the Swedish Research Council (B.S.M.) and the Swedish Cancer Foundation (B.S.M.). We are grateful to the National Clinical Brain Tumor Group and to all of the clinicians and research nurses throughout Sweden who identified all of the cases.

In the UK, funding was provided by Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund (R.S.H., B.K. and P.B.), the Wellcome Trust (R.S.H., B.K. and P.B.) and the DJ Fielding Medical Research Trust (R.S.H., B.K. and P.B.). The National Brain Tumor Study is supported by the National Cancer Research Network, and we acknowledge all clinicians and healthcare professionals who contributed to this initiative. The UK INTERPHONE study was supported by the European Union Fifth Framework Program 'Quality of Life and Management of Living Resources' (QLK4-CT-1999-01563) (A.S., M.J.S. and S.J.F.) and the International Union against Cancer (UICC) (A.S., M.J.S. and S.J.F.). The UICC received funds from the Mobile Manufacturers' Forum and the GSM Association. Provision of funds via the UICC was governed by agreements that guaranteed INTERPHONE's scientific independence (http://www.iarc.fr/ENG/Units/RCAd.html), and the views expressed in the paper are not necessarily those of the funders. The UK centers were also supported by the Mobile Telecommunications and Health Research (MTHR) Programme, and the Northern UK Centre (A.S., M.J.S. and S.J.F.) was supported by the Health and Safety Executive, Department of Health and Safety Executive and the UK Network Operators.

In France, funding was provided by the Ligue Nationale Contre le Cancer (J.-Y.D.), the Fondation ARC (M. Sanson), the Institut National du Cancer (INCa; PL046; (M. Sanson)), the French Ministry of Higher Education and Research and the program "Investissements d'avenir" ANR-10-IAIHU-06 (M. Sanson, J.-Y.D., M. Labussière, A.-L.D.S., P.G., K.M., A.I., K.H.-X. and K.L.). This study was additionally supported by a grant from Génome Québec, le Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie (MESRST) Québec (M. Lathrop) and McGill University (M. Lathrop).

In Germany, funding was provided to M. Simon and J. Schramm by the Deutsche Forschungsgemeinschaft (Si552, Schr285), the Deutsche Krebshilfe (70-2385-Wi2, 70-3163-Wi3, 10-6262) and BONFOR. Funding for the WTCCC was provided by the Wellcome Trust (076113 and 085475; M. Simon and J. Schramm). The KORA Ausburg studies are supported by grants from the German Federal Ministry of Education and Research (BMBF) and were mainly financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg. This work was financed by the German National Genome Research Network (NGFN) (S. Schreiber and H.E.-W.) and supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ (S. Schreiber and H.E.-W.). Generation of the German control data was partially supported by a grant of the German Federal Ministry of Education and Research (BMBF) through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders), under the auspices of the e:Med research and funding concept (01ZX1314A) (M.M.N., S. Herms and S. Heilmann). M.M.N. is a member of the DFG-funded Excellence Cluster ImmunoSensation and received support from the Alfried Krupp von Bohlen und Halbach-Stiftung.

For the UK GWAS, we acknowledge the funders, organizations and individuals who contributed to the blood sample and data collection as listed in Hepworth *et al.*[45]. MD Anderson acknowledges the work of P. Adatto, F. Morice, H. Zhang, V. Levin, A.W.K. Yung, M. Gilbert, R. Sawaya, V. Puduvalli, C. Conrad, F. Lang and J. Weinberg from the Brain and Spine Center for the MDA GWAS. For the French study, we are indebted to A. Rahimian (Onconeurotek), A.M. Lekieffre and M. Brandel for help in collecting data and to Y. Marie for database support. For the German study, we are indebted to B. Harzheim (Bonn), S. Ott and A. Müller-Erkwoh (Bonn) for help with the acquisition of clinical data and to R. Mahlberg (Bonn), who provided technical support. The UK study made use of control genotyping data generated by the Wellcome Trust Case–Control Consortium. A full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk. The MDA GWAS made use of control genotypes from the CGEMS prostate and breast cancer studies. A full list of the investigators who contributed to the generation of the data is available from http://cgems.cancer.gov/. French controls were taken from the SU.VI.MAX study. The German GWAS made use of genotyping data from three population control sources: KORA-gen39, the Heinz-Nixdorf RECALL study and POPGEN. The HNR cohort was established with the support of the Heinz-Nixdorf Foundation. F.D. received support from the BONFOR Programme of the University of Bonn, Germany.

