Text S1

Subjects

UK- GWA study

The UK study was based on 636 cases (401 male, 235 female; mean age 46 years; SD 12) ascertained through the INTERPHONE Study(1). Briefly, the INTERPHONE Study was an international multi-centre case-control study of PBT coordinated by the International Agency for Research on Cancer (IARC), with material collected between September 2000 and February 2004. UK patients with PBT were collected through neurosurgery, neuropathology, oncology, and neurology centers in the Thames regions of Southeast England and the Northern UK including central Scotland, the West Midlands, West Yorkshire, and the Trent area. Cases were patients with glioma [International Classification of Diseases for oncology (ICD-O), 2nd ed., codes 9380-9384, 9390-9411, 9420-9451 and 9505; ICD10 code C71]. Cases with previous brain tumors were excluded. To minimize population stratification cases with self reported non-UK ancestry were excluded from the present study. Individuals from the 1958 Birth Cohort served as source of the controls(2).

US-GWA study

The US study was based on 1,281 cases (786 male, 495 female; mean age 47 years; SD 13) ascertained through the MD Anderson Cancer Center, Texas, between 1990 and 2008. Cases were patients with glioma (ICD10 code C71; ICD-O codes 9380-9384, 9390-9411, 9420-9451, and 9505). Individuals from CGEMS(3, 4) served as controls.

French-GWA study

The French study was based on a systematic series of 1,495 patients with histologically proven glioma (WHO classification AI, AII, AIII, OII, OIII, OAII, OAIII, GBM-IV) ascertained through the Service de Neurologie Mazarin, Groupe Hospitalier Pitié-Salpêtrière Paris. French controls were taken from the SU.VI.MAX study(5).

German-GWA study

The German series was based on 880 patients who underwent surgery for a glioma at the Department of Neurosurgery, University of Bonn Medical Center, between 1996 and 2008. All histological diagnoses were made at the Institute for Neuropathology/German Brain Tumor Reference Center, University of Bonn Medical Center. Control subjects were taken from KORA survey S4 (Co-operative Health Research in the Region of Augsburg; conducted in the years 1999-2001, n = 488 with Illumina HumanHap550-Quad genotyping)(6, 7) from southern Germany; POPGEN (Population Genetic Cohort; n = 678)(8) from Schleswig-Holstein, northern Germany and from the Heinz Nixdorf Recall study (n = 380)(9).

Ethics

Collection of blood samples and clinico-pathological information from patients and controls was undertaken with informed consent and relevant ethical review board approval in accordance with the tenets of the Declaration of Helsinki.

Genotyping

DNA was extracted from samples using conventional methodologies and quantified using PicoGreen (Invitrogen, Carlsbad, USA). A genome-wide scan of tag SNPs was conducted using the Illumina Infinium HD Human610-Quad BeadChips according to the manufacturer's protocols (Illumina, San Diego, USA; Supplementary Methods). DNA samples with GenCall scores <0.25 at any locus were considered "no calls". A DNA sample was deemed to have failed if it generated genotypes at <95% of loci. A SNP was deemed to have failed if fewer than 95% of DNA samples generated a genotype at the locus. To ensure quality of genotyping, a series of duplicate samples were genotyped in the same batches. For all SNP assays >99% concordant results were obtained.

Statistical analysis

Statistical analyses were undertaken using R (v2.6), STATA (v8; State College, Texas, US) and PLINK (v1.05)(10) software. All *P* values reported are two-sided. Genotype data were used to search for duplicates and closely related individuals amongst all samples in each of the GWA studies. Identity by state (IBS) values were calculated for each pair of individuals and for any pair with allele sharing of >80%, the sample generating the lowest call rate was removed from further analysis. To identify individuals who might have non-Western European ancestry, we merged our case and control data with the 60 western European (CEU), 60 Nigerian (YRI), 90 Japanese (JPT) and 90 Han Chinese (CHB) individuals from the HapMap Project. For each pair of individuals we calculated genome-wide IBS distances, on markers shared between HapMap and our SNP panel, and used these as dissimilarity measures upon which to perform principal component analysis. The first two principal components for each individual were plotted and any individual not present in the main CEU cluster (*i.e.* outside 5% from cluster centroids) was excluded from subsequent analyses.

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. Deviation of the genotype frequencies in the controls from those expected under Hardy-Weinberg Equilibrium (HWE) was assessed by a χ^2 test, or Fisher's exact test where an expected cell count was <5.

