Genome-wide association study reveals specific differences in genetic susceptibility to glioblastoma and non-glioblastoma

Beatrice Melin*#, Jill Barnholtz-Sloan#, Margaret R. Wrensch#, Christoffer Johansen#, Dora Il'yasova#, Ben Kinnersley#, Quinn Ostrom, Karim Labreche, Yanwen Chen, Georgina Armstrong, Yanhong Liu, Jeanette E Eckel-Passow, Paul A Decker, Marianne Labussière, Anna-Luisa Di Stefano, Peter Broderick, Pilar Galan, Karima Mokhtari, Jean-Yves Delattre, Konstantinos Gousias, Johannes Schramm, Minouk J. Schoemaker, Sarah J Fleming, Stefan Herms, Stefanie Heilmann, Markus M Nöthen, Heinz-Erich Wichmann, Stefan Schreiber, Anthony Swerdlow, Mark Lathrop, Matthias Simon, Marc Sanson, Ulrika Andersson, Preetha Rajaraman, Stephen Chanock, Martha Linet, Zhaoming Wang, Meredith Yeager,(on behalf on the glioma scan group), John Wiencke, Helen Hansen, Lucie McCoy, Terri Rice, Matthew L Kosel, Thomas M Kollmeyer, Hugues Sicotte, Chris Amos, Jonine Bernstein, Faith Davis, Dan Lachance, Ching Lau, Ryan T Merrell , Joellen Shildkraut, Ali Osman, Siegal Sadetzki, Michael Scheurer, Sanjay Shete, Rose Lai+, Elizabeth B Claus+, Sara Olson+, Robert B Jenkins+, Richard S Houlston+*, Melissa L Bondy+*

*Corresponding authors
shared first authors
+shared last authors

Author affiliations:
Department of Radiation Sciences, Faculty of Medicine, Umeå University, Umeå, Sweden (Beatrice S. Melin, Ulrika Andersson)

Case Comprehensive Cancer Center, School of Medicine, Case Western Reserve University, Cleveland, Ohio (Jill S. Barnholtz-Sloan Quinn Ostrom, Yanwen Chen)

Department of Neurological Surgery, School of Medicine, University of California, San Francisco, San Francisco, California, USA (Margaret R. Wrensch, John Wiencke, Helen Hansen, Lucie McCoy, Terri Rice)

Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark , Rigshospitalet, University of Copenhagen, Copenhagen, Denmark (Christoffer Johansen)

Department of Epidemiology and Biostatistics, School of Public Health, Georgia State University, Atlanta, Georgia< USA (Dora Il'yasova)

Cancer Control and Prevention Program, Depart- ment of Community and Family Medicine, Duke University Medical Center, Durham, North Carolina,USA (Dora Il'yasova, Joellen Schildkraut)

Departments of Neurology, Neurosurgery, and Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA (Rose Lai)

Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK (Richard S. Houlston, Ben Kinnersley, Karim Labreche, Peter Broderick)

Division of Molecular Pathology, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK (Richard S. Houlston)

Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota, USA.(Hugues Sicotte, Paul A Decker, Matthew L Kosel, Jeanette E Eckel-Passow)

Department of Pediatrics, Division of Hematology-Oncology, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas, USA (Georgina N. Armstrong, Yanhong Liu, Ching C. Lau, Michael E. Scheurer, Melissa L. Bondy)

Sorbonne Universités UPMC Univ Paris 06, INSERM CNRS, France (Karim Labreche, Marianne Labussiere, Anna-Luisa De Stefano, Karima Mokhtari, Jean Yves Delattre, Marc Sanson)

Université Paris 13 Sorbonne Paris Cité, Inserm (U557), Inra (U1125), Cnam, France (Pilar Galan)

Department of Neurosurgery, University of Bonn Medical Center, Sigmund-Freud-Str. 25, 53105 Bonn, Germany (Konstantinos Gousias, Johannes Schramm, Matthias Simon)

Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK (Minouk J Schoemaker, Anthony Swerdlow)

Centre for Epidemiology and Biostatistics, Faculty of Medicine and Health, University of Leeds, Leeds LS2 9JT, UK (Sarah Flemings, Stefan Herms)

Institute of Human Genetics, University of Bonn, Germany (Stefanie Heilmann, Marcus M Nöthen)

Génome Québec, Department of Human Genetics, McGill University, Montreal, Quebec, H3A 0G1, Canada (Marc Lathrop)

Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany (Heinz-Erich Wichmann)

1st Medical Department, University Clinic Schleswig-Holstein, Campus Kiel, Germany (Stefan Schreiber)

Centre for Global Health, National Cancer Institute, Bethesda, USA (Preetha Rajaraman, Stephen Chanock, Martha Linet, Zhaoming Wang, Meredith Yeager).

Department of Public Health Services, School of Public Health, University of Alberta, Edmonton, Alberta, Canada (Faith G. Davis)

Department of Surgery, Duke University Medical Center, Durham, North Carolina, USA (Francis Ali-Osman)

Department of Com- munity and Family Medicine, Department of Genetics, Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA (Christopher I. Amos)

Department of Neurology, NorthShore University HealthSystem, Evanston, Illinois, USA (Ryan T. Merrell)

Department of Epidemiology and Public Health, School of Medicine, Yale University, New Haven, Connecticut (Elizabeth B. Claus)

Department of Neurosurgery, Brigham and Women's Hospital, Boston, Massachusetts, USA (Elizabeth B. Claus)

Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA (Sara H. Olson, Jonine L. Bernstein)

Cancer and Radiation Epidemiology Unit, Gertner Institute, Chaim Sheba Medical Center, Tel Hashomer, Israel (Siegal Sadetzki)

Department of Epidemiology and Preventive Medicine, School of Public Health, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel (Siegal Sadetzki)

30 Department of Laboratory Medicine and Pathology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, Minnesota, USA (Robert B. Jenkins, Thomas M Kollmeyer)

Department of Neurology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, Minnesota, USA (Daniel Lachance)

Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA (Sanjay Shete)

Genome-wide association studies (GWAS) have transformed our understanding of glioma susceptibility, but individual studies have had limited power to identify risk loci. We have performed a meta-analysis of these GWAS, a new GWAS and replication comprising 12,496 cases and 18,190 controls. We identified new loci for glioblastoma (GBM) at 1p31.3 (rs12752552; $P=2.04\times10^{-9}$, odds ratio (OR)=1.22), 11q14.1 (rs11233250; $P=9.95\times10^{-10}$, OR=1.24), 16p13.3 (rs2562152; $P=1.93\text{x}10^{-8}$, OR=1.21), 16q12.1 (rs10852606; $P=1.29\times10^{-11}$, OR=1.18), 22q13.1 (rs2235573; $P=1.76\times10^{-10}$, OR=1.15) and for non-GBM at 1q32.1 (rs4252707; $P=3.34\times10^{-9}$, OR=1.19), 1q44 (rs12076373; $P=2.63\times10^{-10}$, OR=1.23), 2q33.3 (rs7572263; $P=2.18\times10^{-10}$, OR=1.20), 3p14.1 (rs11706832; $P=7.66\times10^{-9}$, OR=1.15), 10q24.33 (rs11598018; $P=3.39\times10^{-8}$, OR=1.14), 11q21 (rs7107785; $P=3.87\times10^{-10}$, OR=1.16), 14q12 (rs10131032; $P=5.07\text{x}10^{-11}$, OR=1.33) and 16p13.3 (rs3751667; $P=2.61\times10^{-9}$, OR=1.18) which localize in/near *RAVER2, FAM181B, MPG, HEATR3, SLC16A8, MDM4, AKT3, IDH1, LRIG1*, *OBFC1*, *MAML2, AKAP6*, and *LMF1* respectively. These data substantiate genetic susceptibility to GBM and non-GBM being highly distinct, likely reflecting different etiology.

Glioma accounts for around 27% of all primary brain tumors and is responsible for around 13,000 cancer-related deaths in the US each year[1,2]. Gliomas can be broadly classified into glioblastoma (GBM) and lower-grade non-GBM[3]. Gliomas typically have a poor prognosis irrespective of medical care, with the most common form, GBM, having a median overall survival of only 12-14 months[1].

So far no environmental exposures have robustly been linked to risk of developing glioma except for ionizing radiation, which accounts for a small proportion of cases[4]. 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 neurofibromatosis. Even collectively, however these account for little of the two-fold familial risk of glioma[5]. Our understanding of the heritability of glioma has been transformed by recent genome-wide association studies (GWAS), which have identified single nucleotide polymorphisms (SNPs) at 13 loci influencing risk[6-12].

Due to limited statistical power of the previous individual studies for additional discovery of novel glioma risk loci[13], we performed a meta-analysis to gain a comprehensive insight to glioma etiology of these previously published GWAS and a new GWAS with replication comprising 12,496 cases and 18,190 controls, identifying 13 new risk loci.

We initially analysed GWAS SNP data passing quality control for 10,977 cases (5,665 GBM, 4,827 non-GBM) and 17,386 controls from seven studies of European ancestry: a new GWAS performed by the Glioma International Case Control Consortium (GICC) **(Supplementary Table 1)** and six previously reported GWAS[8,9,14]. To increase genomic resolution, we imputed >10 million SNPs using the 1000 Genomes Project[15] combined with UK10K[16] as reference. Quantile-Quantile (Q-Q) plots for SNPs with minor allele frequency (MAF) >5% post imputation did not show evidence of substantive over-dispersion (λ = 1.00–1.09; **Supplementary Fig. 1).** We derived joint odds ratios (ORs) and 95% confidence intervals (CIs) under a fixed-effects model for each SNP with MAF >0.01 and associated per allele principle component (PCA) corrected *P*-values for all glioma, GBM and non-GBM.

We sought validation of SNP *P*-values from the meta-analysis at $<10^{-6}$ using *in-silico* replication in an additional 1,519 cases (526 GBM, 992 non-GBM) and 804 controls. In the combined meta-analysis, associations at the previously reported risk loci for all glioma at 17p13.1 (*TP53*), GBM at 5p15.33 (*TERT*), 7p11.2 (*EGFR*), 9p21.3 (*CDKN2A*), and 20q13.33 (*RTEL1*) and for non-GBM at 8q24.21 (*CCDC26*), 10q25.2 (*VTI1A*), 11q23.2, 11q23.3 (*PHLDB1*), 12q21.2 and 15q24.2 (*ETFA*) showed a consistent direction of effect and relationship with tumor histological type previously reported ($P<5.0\times10^{-8}$, **Fig. 1**, **Supplementary Table 2, Supplementary Fig. 2**). Associations at the previously reported 3q26.2 (*TERC*)[10] and 12q23.33 (*POLR3B*)[9] loci for GBM did not attain statistical significance (respective *P*-values for the most associated SNPs = 2.25×10^{-6} and 1.60×10^{-5}; **Supplementary Table 2**).

After meta-analysis of the discovery and replication samples, we identified genome-wide significant associations marking novel loci (**Table 1, Fig. 2, Fig.3**) for GBM at 1p31.3 (rs12752552; $P=2.04\times10^{-9}$, odds ratio (OR)=1.22), 11q14.1 (rs11233250; $P=9.95\times10^{-10}$, OR=1.24), 16p13.3 (rs2562152; $P=1.93\times10^{-8}$, OR=1.21), 16q12.1 (rs10852606; $P=1.29\times10^{-11}$, OR=1.18), 22q13.1 (rs2235573; $P=1.76\times10^{-10}$, OR=1.15) and for non-GBM at 1q32.1 (rs4252707; $P=3.34\times10^{-9}$, OR=1.19), 1q44 (rs12076373; $P=2.63\times10^{-10}$, OR=1.23), 2q33.3 (rs7572263; $P=2.18\times10^{-10}$, OR=1.20), 3p14.1 (rs11706832; $P=7.66\times10^{-9}$, OR=1.15), 10q24.33 (rs11598018; $P=3.39\times10^{-8}$, OR=1.14), 11q21 (rs7107785; $P=3.87\times10^{-10}$, OR=1.16), 14q12 (rs10131032; $P=5.07\times10^{-11}$, OR=1.33) and 16p13.3 (rs3751667; $P=2.61\times10^{-9}$, OR=1.18). We also identified a promising association at 9q21.13 for GBM (rs34718722; $P=7.77\times10^{-8}$). Conditional analysis confirmed the existence of two independent

association signals at 7p11.2 (*EGFR*) defined by SNPs rs75061358 and rs723527 previously reported[6] but did not provide evidence for additional signals at any of the other established identified risk loci or the 13 newly identified loci. Collectively our findings provide strong evidence for subtype associations for glioma consistent with their distinctive molecular profiles presumably resulting from different etiological pathways (**Fig. 4**).

