

ARTICLE

Received 8 May 2015 | Accepted 4 Sep 2015 | Published 1 Oct 2015

DOI: 10.1038/ncomms9559

OPEN

Genome-wide association study identifies multiple susceptibility loci for glioma

Ben Kinnersley¹, Marianne Labussière², Amy Holroyd¹, Anna-Luisa Di Stefano^{2,3,4}, Peter Broderick¹, Jayaram Vijayakrishnan¹, Karima Mokhtari^{2,3,5}, Jean-Yves Delattre^{2,3,4}, Konstantinos Gousias⁶, Johannes Schramm⁶, Minouk J. Schoemaker¹, Sarah J. Fleming⁷, Stefan Herms^{8,9}, Stefanie Heilmann⁸, Stefan Schreiber^{10,11}, Heinz-Erich Wichmann^{12,13}, Markus M. Nöthen⁸, Anthony Swerdlow^{1,14}, Mark Lathrop^{5,15,16}, Matthias Simon⁶, Melissa Bondy¹⁷, Marc Sanson^{2,3,4} & Richard S. Houlston¹

Previous genome-wide association studies (GWASs) have shown that common genetic variation contributes to the heritable risk of glioma. To identify new glioma susceptibility loci, we conducted a meta-analysis of four GWAS (totalling 4,147 cases and 7,435 controls), with imputation using 1000 Genomes and UK10K Project data as reference. After genotyping an additional 1,490 cases and 1,723 controls we identify new risk loci for glioblastoma (GBM) at 12q23.33 (rs3851634, near *POLR3B*, $P = 3.02 \times 10^{-9}$) and non-GBM at 10q25.2 (rs11196067, near *VT11A*, $P = 4.32 \times 10^{-8}$), 11q23.2 (rs648044, near *ZBTB16*, $P = 6.26 \times 10^{-11}$), 12q21.2 (rs12230172, $P = 7.53 \times 10^{-11}$) and 15q24.2 (rs1801591, near *ETFA*, $P = 5.71 \times 10^{-9}$). Our findings provide further insights into the genetic basis of the different glioma subtypes.

¹Division of Genetics and Epidemiology, The Institute of Cancer Research, London SM2 5NG, UK. ²Sorbonne Universités UPMC Univ Paris 06, INSERM CNRS, U1127, UMR 7225, ICM, F-75013 Paris, France. ³Onconeurotek, F-75013 Paris, France. ⁴AP-HP, GH Pitié-Salpêtrière, Service de Neurologie 2, F-75013 Paris, France. ⁵AP-HP, GH Pitié-Salpêtrière, Laboratoire de neuropathologie R Escourolle, F-75013 Paris, France. ⁶Department of Neurosurgery, University of Bonn Medical Center, Sigmund-Freud-Straße 25, 53105 Bonn, Germany. ⁷Centre for Epidemiology and Biostatistics, Faculty of Medicine and Health, University of Leeds, Leeds LS2 9JT, UK. ⁸Department of Biomedicine, Institute of Human Genetics, University of Bonn, 53127 Bonn, Germany. ⁹Division of Medical Genetics, Department of Biomedicine, University of Basel, 4056 Basel, Switzerland. ¹⁰1st Medical Department, University Clinic Schleswig-Holstein, Campus Kiel, House 6, Arnold-Heller-Straße 3, Kiel 24105, Germany. ¹¹Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Arnold-Heller-Straße 3, Kiel 24105, Germany. ¹²Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. ¹³Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany. ¹⁴Division of Breast Cancer Research, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. ¹⁵Foundation Jean Dausset-CEPH, 27 Rue Juliette Dodu, 75010 Paris, France. ¹⁶Génome Québec, Department of Human Genetics, McGill University, Montreal, Quebec, Canada H3A 0G1. ¹⁷Division of Hematology-Oncology, Department of Pediatrics, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA. Correspondence and requests for materials should be addressed to R.S.H. (email: richard.houlston@icr.ac.uk).

Gliomas account for ~40% of all primary brain tumours and cause around 13,000 deaths in the United States of America each year¹. Gliomas are heterogeneous and different tumour subtypes, defined in part by malignancy grade (for example, pilocytic astrocytoma World Health Organization (WHO) grade I, diffuse 'low-grade' glioma WHO grade II, anaplastic glioma WHO grade III and glioblastoma (GBM) WHO grade IV) can be distinguished². Gliomas are typically associated with a poor prognosis irrespective of clinical care, with the most common type, GBM, having a median overall survival of only 10–15 months¹.

While the glioma subtypes have distinct molecular profiles resulting from different aetiological pathways³, no environmental exposures have, however, consistently been linked to risk except for ionizing radiation, which only accounts for a very small number of cases¹. Direct evidence for inherited predisposition to glioma is provided by a number of rare inherited cancer syndromes, such as Turcot's and Li–Fraumeni syndromes, and neurofibromatosis⁴. Even collectively, these diseases however account for little of the twofold increased risk of glioma seen in first-degree relatives of glioma patients⁵. Support for polygenic susceptibility to glioma has come from genome-wide association studies (GWASs) that have identified single-nucleotide polymorphisms (SNPs) at eight loci influencing glioma risk—3q26.2 (near *TERC*), 5p15.33 (near *TERT*), 7p11.2 (near *EGFR*), 8q24.21 (near *CCDC26*), 9p21.3 (near *CDKN2A/CDKN2B*), 11q23.3 (near *PHLDB1*), 17p13.1 (*TP53*) and 20q13.33 (near *RTEL1*) (refs 6–10). Perhaps not surprisingly there is variability in genetic effects on glioma by histology with subtype-specific associations at 5p15.33, 20q13.33 and 7p11.2 for GBM and at 11q23.3 and 8q24 for non-GBM glioma^{6,7}.

Recovery of untyped genotypes via imputation has enabled fine mapping and refinement of association signals, for example, in identification of rs55705857 as the basis of the 8q24 association signal in glioma¹¹. Recently, the use of the 1000 Genomes Project and the UK10K projects as a combined reference panel has been shown to improve accuracy compared with using the 1000 Genomes Project data alone, allowing imputation of alleles with frequencies ~0.5% to be viable¹².

Here we report a meta-analysis of four GWASs totalling 4,147 cases and 7,435 controls to identify new glioma susceptibility loci, after imputation using the 1000 Genomes and the UK10K Project data as reference. After genotyping an additional series of 1,490 cases and 1,723 controls we identified new risk loci for GBM at 12q23.33 and non-GBM at 10q25.2, 11q23.2, 12q21.2 and 15q24.2. Our findings provide further insights into the genetic basis of the different glioma subtypes.

Results

Association analysis. To identify additional glioma susceptibility loci we conducted a pooled meta-analysis of four GWASs in populations of European ancestry, the UK-GWAS, the French-GWAS, the German-GWAS and the US-GWAS, that were genotyped using either Illumina HumanHap 317, 317+240S, 370Duo, 550, 610 or 1M arrays (Supplementary Table 1). After filtering, the studies provided genotypes on 4,147 cases and 7,435 controls of European ancestry (Supplementary Table 1, Supplementary Fig. 1). Consistent with our previous analysis⁶, quantile–quantile (Q–Q) plots for the German and the US series showed some evidence of inflation (inflation factor based on the 90% least-significant SNPs, λ_{90} = 1.15 and 1.11, respectively), however after correcting for population substructure using principal-component analyses as implemented in Eigenstrat¹³, λ_{90} for all four studies was ≤ 1.05 (combined λ_{90} = 1.05, Supplementary Fig. 2). To achieve consistent and dense

genome-wide coverage, we imputed unobserved genotypes at > 10 million SNPs using a combined reference panel comprising 1,092 individuals from the 1000 Genomes Project and 3,781 individuals from the UK10K project. Q–Q plots for all SNPs (minor allele frequency (MAF) > 0.5%) post-imputation did not show evidence of substantive over-dispersion introduced by imputation after Eigenstrat adjustment (combined λ_{90} = 1.07, λ_{90} for individual studies = 1.04–1.06; Supplementary Fig. 2).