The UCSF Adult Glioma Study was supported by the NIH (grant numbers R01CA52689 (M.R.W. and J.K.W.), P50CA097257 (M.R.W. and J.K.W.), R01CA126831 (J.K.W.) and R01CA139020 (M.R.W.)), the Loglio Collective (M.R.W. and J.K.W.), the National Brain Tumor Foundation (M.R.W.), the Stanley D. Lewis and Virginia S. Lewis Endowed Chair in Brain Tumor Research (M.R.W.), the Robert Magnin Newman Endowed Chair in Neuro-oncology (J.K.W.) and by donations from the families and friends of J. Berardi, H. Glaser, E. Olsen, R.E. Cooper and W. Martinusen. This project also was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH, through UCSF–CTSI grant UL1 RR024131 (UCSF CTSI). The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. 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).

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. Other significant contributors for the UCSF Adult Glioma Study include M. Berger, P. Bracci, S. Chang, J. Clarke, A. Molinaro, A. Perry, M. Pezmecki, M. Prados, I. Smirnov, T. Tihan, K. Walsh, J. Wiemels and S. Zheng.

At Mayo, the authors wish to acknowledge the study participants and the clinicians and research staff at the participating medical centers, the Mayo Clinic Biobank and Biospecimens Accessioning and Processing Shared Resource (in particular its manager, M. Cicek). Work at the Mayo Clinic beyond the GICC was also supported by the NIH (grants P50CA108961 (B. O'Neill) and P30CA15083 (R. Diasio)), the National Institute of Neurological Disorders and Stroke (grant RC1NS068222Z (R.B.J.)), the Bernie and Edith Waterman Foundation (R.B.J.) and the Ting Tsung and Wei Fong Chao Family Foundation (R.B.J.).

The GliomaScan Consortium comprised (apart from authors listed in the author list): L.E.B. Freeman (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA), S. Koutros (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA), D. Albanes (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA), K. Visvanathan (Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA and Johns Hopkins Sidney Kimmel Comprehensive Cancer Center, Baltimore, Maryland, USA), V.L. Stevens (Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA), R. Henriksson (Department of Radiation Sciences, Oncology, Umea University, Umea, Sweden), D.S. Michaud (Department of Public Health and Community Medicine, Tufts University Medical School, Boston, Massachusetts, USA), M. Feychting (Unit of Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden), A. Ahlbom (Unit of Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden), G.G. Giles (Cancer Epidemiology Centre, Cancer Council of Victoria, Melbourne, Victoria, Australia and Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Victoria, Australia), R. Milne (Cancer Epidemiology Centre, Cancer Council of Victoria, Melbourne, Victoria, Australia and Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Victoria, Australia), R. McKean-Cowdin (Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA), L. Le Marchand (Cancer Research Center, University of Hawaii, Honolulu, Hawaii, USA), M. Stampfer (Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA and Departments of Epidemiology and Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA), A.M. Ruder (National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Cincinnati, Ohio, USA), T. Carreon (National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Cincinnati, Ohio, USA), G. Hallmans (Department of Public Health and Clinical Medicine/Nutritional Research, Umea University, Umea, Sweden), A. Zeleniuch-Jacquotte (Division of Epidemiology, Department of Environmental Medicine, New York University School of Medicine, New York, New York, USA), J.M. Gaziano (Massachusetts Veteran's Epidemiology, Research and Information Center, Geriatric Research Education and Clinical Center, VA Boston Healthcare System, Boston, Massachusetts, USA), H.D. Sesso (Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA), M.P. Purdue (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA), E. White (Fred Hutchinson Cancer Research Center, Seattle, Washington, USA and Department of Epidemiology, University of Washington, Seattle, Washington, USA) and J. Buring (Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA).

