Chromosome 7p11.2 (EGFR) variation influences glioma risk

Marc Sanson1,2,†, Fay J. Hosking3,†, Sanjay Shete5,†, Diana Zelenka6,†, Sara E. Dobbins3, Yussanne Ma3, Victor Enciso-Mora3, Ahmed Idbaih1,2, Jean-Yves Delattre1,2, Khe Hoang-Xuan1,2, Yannick Marie1, Blandine Boisselier1, Catherine Carpentier1, Xiao-Wei Wang1, Anna Luisa Di Stefano1, Marianne Labussière1, Konstantinos Gousias7, Johannes Schramm7, Anne Boland6, Doris Lechner6, Ivo Gut6, Georgina Armstrong5, Yanhong Liu5, Robert Yu5, Ching Lau8, Maria Chiara Di Bernardo3, Lindsay B. Robertson3, Kenneth Muir9, Sarah Hepworth10, Anthony Swerdlow4, Minouk J Schoemaker4, H.-Erich Wichmann11,12,13, Martina Müller11,12,14, Stefan Schreiber15,16, Andre Franke16, Susanne Moebus17, Lewin Eisele18, Asta Förstl19, Kari Hemminki19, Mark Lathrop6,20, Melissa Bondy5,‡, Richard S. Houlston3,*,‡ and Matthias Simon7,‡

1Université Pierre et Marie Curie-Paris 6, Centre de Recherche de l’Institut du Cerveau et de la Moelle épinière (CRICM) UMR-S975, GH Pitié-Salpêtrière, 47 bd de l’Hôpital, 75651 Paris, France, 2AP-HP, GH Pitié-Salpêtrière, Service de Neurologie Mazarin, 47 bd de l’Hôpital, 75651 Paris, France, 3Section of Cancer Genetics, 4Section of Epidemiology, Institute of Cancer Research, 5Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, PO Box 301439, Houston, TX 77230, USA, 6Centre National de Génotypage, IG/CEA, 2 rue Gaston Crémieux, CP 5721, 91057 Evry, France, 7Neurochirurgische Universitätsklinik, Universitätskliniken Bonn, Sigmund-Freud-Str 25, 53105 Bonn, Germany, 8Texas Children’s Cancer Center, Baylor College of Medicine, 6701 Fannin St., Ste. 142, Houston, TX 77030, USA, 9Health Sciences Research Institute, Warwick Medical School, University of Warwick, Gibbett Hill Road, Coventry CV4 7AL, UK, 10Centre for Epidemiology and Biostatistics, Faculty of Medicine and Health, University of Leeds, Leeds LS2 9JT, UK, 11Institute of Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany, 12Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 Munich, Germany, 13Klinikum Grosshadern and 14Department of Medicine I, Klinikum Grosshadern, Ludwig-Maximilians-Universität Marchioninistraße 15, 81377 Munich, Germany, 151st Medical Department, University Clinic Schleswig-Holstein, Campus Kiel, House 6, Arnold-Heller-Str. 3, 24105 Kiel, Germany, 16Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Arnold-Heller-Straße 3, 24105 Kiel, Germany, 17Institute for Medical Informatics, Biometry and Epidemiology and 18Department of Haematology, University Hospital of Essen, University Duisburg-Essen, Hufelandstr. 55, 45122 Essen, Germany, 19Division of Molecular Genetic Epidemiology, German Cancer Research Centre (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany, 20Foundation Jean Dausset-CEPH, 27 Rue Juliette Dodu, 75010 Paris, France

Received January 20, 2011; Revised April 13, 2011; Accepted April 26, 2011
While gliomas are the most common primary brain tumors, their etiology is largely unknown. To identify novel risk loci for glioma, we conducted genome-wide association (GWA) analysis of two case–control series from France and Germany (2269 cases and 2500 controls). Pooling these data with previously reported UK and US GWA studies provided data on 4147 glioma cases and 7435 controls genotyped for 424 460 common tagging single-nucleotide polymorphisms. Using these data, we demonstrate two statistically independent associations between glioma and rs11979158 and rs2252586, at 7p11.2 which encompasses the EGFR gene (population-corrected statistics, $P = 7.72 \times 10^{-8}$ and $2.09 \times 10^{-8}$, respectively). Both associations were independent of tumor subtype, and were independent of EGFR amplification, $p16INK4a$ deletion and IDH1 mutation status in tumors; compatible with driver effects of the variants on glioma development. These findings show that variation in 7p11.2 is a determinant of inherited glioma risk.