Pooling data from each GWAS into a joint discovery data set, we derived joint odds ratios (ORs) and 95% confidence intervals (CIs) under a fixed-effects model for each SNP with MAF > 0.005 and associated per allele Eigenstrat-corrected *P* values. Overall and histology-specific ORs were derived for all glioma, GBM and non-GBM. In the pooled data set, associations at the established risk loci for glioma at 5p15.33, 7p11.2, 8q24.21, 9p21.3, 11q23.3, 17p13.1 and 20q13.33 showed a consistent direction of effect with previously reported studies ($P < 5.0 \times 10^{-8}$, Fig. 1 and Supplementary Table 2). In contrast we found no significant support for the association between rs1920116 near *TERC* (3q26.2) and risk of high-grade glioma recently reported by Walsh *et al.*¹⁰ (combined *P* value for GBM = 0.179; Supplementary Table 2 and Supplementary Fig. 3). While the UK-GWAS and the study of Walsh *et al.* share use of UK 1958 Birth Cohort controls, the other three GWAS we analysed are fully independent.

After filtering at $P < 5.0 \times 10^{-6}$ in either all glioma, GBM or non-GBM, we selected 14 SNPs for follow-up, mapping to distinct loci not previously associated with glioma risk (Fig. 1 and Supplementary Table 2). rs141035288, rs117527984, rs138170678 were not taken forward as there was poor concordance between imputed and sequenced genotypes (Supplementary Table 3), and rs145034266 could not be genotyped as it mapped within a highly repetitive region.

The 10 remaining SNPs underwent replication genotyping in an additional set of 1,490 glioma cases and 1,723 controls (UK replication series, Supplementary Table 4). Meta-analysis was then conducted across discovery and replication stages, with genotype data available on 5,637 cases and 9,158 controls. In the combined analysis five SNPs showed an association with tumour risk, which was genome-wide significant (Table 1)—rs3851634 (12q23.3, $P_{\text{GBM}} = 3.02 \times 10^{-9}$), rs11196067 (10q25.2, $P_{\text{Non-GBM}} = 4.32 \times 10^{-8}$), rs648044 (11q23.2, $P_{\text{Non-GBM}} = 6.26 \times 10^{-11}$), rs12230172 (12q21.2, $P_{\text{Non-GBM}} = 7.53 \times 10^{-11}$) and rs1801591 (15q24.2, $P_{\text{Non-GBM}} = 5.71 \times 10^{-9}$). We tested for secondary signals at each locus by adjusting for the sentinel SNP in each region, but found no evidence for independent associations (Supplementary Fig. 4).

The association signal at 12q23.3 defined by rs3851634 was specific for GBM. The rs3851634 maps to intron 12 of the gene encoding polymerase III, RNA, subunit b (*POLR3B*; Fig. 2a) within a ~350-kb block of linkage disequilibrium (LD) at 12q23.3, which also contains the genes *CKAP4* and *TCPIL2*. The other four SNP associations defined by rs11196067, rs648044, rs12230172 and rs1801591 were specific to non-GBM glioma. rs11196067 (10q25.2) is located in intron 7 of *VTGIA* (vesicle transport through interaction with t-SNAREs 1A, Fig. 2b). Similarly rs648044 (11q23.2) is also intronic mapping within ZBTB16 (zinc finger and BTB domain-containing protein 16, alias PLZF; Fig. 2c). The rs12230172 (12q21.2) maps within the lincRNA *RP11-114H23.1* and is centromeric to the gene encoding *PHLDA1* (centromeric pleckstrin homology-like domain, family a, MEMBER 1, Fig. 2d). rs1801591 (15q24.2) is responsible for the p.Thr171Ile substitution in *ETFA* (electron transfer flavoprotein, alpha polypeptide gene, which resides within a 500-kb region of LD to which *ISL2*, *TYRO3P* and *SCAPPER* genes also map Fig. 2e).

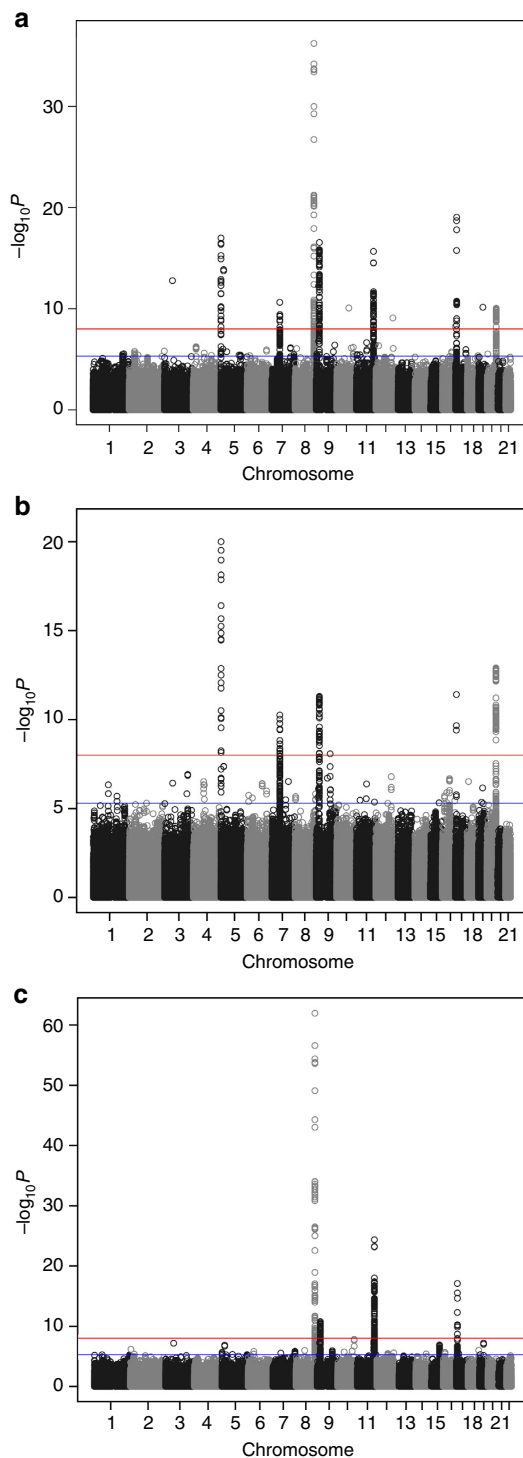


Figure 1 | Genome-wide meta-analysis P values ($-\log_{10}P$, y axis) plotted against their chromosomal positions (x axis). (a) All glioma, (b) GBM (c) non-GBM. The red and blue horizontal lines represent significance thresholds of $P = 5.0 \times 10^{-8}$ and $P = 5.0 \times 10^{-6}$, respectively.

Relationship between the new glioma SNPs and tumour profile.

To investigate the impact of the new risk SNPs on glioma subtype we examined rs11196067, rs648044, rs12230172, rs1801591 and rs3851634 genotypes in the French case series for which comprehensive histology and molecular phenotyping had been performed (Supplementary Data 1). The GBM SNP rs3851634 was associated with 10q-deleted glioma ($P = 0.016$).