UK10K data generation and access was organized by the UK10K consortium and funded by the Wellcome Trust.

References

1. Bondy ML, et al. Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium. Cancer. 2008; 113:1953–1968. [PubMed: 18798534]
2. Ostrom QT, 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. [PubMed: 26511214]
3. Louis DN, et al. The 2007 WHO classification of tumors of the central nervous system. Acta Neuropathol. 2007; 114:97–109. [PubMed: 17618441]
4. Ostrom QT, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. Neuro-oncol. 2014; 16:iv1–iv63. [PubMed: 25304271]
5. Ostrom QT, et al. The epidemiology of glioma in adults: a 'state of the science' review. Neuro-oncol. 2014; 16:896–913. [PubMed: 24842956]

6. Hemminki K, Tretli S, Sundquist J, Johannesen TB, Granstrom C. Familial risks in nervous-system tumors: a histology-specific analysis from Sweden and Norway. Lancet Oncol. 2009; 10:481–488. [PubMed: 19356978]
7. Sanson M, et al. Chromosome 7p11.2 (*EGFR*) variation influences glioma risk. Hum. Mol. Genet. 2011; 20:2897–2904. [PubMed: 21531791]
8. Shete S, et al. Genome-wide association study identifies five susceptibility loci for glioma. Nat. Genet. 2009; 41:899–904. [PubMed: 19578367]
9. Wrensch M, et al. Variants in the *CDKN2B* and *RTEL1* regions are associated with high-grade glioma susceptibility. Nat. Genet. 2009; 41:905–908. [PubMed: 19578366]
10. Kinnersley B, et al. Genome-wide association study identifies multiple susceptibility loci for glioma. Nat. Commun. 2015; 6:8559. [PubMed: 26424050]
11. Walsh KM, et al. Variants near *TERT* and *TERC* influencing telomere length are associated with high-grade glioma risk. Nat. Genet. 2014; 46:731–735. [PubMed: 24908248]
12. Jenkins RB, 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; 44:1122–1125. [PubMed: 22922872]
13. Rajaraman P, et al. Genome-wide association study of glioma and meta-analysis. Hum. Genet. 2012; 131:1877–1888. [PubMed: 22886559]
14. Stacey SN, et al. A germline variant in the *TP53* polyadenylation signal confers cancer susceptibility. Nat. Genet. 2011; 43:1098–1103. [PubMed: 21946351]
15. Kinnersley B, et al. Quantifying the heritability of glioma using genome-wide complex trait analysis. Sci. Rep. 2015; 5:17267. [PubMed: 26625949]
16. Lonsdale J, et al. The Genotype–Tissue Expression (GTEx) project. Nat. Genet. 2013; 45:580–585. [PubMed: 23715323]
17. Westra HJ, et al. Systematic identification of *trans* eQTLs as putative drivers of known disease associations. Nat. Genet. 2013; 45:1238–1243. [PubMed: 24013639]
18. Zhu Z, et al. Integration of summary data from GWAS and eQTL studies predicts complex-trait gene targets. Nat. Genet. 2016; 48:481–487. [PubMed: 27019110]
19. Andersson R, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507:455–461. [PubMed: 24670763]
20. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. [PubMed: 24119843]
21. Ruark E, et al. Identification of nine new susceptibility loci for testicular cancer, including variants near *DAZL* and *PRDM14*. Nat. Genet. 2013; 45:686–689. [PubMed: 23666240]
22. Wu C, et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene–environment interactions. Nat. Genet. 2012; 44:1090–1097. [PubMed: 22960999]
23. Riemenschneider MJ, et al. Amplification and overexpression of the *MDM4* (*MDMX*) gene from 1q32 in a subset of malignant gliomas without *TP53* mutation or *MDM2* amplification. Cancer Res. 1999; 59:6091–6096. [PubMed: 10626796]
24. Boland E, et al. Mapping of deletion and translocation breakpoints in 1q44 implicates the serine–threonine kinase AKT3 in postnatal microcephaly and agenesis of the corpus callosum. Am. J. Hum. Genet. 2007; 81:292–303. [PubMed: 17668379]
25. Turner KM, et al. Genomically amplified Akt3 activates DNA repair pathway and promotes glioma progression. Proc. Natl. Acad. Sci. USA. 2015; 112:3421–3426. [PubMed: 25737557]
26. Gur G, et al. LRIG1 restricts growth factor signaling by enhancing receptor ubiquitylation and degradation. EMBO J. 2004; 23:3270–3281. [PubMed: 15282549]
27. Yang JA, et al. LRIG1 enhances the radio-sensitivity of radio-resistant human glioblastoma U251 cells via attenuation of the EGFR–AKT signaling pathway. Int. J. Clin. Exp. Pathol. 2015; 8:3580–3590. [PubMed: 26097540]
28. Wei J, et al. miR-20a mediates temozolomide resistance in glioblastoma cells via negatively regulating LRIG1 expression. Biomed. Pharmacother. 2015; 71:112–118. [PubMed: 25960225]