Across the new and known risk loci, we confirmed a significant enrichment of the enhancer associated histone marks, including H3K4me1, H3K4me4, and H3K27ac in neural-progenitor cells ($P<1x10^{-4}$). These observations support the assertion that the GWAS loci influence glioma risk through effects on neural cis-regulatory networks, and 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 glioma from The Cancer Genome Atlas (TCGA) and normal human brain using the GTEx portal. We examined for an association between SNP genotype and expression of genes mapping within 1 Mb of the sentinel SNP (**Supplementary Data 1**), noting significant associations in both normal brain and tumor at 16q12.1 for rs10852606 and *HEATR3* ($P_{TCGA\text{-}all\text{-}glioma}=8.22x10^{-7}$; $P_{GTEx\text{-}cerebellum}=2.1x10^{-30}$), as well as in glioma tumors at 11q14.1 for rs11233250 and *RP11-179A16.1* ($P_{TCGA\text{-}all\text{-}glioma}=7.30x10^{-7}$) and normal brain at 1p31.3 for rs12752552 with *JAK1* ($P_{GTEx\text{-}cerebellum}=2.1x10^{-7}$). Additionally for each of the risk SNPs at the 13 new loci (as well as correlated variants, $r^2>0.8$) we examined published data[17,18] and made use of the online resources, HaploRegv3, RegulomeDB, and SeattleSeq for evidence of functional effect (**Supplementary Table 3**).

The 1q32.1 association marked by rs4252707 (**Fig. 2**) maps to intron eight of the gene encoding *MDM4* (mouse double minute 4 homolog) a p53-binding protein. Over-expression of MDM4 is a feature in *TP53*-mutation and *MDM2*-amplification negative glioma, consistent with MDM4 amplification being a mechanism by which the p53-dependent growth control is inactivated[19]. The 1q44 association marked by rs12076373 maps intronic to *AKT3* (v-akt murine thymoma viral oncogene homolog 3) one of the major downstream effectors of phosphatidylinositol 3-kinase which is highly expressed during active neurogenesis, with haploinsufficiency causing postnatal microcephaly and agenesis of the corpus callosum[20]. Importantly *AKT3* is hyper-expressed in glioma playing an important role in tumor viability by activating DNA repair[21]. The 3p14.1 association marked by rs11706832 localizes to intron 2 of *LRIG1* (leucine-rich repeats- and immunoglobulin-like domains-containing protein 1). *LRIG1* is highly expressed in the brain and is a pan-negative regulator

of the *EGFR* signaling pathway which inhibits hypoxia-induced vasculogenic mimicry via *EGFR*/*PI3K*/*AKT* pathway suppression and epithelial-to-mesenchymal transition[22]. Reduced *LRIG1* expression is linked tumor aggressiveness, temozolomide-resistance and radio-resistance[23,24]. While we have previously shown an association for glioma at *EGFR* (7p11.2)[6], which is well established to be pivotal in both initiation of primary GBM and progression of lower-grade glioma to grade IV we now suggest a more extensive pathway involving variation in *LRIG1* and *AKT3*.

Of particular interest is rs7572263 mapping to 2q33.3 which localises ~50 kb telomeric to the gene encoding *IDH1* (isocitrate dehydrogenase 1). Mutation of *IDH1* is a driver for gliomagenesis[25,26] and is responsible for the CpG island methylator (G-CIMP) phenotype[27,28]. Since *IDH* mutation predominates in non-GBM glioma[29,30] the association at 2q33.3 is entirely plausible as a basis for susceptibility to non-GBM glioma.

Maintenance of telomeres is central to cell immortalization and plays a central role in gliomagenesis[31]. We have previously shown the risk of GBM is strongly linked to genetic variation in the telomere-related genes *TERT* (5p15.33) and *RTEL1* (20q13.33), but probably also *TERC* (3q26.2)[7,8,10]. To fully decipher the biological impact of these SNP associations on the glioma development require additional functional analyses. The glioma risk alleles at *TERT*, *RTEL1*, *TERC* and *OBFC1* are however associated with increased leukocyte telomere length thereby providing a direct relationship between genotype and biology[31-33].

The 10q24.33 association marked by rs11598018 lies intronic to *OBFC1* (oligonucleotide/oligosaccharide-binding fold-containing protein 1), which functions in a telomere-associated complex protecting telomeres independently of POT1. The CST complex encoded by *OBFC1*, *CTC1*, and *TEN1* competes with shelterin for telomeric DNA inhibiting telomerase-based telomere extension. The significant association between risk of non-GBM and *OBFC1* variation is particularly intriguing in light of a recent report demonstrating that germline loss-of-function mutations in shelterin-complex genes are a rare cause of familial oligodendroglioma[34].

Deregulation of pathways involved in telomere length and *EGFR* signalling are thus consistent with glioma risk being governed by pathways important in the longevity of glial cells and substantiate early observations that genetic susceptibility to GBM and non-GBM is highly distinct, presumably reflecting different aetiologies between GBM and non-GBM tumors (**Fig. 4**).

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*; a N-methylpurine DNA glycosylase that removes a diverse group of damaged bases from DNA including cytotoxic and mutagenic alkylation adducts of purines. *MPG* expression is linked to temozolomide resistance[35]. The association at 1p31.3 implicates Raver2; a modulator of the splicing repressor PTB (Polypyrimidine Tract Binding Protein) that is highly expressed during neurogenesis. However the region of LD encompassing rs12752552 also contains *JAK1.* Since the *JAK1-STAT6* signaling is increasing being recognized to be relevant to glioma progression it remains to be established which gene is the functional basis for the 1p31.3 association. Similarly, the genetic and functional basis for associations at 11q14.1 (GBM), 16q12.1 (*HEATR3*; GBM), 22q13.1 (*SLC16A8*; GBM), 11q21 (*MAML2*; non-GBM), 14q12 (*AKAP6*; non-GBM), and 16p13.3 (*LMF1*; non-GBM) remain to be established. However, the observation that rs10852606 variation at 16q12.1 has previously been associated with risk of both testicular (rs8046148) and esophageal (rs4785204) cancer (pairwise r^2 and D' with rs10852606, 0.67, 1.0 and 0.16, 1.0 respectively) is compatible with this locus having pleiotropic effects on tumor risk.

In conclusion we have performed the largest glioma GWAS to date identifying 13 new glioma risk loci taking the total count to 26. Our findings provide further evidence for a polygenic basis of genetic susceptibility to glioma however, and it is important to understand the biology behind these risk variants. Currently identified risk SNPs for glioma account for at best around 27% and 34% of the familial risk of GBM and non-GBM tumours respectively (Supplementary Table 4).therefore further GWAS-based studies in concert with functional analyses should lead to additional insights into the biology and etiological basis of the different glioma histologies. Importantly, such information can inform gene discovery initiatives and thus have a measurable impact on the successful development of new therapeutic agents.

Note: Supplementary information is available on online

ACKNOWLEDGEMENTS

Glioma scan group;

Laura E. Beane Freeman , Stella Koutros, Demetrius Albanes, Kala Visvanathan, Victoria L. Stevens , Roger Henriksson, Dominique S. Michaud , Maria Feychting, Anders Ahlbom , Graham G. Giles Roger Milne , Roberta McKean-Cowdin, Loic Le Marchand, Meir Stampfer, Avima M. Ruder, Tania Carreon, Goran Hallmans, Anne Zeleniuch-Jacquotte, J. Michael Gaziano , Howard D. Sesso, Mark P. Purdue, Emily White , Ulrike Peters, Howard D. Sesso, Julie Buring.

The GICC was supported by grants from the National Institutes of Health, Bethesda, Maryland (R01CA139020, R01CA52689, P50097257, P30CA125123). Additional support was provided by the McNair Medical Institute and the Population Sciences Biorepository at Baylor College of Medicine.

In Sweden work was additionally supported by Acta Oncologica through the Royal Swedish Academy of Science (BM salary) and The Swedish Research council and Swedish Cancer foundation. We are grateful to all clinicians and research nurses throughout Sweden that identified all cases.

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. The National Brain Tumour Study is supported by the National Cancer Research Network and we acknowledge the contribution of all clinicians and heath 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 Génome 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 (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&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.

For the UK study, we acknowledge the funders who contributed to the blood sample and data collection for this study as listed in Hepworth et al. (BMJ 2006, 332, 883). 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 Puduvalli, 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 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-gen39, The Heinz-Nixdorf RECALL studyand 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.

The UCSF Adult Glioma Study was supported by the National Institutes of Health (grant numbers R01CA52689, P50CA097257, R01CA126831, and R01CA139020), the Loglio Collective, the National Brain Tumor Foundation, the Stanley D. Lewis and Virginia S. Lewis Endowed Chair in Brain Tumor Research, the Robert Magnin Newman Endowed Chair in Neuro-oncology, and by donations from families and friends of John Berardi, Helen Glaser, Elvera Olsen, Raymond E. Cooper, and William Martinusen. This project also was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 RR024131. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The 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, A Molinaro, A Perry, M Pezmecki, M Prados, I Smirnov, T Tihan, K Walsh, J Wiemels, S Zheng.

At Mayo the authors wish to acknowledge study participants, 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 Mine Cicek). Work at the Mayo Clinic beyond the GICC was also supported by the US National Institutes of Health (NIH; grants P50CA108961 and P30CA15083), the National Institute of Neurological Disorders and Stroke (grant RC1NS068222Z), the Bernie and Edith Waterman Foundation and the Ting Tsung and Wei Fong Chao Family Foundation.

UK10K data generation and access was organised by the UK10K consortium and funded by the Wellcome Trust. The results here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

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 centers who contributed to the blood sample and data collection.

AUTHOR CONTRIBUTIONS

M.B., B.M., R.S.H. and J.B-S. performed project management; R.S.H., M.B., B.M., J.B-S., R.B., Q.O., B.K. and M.W. drafted the manuscript; Q.O., K.L., B.K. and J.E.E-P., P.D performed statistical analyses; YW, K.L., and B.K. performed bioinformatic analyses; xx B.M., J.B-S, M.R.W., C.J., D.Il'y, R.L., , G.A., , Paul A Decker, U.A., T.R., M.L.K., T.M., H.S., J.B., F.D., D.L., R.T.M., J.S., A.O., S.S., M.S, S.S, E.B.C., S.O., R.B.J., R.S.H., M.L.B. performed sample acquisition; P.R., S.C., M.L, Z.W. and M.Y., provided NCI data; all authors contributed to the final manuscript.

COMPETING INTERESTS STATEMENT

The remaining authors declare no competing financial interests.

METHODS

Ethics

Collection of patient samples and associated clinico-pathological information was undertaken with written informed consent and relevant ethical review board approval at respective study centers in accordance with the tenets of the Declaration of Helsinki. Specifically, UK: South-East Multicentre Research Ethics Committee (MREC) and the Scottish MREC; France: APHP ethical committee-CPP (comité de Protection des Personnes); Germany: Ethics Commission of the Medical Faculty of the University of Bonn and USA: US: University of Texas MD Anderson Cancer Institutional Review Board, the Mayo Clinic Office for Human Research Protection, the UCSF Committee on Human Research, the University Hospitals of Cleveland Institutional Review Board and the Cleveland Clinic Institutional Review Board (board for the Case Comprehensive Cancer Center). 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.

Primary GWAS

Studies participating in GICC are described in Amirian *et al*.[36] and in **Supplementary Table 1**, and comprise 5,189 glioma cases and 3,827 controls ascertained through centers in the US, Denmark, Sweden and the UK. Cases had newly diagnosed glioma and controls had no personal history of cancer at ascertainment. Detailed information regarding recruitment protocol is given in Amirian *et al.*[36]. Cases and controls were genotyped using the Illumina Oncoarray according to the manufacturer's recommendations (Illumina Inc.). Individuals with call rate <99% as well as all individuals evaluated to be of non-European ancestry (<80% estimated European ancestry using the FastPop procedure developed by the GAMEON consortium HapMap version 2 CEU, JPT/CHB and YRI populations as a reference, **Supplementary Fig. 3** were excluded. For apparent first-degree relative pairs, 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 displaying significant deviation from Hardy-Weinberg equilibrium (HWE) (*i.e.* $P<10^{-5}$).