In the case of non-GBM SNPs rs11196067 showed the strongest association with grade II glioma ($P = 3.2 \times 10^{-5}$) and TP53 non-mutated glioma ($P = 5.82 \times 10^{-5}$); rs648044 with grade II oligodendroglioma ($P = 0.026$) and 10q non-deleted glioma ($P = 0.006$); rs1801591 with grade II astrocytoma ($P = 0.001$) and IDH1/IDH2 mutated glioma ($P = 0.005$) and rs12230172 with grade II oligodendroglioma ($P = 0.009$), IDH1/IDH2 mutated ($P = 0.009$) and 10q non-deleted glioma ($P = 0.003$).

Functional annotation of risk variants. For each of the sentinel risk SNPs at the five risk loci (as well as correlated variants, $r^2 > 0.8$) we examined published data^{14,15} and made use of the online resources HaploReg v3, RegulomeDB and SeattleSeq for evidence of functionality and regulatory motifs at genomic regions (Supplementary Table 5). rs1801591, which is responsible for the *ETFA* p.Thr171Ile substitution, resides within a highly conserved region of the genome (genomic evolutionary rate profiling (GERP) = 5.65) and the amino-acid change is predicted to be damaging (PolyPhen = 1). Although rs648044 exhibits low evolutionary conservation (GERP = -9.32) it maps within a strong DNase hypersensitivity site and predicted enhancer/super-enhancer element for multiple tissues including the brain. The region surrounding rs648044 is also predicted to interact with the *ZBTB16* promoter, which combined with alteration of a Pax-5 motif is suggestive of direct functional impact. rs12230172 localizes within a moderately conserved region (GERP = 3.41) and occupies promoter histone marks in the brain as well as enhancers predicted to associate with transcriptional start sites for *PHLDA1* and *GLIPR1*. rs11196067 in *VTG1A*, while having a low conservation score (GERP = 0.719), occupies enhancer histone marks in embryonic stem cells although not in brain cells. Similarly, rs3851634 maps to a moderately conserved region (GERP = 2.37) and occupies enhancer histone marks in 18 organs including the brain.

eQTL analysis of the five new glioma SNPs. To gain further insight into the functional basis of rs11196067, rs648044, rs12230172, rs1801591 and rs3851634 associations we performed an expression quantitative trait loci (eQTL) analysis using RNA-Seq expression data on 389 low-grade gliomas (LGGs) and 138 GBMs from The Cancer Genome Atlas (TCGA), together with lymphoblastoid cell line RNA-Seq data on 363 samples from GEUVADIS¹⁶. We examined for an association between SNP genotype and expression of genes mapping within 1 Mb of the sentinel SNP (Supplementary Data 2). After adjusting for multiple testing within each region no statistically significant eQTL was seen for rs11196067, rs12230172, rs1801591 or rs3851634. The strongest association between rs648044 genotype and gene expression was with *ZW10* in LGG ($P = 5.7 \times 10^{-5}$), with the risk allele (T) associated with lower expression, remaining significant after adjustment for multiple testing. To explore the possibility that rs648044 is correlated with a SNP exhibiting a stronger association with *ZW10*, we examined associations with *ZW10* expression in LGG tumours in all SNPs in LD ($r^2 > 0.4$) with rs648044. All of the proxy SNPs examined were more weakly associated with *ZW10* than rs648044 (Supplementary Table 6). Following on from these analyses we made use of publicly available eQTL mRNA expression array data on adipose tissue, lymphoblastoid cell lines and skin from 856 twins (MuTHER¹⁷) and 5,311 non-transformed peripheral blood samples using the blood eQTL browser¹⁸. The risk allele (C) of rs3851634 was associated with significantly lower levels of *POLR3B* ($P = 7.49 \times 10^{-6}$) in peripheral blood analysis with a nominally significant association in skin ($P = 0.0052$). The risk allele (T) of rs1801591, was associated with significantly lower

Table 1 | Association between SNP and glioma risk in discovery and replication data sets for rs1196067, rs648044, rs12230172, rs3851634 and rs1801591.

SNP	MAF			All glioma		GBM		Non-GBM	
	Case	Control	Study	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)
rs111696067 (<i>VTT11A</i>) 10q25.2 A/T	0.38	0.41	FRE	5.00×10^{-5}	0.79 (0.71-0.89)	0.26	0.91 (0.78-1.07)	2.54×10^{-6}	0.74 (0.66-0.84)
			GER	0.44	0.95 (0.83-1.08)	0.88	1.01 (0.86-1.19)	0.15	0.88 (0.75-1.04)
			UK	0.012	0.85 (0.75-0.96)	0.34	0.91 (0.76-1.10)	8.11×10^{-3}	0.81 (0.69-0.95)
			USA	0.016	0.88 (0.80-0.98)	0.097	0.90 (0.79-1.02)	0.033	0.87 (0.76-0.99)
			Replication	0.56	0.97 (0.88-1.07)	0.91	0.99 (0.89-1.11)	0.33	0.93 (0.81-1.07)
			Combined	4.32×10^{-6}	0.89 (0.85-0.93)	0.11	0.95 (0.89-1.01)	4.32×10^{-8}	0.84 (0.79-0.89)
rs648044 (<i>ZBTB16</i>) 11q23.2 C/T	0.40	0.38	FRE	0.019	1.15 (1.02-1.30)	0.93	0.99 (0.83-1.18)	1.76×10^{-3}	1.23 (1.08-1.41)
			GER	0.043	1.16 (1.00-1.34)	0.39	1.08 (0.90-1.30)	0.016	1.25 (1.04-1.50)
			UK	0.78	1.02 (0.89-1.16)	0.044	0.82 (0.67-0.99)	0.037	1.20 (1.01-1.42)
			USA	0.088	1.10 (0.99-1.23)	0.62	0.97 (0.86-1.09)	1.02×10^{-3}	1.27 (1.10-1.46)
			Replication	0.97	1.08 (0.97-1.19)	0.59	0.97 (0.86-1.09)	4.16×10^{-4}	1.29 (1.12-1.48)
			Combined	5.29×10^{-4}	1.10 (1.04-1.16)	0.32	0.97 (0.90-1.03)	6.26×10^{-11}	1.25 (1.17-1.34)
rs12230172 (intergenic) 12q21.2 G/A	0.45	0.46	FRE	0.054	0.90 (0.81-1.00)	0.72	1.03 (0.88-1.20)	4.40×10^{-3}	0.84 (0.74-0.95)
			GER	0.043	0.88 (0.77-1.00)	0.84	0.98 (0.84-1.16)	2.17×10^{-3}	0.78 (0.66-0.91)
			UK	0.44	0.95 (0.84-1.09)	0.77	0.97 (0.85-1.11)	0.42	0.94 (0.80-1.10)
			USA	0.30	0.95 (0.86-1.05)	0.55	1.04 (0.92-1.18)	0.018	0.85 (0.75-0.97)
			Replication	1.84×10^{-6}	0.79 (0.70-0.86)	7.00×10^{-3}	0.85 (0.76-0.96)	3.59×10^{-8}	0.67 (0.58-0.77)
			Combined	1.57×10^{-6}	0.88 (0.84-0.93)	0.22	0.96 (0.91-1.02)	7.53×10^{-11}	0.81 (0.76-0.86)
rs3851634 (<i>POLR3B</i>) 12q23.3 T/C	0.27	0.30	FRE	0.053	0.89 (0.79-1.00)	0.020	0.81 (0.69-0.97)	0.25	0.93 (0.81-1.06)
			GER	0.18	0.91 (0.73-1.04)	0.12	0.87 (0.73-1.04)	0.59	0.95 (0.85-1.14)
			UK	0.058	0.88 (0.60-0.89)	1.56×10^{-3}	0.73 (0.60-0.89)	0.92	1.01 (0.80-1.20)
			USA	2.84×10^{-4}	0.81 (0.73-0.91)	7.21×10^{-4}	0.79 (0.68-0.90)	0.021	0.84 (0.73-0.98)
			Replication	0.022	0.88 (0.79-0.98)	5.00×10^{-3}	0.83 (0.74-0.95)	0.57	0.96 (0.83-1.11)
			Combined	4.07×10^{-7}	0.87 (0.82-0.92)	3.02×10^{-9}	0.81 (0.76-0.87)	0.037	0.93 (0.87-1.00)
rs1801591 (<i>ETFA</i>) 15q24.2 G/A	0.10	0.09	FRE	6.67×10^{-3}	1.32 (1.08-1.61)	0.29	1.17 (0.87-1.58)	2.51×10^{-3}	1.40 (1.13-1.74)
			GER	0.037	1.25 (1.01-1.53)	0.17	1.20 (0.93-1.56)	0.052	1.31 (1.00-1.72)
			UK	0.44	1.08 (0.88-1.33)	0.93	0.99 (0.73-1.33)	0.23	1.17 (0.90-1.53)
			USA	0.016	1.23 (1.04-1.46)	0.97	1.00 (0.80-1.24)	5.13×10^{-5}	1.56 (1.26-1.94)
			Replication	0.16	1.13 (0.95-1.33)	0.89	1.01 (0.83-1.23)	0.013	1.31 (1.06-1.63)
			Combined	2.75×10^{-5}	1.20 (1.10-1.30)	0.32	1.06 (0.95-1.18)	5.71×10^{-9}	1.36 (1.23-1.51)