29. Watanabe T, Nobusawa S, Kleihues P, Ohgaki H. *IDH1* mutations are early events in the development of astrocytomas and oligodendrogliomas. Am. J. Pathol. 2009; 174:1149–1153. [PubMed: 19246647]
30. Yan H, et al. *IDH1* and *IDH2* mutations in gliomas. N. Engl. J. Med. 2009; 360:765–773. [PubMed: 19228619]
31. Noushmehr H, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell. 2010; 17:510–522. [PubMed: 20399149]
32. Christensen BC, et al. DNA methylation, isocitrate dehydrogenase mutation and survival in glioma. J. Natl. Cancer Inst. 2011; 103:143–153. [PubMed: 21163902]
33. Sanson M, et al. Isocitrate dehydrogenase 1 codon 132 mutation is an important prognostic biomarker in gliomas. J. Clin. Oncol. 2009; 27:4150–4154. [PubMed: 19636000]
34. Eckel-Passow JE, et al. Glioma groups based on 1p/19q, *IDH* and *TERT* promoter mutations in tumors. N. Engl. J. Med. 2015; 372:2499–2508. [PubMed: 26061753]
35. Walsh KM, et al. Telomere maintenance and the etiology of adult glioma. Neuro-oncol. 2015; 17:1445–1452. [PubMed: 26014050]
36. Miyake Y, et al. RPA-like mammalian Ctc1–Stn1–Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. Mol. Cell. 2009; 36:193–206. [PubMed: 19854130]
37. Chen LY, Redon S, Lingner J. The human CST complex is a terminator of telomerase activity. Nature. 2012; 488:540–544. [PubMed: 22763445]
38. Bainbridge MN, et al. Germline mutations in shelterin complex genes are associated with familial glioma. J. Natl. Cancer Inst. 2014; 107:384. [PubMed: 25482530]
39. Zhang C, et al. Genetic determinants of telomere length and risk of common cancers: a mendelian randomization study. Hum. Mol. Genet. 2015; 24:5356–5366. [PubMed: 26138067]
40. Walsh KM, et al. Longer genotypically estimated leukocyte telomere length is associated with increased adult glioma risk. Oncotarget. 2015; 6:42468–42477. [PubMed: 26646793]
41. Ceccarelli M, et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. Cell. 2016; 164:550–563. [PubMed: 26824661]
42. Xipell E, et al. Endoplasmic-reticulum-stress-inducing drugs sensitize glioma cells to temozolomide through downregulation of MGMT, MPG and Rad51. Neuro-oncol. 2016; 18:1109–1119. [PubMed: 26951384]
43. Nicolas CS, et al. The role of JAK–STAT signaling within the CNS. JAK-STAT. 2013; 2:e22925. [PubMed: 24058789]
44. Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N. Engl. J. Med. 2015; 372:2481–2498. [PubMed: 26061751]
45. Hepworth SJ, et al. Mobile phone use and risk of glioma in adults: case–control study. BMJ. 2006; 332:883–887. [PubMed: 16428250]
46. Amirian ES, et al. The Glioma International Case–Control Study: a report from the Genetic Epidemiology of Glioma International Consortium. Am. J. Epidemiol. 2016; 183:85–91. [PubMed: 26656478]
47. Li Y, et al. FastPop: a rapid principal component–derived method to infer intercontinental ancestry using genetic data. BMC Bioinformatics. 2016; 17:122. [PubMed: 26961892]
48. Cardis E, et al. The INTERPHONE study: design, epidemiological methods and description of the study population. Eur. J. Epidemiol. 2007; 22:647–664. [PubMed: 17636416]
49. Hercberg S, et al. The SU.VI.MAX study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. Arch. Intern. Med. 2004; 164:2335–2342. [PubMed: 15557412]
50. Wichmann HE, Gieger C, Illig T. MONICA–KORA Study Group. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen. 2005; 67(Suppl. 1):S26–S30. [PubMed: 16032514]
51. Krawczak M, et al. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype–phenotype relationships. Community Genet. 2006; 9:55–61. [PubMed: 16490960]