Published GWAS

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; Briefly: (1) The UK-GWAS[6,7,9] was based on 636 cases (401 males; mean age 46 years) ascertained through the INTERPHONE

study[37]. Individuals from the 1958 Birth Cohort (n=2,930) served as a source of controls; (2) The French-GWAS[6,9] 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 (SUpplementation en VItamines et MineralsAntioXydants) study of 12,735 healthy subjects (women aged 35–60 years; men aged 45–60 years)[38]; (3) The German-GWAS[9] 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)[39]; POPGEN (Population Genetic Cohort; n=678)[40] and from the Heinz Nixdorf Recall study (n=380)[41]. Standard, quality control measures were applied to the UK, French and German GWAS and have previously been reported. (4) The MDA-GWAS[7] 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[42,43]. Quality control measures were applied as per the Primary GWAS. (5) The UCSF GWAS. The UCSF adult glioma case-control study includes participants of the San Francisco Bay Area Adult Glioma Study (AGS). Details of subject recruitment for AGS have been reported previously[8,12,30,44,45]. Briefly, cases were adults (>18 years of age) with newly diagnosed histologically confirmed glioma. Population-based cases diagnosed between 1991 and 2009 (Series 1-4) and 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 diagnosed 2002-2012, (Series 3-5) were recruited from the UCSF Neuro-oncology Clinic, regardless of 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 dialling and were frequency matched to population-based cases on 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 on 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 iControls) were used in the current analysis.(6) The GliomaScan GWAS[14] – in addition to the published analysis we excluded samples from the ATBC (Finnish study) and controls from NSHDS which were excluded 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 imputation

GWAS data were imputed to >10 million SNPs with IMPUTE2 v2.3[46] software using a merged reference panel consisting of data from 1000 Genomes Project (phase 1 integrated release 3, March 2012)[15] and UK10K (ALSAPAC, EGAS00001000090 / EGAD00001000195 and TwinsUK EGAS00001000108/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>0.005. Poorly imputed SNPs defined by an information measure <0.80 with IMPUTE2 were excluded. Test of association between imputed SNPs and glioma was performed using logistic regression under an additive genetic model in SNPTESTv2.5[47]. The adequacy of the case-control matching and possibility of differential genotyping of cases and controls were formally evaluated using Q-Q plots of test statistics (**Supplementary Fig.1**). The inflation factor λ was based on the 90% least-significant SNPs. Where appropriate, principle 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. Principle components, based on genotyped SNPs were generated for GICC, GliomaScan, MDA-GWAS and SFAGS using PLINK[48]. Eigenvectors for the German GWAS dataset was inferred using smartpca (part of EIGENSOFTv2.4)[49,50] by merging cases and controls with Phase II HapMap samples[9].

***In silico* replication**

For replication of promising associations we analysed Mayo case-control and UCSF case-control and Mayo Clinic Biobank control data. The Mayo Clinic case-control study has been described previously[8,30,51]. 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), were at least 18 years of age, and had a surgical resection or biopsy between 1973 and 2014. Controls were at least 18 years of age, underwent a general medical examination at Mayo Clinic between 2005 and 2012, and had no previous history of a brain tumor. Controls were matched to cases by gender, age, ethnicity, and residence. 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. The Mayo Clinic Biobank controls comprised volunteers who donated biological specimens, provide risk factor data, access to clinical data obtained from the medical record and provide 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. While 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, MN; Jacksonville, FL; and the Mayo Clinic Health System sites in La Crosse and Onalaska, WI. All were aged 18 years and older at 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; 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[48]. STRUCTURE[52] was used to assess population admixture with 1000 Genomes as reference. Subjects indicated to be non-Caucasian were excluded. Prior to imputation, SNPs were tested for HWE and SNPs with HWE $P<10^{-6}$ 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 genome 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. 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 coded as 0, 1, or 2 copies of the minor allele, adjusted for age, sex, and the first three principal components.

Meta-analysis

Meta-analyses were performed using the fixed-effects inverse-variance method based on the β estimates and standard errors from each study using META v1.6[53]. 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 calculated[54]. Using the meta-analysis summary statistics and LD correlations from a reference panel of 1000 Genomes Project combined with UK10K we used GCTA[55,56] to perform conditional association analysis. Association statistics were calculated for all SNPs

conditioning on the top SNP in each loci showing genome-wide significance. This is carried out in a step-wise fashion.

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[57], RegulomeDB[58] and SeattleSeq Annotation[59]. These servers make use of data from ENCODE, genomic evolutionary rate profiling (GERP) conservation metrics, combined annotation dependent depletion (CADD) scores and PolyPhenscores. We searched for overlap of associated SNPs with enhancers defined by the FANTOM5 enhancer atlas[18], annotating by ubiquitous enhancers as well as 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[17] 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 project[60].

Expression quantitative trait loci analysis

To examine the relationship between SNP genotype and gene expression, we made use of RNA sequence data for 389 low-grade and 138 GBM tumors of European ancestry from TCGA (accession number phs000178.v9.p8). 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 featureCounts[61] as part of the Rsubread Bioconductor package[62]. For TCGA samples, European ancestry was assessed through visualization of clustering with CEU samples after principal components analysis. Untyped genotypes were imputed from Affymetrix 6 array data 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 publically available eQTL mRNA expression data from the GTEx portal (http://www.gtexportal.org).

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*., 2008[63] assuming the familial risk of glioma to be 1.77[64].

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 towards 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:

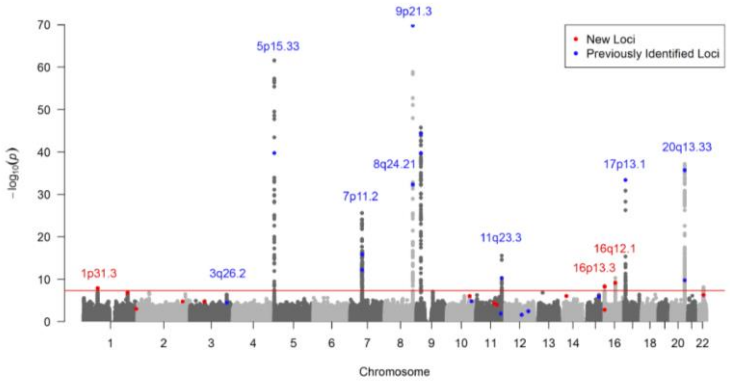
$$V_i/V$$

LD metrics were calculated in vcftools v0.1.12b[65] using UK10K data and plotted using visPIG[66]. 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*.[67]

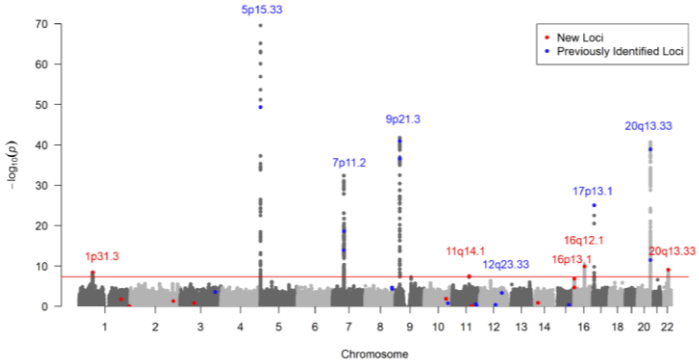
FIGURE AND TABLE LEGENDS

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 horizontal line corresponds to a significance threshold of $P = 5.0 \times 10^{-8}$. New and known loci are labelled in red and blue respectively.

a) All glioma



b) GBM



c) Non-GBM

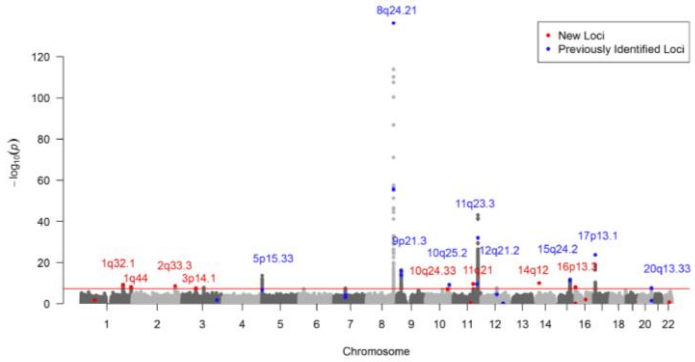
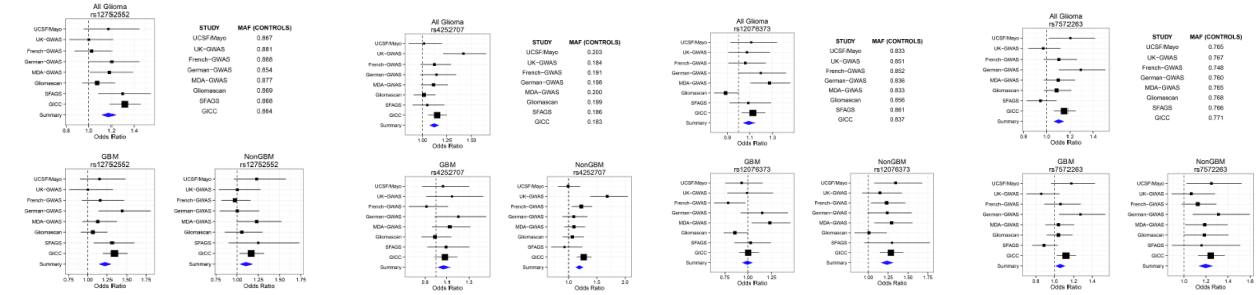
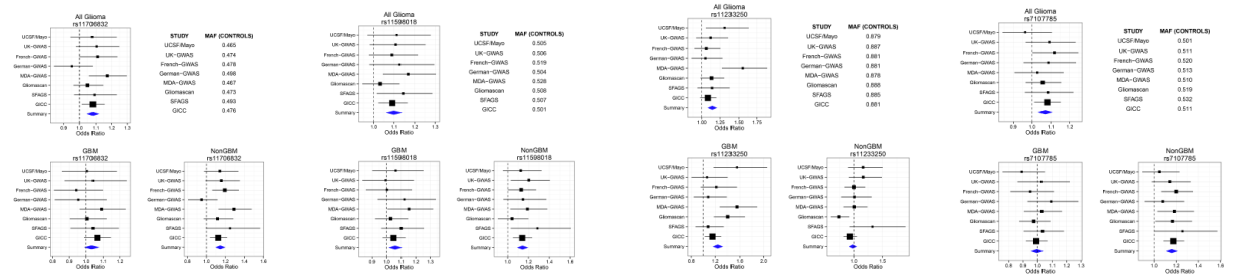


Figure 2: Forest plots of effect size and direction for the new SNPs associated with glioma risk. Associations are for SNPs as reported in Table 1. Boxes denote OR point estimates, with their areas proportional to the inverse variance weight of the estimate. Horizontal lines represent 95% confidence intervals. The diamond (and broken line) represents the summary OR computed under a fixed effects model, with 95% C.I. given by the width of the diamond. The unbroken vertical line is at the null value (OR = 1.0).