ORs derived with respect to the minor allele, highlighted in bold. The SNPs rs3851634 and rs1801591 were directly genotyped while rs1196067, rs648044 and rs12230172 were imputed with imputation information scores (Is) of 0.99, 0.87 and 1.00, respectively. Sample sizes in the individual data sets are as follows: FRE (French-GWAS), 1,423 cases and 1,190 controls; GER (German-GWAS), 846 cases and 1,310 controls; UK (UK-GWAS), 631 cases and 2,699 controls; USA (USA-GWAS), 1,247 cases and 2,236 controls; replication, 1,490 cases and 1,723 controls; combined, 5,637 cases and 9,158 controls. MAF, minor allele frequency in discovery series.

ETFA levels in peripheral blood ($P = 7.90 \times 10^{-12}$); there was a nominally significant association in MuTHER lymphoblastoid cell lines ($P = 0.037$).

Somatic mutation of newly implicated risk genes in glioma. We examined mutation data from TCGA for evidence of recurrent mutation in genes annotated by the new GWAS signals. Collectively *POLR3B*, *ETFA*, *VTT11A*, *ZBTB16* and *PHLDA1* are altered in 8% (22/286) of LGG as compared with 3% (8/273) of GBM ($P = 0.014$, Supplementary Table 7) providing support for these genes having a role in glioma tumorigenesis.

Individual variance in risk associated with glioma SNPs. To explore the relative contributions of previously reported and newly described loci to glioma risk, we applied the method of Pharoah *et al.*¹⁹ to eight previously reported SNPs as well as the five new risk SNPs (Supplementary Table 8). The variance in risk attributable to all 12 SNPs is 26%, 27% and 43% for all glioma, GBM and non-GBM, respectively.

Pathway enrichment of glioma GWAS SNPs. To gain further insights into the biological basis of associations we performed a pathway analysis on GWAS associations in all glioma, GBM and non-GBM. Applying a false discovery rate (FDR) threshold of <0.1 revealed enrichment for 14 pathways in all glioma, 8 in GBM and 9 in non-GBM tumours (Supplementary Table 9).

Pathways implicated in GBM tumours primarily include DNA repair and Notch-signalling, whereas for non-GBM tumours pathways were primarily associated with cell-cycle progression and energy metabolism (Supplementary Table 9).

Discussion

To our knowledge we have performed the largest GWAS of glioma to date, identifying five novel glioma susceptibility loci at 12q23.33, 10q25.2, 11q23.2, 12q21.2 and 15q24.2 and taking the total count of risk loci to 12. Through making use of a combined reference panel from the UK10K and the 1000 Genomes Projects we were able to recover genotypes from ~8 million SNPs for association analysis, a significant increase from using array SNPs alone. In addition, we have provided further evidence that genetic susceptibility to glioma can be subtype specific, emphasising the importance of searching for histology-specific risk variants.

While deciphering the functional impact of these SNP associations on glioma development requires additional analyses, a number of the genes implicated have relevance to the biology of this cancer *a priori*. As well as participating in regulating insulin-stimulated trafficking of secretory vesicles²⁰, *VTT11A* plays a key role in neuronal development and in selectively maintaining spontaneous neurotransmitter release²¹. Intriguingly recent GWAS have identified associations between the *VTT11A* SNPs rs7086803 and lung cancer²² and between rs12241008 and colorectal cancer²³; rs7086803 and rs12241008 are not