52. Schmermund A, et al. Assessment of clinically silent atherosclerotic disease, and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL study. Risk factors, evaluation of coronary calcium and lifestyle. Am. Heart J. 2002; 144:212–218. [PubMed: 12177636]
53. Hunter DJ, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. 2007; 39:870–874. [PubMed: 17529973]
54. Yeager M, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat. Genet. 2007; 39:645–649. [PubMed: 17401363]
55. Wiemels JL, et al. History of allergies among adults with glioma and controls. Int. J. Cancer. 2002; 98:609–615. [PubMed: 11920623]
56. Felini MJ, et al. Reproductive factors and hormone use risk of adult gliomas. Cancer Causes Control. 2009; 20:87–96. [PubMed: 18766447]
57. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]
58. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–1073. [PubMed: 20981092]
59. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 2007; 39:906–913. [PubMed: 17572673]
60. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007; 81:559–575. [PubMed: 17701901]
61. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]
62. Jenkins RB, et al. Distinct germline polymorphisms underlie glioma morphologic heterogeneity. Cancer Genet. 2011; 204:13–18. [PubMed: 21356187]
63. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–959. [PubMed: 10835412]
64. Liu JZ, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat. Genet. 2010; 42:436–440. [PubMed: 20418889]
65. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br. Med. J. 2003; 327:557–560. [PubMed: 12958120]
66. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex-trait analysis. Am. J. Hum. Genet. 2011; 88:76–82. [PubMed: 21167468]
67. Yang J, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 2012; 44:369–375. [PubMed: 22426310]
68. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–D934. [PubMed: 22064851]
69. Boyle AP, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012; 22:1790–1797. [PubMed: 22955989]
70. Ng SB, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009; 461:272–276. [PubMed: 19684571]
71. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. [PubMed: 25693563]
72. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction and targeted prevention of breast cancer. N. Engl. J. Med. 2008; 358:2796–2803. [PubMed: 18579814]
73. Scheurer ME, et al. Familial aggregation of glioma: a pooled analysis. Am. J. Epidemiol. 2010; 172:1099–1107. [PubMed: 20858744]
74. Danecek P, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156–2158. [PubMed: 21653522]

75. Scales M, Jager R, Migliorini G, Houlston RS, Henrion MY. visPIG—a web tool for producing multiregion, multitrack, multiscale plots of genetic data. PLoS One. 2014; 9:e107497. [PubMed: 25208325]

76. Gabriel SB, et al. The structure of haplotype blocks in the human genome. Science. 2002; 296:2225–2229. [PubMed: 12029063]