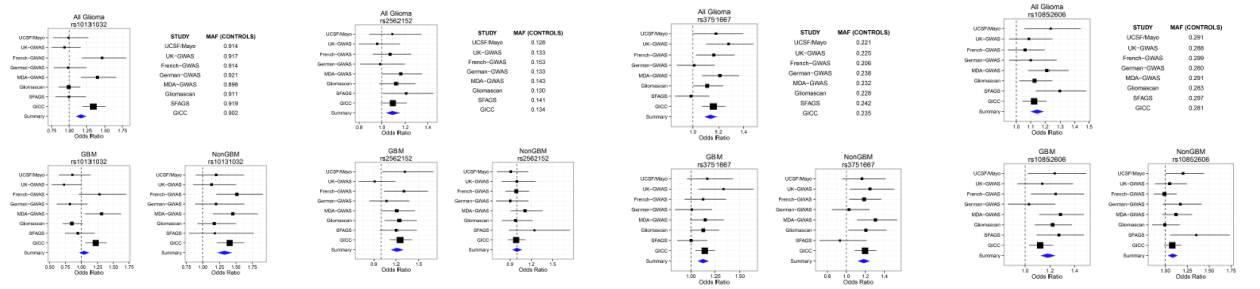
rs12752552 (1p31.3) rs4252707 (1q32.1) rs12076373 (1q44) rs7572263 (2q33.3)



rs11706832 (3p14.1) rs11598018 (10q24.33) rs11233250 (11q14.1) rs7107785 (11q21)



rs10131032 (14q12) rs2562152 (16p13.3) rs3751667 (16p13.3) rs10852606 (16q12.1)



rs2235573 (22q13.1)

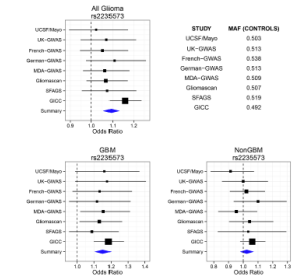
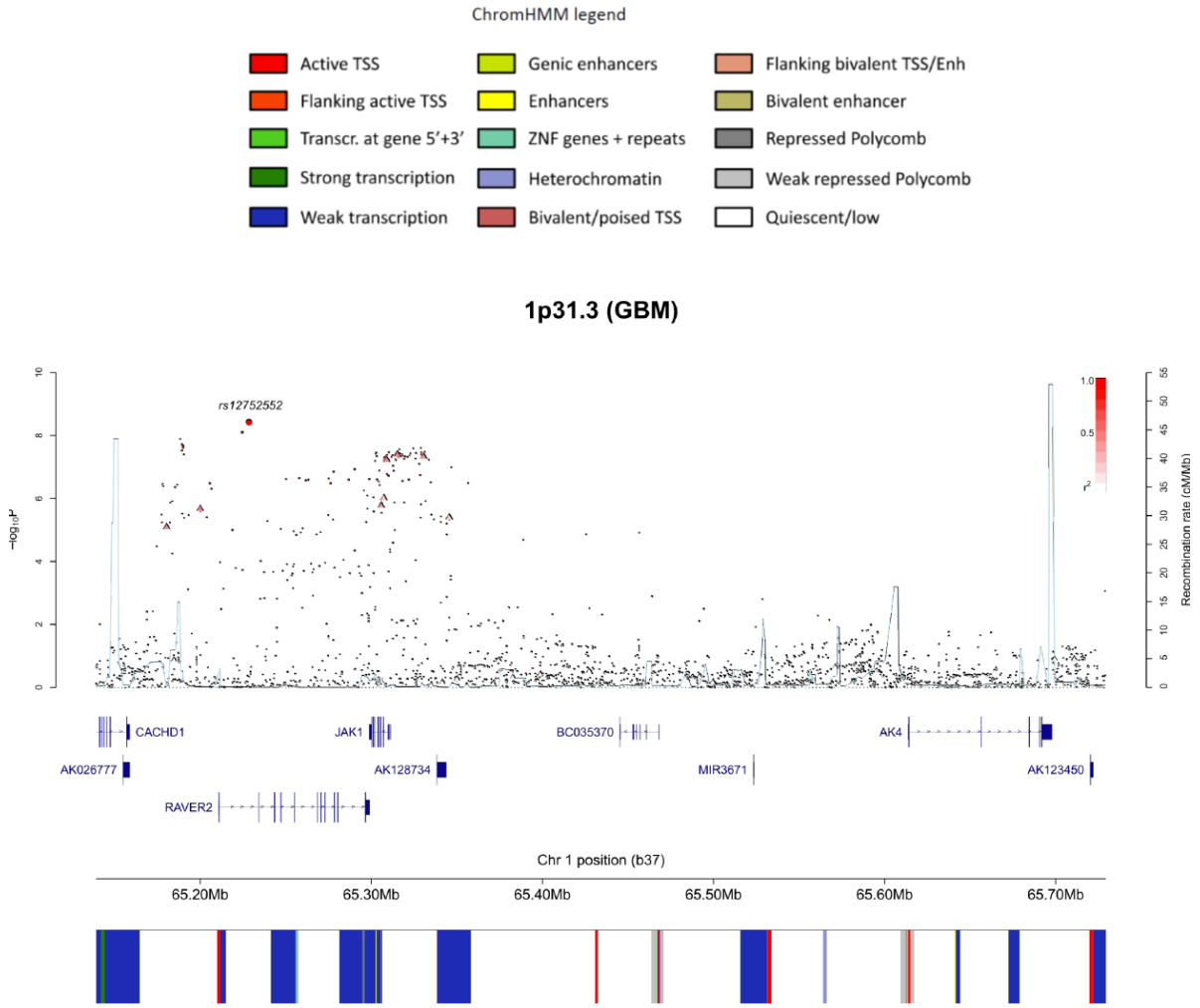
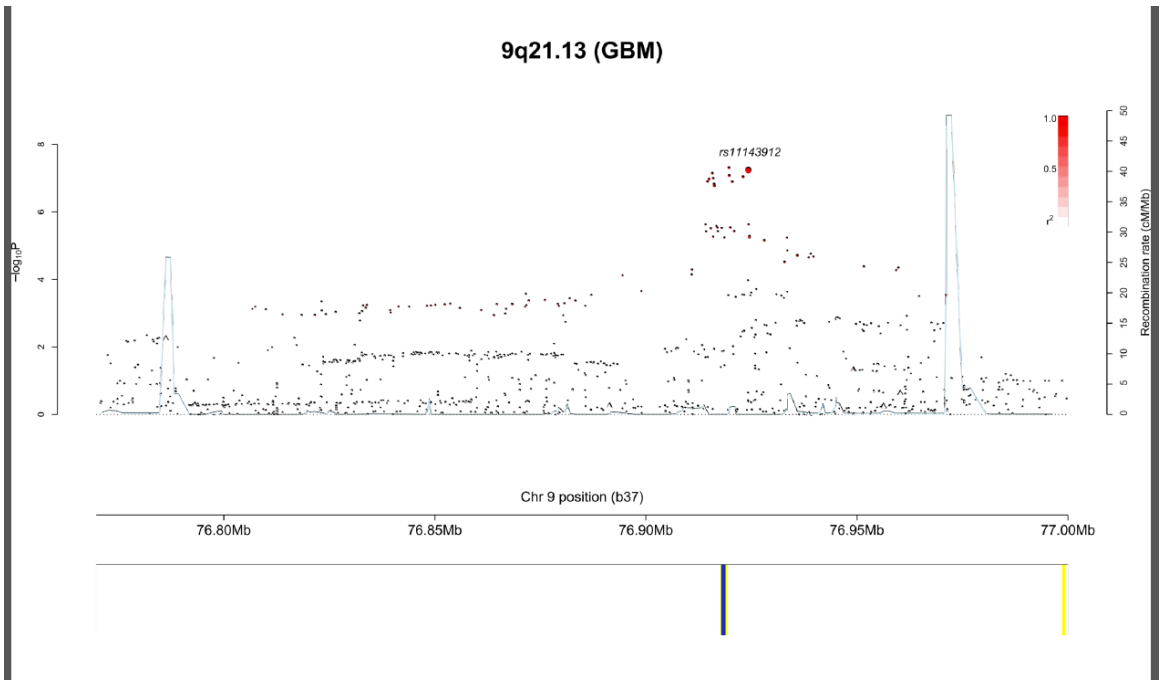
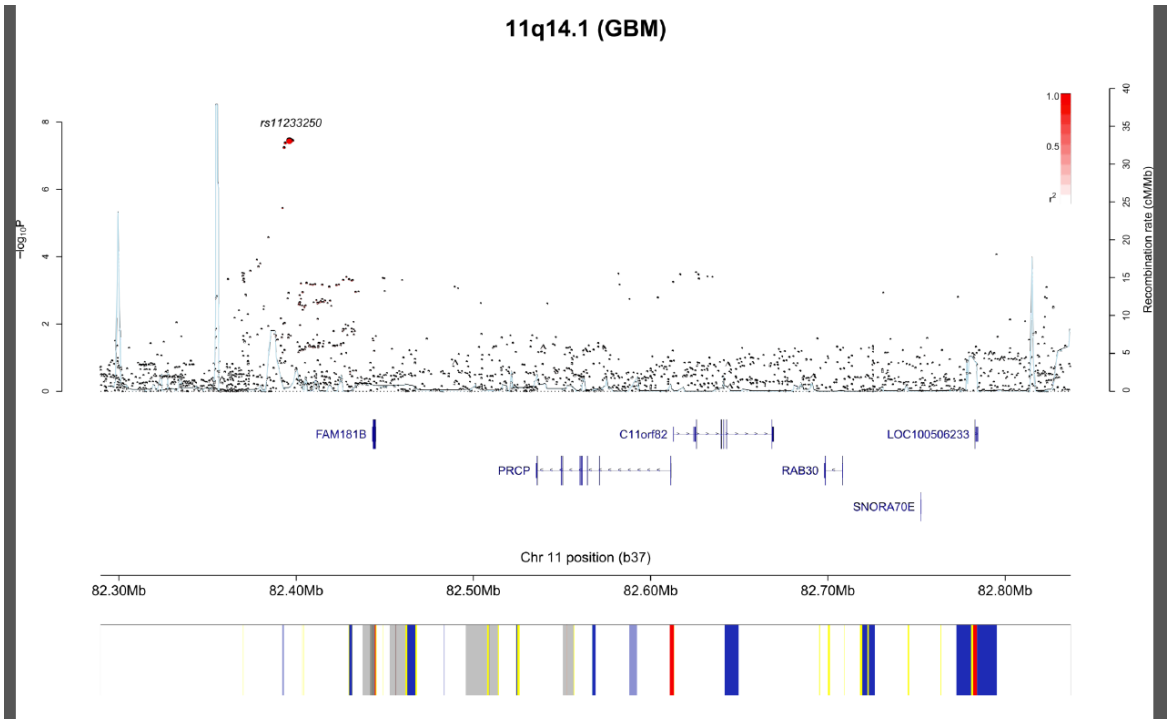
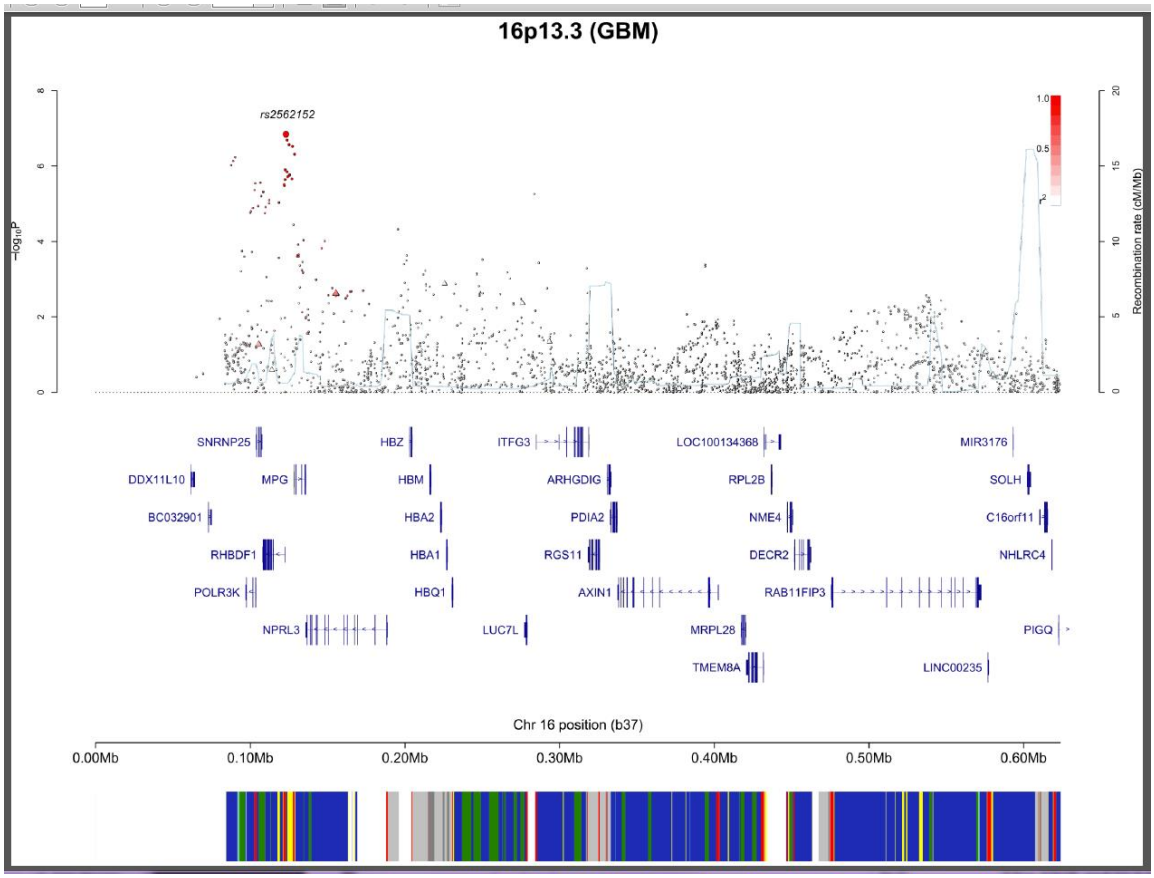