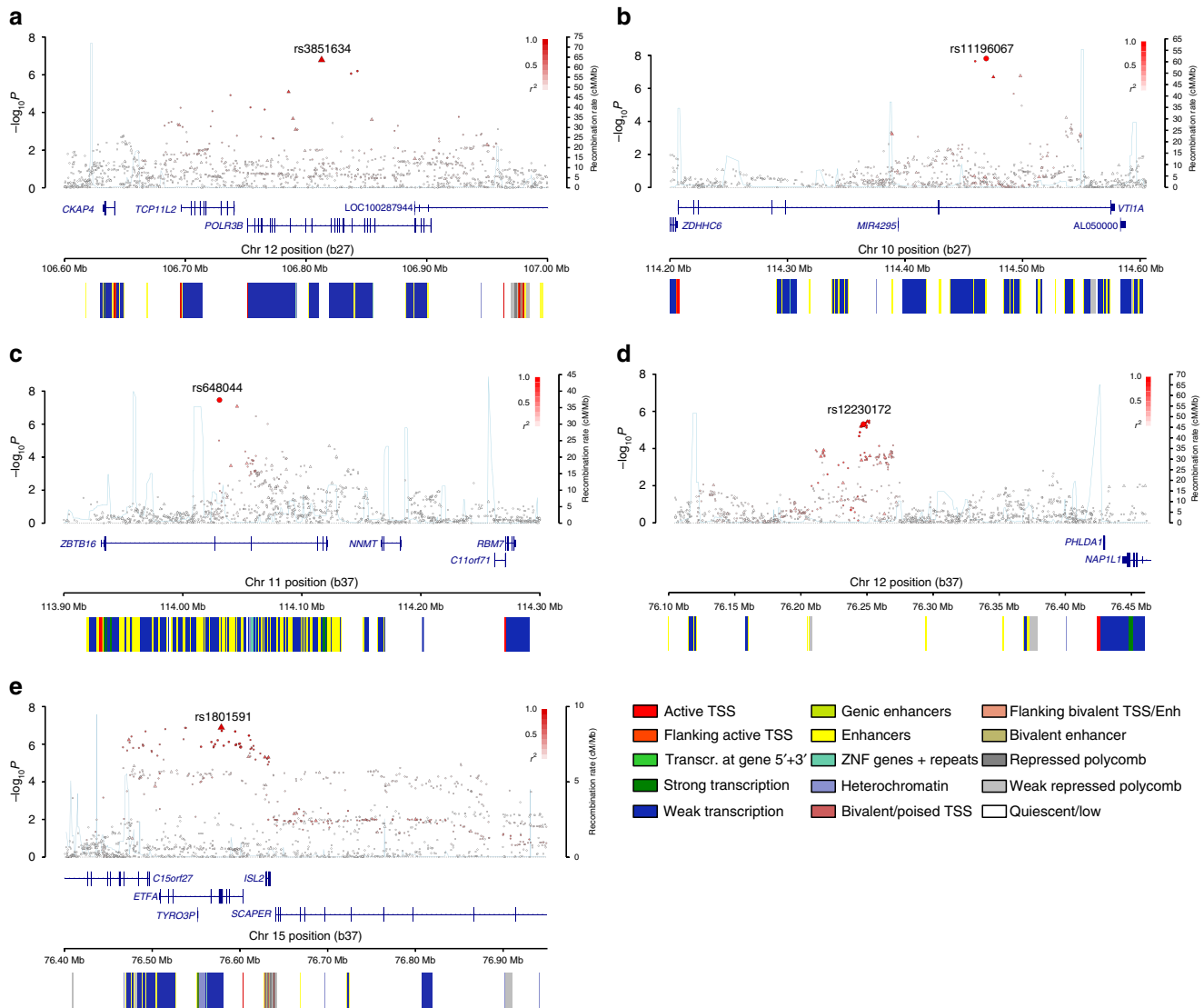


Figure 2 | Regional plots of discovery-phase association results, recombination rates and chromatin state segmentation tracks for five glioma-risk loci. Results for: (a) 12q23.33, rs3851634 (GBM); (b) 10q25.2, rs11196067 (non-GBM); (c) 11q23.2, rs648044 (non-GBM); (d) 12q21.2 rs12230172 (non-GBM); and (e) 15q24.2, rs1801591 (non-GBM). Plots show discovery association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates. The $-\log_{10} P$ values (y axes) of the SNPs are shown according to their chromosomal positions (x axes). The lead SNP in each combined analysis is shown as a large circle or triangle (if imputed or directly genotyped, respectively) and is labelled by its rsID. The colour intensity of each symbol reflects the extent of LD with the top genotyped SNP, white ($r^2 = 0$) to dark red ($r^2 = 1.0$). Genetic recombination rates, estimated using HapMap samples from Utah residents of western and northern European ancestry (CEU), are shown with a light blue line. Physical positions are based on NCBI build 37 of the human genome. Also shown are the relative positions of University of Carolina, Santa Cruz (UCSC) genes and transcripts mapping to the region of association. Genes have been redrawn to show their relative positions; therefore, maps are not to physical scale. Below each plot is a diagram of the exons and introns of the genes of interest, the associated SNPs and the chromatin state segmentation track (ChromHMM) for H1 neural progenitor cells derived from the epigenome roadmap project, as per legend. TSS, transcriptional start sites.

correlated with each other ($r^2 = 0.22$, $D' = 0.72$) and are also not correlated with rs11196067 ($r^2 = 0.03/0.00$, $D' = 1.00/0.22$, respectively), suggesting the existence of multiple risk loci within the region with different tumour specificities.

ZBTB16 is highly expressed in undifferentiated, multipotential progenitor cells and its expression has been shown to influence resistance to retinoid-mediated re-differentiation in *t(11;17)(q23;21)* acute promyelocytic leukaemia²⁴. The BTB domain of *ZBTB16* has transcriptional repression activity and interacts with components of the histone deacetylase complex thereby linking the transcription factor with regulation of chromatin conformation²⁵. Although rs648044 lies within an enhancer active in brain and is predicted to interact with the

ZBTB16 promoter, providing an attractive functional basis for the 11q23.2 association through differential *ZBTB16* expression, we found a strong association between rs648044 and *ZW10* expression in LGG ($P = 5.7 \times 10^{-5}$). Since *ZW10* plays a role in chromosome segregation²⁶ it also represents a plausible candidate for the 11q23.2 association.

We also observed a strong association between *ETFA* expression and rs1801591 in peripheral blood ($P = 7.90 \times 10^{-12}$). *ETFA* participates in mitochondrial fatty acid beta oxidation; shuttling electrons between flavoprotein dehydrogenases and the membrane-bound electron transfer flavoprotein ubiquinone oxidoreductase²⁷. Mutations of *ETFA* have been reported to be a cause of recessive glutaric acidemia IIA (refs 28,29), which

features gliosis. While the p.Thr171Ile change is reported to decrease thermal stability of *ETFA*³⁰ thereby providing evidence for a direct functional effect the strong eQTL data is consistent with the functional basis for the 15q24.2 association being mediated through differential expression.

RNA polymerase III (*POLR3B*) is involved in the transcription of small noncoding RNAs and short interspersed nuclear elements, as well as all transfer RNAs³¹. Although mutations in *POLR3B* have been shown to cause recessive hypomyelinating leukoencephalopathy³² thus far there is no evidence implicating the gene in the development of glioma. Albeit in peripheral blood there was a strong association between *POLR3B* expression and rs3851634 ($P=7.49 \times 10^{-6}$), providing a possible functional basis of the 12q23.2 association.

At 12q21.2 rs12230172 maps within *RP11-114H23.1*, a lincRNA of currently unknown function. Although only lying adjacent to *PHLDA1*, the known 11q23.3 association maps to the related gene *PHLDB1*, which is also specific to non-GBM tumours⁷. Although a role for *PHLDA1* in glioma has yet to be established downregulation of *PHLDA1* in neuronal cells has been shown to enhance cell death without Fas induction³³, additionally *PHLDA1* expression may be involved in regulation of anti-apoptotic effects of *IGF1* (ref. 34).

Intriguingly across all of the four GWAS data sets we analysed we did not replicate the association between rs1920116 (near *TERC*) at 3q26.2 and risk of high-grade glioma recently reported by Walsh *et al.*¹⁰ ($P=8.3 \times 10^{-9}$, OR = 1.30 versus $P=0.18$, OR = 1.06 relative to the G-allele in our GBM data set), despite our study having a similar power to demonstrate a relationship (1,783 GBM cases, 7,435 controls in our study as compared with 1,644 cases, 7,736 controls). It is, however noteworthy that the Walsh *et al.* analysed both anaplastic astrocytoma and GBM. While we could not demonstrate a significant association with either subtype we did see an association between rs1920116 and *TP53*-mutated glioma ($P=0.016$, Supplementary Data 1) suggesting that the association might be restricted to a specific molecularly defined subtype of glioma.

Our findings provide further evidence for an inherited genetic susceptibility to glioma. Future investigation of the genes targeted by the risk SNPs we have identified is likely to yield increased insight into the development of this malignancy. We estimate that the risk loci so far identified for glioma account for 27 and 43% of the familial risk of GBM and non-GBM tumours, respectively, of which 0.8% and 7.6% can be explained by the loci newly reported in this study (Supplementary Table 8). Although the power of our study to detect the major common loci (MAF > 0.2) conferring risk ≥ 1.2 was high ($\sim 80\%$), we had low power to detect alleles with smaller effects and/or MAF < 0.1. By implication, variants with such profiles probably represent a much larger class of susceptibility loci for glioma because of the truly small effect sizes or submaximal LD with tagging SNPs. Thus, it is probable that a large number of variants remain to be discovered. In addition, as we have recently shown, stratified analysis of glioma by molecular profile may lead to the discovery of additional subtype-specific risk variants. However, such subtype analyses can increase the statistical burden of adjusting for multiple testing. For example, if applying an additional Bonferroni correction for GBM and non-GBM subtypes, the rs11196067 (*VTTIA*) association at $P=8.64 \times 10^{-8}$ would not be declared genome-wide significant. An issue in future subtype analyses of glioma will therefore be to have sufficient study power to mitigate type II error given the additional constraints of multiple testing. Further efforts to expand the scale of GWAS meta-analyses through international consortia and increasing the number of SNPs taken forward to large-scale replication will be required to address this challenge.