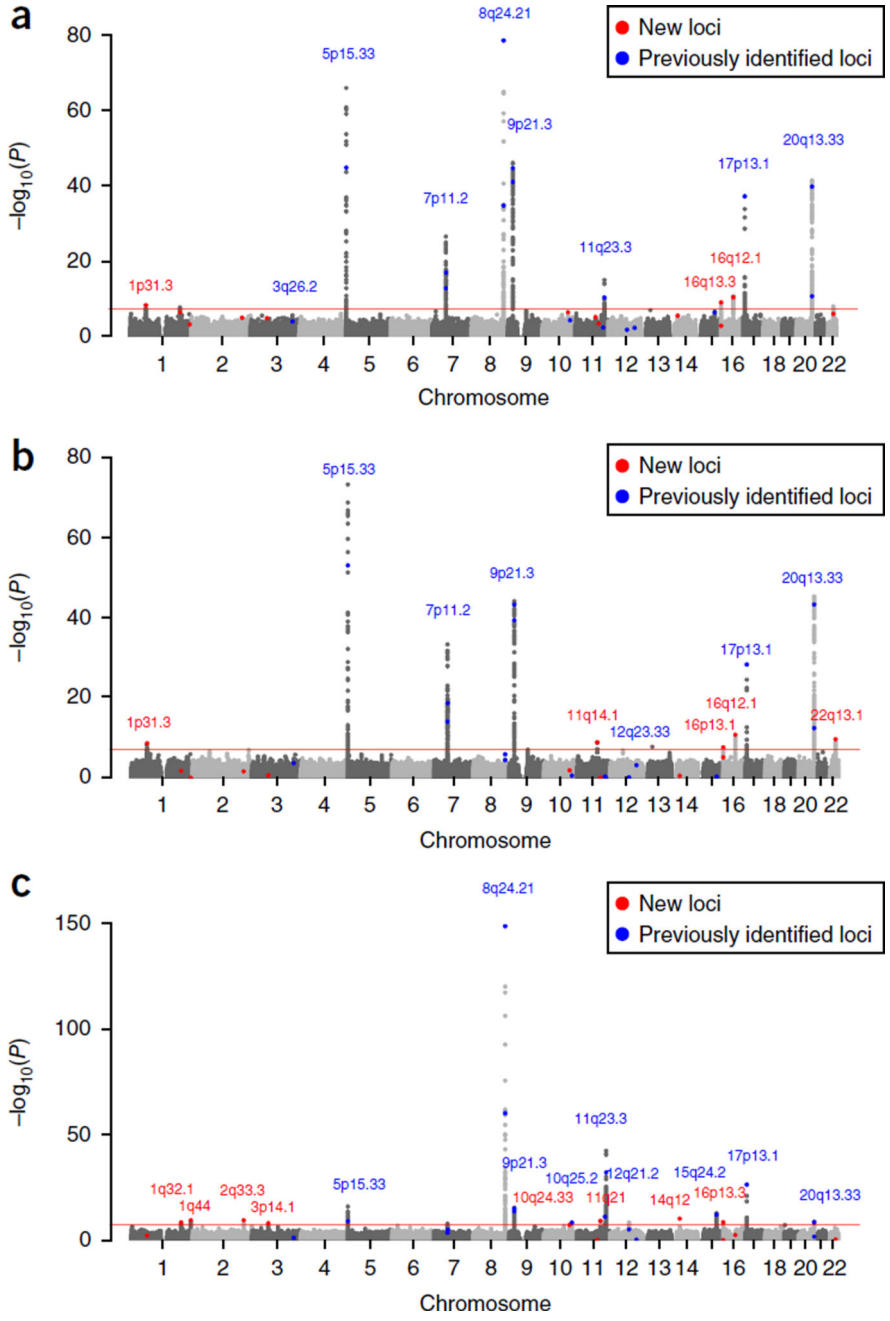

Figure 1.

Genome-wide discovery-phase meta-analysis *P*-values ($-\log_{10}P$) plotted against their chromosomal positions. (**a**) All glioma. (**b**) GBM. (**c**) Non-GBM tumors. The red horizontal line corresponds to a significance threshold of $P = 5.0 \times 10^{-8}$. New and known loci are labeled in red and blue, respectively.