The association between each SNP and risk of glioma was assessed by the Cochran-Armitage trend test. Odds ratios and associated 95% CIs were calculated by unconditional logistic regression. 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.

The combined effect of each pair of risk locus was investigated by logistic regression modeling with evidence for interactive effects between SNPs assessed by a likelihood ratio test. The OR and trend test for increasing numbers of deleterious alleles was estimated by counting two for a homozygote and one for a heterozygote.

Meta-analysis was conducted using standard methods based on weighted average of study-specific estimates of the ORs, using inverse variance weights. 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.

Bioinformatics

LD metrics between SNPs reported in HapMap were based on Data Release 2/phaseIII Feb09 on NCBI B35 assembly, dbSNPb125. We used Haploview software (v3.2) to infer the LD structure of the genome in the regions containing loci associated with disease risk.

Genotyping using Illumina arrays

Genotyping of both GWA studies was conducted by Illumina Service laboratory (San Diego, USA; www.illumina.com/*)* using the Illumina Infinium Human610-Quad BeadChips according to Illumina protocols. DNA samples with GenCall scores <0.25 at any locus were considered "no calls".

Evaluating and editing Cluster Positions

Intensity data from arrays was imported into Illumina's BeadStudio clustering and calling software application. For the small subset of loci that were not clustered properly by the automated algorithm the data were reviewed to identify loci that needed to be removed, manually edited, or left unchanged. Clustered SNPs were evaluated using the metrics listed in the SNP Table of the BeadStudio software. These metrics are based on all samples for each locus, and thus provide overall performance information for each locus. To identify loci potentially needing to be edited or removed each quality metric column in the SNP table was sequentially sorted. Metrics used for identifying poorly or incorrectly clustered data included intensity, cluster separation, position of each cluster (AA, AB, BB), Hardy-Weinberg equilibrium, call frequency, and variation of cluster width. The reproducibility of control samples on each plate as well as replicates were also used to identify mis-clustered loci. While not all cluster plots were assessed ~10% of the lowest performing loci were examined. Of these 10%, ~20% were edited or annotated (to indicate loci with nearby polymorphisms or hemizygous deletions), and ~2% were excluded. Data review was conducted by a second individual to determine whether any metrics were missed or if further editing was required. Overall this process provided for significantly increased genotyping accuracy.

URLs

The R suite can be found at http://www.r-project.org/
Detailed information on the tagSNP panel can be found at http://www.illumina.com/
dbSNP: http://www.ncbi.nlm.nih.gov/projects/SNP/
HAPMAP: http://www.hapmap.org/
1958 Birth Cohort:
http://www.cls.ioe.ac.uk/studies.asp?section=000100020003
CGEMS: http://cgems.cancer.gov/
SNPTEST: http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html
KORA: http://epi.helmholtz-muenchen.de/kora-gen/index_e.php
POPGEN : http://www.popgen.de/
EIGENSTRAT: http://genepath.med.harvard.edu/~reich/Software.htm

REFERENCES

1 Cardis, E., Richardson, L., Deltour, I., Armstrong, B., Feychting, M., Johansen, C., Kilkenny, M., McKinney, P., Modan, B., Sadetzki, S. *et al.* (2007) The INTERPHONE study: design, epidemiological methods, and description of the study population. *Eur J Epidemiol*, **22**, 647-664.

2 Power, C. and Elliott, J. (2006) Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol*, **35**, 34-41.

3 Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*, **39**, 870-874.

4 Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*, **39**, 645-649.

5 Hercberg, S., Galan, P., Preziosi, P., Bertrais, S., Mennen, L., Malvy, D., Roussel, A.M., Favier, A. and Briancon, S. (2004) The SU.VI.MAX Study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. *Arch Intern Med*, **164**, 2335-2342.

6 Holle, R., Happich, M., Lowel, H. and Wichmann, H.E. (2005) KORA--a research platform for population based health research. *Gesundheitswesen*, **67 Suppl 1**, S19-25.

7 Wichmann, H.E., Gieger, C. and Illig, T. (2005) KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*, **67 Suppl 1**, S26-30.

8 Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P.J., El Mokhtari, N.E. and Schreiber, S. (2006) PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet*, **9**, 55-61.