Methods

Ethics. Collection of blood samples and clinico-pathological information from patients and controls was undertaken with informed consent and relevant ethical review board approval in accordance with the tenets of the Declaration of Helsinki. Ethical committee approval for this study was obtained from relevant study centres (UK: South East Multicentre Research Ethics Committee (MREC) and the Scottish Multicentre Research Ethics Committee; France: APHP Ethical Committee-CPP (comité de Protection des Personnes); Germany: Ethics Commission of the Medical Faculty of the University of Bonn; and USA: University of Texas MD Anderson Cancer Institutional Review Board).

Genome-wide association studies. We used GWAS data previously generated on four non-overlapping case-control series of Northern European ancestry, which have been the subject of previous studies^{6,7}; summarized in Supplementary Table 1. Briefly, the UK-GWAS was based on 636 cases (401 males; mean age 46 years) ascertained through the INTERPHONE study³⁵. Individuals from the 1958 Birth Cohort ($n=2,930$) served as a source of controls. The US-GWAS was based on 1,281 cases (786 males; mean age 47 years) 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^{36,37}. The French-GWAS study comprised 1,495 patients with glioma ascertained through the Service de Neurologie Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris. The controls ($n=1,213$) were ascertained from the SU.VI.MAX (Supplémentation en Vitamines et Minéraux Antioxydants) study of 12,735 healthy subjects (women aged 35–60 years; men aged 45–60 years)³⁸. The German-GWAS comprised 880 patients who underwent 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; $n=488$) (ref. 39); POPGEN (Population Genetic Cohort; $n=678$) (ref. 40) and from the Heinz Nixdorf Recall study ($n=380$) (ref. 41).

Replication genotyping. For replication we made use of DNA from 1,490 glioma cases recruited to an ongoing UK study of primary brain tumours (National Brain Tumour Study). Controls were healthy individuals that had been recruited to the National Study of Colorectal Cancer Genetics⁴² and the Genetic Lung Cancer Predisposition Study⁴³. All cases and controls were UK residents and had self-reported European ancestry. Controls reported no personal history of cancer at the time of ascertainment. Genotyping of rs76178334, rs4432939, rs182521816, rs12780046, rs11196067, rs648044, rs12230172, rs3851634, rs1801591 and rs78543262 was performed using competitive allele-specific PCR KASPar chemistry (LGC, Hertfordshire, UK, primer sequences detailed in Supplementary Table 10). Conditions used are available on request. Call rates for SNP genotypes were > 95%. To ensure quality of genotyping in all assays, at least two negative controls and 1–10% duplicates (showing a concordance > 99%) were genotyped. For SNPs with MAF < 5%, at least two known heterozygotes were included per genotyping plate, to aid clustering.

Statistical and bioinformatic analysis. Data were imputed for all scans for over 10 million SNPs using IMPUTE2 v2.3.0 (ref. 44) software and the 1000 Genomes Project (Phase 1 integrated release 3, March 2012 (ref. 45)) and the UK10K data (ALSPAC, EGAS00001000090/EGAD00001000195, and TwinsUK, EGAS00001000108/EGAD00001000194, studies only) as reference panels (Supplementary Table 1). Genotypes were aligned to the positive strand in both imputation and genotyping. Imputation was conducted separately for each scan in which before imputation each GWAS data set was pruned to a common set of 425,190 SNPs. Poorly imputed SNPs defined by an information score (Is) < 0.70 and Hardy-Weinberg equilibrium $P < 1.0 \times 10^{-5}$ were excluded from the analyses. Tests of association between imputed SNPs and glioma was performed under a probabilistic dosage model in SNPTEST v2.5 (ref. 46).

Eigenvectors for the GWAS data sets were inferred using *smartpca* (part of EIGENSOFT v2.4 (refs 13,47)) using $\sim 100,000$ ld-pruned SNPs. Eigenstrat adjustment was carried out in SNPTEST by including the first 10 eigenvectors as covariates. The adequacy of the case-control matching and possibility of differential genotyping of cases and controls was evaluated using Q-Q plots of test statistics. The inflation factor λ was based on the 90% least-significant SNPs as previously advocated⁴⁸. Testing for secondary signals was carried out in SNPTEST, adjusting for the sentinel SNP using the '-condition_on' option. Visualization of population ancestry was carried out in *smartpca* by projecting query samples onto eigenvectors inferred from the 1000 Genomes Project populations (Supplementary Fig. 1). Meta-analysis of GWAS data sets under a fixed-effects model was undertaken in META v1.6 (ref. 49) using the inverse-variance approach. Cochran's Q-statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity were calculated⁵⁰. P_{het} values < 0.05 are considered characteristic of large heterogeneity⁵⁰. In addition, analyses stratified by glioma tumour histology and molecular characteristics were performed. All statistical P values were two sided.

Estimates of individual variance in risk associated with glioma-risk SNPs was carried out using the method described in Pharoah *et al.*¹⁹ assuming the familial

risk of glioma to be 1.77 (ref. 51). Briefly, for a single allele (i) of frequency p , relative risk R and in risk r , the variance (V_i) of the risk distribution due to that allele is given by:

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

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 towards the normal with variance:

$$V = \sum V_i$$

The total genetic variance (V) for all susceptibility alleles has been estimated to be $\sqrt{1.77}$. 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 (ref. 52) using UK10K data and plotted using visPig (ref. 53). 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.*⁵⁴

SNPs were annotated for putative functional effect using RegulomeDB⁵⁵, HaploReg v3 (ref. 56) and SeattleSeq Annotation 138 (ref. 57). These servers make use of data from ENCODE⁵⁸, GERP⁵⁹ conservation metrics, combined annotation-dependent depletion (CADD)⁶⁰ scores and PolyPhen 2 (ref. 61) scores. We searched for overlap of associated SNPs with enhancers defined by the FANTOM5 enhancer atlas¹⁵, annotating by ubiquitous enhancers as well as enhancers specifically expressed in astrocytes, neurons, neuronal stem cells and brain tissue. Similarly, we searched for overlap with 'super-enhancer' regions as defined by Hnisz *et al.*¹⁴, restricting analysis to U87 GBM cells, astrocyte cells and brain tissue. We additionally made use of 15-state chromHMM data from H1-derived neuronal progenitor cells available from the Epigenome roadmap project⁶². Mutation data in LGG and GBM tumours from TCGA was assessed using the cBioPortal for cancer genomics⁶³.

To search for biological pathways enriched for glioma SNP associations we made use of Improved Gene Set Enrichment Analysis for Genome-wide Association Study (i-GSEA4GWAS v1.1) (ref. 64). SNPs up to 5 kb upstream and downstream of a given gene were mapped to that gene, with the maximum P value of all SNPs mapping to a gene used to represent the gene. Gene sets used were: canonical pathways, gene ontology (GO) biological process, GO molecular function, GO cellular component. As recommended we applied an FDR cutoff of <0.10 on all reported gene sets. In the case of multiple identical pathways, that with the lower FDR value is retained.