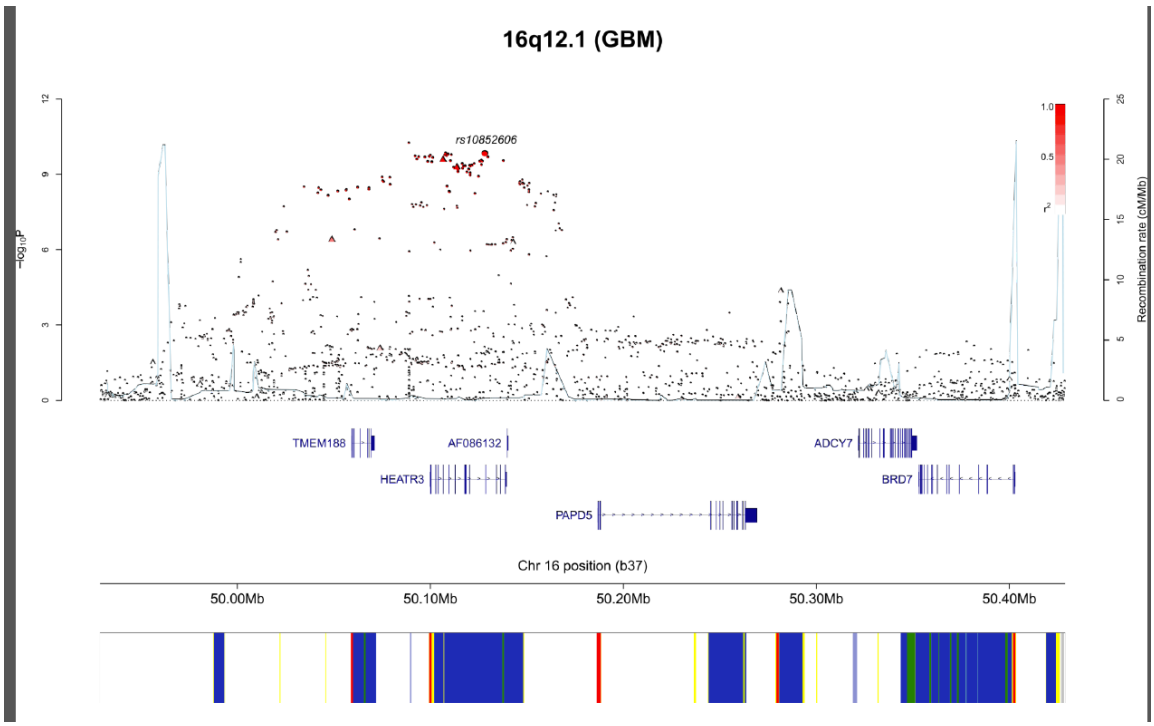
Figure 3: Regional plots of discovery-phase association results, recombination rates and chromatin state segmentation tracks for the new glioma risk loci. Plots show discovery association results of both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates. −log10 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 labeled by its rsID. The color intensity of each symbol reflects the extent of LD with the top genotyped SNP, white (r2 = 0) through 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 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 (f) – Legend file depicting chromatin states for chromHMM track.

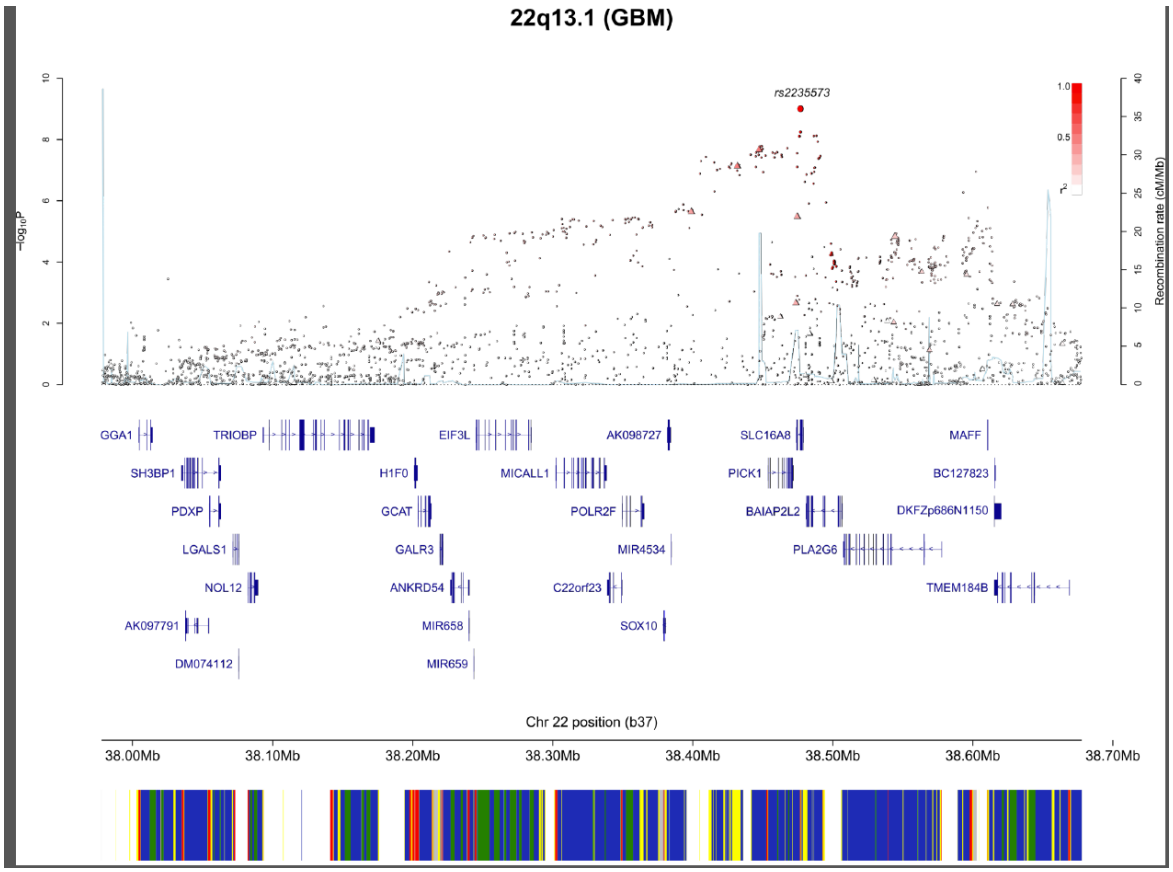


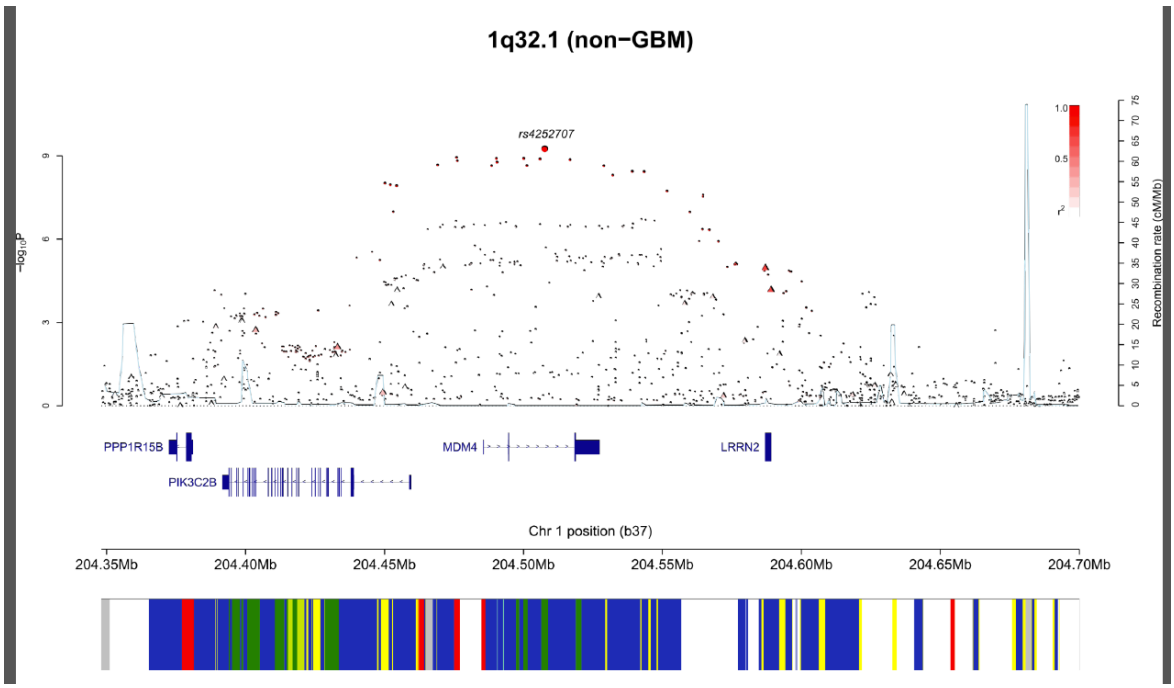

9q21.13 (GBM)
rs11143912
1.0
0.5
r^2
$-\log_{10}P$
Recombination rate (cM/Mb)
Chr 9 position (b37)
76.80Mb
76.85Mb
76.90Mb
76.95Mb
77.00Mb

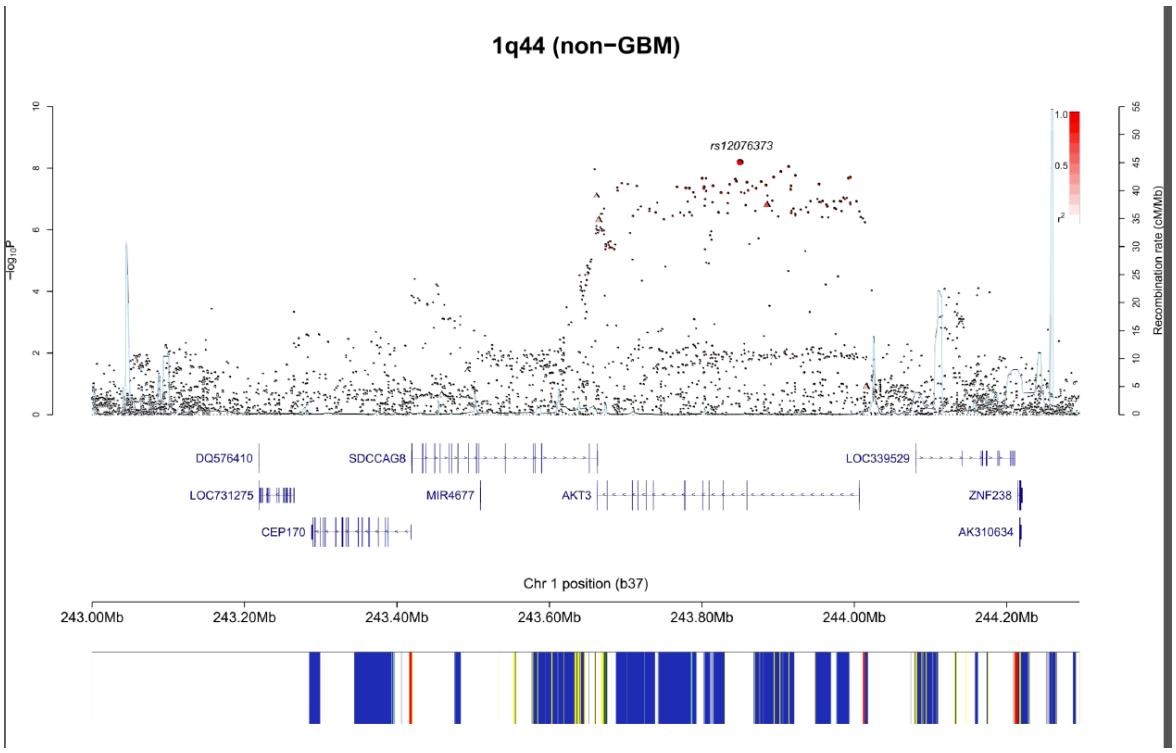

11q14.1 (GBM)
rs11233250
1.0
0.5
r^2
$-\log_{10}P$
Recombination rate (cM/Mb)
FAM181B
C11orf82
LOC100506233
PRCP
RAB30
SNORA70E
Chr 11 position (b37)
82.30Mb
82.40Mb
82.50Mb
82.60Mb
82.70Mb
82.80Mb