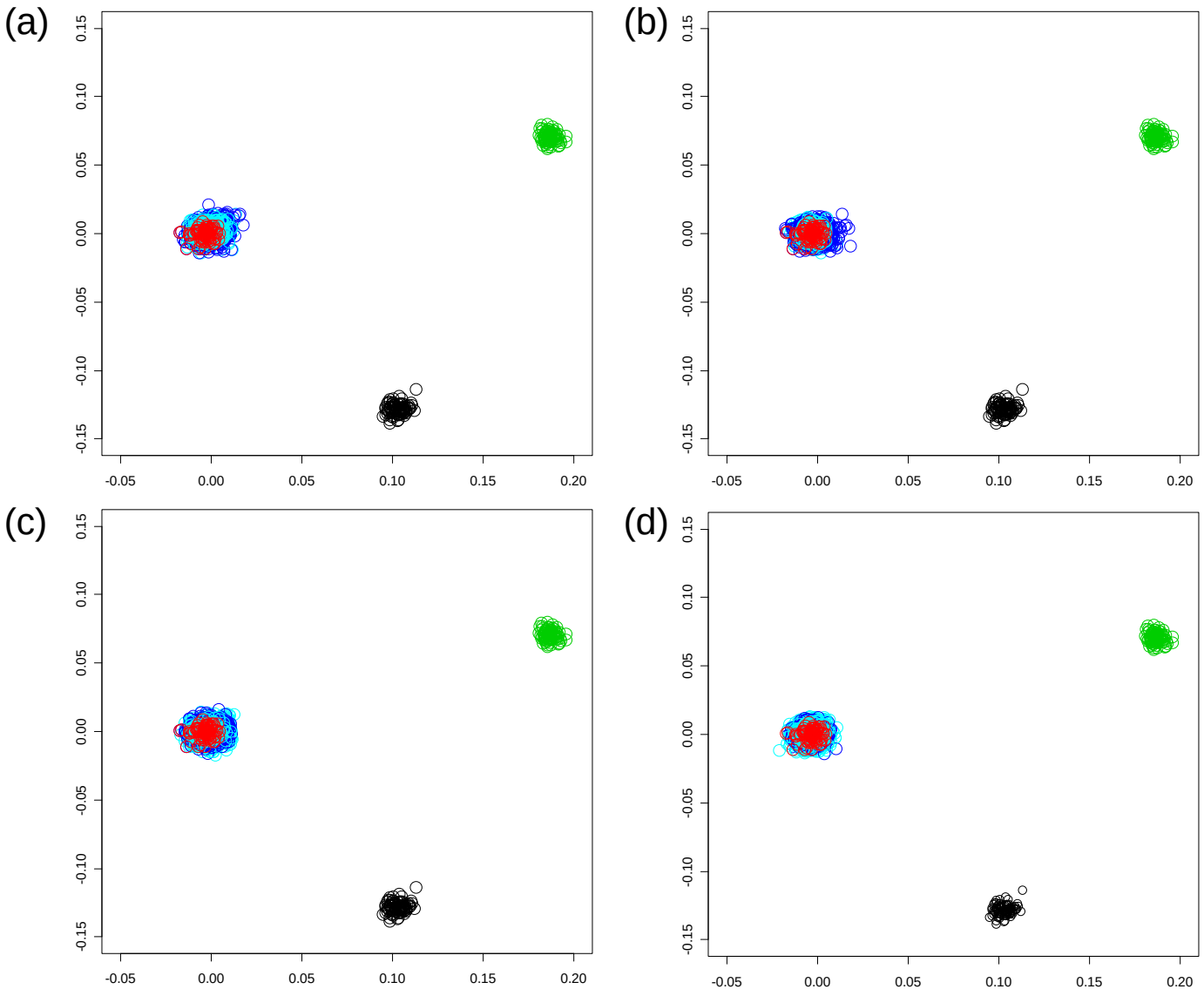
9 Schmermund, A., Mohlenkamp, S., Stang, A., Gronemeyer, D., Seibel, R., Hirche, H., Mann, K., Siffert, W., Lauterbach, K., Siegrist, J. *et al.* (2002)