Imputation concordance assessment. The fidelity of imputation as assessed by the concordance between imputed and directly genotyped SNPs was examined in 192 cases and 187 controls from the UK-GWAS discovery series (Supplementary Table 3). Targeted sequencing for the SNPs rs141035288, rs117527984, rs76178334, rs4432939, rs182521816, rs138170678, rs145034266, rs12780046, rs11196067, rs648044, rs12230172 and rs78543262 was performed by Sanger on an ABI3700 analyser (Applied Biosystems; Supplementary Table 10, conditions are available on request). For SNPs with MAF <0.05 , samples were included to ensure at least 10 predicted heterozygotes were sequenced. Imputed genotypes were considered for concordance assessment if exhibiting probability >0.9 .

Tumour genotyping. Tumour samples were available from a subset of the patients ascertained through the Service de Neurologie Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris. Tumours were snap frozen in liquid nitrogen and DNA was extracted using the QIAmp DNA minikit, according to the manufacturer's instructions (Qiagen, Venlo, LN, USA). DNA was analysed for large-scale copy number variation by CGH array as previously described^{65,66}. In the cases not analysed by CGH array, 9p, 10q, 1p and 19q status was assigned using PCR microsatellites, and *EGFR* amplification and *CDKN2A-p16-INK4a* homozygous deletion by quantitative PCR. *IDH1*, *IDH2* and *TERT* promoter mutation status was determined by sequencing as previously described^{67,68}.

Expression quantitative trait loci analysis. To examine the relationship between SNP genotype and gene expression, we made use of tumour RNA sequence data and blood Affymetrix 6.0 SNP Array data for 389 low-grade and 138 GBM tumours of European ancestry from TCGA (accession number phs000178.v9.p8), as well as RNA sequence data from lymphoblastoid cells (GEUVADIS project¹⁶) and genotype data for 363 European individuals from the 1000 Genomes Project⁴⁵. Sequence reads from downloaded FASTQ files were aligned to the human hg19 reference genome and GRCh37 Ensembl transcriptome using TopHat v2.0.7 and Bowtie v2.0.6. Read counts per gene were generated for 62,069 Ensembl genes using featureCounts⁶⁹ as part of the Rsubread Bioconductor package⁷⁰. For TCGA samples, European ancestry was assessed through visualization of clustering with CEU samples after principal components analysis (data not shown). Untyped genotypes were imputed from the Affymetrix 6 array using similar methods to those discussed previously. Genotypes with probability >0.9 were taken forward

for eQTL analysis. The association between SNP and gene expression was quantified using the Kruskal–Wallis trend test.

We additionally queried publicly available eQTL mRNA expression data using MuTHER, and the Blood eQTL browser. MuTHER contains expression adipose tissue, lymphoblastoid cells and skin expression data from 856 healthy twins¹⁷. rs500629 was used as a proxy for rs648044 ($r^2 = 0.52$, $D' = 0.85$). The blood eQTL browser contains expression data from 5,311 non-transformed peripheral blood samples¹⁸. Putative eQTLs were thresholded at FDR <0.1 .

References

- Bondy, M. L. *et al.* Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium. *Cancer* **113**, 1953–1968 (2008).
- Louis, D. N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**, 97–109 (2007).
- Ostrom, Q. T. *et al.* The epidemiology of glioma in adults: a "state of the science" review. *Neuro Oncol.* **16**, 896–913 (2014).
- Hodgson, S. V., Maher, E. R. & Hodgson, S. *A Practical Guide to Human Cancer Genetics* (Springer, 1999).
- Hemminki, K., Tretli, S., Sundquist, J., Johannesen, T. B. & Granstrom, C. Familial risks in nervous-system tumours: a histology-specific analysis from Sweden and Norway. *Lancet Oncol.* **10**, 481–488 (2009).
- Sanson, M. *et al.* Chromosome 7p11.2 (EGFR) variation influences glioma risk. *Hum. Mol. Genet.* **20**, 2897–2904 (2011).
- Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.* **41**, 899–904 (2009).
- Wrensch, M. *et al.* Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat. Genet.* **41**, 905–908 (2009).
- Enciso-Mora, V. *et al.* Low penetrance susceptibility to glioma is caused by the TP53 variant rs78378222. *Br. J. Cancer* **108**, 2178–2185 (2013).
- Walsh, K. M. *et al.* Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk. *Nat. Genet.* **46**, 731–735 (2014).
- Enciso-Mora, V. *et al.* Deciphering the 8q24.21 association for glioma. *Hum. Mol. Genet.* **22**, 2293–2302 (2013).
- Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 8111 (2015).
- Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
- Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
- Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
- Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
- Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
- Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
- Pharoah, P. D., Antoniou, A. C., Easton, D. F. & Ponder, B. A. Polygenes, risk prediction, and targeted prevention of breast cancer. *N. Engl. J. Med.* **358**, 2796–2803 (2008).
- Bose, A. *et al.* The v-SNARE Vti1a regulates insulin-stimulated glucose transport and Acrp30 secretion in 3T3-L1 adipocytes. *J. Biol. Chem.* **280**, 36946–36951 (2005).
- Ramirez, D. M., Khvotchev, M., Trauterman, B. & Kavalali, E. T. Vti1a identifies a vesicle pool that preferentially recycles at rest and maintains spontaneous neurotransmission. *Neuron* **73**, 121–134 (2012).
- Lan, Q. *et al.* Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat. Genet.* **44**, 1330–1335 (2012).
- Wang, H. *et al.* Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. *Nat. Commun.* **5**, 4613 (2014).
- Chen, S. J. *et al.* Rearrangements of the retinoic acid receptor alpha and promyelocytic leukemia zinc finger genes resulting from t(11;17)(q23;q21) in a patient with acute promyelocytic leukemia. *J. Clin. Invest.* **91**, 2260–2267 (1993).
- Ahmad, K. F., Engel, C. K. & Prive, G. G. Crystal structure of the BTB domain from PLZF. *Proc. Natl Acad. Sci. USA* **95**, 12123–12128 (1998).
- Vallee, R. B., Varma, D. & Dujardin, D. L. ZW10 function in mitotic checkpoint control, dynein targeting, and membrane trafficking: is dynein the unifying theme? *Cell Cycle* **5**, 2447–2451 (2014).
- Frerman, F. E. Acyl-CoA dehydrogenases, electron transfer flavoprotein and electron transfer flavoprotein dehydrogenase. *Biochem. Soc. Trans.* **16**, 416–418 (1988).
- Indo, Y., Glassberg, R., Yokota, I. & Tanaka, K. Molecular characterization of variant alpha-subunit of electron transfer flavoprotein in three patients with glutaric acidemia type II—and identification of glycine substitution for valine-157 in the sequence of the precursor, producing an unstable mature protein in a patient. *Am. J. Hum. Genet.* **49**, 575–580 (1991).