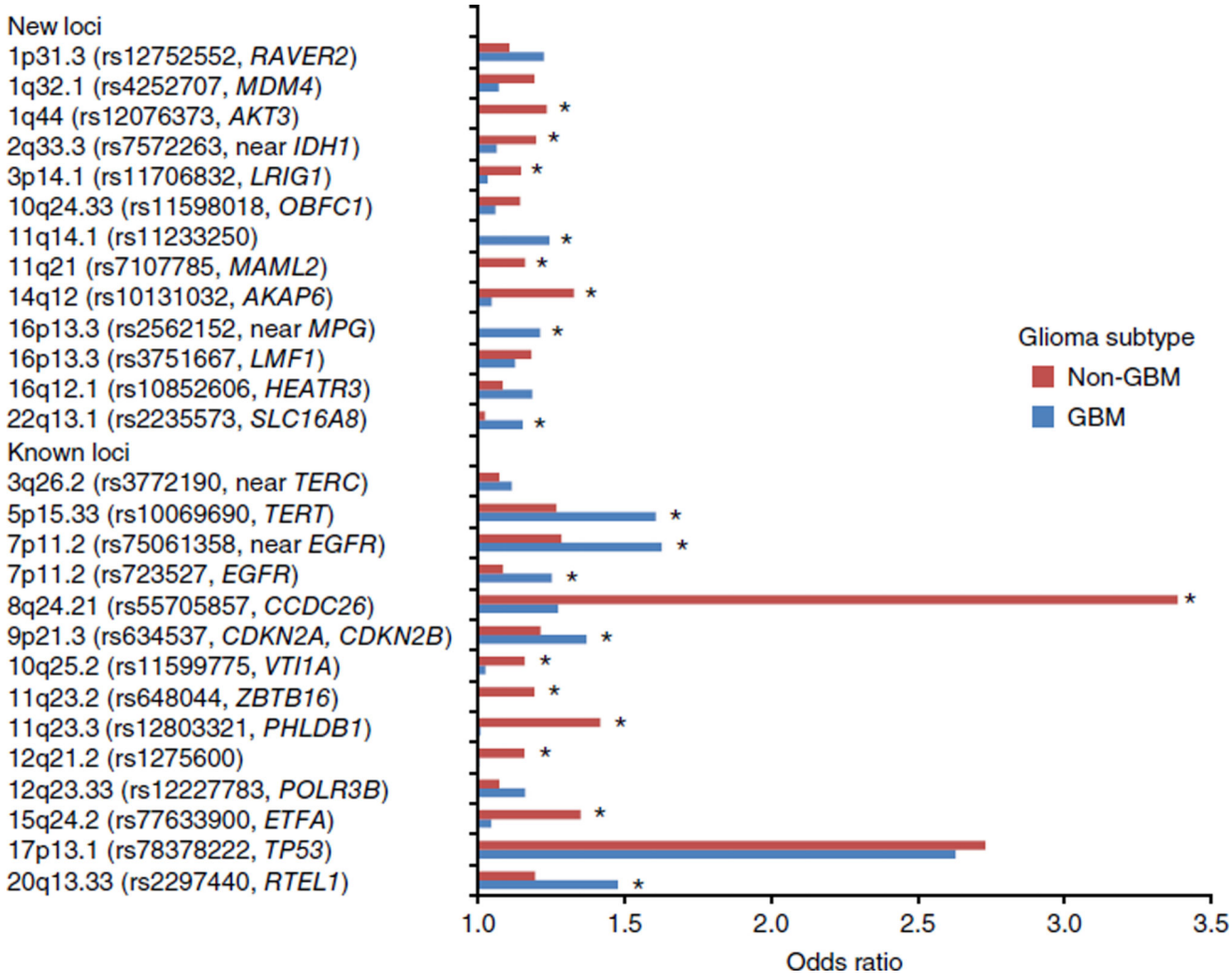

Figure 2.

Relative impact of SNP associations at known and newly identified risk loci for GBM and non-GBM tumors. Odds ratios (ORs) derived with respect to the risk allele. Asterisks denote SNPs showing a significant difference between GBM and non-GBM tumors from the case-only analysis as detailed in Supplementary Table 4.