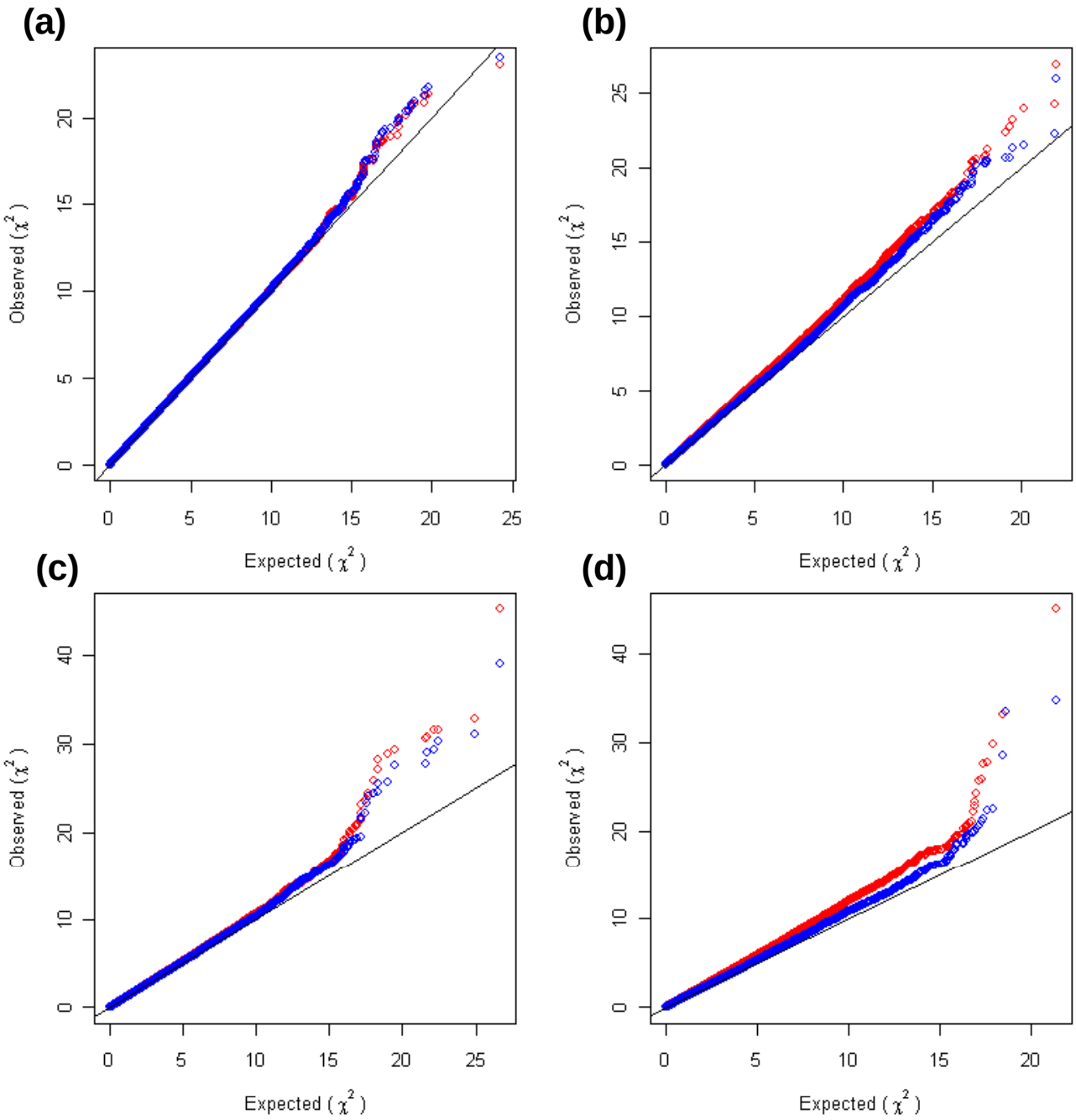
Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am Heart J*, **144**, 212-218.
10 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**, 559-575.

Supplementary Material Figure Legends

Figure S1: Ancestry of individuals after sample QC. The first two principal components are plotted from analysis based on 5,000 randomly selected SNPs. HapMap CEU individuals are plotted in red; CHB+JPT are plotted in black; YRI individuals are plotted in green; cases are plotted in light blue and controls in dark blue. The figure panels correspond to (a) France, (b) Germany, (c) US and (d) UK GWA studies.