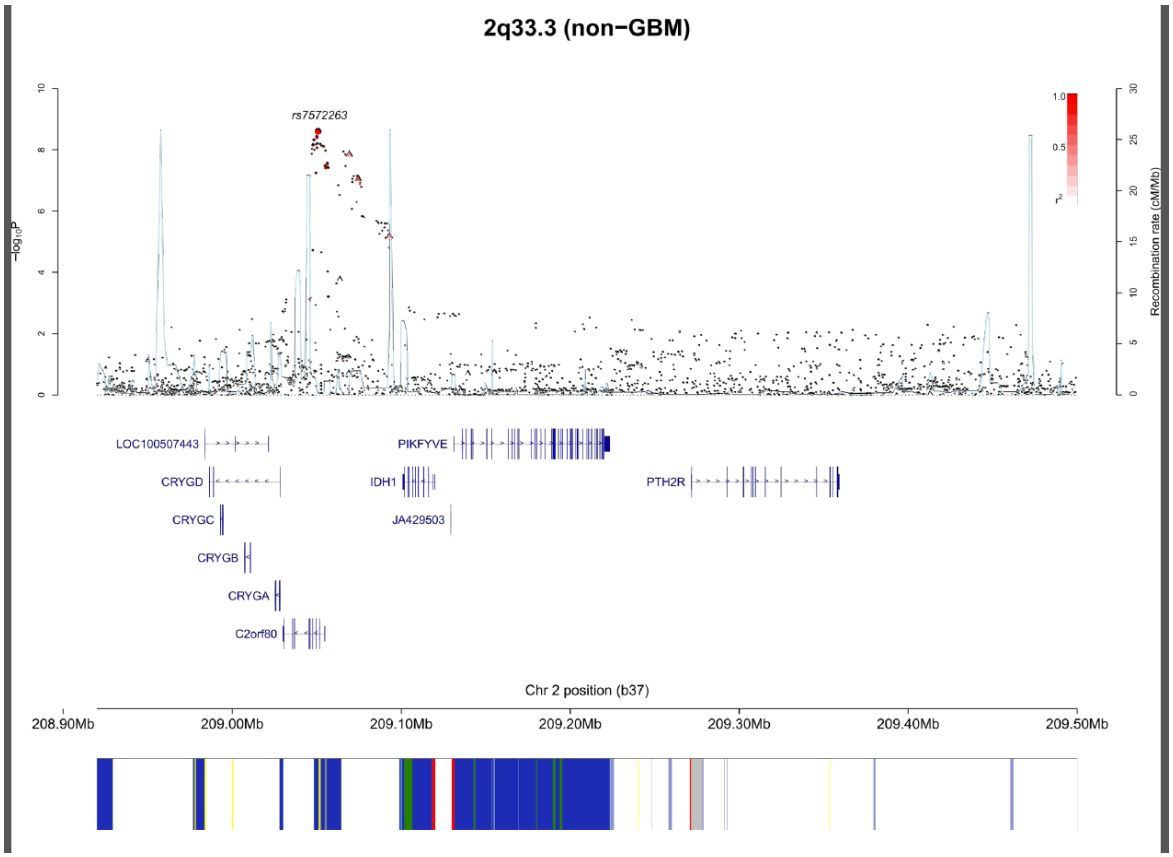

16p13.3 (GBM)
rs2562152
−log10P
Recombination rate (cM/Mb)
r2
SNRNP25
HBZ
ITFG3
LOC100134368
MIR3176
DDX11L10
MPG
HBM
ARHGDIG
RPL2B
SOLH
BC032901
HBA2
PDIA2
NME4
C16orf11
RHBDF1
HBA1
RGS11
DECR2
NHLRC4
POLR3K
HBQ1
AXIN1
RAB11FIP3
NPRL3
LUC7L
MRPL28
PIGQ
TMEM8A
LINC00235
Chr 16 position (b37)
0.00Mb
0.10Mb
0.20Mb
0.30Mb
0.40Mb
0.50Mb
0.60Mb

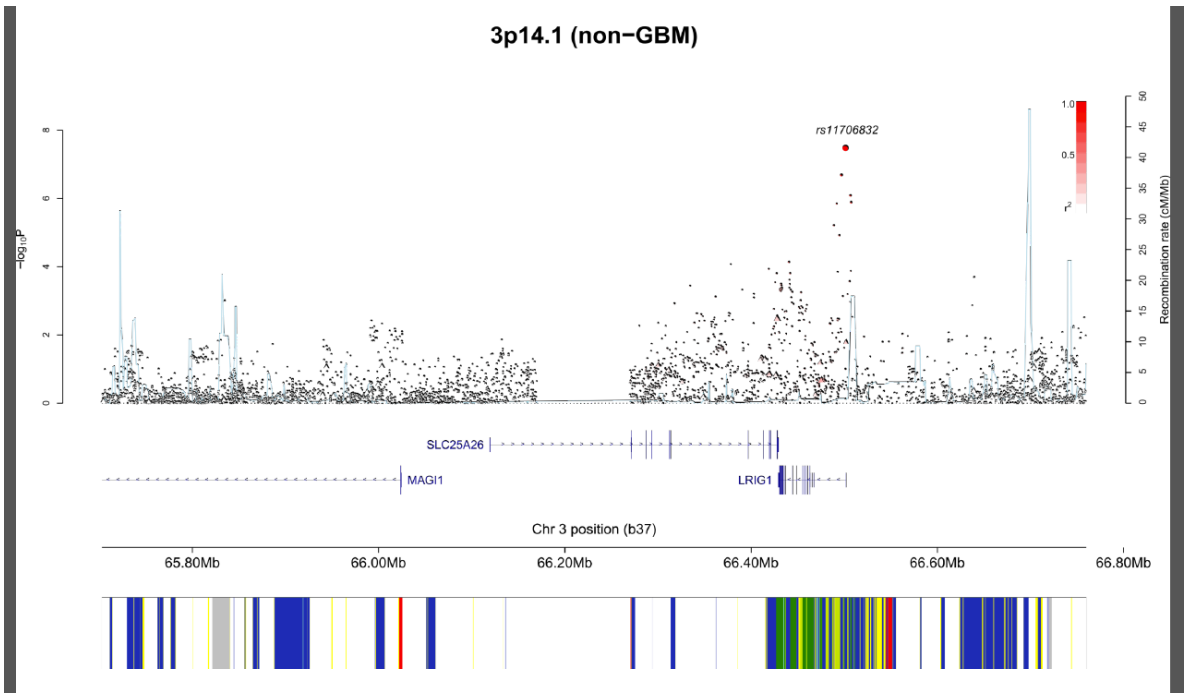

16q12.1 (GBM)
rs10852606
−log10P
Recombination rate (cM/Mb)
r2
TMEM188
AF086132
ADCY7
HEATR3
BRD7
PAPD5
Chr 16 position (b37)
50.00Mb
50.10Mb
50.20Mb
50.30Mb
50.40Mb

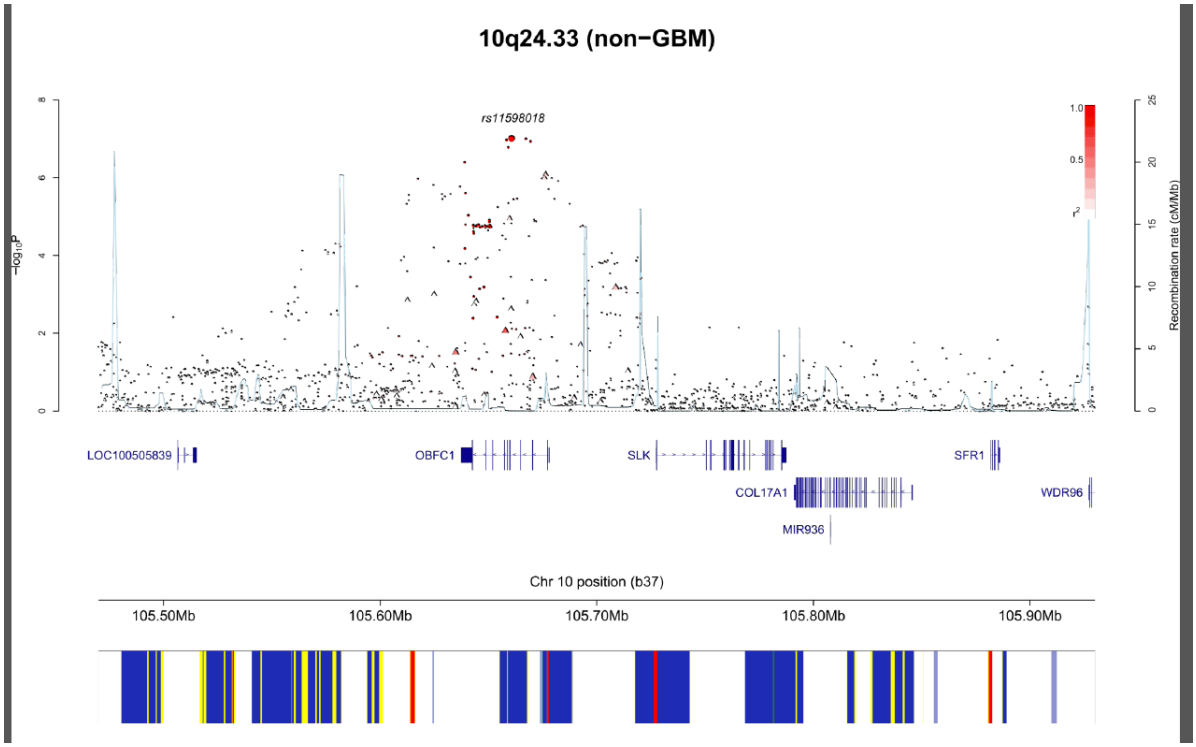

22q13.1 (GBM)
rs2235573
−log10P
Recombination rate (cM/Mb)
r2
GGA1
TRIOBP
EIF3L
AK098727
SLC16A8
MAFF
SH3BP1
H1F0
MICALL1
PICK1
BC127823
PDXP
GCAT
POLR2F
BAIAP2L2
DKFZp686N1150
LGALS1
GALR3
MIR4534
PLA2G6
NOL12
ANKRD54
C22orf23
TMEM184B
AK097791
MIR658
SOX10
DM074112
MIR659
Chr 22 position (b37)
38.00Mb
38.10Mb
38.20Mb
38.30Mb
38.40Mb
38.50Mb
38.60Mb
38.70Mb