Table 1

Association statistics for the top SNP at each of the newly-reported glioma risk loci

							All glioma		GBM glioma		Non-GBM glioma	
Locus	**Subtype**	**SNP**	**Position**	**Alleles**	**RAF**	**INFO**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**
1p31.3	GBM	rs12752552	65229299	**T**/C	0.870	0.992	$\mathbf{4.07 \times 10^{-9}}$	1.18 (1.11–1.24)	$\mathbf{2.04 \times 10^{-9}}$	1.22 (1.15–1.31)	4.78×10^{-3}	1.11 (1.03–1.18)
1q32.1	Non-GBM	rs4252707	204508147	G/**A**	0.220	0.992	2.97×10^{-7}	1.12 (1.07–1.17)	0.015	1.07 (1.01–1.13)	$\mathbf{3.34 \times 10^{-9}}$	1.19 (1.12–1.26)
1q44	Non-GBM	rs12076373	243851947	**G**/C	0.837	0.996	4.97×10^{-4}	1.09 (1.04–1.15)	0.846	0.99 (0.94–1.06)	$\mathbf{2.63 \times 10^{-10}}$	1.23 (1.16–1.32)
2q33.3	Non-GBM	rs7572263	209051586	**A**/G	0.756	0.997	2.58×10^{-6}	1.11 (1.06–1.15)	0.019	1.06 (1.01–1.12)	$\mathbf{2.18 \times 10^{-10}}$	1.20 (1.13–1.26)
3p14.1	Non-GBM	rs11706832	66502981	A/**C**	0.456	0.997	1.06×10^{-5}	1.08 (1.05–1.12)	0.158	1.03 (0.99–1.08)	$\mathbf{7.66 \times 10^{-9}}$	1.15 (1.09–1.20)
10q24.33	Non-GBM	rs11598018	105661315	**C**/A	0.462	0.960	3.07×10^{-7}	1.10 (1.06–1.14)	0.0103	1.06 (1.01–1.11)	$\mathbf{3.39 \times 10^{-8}}$	1.14 (1.09–1.20)
11q14.1	GBM	rs11233250	82397014	**C**/T	0.868	0.990	5.40×10^{-6}	1.14 (1.08–1.21)	$\mathbf{9.95 \times 10^{-10}}$	1.24 (1.16–1.33)	0.592	0.98 (0.91–1.05)
11q21	Non-GBM	rs7107785	95747337	**T**/C	0.479	0.997	2.96×10^{-4}	1.07 (1.03–1.11)	0.844	1.00 (0.95–1.04)	$\mathbf{3.87 \times 10^{-10}}$	1.16 (1.11–1.21)
14q12	Non-GBM	rs10131032	33250081	**G**/A	0.916	0.991	2.33×10^{-6}	1.17 (1.09–1.24)	0.247	1.05 (0.97–1.13)	$\mathbf{5.07 \times 10^{-11}}$	1.33 (1.22–1.44)
16p13.3	GBM	rs2562152	123896	A/**T**	0.850	0.937	1.18×10^{-3}	1.09 (1.04–1.15)	$\mathbf{1.93 \times 10^{-8}}$	1.21 (1.13–1.29)	0.948	1.00 (0.93–1.07)
16p13.3	Non-GBM	rs3751667	1004554	C/**T**	0.208	0.985	$\mathbf{8.75 \times 10^{-10}}$	1.14 (1.09–1.19)	5.95×10^{-6}	1.13 (1.07–1.19)	$\mathbf{2.61 \times 10^{-9}}$	1.18 (1.12–1.25)
16q12.1	GBM	rs10852606	50128872	T/**C**	0.713	0.990	$\mathbf{3.66 \times 10^{-11}}$	1.14 (1.10–1.19)	$\mathbf{1.29 \times 10^{-11}}$	1.18 (1.13–1.24)	2.42×10^{-3}	1.08 (1.03–1.14)
22q13.1	GBM	rs2235573	38477930	**G**/A	0.507	0.995	8.64×10^{-7}	1.09 (1.06–1.13)	$\mathbf{1.76 \times 10^{-10}}$	1.15 (1.10–1.20)	0.325	1.02 (0.97–1.07)

Associations at $P < 5 \times 10^{-8}$ are highlighted in bold. Odds ratios (ORs) were derived with respect to the risk allele underlined and highlighted in bold. Risk allele frequency (RAF) is according to European samples from the 1000 Genomes project. The INFO column indicates the average imputation info score across all studies, with a score of 1 indicating that the SNP is directly genotyped in all studies. CI, confidence interval.

Nat Genet. Author manuscript; available in PMC 2017 August 16.