INTRODUCTION

Gliomas account for ~40% of all primary brain tumors and are responsible for around 13 000 deaths in the USA each year. Gliomas are heterogeneous and different tumor subtypes defined in part by malignancy grade [e.g. pilocytic astrocytoma WHO grade I, diffuse ‘low grade’ glioma WHO grade II, anaplastic glioma WHO grade III and glioblastoma (GBM) WHO grade IV] can be distinguished. Most gliomas are associated with a poor prognosis irrespective of clinical care, with the most common type of glioma, GBM, having a median overall survival of 10–15 months (1–3).

While glioma subtypes have distinct molecular profiles resulting from different etiological pathways, no lifestyle exposure have, however, consistently been linked to glioma risk except for ionizing radiation, which only accounts for a very small number of cases (4). 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 neurofibromatosis (4,5). However, collectively, these diseases account for little of the 2-fold increased risk of glioma seen in first-degree relatives of glioma patients (6), and much of the excess familial risk is likely to be a consequence of the co-inheritance of multiple low-risk variants.

To search for genetic risk variants influencing glioma, we have previously conducted genome-wide association (GWA) studies of UK and US glioma patients with replication of the most significantly associated single-nucleotide polymorphisms (SNPs) in independent case–control series from France, Germany and Sweden. This analysis robustly demonstrated that common variants mapping to 5p15.33 ($TERT$), 8q24.21 ($CCDC26$), 9p21.3 ($CDKN2A/CDKN2B$), 20q13.33 ($RTEL1$) and 11q23.3 ($PHLDB1$) contribute toheritable risk of glioma (7). Confirmation of the 9p21.3 and 20q13.33 loci as risk factors for GBM was provided by a contemporaneous study (8).

GWA studies are not contingent on prior information concerning candidate genes or pathways, and thereby have the ability to identify important variants in hitherto unstudied genes. However, the effect sizes of individual variants and the need for stringent thresholds for establishing statistical significance inevitably constrain study power. To increase our power to identify novel genetic risk loci for glioma, we have performed GWA scans on augmented series based on the aforementioned French and German cases and combined these data with our GWA scans of US and UK glioma patients. This analysis identifies genetic variation at 7p11.2 as risk factors for the development of glioma and provides further insight into the biological basis of tumor development.

RESULTS

The four GWA studies collectively provided data on 4147 glioma cases and 7435 controls genotyped for 424 460 common tagging SNPs. As the quantile–quantile (Q–Q) plots for the German and US series showed some evidence of inflation [inflation factor $\lambda = 1.16$ and 1.11, respectively, based on the 90% least significant SNPs (9)], we corrected for population substructure using principal-components analyses as implemented in Eigenstrat (Supplementary Material, Fig. S1). This analysis reduced all four inflation factors to <1.05.

Meta-analyses of these data resulted in 37 SNPs showing an association with glioma risk at $P < 5.0 \times 10^{-7}$; 35 of these map to the five loci we have previously shown (7) to influence disease risk namely, 5p15.33, 8q24.21, 9p21.3, 20q13.33 and 11q23.3 (Supplementary Material, Table S1). The strongest evidence for a novel association was provided by two SNPs mapping to 7p11.2.

The two SNPs on 7p11.2, rs11979158 and rs2252586, mapping to 55 126 843 and 54 946 418 bp, yielded $P$-values of $7.03 \times 10^{-8}$ and $7.89 \times 10^{-8}$, respectively; Table 1, Fig. 1). For completeness, we also derived population-corrected $P$-values using PLINK software ($P$-values for rs11979158 and rs2252586 were $3.11 \times 10^{-7}$ and $4.55 \times 10^{-7}$), although only one of the associations when corrected for population stratification attained the conventional threshold for genome-wide significance (i.e. $P \leq 5.0 \times 10^{-8}$), both associations were consistent across the four GWA studies ($P_{het} = 0.44$ and 0.27, respectively; Table 1). The SNP rs11979158 maps within intron 1 of the Epidermal Growth Factor Receptor 1 ($EGFR$) gene and rs2252586 lies 107 kb telomeric from $EGFR$ (Fig. 1). Several lines of evidence support the two SNPs defining independent disease loci. First, there is low linkage disequilibrium (LD) between the two SNPs ($D' = 0.62$, $r^2 = 0.03$).
Secondly, adjusting rs11979158 through conditional logistic regression for rs2252586 provided evidence of an association (OR = 1.20, 95% CI: 1.10–1.30; \( P = 1.36 \times 10^{-5} \)). Similarly, adjusting rs2252586 for rs11979158 also provided evidence of an association (OR = 1.14, 95% CI: 1.07–1.22; \( P = 2.04 \times 10^{-5} \)). Thirdly, there was also an increasing trend in OR with an increasing number of risk alleles (Fig. 2; Supplementary Material, Table S2). To examine the possibility that these SNPs may be correlated with an untyped variant, we made use of HapMap3 and 1000 Genomes data to impute additional variants localizing to 54 850 000–55 200 000 bp region. This analysis provided no SNPs with significantly superior evidence for an association and provides further support for the existence of two independent risk loci at 7p11.2 (Supplementary Material, Fig. S2).