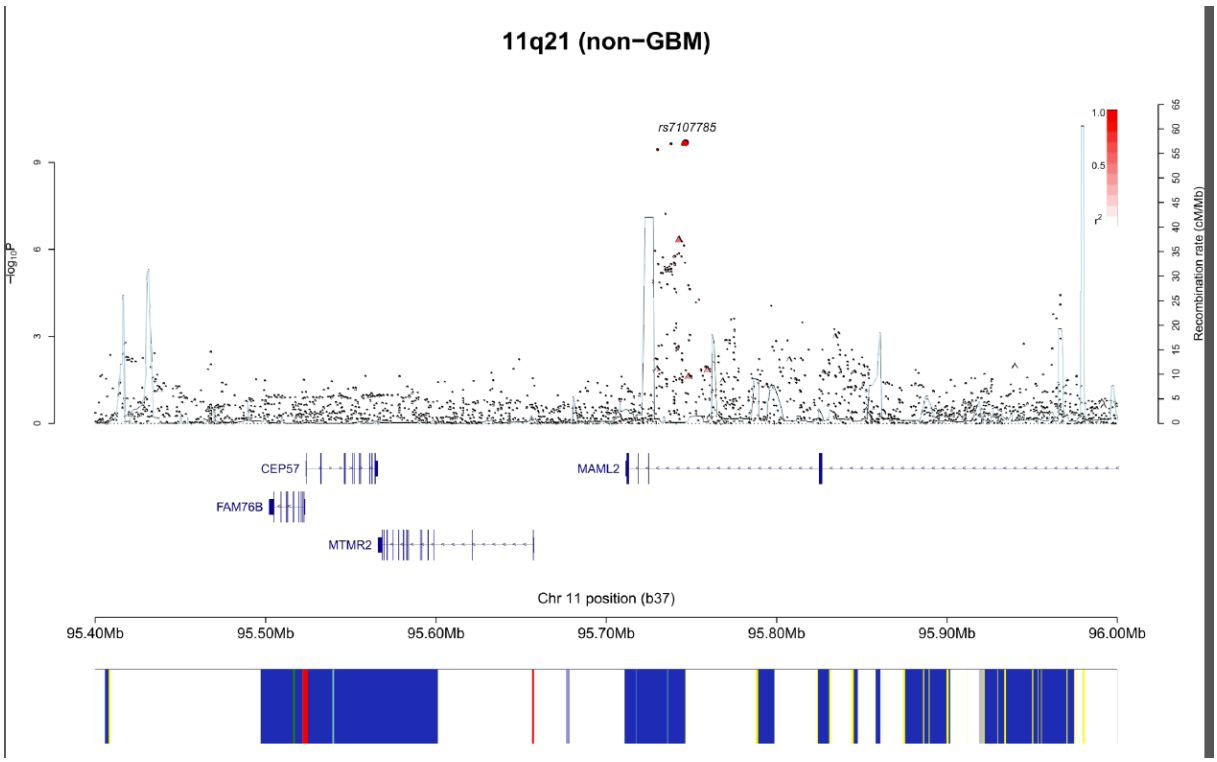

1q32.1 (non−GBM)
rs4252707
−log10P
Recombination rate (cM/Mb)
r2
PPP1R15B
MDM4
LRRN2
PIK3C2B
Chr 1 position (b37)
204.35Mb
204.40Mb
204.45Mb
204.50Mb
204.55Mb
204.60Mb
204.65Mb
204.70Mb

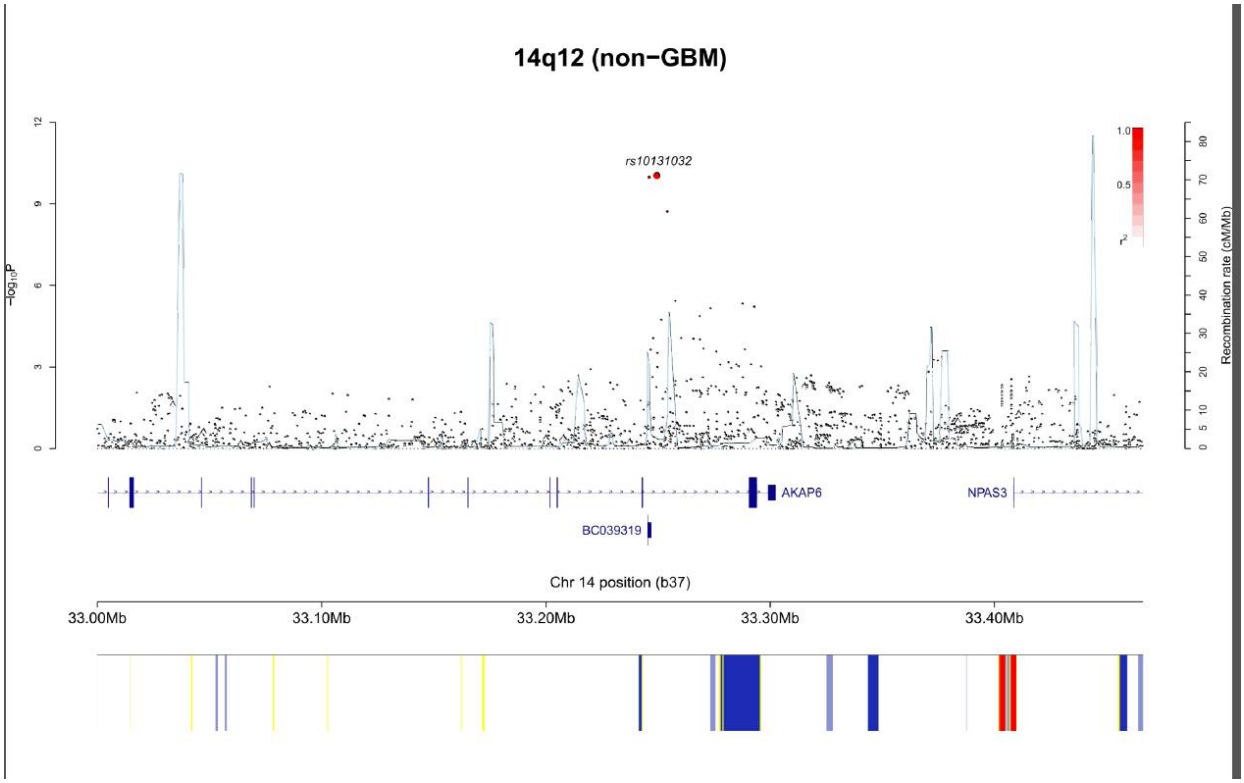

1q44 (non−GBM)
rs12076373
−log10P
Recombination rate (cM/Mb)
r2
DQ576410
SDCCAG8
LOC339529
LOC731275
MIR4677
AKT3
ZNF238
CEP170
AK310634
Chr 1 position (b37)
243.00Mb
243.20Mb
243.40Mb
243.60Mb
243.80Mb
244.00Mb
244.20Mb


2q33.3 (non−GBM)
rs7572263
−log10P
Recombination rate (cM/Mb)
r2
LOC100507443
PIKFYVE
CRYGD
IDH1
PTH2R
CRYGC
JA429503
CRYGB
CRYGA
C2orf80
Chr 2 position (b37)
208.90Mb
209.00Mb
209.10Mb
209.20Mb
209.30Mb
209.40Mb
209.50Mb


3p14.1 (non−GBM)
rs11706832
−log10P
Recombination rate (cM/Mb)
r2
SLC25A26
MAGI1
LRIG1
Chr 3 position (b37)
65.80Mb
66.00Mb
66.20Mb
66.40Mb
66.60Mb
66.80Mb


10q24.33 (non−GBM)
rs11598018
−log10P
Recombination rate (cM/Mb)
r2
LOC100505839
OBFC1
SLK
SFR1
COL17A1
WDR96
MIR936
Chr 10 position (b37)
105.50Mb
105.60Mb
105.70Mb
105.80Mb
105.90Mb


11q21 (non−GBM)
rs7107785
−log10P
Recombination rate (cM/Mb)
r2
CEP57
MAML2
FAM76B
MTMR2
Chr 11 position (b37)
95.40Mb
95.50Mb
95.60Mb
95.70Mb
95.80Mb
95.90Mb
96.00Mb


14q12 (non−GBM)
rs10131032
−log10P
Recombination rate (cM/Mb)
r2
AKAP6
NPAS3
BC039319
Chr 14 position (b37)
33.00Mb
33.10Mb
33.20Mb
33.30Mb
33.40Mb

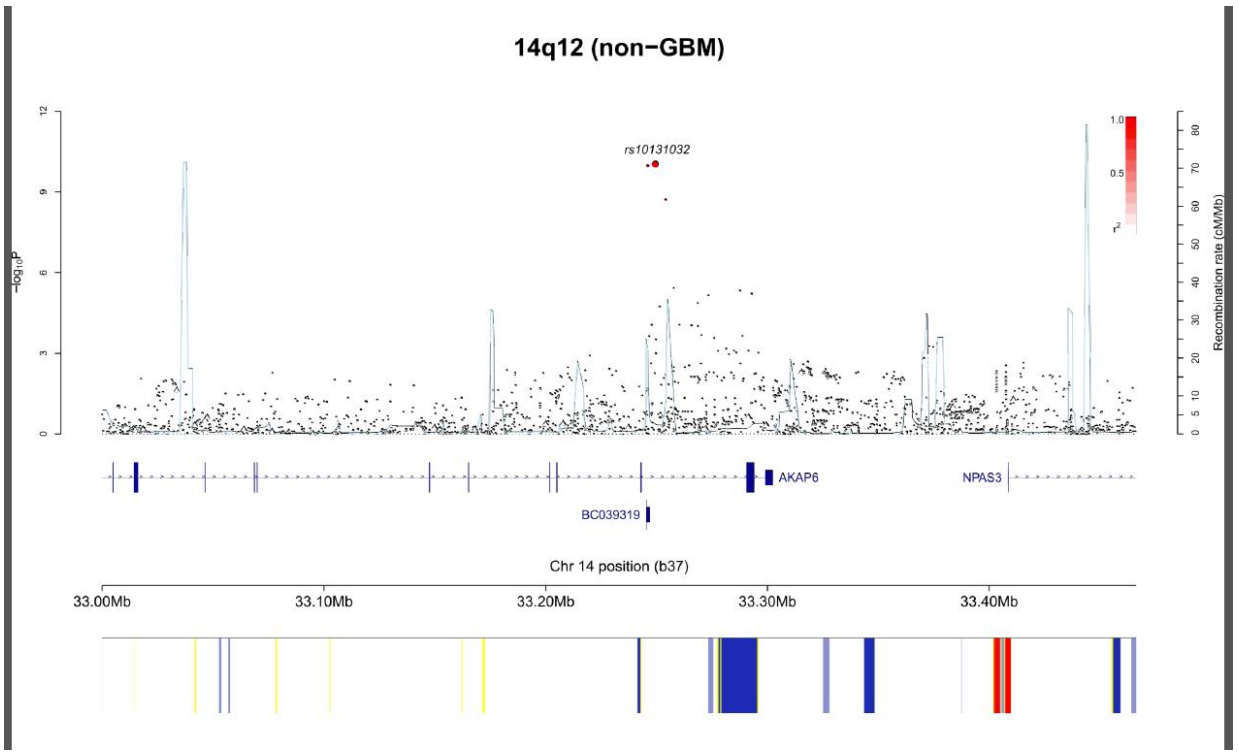

14q12 (non−GBM)
rs10131032
1.0
0.5
r2
−log10P
Recombination rate (cM/Mb)
AKAP6
NPAS3
BC039319
Chr 14 position (b37)
33.00Mb
33.10Mb
33.20Mb
33.30Mb
33.40Mb

Figure 4: 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.