29. Freneau, E., Sheffield, V. C., Molin, L., Shires, A. & Rhead, W. J. Glutaric acidemia type II. Heterogeneity in beta-oxidation flux, polypeptide synthesis, and complementary DNA mutations in the alpha subunit of electron transfer flavoprotein in eight patients. *J. Clin. Invest.* **90**, 1679–1686 (1992).
30. Bross, P. *et al.* A polymorphic variant in the human electron transfer flavoprotein α -chain (α -T171) displays decreased thermal stability and is overrepresented in very-long-chain acyl-CoA dehydrogenase-deficient patients with mild childhood presentation. *Mol. Genet. Metab.* **67**, 138–147 (1999).
31. Dieci, G., Fiorino, G., Castelnovo, M., Teichmann, M. & Pagano, A. The expanding RNA polymerase III transcriptome. *Trends Genet.* **23**, 614–622 (2007).
32. Saitsu, H. *et al.* Mutations in POLR3A and POLR3B encoding RNA Polymerase III subunits cause an autosomal-recessive hypomyelinating leukoencephalopathy. *Am. J. Hum. Genet.* **89**, 644–651 (2011).
33. Johnson, E. O. *et al.* PHLDA1 is a crucial negative regulator and effector of Aurora A kinase in breast cancer. *J. Cell Sci.* **124**, 2711–2722 (2011).
34. Sellheyer, K., Nelson, P., Kutzner, H. & Patel, R. M. The immunohistochemical differential diagnosis of microcystic adnexal carcinoma, desmoplastic trichoepithelioma and morpheiform basal cell carcinoma using BerEP4 and stem cell markers. *J. Cutan. Pathol.* **40**, 363–370 (2013).
35. Cardis, E. *et al.* The INTERPHONE study: design, epidemiological methods, and description of the study population. *Eur. J. Epidemiol.* **22**, 647–664 (2007).
36. Hunter, D. J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39**, 870–874 (2007).
37. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).
38. 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.* **164**, 2335–2342 (2004).
39. Wichmann, H. E., Gieger, C. & Illig, T. MONICA/KORA Study Group. KORA—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**(Suppl 1): S26–S30 (2005).
40. Krawczak, M. *et al.* PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* **9**, 55–61 (2006).
41. Schermund, 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.* **144**, 212–218 (2002).
42. Penegar, S. *et al.* National study of colorectal cancer genetics. *Br. J. Cancer* **97**, 1305–1309 (2007).
43. Eisen, T., Matakidou, A., Houlston, R. & Consortium, G. Identification of low penetrance alleles for lung cancer: the GENetic Lung CAncer Predisposition Study (GELCAPS). *BMC Cancer* **8**, 244 (2008).
44. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
45. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
46. 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.* **39**, 906–913 (2007).
47. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
48. Clayton, D. G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **37**, 1243–1246 (2005).
49. Liu, J. Z. *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* **42**, 436–440 (2010).
50. Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
51. Scheurer, M. E. *et al.* Familial aggregation of glioma: a pooled analysis. *Am. J. Epidemiol.* **172**, 1099–1107 (2010).
52. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
53. Scales, M., Jager, R., Migliorini, G., Houlston, R. S. & Henion, M. Y. visPIG—a web tool for producing multi-region, multi-track, multi-scale plots of genetic data. *PLoS ONE* **9**, e107497 (2014).
54. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
55. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
56. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
57. Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
58. de Souza, N. The ENCODE project. *Nat. Methods* **9**, 1046 (2012).
59. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
60. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
61. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
62. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
63. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pii (2013).
64. Zhang, K., Cui, S., Chang, S., Zhang, L. & Wang, J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.* **38**, W90–W95 (2010).
65. Idbaih, A. *et al.* BAC array CGH distinguishes mutually exclusive alterations that define clinicogenetic subtypes of gliomas. *Int. J. Cancer* **122**, 1778–1786 (2008).
66. Gonzalez-Aguilar, A. *et al.* Recurrent mutations of MYD88 and TBL1XR1 in primary central nervous system lymphomas. *Clin. Cancer Res.* **18**, 5203–5211 (2012).
67. Sanson, M. *et al.* Isocitrate dehydrogenase 1 codon 132 mutation is an important prognostic biomarker in gliomas. *J. Clin. Oncol.* **27**, 4150–4154 (2009).
68. Labussiere, M. *et al.* TERT promoter mutations in gliomas, genetic associations and clinico-pathological correlations. *Br. J. Cancer* **111**, 2024–2032 (2014).
69. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
70. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).

Acknowledgements

In the UK, funding was provided by Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund), the Wellcome Trust and the DJ Fielding Medical Research Trust. B.K. is supported by a PhD studentship funded by the Sir John Fisher Foundation. The National Brain Tumour Study is supported by the National Cancer Research Network and we acknowledge the contribution of all clinicians and health-care professionals 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) and the International Union against Cancer (UICC). The UICC received funds from the Mobile Manufacturers' Forum and 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 centres were also supported by the Mobile Telecommunications and Health Research (MTHR) Programme and the Northern UK Centre 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, the fondation ARC, the Institut National du Cancer (INCa; PL046), the French Ministry of Higher Education and Research and the program 'Investissements d'avenir' ANR-10-IAIHU-06. This study was additionally supported by a grant from Génomique Québec, le Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie (MESRST) Québec and McGill University. In Germany, funding was provided to M.S. and J.S. by the Deutsche Forschungsgemeinschaft (Sf552, 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&085475). 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) and supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. 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). Markus M. Nöthen received support from the Alfried Krupp von Bohlen und Halbach-Stiftung and is a member of the DFG-funded Excellence Cluster ImmunoSensation. We are grateful to all the patients and individuals for their participation and we would also like to thank the clinicians and other hospital staff, cancer registries and study staff in respective centres who contributed to the blood sample and data collection. For the UK study, we acknowledge the funders who contributed to the blood sample and data collection for this study. We also thank colleagues from the UK National Cancer Research Network (for NSCCG). MD Anderson acknowledges the work on the USA GWA study of Phyllis Adatto, Fabian Morice, Hui Zhang, Victor Levin, Alfred W.K. Yung, Mark Gilbert, Raymond Sawaya, Vinay Puduvali, Charles Conrad, Fredrick Lang and Jeffrey Weinberg from the Brain and Spine Center. For the French study, we are indebted to A. Rahimian (Onconeurotek),

A.M. Lekieffre and M Brandel for help in collecting data, and Y Marie for database support. For the German study, we are indebted to B. Harzheim (Bonn), S. Ott and Dr A. Müller-Erkwoh (Bonn) for help with the acquisition of clinical data, and 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 www.wtccc.org.uk. The US-GWA study made use of control genotypes from the Cancer Genetic Markers of Susceptibility (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 GWA study made use of genotyping data from three population control sources: KORA-gen, The Heinz Nixdorf RECALL study and POPGEN. The HNR cohort was established with the support of the Heinz Nixdorf Foundation. Franziska Degenhardt received support from the BONFOR Programme of the University of Bonn, Germany. We are grateful to all investigators who contributed to the generation of this data set. UK10K data generation and access was organized by the UK10K consortium and funded by the Wellcome Trust. The results here are in part based on data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

Author contributions

R.S.H. and B.K. designed the study, drafted the manuscript with contributions from all other authors; B.K. performed statistical and bioinformatics analyses; P.B. performed laboratory management and oversaw replication genotyping and sequencing of UK cases and controls by B.K., A.H. and J.V.; in the UK, A.S., M.J.S., S.J.F. and R.S.H. developed patient recruitment, sample acquisition and performed sample collection of cases; in the US, M.B. developed protocols and patient recruitment; in Germany, M.S., M.M.N.,

H.-E.W., S.S. and J.S. developed patient recruitment and blood sample collection, M.S. oversaw DNA isolation and storage and performed case and control ascertainment and supervision of DNA extractions, S. Herms, S. Heilmann and K.G. performed experimental work; in France, M.S. and J.-Y.D. developed patient recruitment, M.L. and A.-L.D.S performed tumour genotyping and K.M. performed the neuropathological analysis; M. Lathrop performed laboratory management and oversaw genotyping of the French samples; all authors contributed to the final paper.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Kinnersley, B. *et al.* Genome-wide association study identifies multiple susceptibility loci for glioma. *Nat. Commun.* 6:8559 doi: 10.1038/ncomms9559 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>