We found no evidence for a pairwise interaction between rs2252586, rs11979158 and the five previously identified risk variants for glioma at 5p15.33 (rs2736100), 8q24.21 (rs4295627), 9p21.3 (rs4977756), 20q13.33 (rs6010620) and 11q23.3 (rs498872) (\( P > 0.1 \) after correction for multiple testing; Supplementary Material, Table S3). These data are consistent with each variant having an independent role in defining glioma risk whereby the risk increases with increasing number of variant risk alleles (OR\(_{\text{per allele}} = 1.24, 95\% \text{ CI: } 1.21–1.27, P = 2.89 \times 10^{-72}\); Supplementary Material, Table S2).

To examine whether the implicated SNPs at 7p11.2 loci were differentially associated with tumor subgroup, we analyzed their prevalence according to histology across all cases where such data were available (\( n = 4002 \)). The carrier frequencies of the risk alleles for the 7p11.2 SNPs showed no correlation with stratification of tumors by WHO grade in any of the four patient cohorts (Supplementary Material, Table S4). These observations contrast with the strong relationship between genotype and histology seen for 5p15.33 (\( TERT;\) rs2736100), 8q24.21 (\( CCDC26;\) rs4295627),
20q13.33 (RTEL1; rs6010620) and 11q23.3 (PHLDB1; rs498872) risk variants (Fig. 3).

Figure 2. Cumulative effect of glioma risk alleles. Plot of increasing ORs for glioma for increasing number of risk alleles. The ORs are relative to the median number of eight risk alleles; vertical bars correspond to 95% CIs. Horizontal line marks the null value (OR = 1.0).

Figure 3. Relationship between 7p11.2 genotype and WHO tumor grade. Also shown are the relationships between histology and 5p15.33 (rs2736100), 8q24.21 (rs4295627), 9p21.3 (rs4977756), 20q13.33 (rs6010620) and 11q23.3 (rs498872) genotypes. *Significant association at P < 0.05.

DISCUSSION

This analysis has shown that the risk of developing glioma is influenced by genetic variation at 7p11.2. Our data support a previous report of an association between rs11979158 and GBM risk based on analysis of a smaller data set (12). These associations are a priori biologically plausible, since EGFR is the prototypical member of the ErbB/EGFR family, which is involved in multiple cellular processes, including cell division, migration, adhesion, differentiation and apoptosis. EGFR is well established to be pivotal in both initiation of primary GBM and progression of lower-grade glioma to grade IV (13); hence genetic variation in EGFR is an attractive basis for predisposition to glioma. Furthermore, the Cancer Genome Atlas (TCGA) consortium reported EGFR as the fourth most highly mutated gene in a compendium of common cancer genes sequenced in GBM (14). While EGFR amplification characterizes 40–70% of primary GBM, it is rarely seen in lower-grade astrocytomas, thereby implicating EGFR activation as a driver of glial tumorigenesis (13). There are multiple mechanisms by which EGFR mediates tumor initiation and progression, all of which are seen in primary glioma. Increased EGFR is common in primary GBM as a consequence of gene amplification and is often associated with exon 2–7 truncation resulting in constitutive receptor activation (EGFRVIII), or less frequently, an activating missense mutation of the EGFR extracellular domain (15–17). While speculative at this juncture, it is entirely plausible that the influence of SNP-mediated risk on glioma development is a consequence of a subtle effect on EGFR expression, thereby having the potential to impact on all histological forms of glioma.