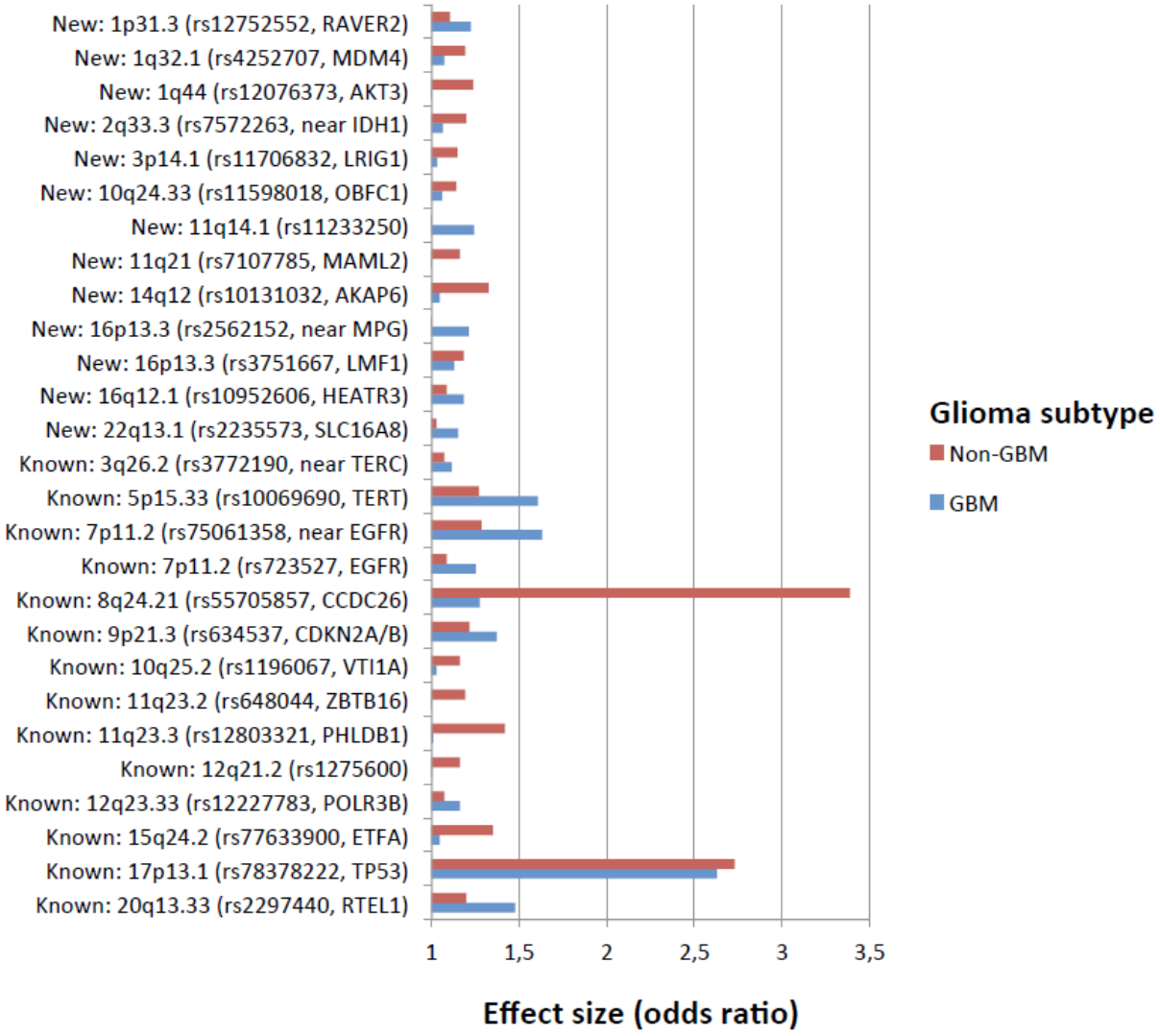


Table 1: Association statistics for top SNP at each of the newly-reported glioma risk loci. Associations at $P<5x10^{-8}$ are highlighted in bold. Odds ratios were derived with respect to the risk allele underlined and highlighted in bold. Minor allele frequency (MAF) is according to European samples from 1000 genomes project. The INFO column indicates the average imputation info score across the studies in the discovery phase, with a score of 1 indicating the SNP is directly genotyped in all studies.

							All glioma		GBM glioma		Non-GBM glioma	
Locus	**Subtype**	**SNP**	**Position**	**Alleles**	**RAF**	**INFO**	***P***	**Odds ratio**	***P***	**Odds ratio**	***P***	**Odds ratio**
1p31.3	GBM	rs12752552	65229299	**T**/C	0.870	0.994	**$4.07x10^{-9}$**	1.18 (1.11-1.24)	**$2.04x10^{-9}$**	1.22 (1.15-1.31)	$4.78x10^{-3}$	1.11 (1.03-1.18)
1q32.1	non-GBM	rs4252707	204508147	G/**A**	0.220	0.990	$2.97x10^{-7}$	1.12 (1.07-1.17)	0.015	1.07 (1.01-1.13)	**$3.34x10^{-9}$**	1.19 (1.12-1.26)
1q44	non-GBM	rs12076373	243851947	**G**/C	0.837	0.992	$4.97x10^{-4}$	1.09 (1.04-1.15)	0.846	0.99 (0.94-1.06)	**$2.63x10^{-10}$**	1.23 (1.16-1.32)
2q33.3	non-GBM	rs7572263	209051586	**A**/G	0.756	0.998	$2.58x10^{-6}$	1.11 (1.06-1.15)	0.019	1.06 (1.01-1.12)	**$2.18x10^{-10}$**	1.20 (1.13-1.26)
3p14.1	non-GBM	rs11706832	66502981	A/**C**	0.456	0.991	$1.06x10^{-5}$	1.08 (1.05-1.12)	0.158	1.03 (0.99-1.08)	**$7.66x10^{-9}$**	1.15 (1.09-1.20)
10q24.33	non-GBM	rs11598018	105661315	**C**/A	0.462	0.957	$3.07x10^{-7}$	1.10 (1.06-1.14)	0.0103	1.06 (1.01-1.11)	**$3.39x10^{-8}$**	1.14 (1.09-1.20)
11q14.1	GBM	rs11233250	82397014	**C**/T	0.868	0.988	$5.40x10^{-6}$	1.14 (1.08-1.21)	**$9.95x10^{-10}$**	1.24 (1.16-1.33)	0.592	0.98 (0.91-1.05)
11q21	non-GBM	rs7107785	95747337	**T**/C	0.479	0.995	$2.96x10^{-4}$	1.07 (1.03-1.11)	0.844	1.00 (0.95-1.04)	**$3.87x10^{-10}$**	1.16 (1.11-1.21)
14q12	non-GBM	rs10131032	33250081	**G**/A	0.916	0.996	$2.33x10^{-6}$	1.17 (1.09-1.24)	0.247	1.05 (0.97-1.13)	**$5.07x10^{-11}$**	1.33 (1.22-1.44)
16p13.3	GBM	rs2562152	123896	A/**T**	0.850	0.930	$1.18x10^{-3}$	1.09 (1.04-1.15)	**$1.93x10^{-8}$**	1.21 (1.13-1.29)	0.948	1.00 (0.93-1.07)
16p13.3	non-GBM	rs3751667	1004554	C/**T**	0.208	0.992	**$8.75x10^{-10}$**	1.14 (1.09-1.19)	$5.95x10^{-6}$	1.13 (1.07-1.19)	**$2.61x10^{-9}$**	1.18 (1.12-1.25)
16q12.1	GBM	rs10852606	50128872	T/**C**	0.713	0.993	**$3.66x10^{-11}$**	1.14 (1.10-1.19)	**$1.29x10^{-11}$**	1.18 (1.13-1.24)	$2.42x10^{-3}$	1.08 (1.03-1.14)
22q13.1	GBM	rs2235573	38477930	**G**/A	0.507	0.995	$8.64x10^{-7}$	1.09 (1.06-1.13)	**$1.76x10^{-10}$**	1.15 (1.10-1.20)	0.325	1.02 (0.97-1.07)

REFERENCES

1. Bondy, M.L. *et al.* Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium. *Cancer* **113**, 1953-68 (2008).
2. Ostrom, Q.T. *et al.* CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro Oncol* **17 Suppl 4**, iv1-iv62 (2015).
3. Louis, D.N. *et al.* The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* **114**, 97-109 (2007).
4. Ostrom, Q.T. *et al.* The epidemiology of glioma in adults: a "state of the science" review. *Neuro Oncol* **16**, 896-913 (2014).
5. 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**, 481-8 (2009).
6. Sanson, M. *et al.* Chromosome 7p11.2 (EGFR) variation influences glioma risk. *Hum Mol Genet* **20**, 2897-904 (2011).
7. Shete, S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* **41**, 899-904 (2009).
8. Wrensch, M. *et al.* Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat Genet* **41**, 905-8 (2009).
9. Kinnersley, B. *et al.* Genome-wide association study identifies multiple susceptibility loci for glioma. *Nat Commun* **6**, 8559 (2015).
10. Walsh, K.M. *et al.* Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk. *Nat Genet* **46**, 731-5 (2014).
11. Enciso-Mora, V. *et al.* Low penetrance susceptibility to glioma is caused by the TP53 variant rs78378222. *Br J Cancer* **108**, 2178-85 (2013).
12. Jenkins, R.B. *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* **44**, 1122-5 (2012).
13. Kinnersley, B. *et al.* Quantifying the heritability of glioma using genome-wide complex trait analysis. *Sci Rep* **5**, 17267 (2015).
14. Rajaraman, P. *et al.* Genome-wide association study of glioma and meta-analysis. *Hum Genet* **131**, 1877-88 (2012).
15. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
16. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* **6**, 8111 (2015).
17. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
18. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-61 (2014).
19. Riemenschneider, M.J. *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* **59**, 6091-6 (1999).
20. 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* **81**, 292-303 (2007).

21. Turner, K.M. *et al.* Genomically amplified Akt3 activates DNA repair pathway and promotes glioma progression. *Proc Natl Acad Sci U S A* **112**, 3421-6 (2015).
22. Gur, G. *et al.* LRIG1 restricts growth factor signaling by enhancing receptor ubiquitylation and degradation. *EMBO J* **23**, 3270-81 (2004).
23. Yang, J.A. *et al.* LRIG1 enhances the radiosensitivity of radioresistant human glioblastoma U251 cells via attenuation of the EGFR/Akt signaling pathway. *Int J Clin Exp Pathol* **8**, 3580-90 (2015).
24. Wei, J. *et al.* miR-20a mediates temozolomide-resistance in glioblastoma cells via negatively regulating LRIG1 expression. *Biomed Pharmacother* **71**, 112-8 (2015).
25. Watanabe, T., Nobusawa, S., Kleihues, P. & Ohgaki, H. IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. *Am J Pathol* **174**, 1149-53 (2009).
26. Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-73 (2009).
27. Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510-22 (2010).
28. Christensen, B.C. *et al.* DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma. *J Natl Cancer Inst* **103**, 143-53 (2011).
29. Sanson, M. *et al.* Isocitrate dehydrogenase 1 codon 132 mutation is an important prognostic biomarker in gliomas. *J Clin Oncol* **27**, 4150-4 (2009).
30. Eckel-Passow, J.E. *et al.* Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. *N Engl J Med* **372**, 2499-508 (2015).
31. Walsh, K.M. *et al.* Telomere maintenance and the etiology of adult glioma. *Neuro Oncol* **17**, 1445-52 (2015).
32. Zhang, C. *et al.* Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. *Hum Mol Genet* **24**, 5356-66 (2015).
33. Walsh, K.M. *et al.* Longer genotypically-estimated leukocyte telomere length is associated with increased adult glioma risk. *Oncotarget* **6**, 42468-77 (2015).
34. Bainbridge, M.N. *et al.* Germline mutations in shelterin complex genes are associated with familial glioma. *J Natl Cancer Inst* **107**, 384 (2015).
35. Xipell, E. *et al.* Endoplasmic reticulum stress-inducing drugs sensitize glioma cells to temozolomide through downregulation of MGMT, MPG, and Rad51. *Neuro Oncol* (2016).
36. Amirian, E.S. *et al.* The Glioma International Case-Control Study: A Report From the Genetic Epidemiology of Glioma International Consortium. *Am J Epidemiol* **183**, 85-91 (2016).
37. Cardis, E. *et al.* The INTERPHONE study: design, epidemiological methods, and description of the study population. *Eur J Epidemiol* **22**, 647-64 (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**, 2335-42 (2004).
39. Wichmann, H.E., Gieger, C., Illig, T. & Group, M.K.S. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, S26-30 (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. 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* **144**, 212-8 (2002).
42. 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**, 870-4 (2007).
43. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* **39**, 645-9 (2007).

44. Wiemels, J.L. *et al.* History of allergies among adults with glioma and controls. *Int J Cancer* **98**, 609-15 (2002).
45. Felini, M.J. *et al.* Reproductive factors and hormone use and risk of adult gliomas. *Cancer Causes Control* **20**, 87-96 (2009).
46. 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).
47. 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**, 906-13 (2007).
48. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
49. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
50. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
51. Jenkins, R.B. *et al.* Distinct germ line polymorphisms underlie glioma morphologic heterogeneity. *Cancer Genet* **204**, 13-8 (2011).
52. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59 (2000).
53. Liu, J.Z. *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* **42**, 436-40 (2010).
54. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557-60 (2003).
55. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
56. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
57. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).
58. 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**, D930-4 (2012).
59. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272-6 (2009).
60. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
61. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
62. 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).
63. 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**, 2796-803 (2008).
64. Scheurer, M.E. *et al.* Familial aggregation of glioma: a pooled analysis. *Am J Epidemiol* **172**, 1099-107 (2010).
65. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
66. Scales, M., Jager, R., Migliorini, G., Houlston, R.S. & Henrion, M.Y. visPIG--a web tool for producing multi-region, multi-track, multi-scale plots of genetic data. *PLoS One* **9**, e107497 (2014).
67. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225-9 (2002).