We have previously shown that the TERT and RTEL1 risk variants are primarily associated with high-grade disease and CCDC26 and PHLDB1 with the development of less aggressive glial tumors (18). Moreover, we have now shown that there are differences in the molecular phenotype with respect to p16INK4a and IDH1 status for some of these associations. In contrast, CDKN2A/CDKN2B variation appears independent of the expression of a malignant phenotype,
compatible with a driver impact on tumorigenesis (18). Here
we have implicated variation at EGFR as risk factors for all
forms of glioma which is also reflected in no differences in
molecular phenotype as defined by EGFR, p16INK4a or
IDH1 status. While these risk variants do not impact on
disease outcome, they provide evidence that the glioma sub-
types result from different etiological pathways.

As the SNPs genotyped during GWA studies are generally
not themselves strong candidates for causality, enumeration
of the causal variant at 7p11.2 will involve fine mapping and
functional analyses to elucidate the causal basis of the
association. While partly speculative as these SNPs are not
correlated with known polymorphisms in the coding sequence
of EGFR, the glioma association may be mediated through a
change in gene expression rather than a sequence change in
the expressed protein or through LD with low-frequency vari-
ants not catalogued. Although the risk of glioma associated
with the 7p11.2 SNPs is modest, the carrier frequency of
these risk alleles is high in the European population and there-
fore, irrespective of the causal basis of the association, these
loci make a substantial contribution to the overall develop-
ment of glioma.

In conclusion, this large study provides unambiguous evi-
dence that common genetic variation in 7p11.2 influences
the risk of developing glioma. Furthermore, our findings
provide support for the notion that the glioma subtypes result from different etiologic pathways, rather than different
stages of tumor evolution within a common carcinogenic
pathway. The impact of the currently identified SNPs on
glioma risk is small in isolation and hence individually they
do not have immediate clinical application. However, the
observed differences provide insight into the biological mech-
nisms that underscore inherited susceptibility to glioma.

MATERIALS AND METHODS

Ethics

Collection of blood samples and clinico-pathological infor-
mation 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 centers [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 And-
erson Cancer Institutional Review Board].

Subjects

The present study is based on pooling data from GWA studies
of UK, US, French and German case–control series. The UK
and US GWA studies have been the subject of a previous pub-
llication (7). Briefly, the UK study was based on 636 cases (401
male; mean age 46 years) ascertained through the INTER-
PHONE Study (19). Individuals from the 1958 Birth Cohort
served as a source of controls. The US study was based on
1281 cases (786 male; 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) studies served as controls
(20,21). The French GWA study comprised 1495 patients with
glioma ascertained through the Service de Neurologie
Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris. The con-
trols were ascertained from the SU.VI.MAX (SUPplemen-
tation en Vitamines et MinerauxAntioXydants) study of
12 735 healthy subjects (women aged 35–60 years; men
aged 45–60 years) (22). The German GWA study 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 publicly available studies: KORA (Co-operative
Health Research in the Region of Augsburg; n = 488)
(23,24), POPGEN (Population Genetic Cohort; n = 678)
(25) and from the Heinz Nixdorf Recall study (n = 380) (26).

Genotyping and molecular analysis

Genotyping of cases were conducted using Illumina Infinium
HD Human610-Quad BeadChips according to the manufac-
turer’s protocols (Illumina, San Diego, CA, USA; Supple-
mentary Material S1). Descriptions of genotyping are included
in Supplementary Material S1. To determine amplification,
p16INK4a deletion and IDH1 mutation status in gliomas,
DNA was extracted from frozen tumors using a standard pro-
tocol. CGH-array analysis, EGFR amplification, p16INK4a
deletion assessment and sequencing of IDH1 were performed
as previously described (2,27).

Statistical analysis

Genotype data from each of the four studies were filtered on
the basis of pre-specified quality-control measures (Sup-
plementary Material, Text S1). Individual SNPs were
excluded from further analysis if they showed deviation
from the Hardy–Weinberg equilibrium with a P-value of
\( < 1 \times 10^{-5} \), an individual SNP genotype yield of <98% or
a minor allele frequency of <5%. This filtering resulted in
the use of 424 460 SNPs, common to the four case–control
series. A total of 363 samples were removed during quality-
control steps for reasons including a failure to genotype,
unknown duplicates, closely related individuals and sex dis-
crepancies (Fig. 4). Furthermore, for the UK control group,
we made the exclusions as proposed by the Wellcome Trust
Case Control Consortium (WTCCC; n = 65). For the other
controls, we removed people with a history of cancer as
well as those whose parents/grandparents were not born in
the study regions (n = 216; Fig. 4). We then performed
principal-components analyses to identify outlier samples to
reduce the effects of population stratification (Supplemen-
tary Material, Fig. S1). These analyses resulted in the removal
of 644 samples (145 cases and 499 controls), resulting in a
total of 4147 patients and 7435 control subjects (Fig. 4).

The association between each SNP and risk of glioma was
assessed by the Cochran–Armitage trend test. ORs and associ-
ated 95% CIs were calculated by unconditional logistic
regression using R software. Relationships between multiple
SNPs showing association with glioma risk in the same region were investigated using logistic regression analysis, and the impact of additional SNPs from the same region was assessed by a likelihood ratio test.

To further control for the potentially confounding influence of population stratification, we utilized the program Eigenstrat (28). The Eigenstrat algorithm adjusts genotypes and phenotypes by amounts attributable to ancestry using the calculated principal components of variation. These adjustments were applied to each of the four GWA studies correcting for the top 10 principal components of variation. Figure S3 (Supplementary Material) shows Q–Q plots before and after correction. Adjusted data from each of the four data sets were used to perform a meta-analysis using a modified Cochran–Armitage trend test. The results are reported as Eigenstrat $P$-values ($P_c$). The Eigenstrat algorithm involves the adjustment of genotypes and phenotypes by amounts attributable to ancestry using the calculated axes of variation. We also corrected for population structure using PLINK software (http://pngu.mgh.harvard.edu/~purcell/plink/) which implements a logistic regression analysis using 10 principal components as covariates.

Prediction of the untyped SNPs was carried out using IMPUTEv2, based on HapMapIII Release27 (February 2009, NCBI B36, dbSNP26) and the 1000 Genomes Project. Imputed data were analyzed using SNPTESv2 to account for uncertainties in SNP prediction. LD metrics between HapMap SNPs were based on HapMapIII Release27, viewed using Haploview (v4.2) and plotted using SNAP.

Of the 4147 patients, clinical (age at diagnosis, sex, preoperative KPI, degree of resection, chemotherapy and radiotherapy) and histological covariate data obtained at diagnosis were available for most. Complete survival data were available for 1740 patients (77%) in the French and German cohorts, with a median follow-up interval of 8.9 years for patients without an event. Association analyses of 7p11.2 SNPs with clinical characteristics were performed with the $\chi^2$ test on allele and genotype counts. Association with
overall survival was performed by comparing the Kaplan–Meier survival curves by means of the log-rank test in a pairwise fashion.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

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. For the UK GWA study, we acknowledge the participation of the clinicians and other hospital staff, cancer registries, study staff and funders who contributed to the blood sample and data collection for this study as listed in Hepworth et al. (BMJ 2006, 332, 883). 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 German GWA study, we are indebted to B. Harzheim (Bonn) and Dr A. Müller-Erkwoh (Bonn) for help with the acquisition of clinical data and R. Mahlberg (Bonn) who provided technical support. The UK GWA 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 (22). The German GWA study made use of genotyping data from three population control sources: KORA-gen (23,24), The Heinz-Nixdorf RECALL study (26) and POPS-GEN (25). We are extremely grateful to all investigators who contributed to the generation of these data sets.

**Conflict of Interest statement.** None declared.

**FUNDING**

In the UK, funding was provided by the Wellcome Trust and Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund) and the DJ Fielding Medical Research Trust. The UK INTERPHONE study was supported by the European Commission Fifth Framework Program ‘Quality of life and Management of Living Resources’ (QLK4-CT-1999-01563) and the International Union against Cancer (UICC). The UIICC received funds from the Mobile Manufacturers’ Forum and GSM Association. Provision of funds via the UIICC was governed by agreements that guaranteed INTERPHONE’s scientific independence (http://www.iafc.org/ENG/Units/RCAd.html). The UK centers 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 the US, funding was provided by NIH grants SRO1(CA119215&5SR01 CA070917). Support was also obtained from the American Brain Tumor Association and the National Brain Tumor Society. In France, funding was provided by the Délégation à la Recherche Clinique (MULO3012), the Association pour la Recherche sur les Tumeurs Cérébrales (ARTC), the Institut National du Cancer (INCa;PL046) and the French Ministry of Higher Education and Research. 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 Augsburg 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.

**REFERENCES**