Figure S2: Corrected and uncorrected quantile-quantile plots for results from the four GWA studies. Original chi-squared values are shown in red and values adjusted by Eigenstrat are shown in blue. Uncorrected and corrected inflation statistics are in (a) UK 1.003,1.002 (b) US 1.106,1.037 (c) France 1.046, 1.021 and (d) Germany 1.160, 1.045 respectively.


(a)
(b)
(c)
(d)
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.05
0.00
0.05
0.10
0.15
0.20


(a)
(b)
(c)
(d)
Observed (χ^2)
Expected (χ^2)

SNP	Chr.	Location (bp)	Nearby gene(s)	*Combined OR	Combined *P* value	Q	*P* adjusted
rs4295627	8	130,754,639	*CCDC26*	1.40	4.60×10^{-21}	6.35	3.49×10^{-20}
rs6470745	8	130,711,103	*CCDC26*	1.33	4.02×10^{-17}	2.14	4.48×10^{-16}
rs16904140	8	130,734,825	*CCDC26*	0.77	1.87×10^{-14}	2.51	1.49×10^{-13}
rs891835	8	130,560,934	*CCDC26*	1.26	1.543×10^{-12}	4.03	2.06×10^{-11}
rs10464870	8	130,547,005	*CCDC26*	1.26	1.49×10^{-11}	0.83	1.93×10^{-10}
rs9656979	8	130,733,589	*CCDC26*	1.21	6.99×10^{-11}	0.94	1.73×10^{-10}
rs4636162	8	130,708,722	*CCDC26*	0.84	2.89×10^{-9}	6.23	1.62×10^{-9}
rs12544799	8	130,732,692	*CCDC26*	0.84	1.06×10^{-8}	1.30	3.39×10^{-9}
rs6985166	8	130,748,358	*CCDC26*	0.84	4.21×10^{-8}	2.96	1.11×10^{-8}
rs2157719	9	22,023,366	*CDKN2A/B*	1.26	4.83×10^{-16}	4.26	5.07×10^{-18}
rs1063192	9	21,993,367	*CDKN2A/B*	1.25	4.91×10^{-15}	4.66	7.82×10^{-17}
rs1412829	9	22,033,926	*CDKN2A/B*	1.25	5.00×10^{-15}	4.40	4.49×10^{-17}
rs4977756	9	22,058,652	*CDKN2A/B*	1.24	7.19×10^{-14}	2.06	5.33×10^{-16}
rs573687	9	22,001,642	*CDKN2A/B*	0.81	1.28×10^{-12}	6.08	3.61×10^{-14}
rs10120688	9	22,046,499	*CDKN2A/B*	1.21	3.07×10^{-11}	8.44	2.35×10^{-13}
rs2151280	9	22,024,719	*CDKN2A/B*	1.20	6.79×10^{-11}	6.74	4.20×10^{-12}
rs4977574	9	22,088,574	*CDKN2A/B*	0.83	1.08×10^{-10}	0.48	4.07×10^{-12}
rs1537375	9	22,106,071	*CDKN2A/B*	0.83	1.56×10^{-10}	1.31	9.36×10^{-12}
rs2383207	9	22,105,959	*CDKN2A/B*	0.84	1.99×10^{-10}	0.92	8.84×10^{-12}
rs7049105	9	22,018,801	*CDKN2A/B*	0.83	2.16×10^{-10}	6.52	1.21×10^{-11}
rs3217992	9	21,993,223	*CDKN2A/B*	1.21	2.86×10^{-10}	6.09	2.09×10^{-10}
rs944797	9	22,105,286	*CDKN2A/B*	0.84	2.99×10^{-10}	1.10	1.60×10^{-11}
rs10116277	9	22,071,397	*CDKN2A/B*	1.19	4.48×10^{-10}	1.35	2.88×10^{-11}
rs1412832	9	22,067,543	*CDKN2A/B*	1.18	2.30×10^{-8}	2.29	1.99×10^{-10}
rs498872	11	117,982,577	*PHLDB1*	0.82	4.96×10^{-11}	1.63	3.93×10^{-11}
rs494560	11	118,026,759	*PHLDB1*	1.21	8.38×10^{-11}	6.01	8.91×10^{-12}
rs11603023	11	117,991,277	*PHLDB1*	1.20	1.45×10^{-10}	6.15	1.61×10^{-11}
rs17748	11	118,033,634	*PHLDB1*	0.82	8.02×10^{-10}	0.35	5.40×10^{-10}
rs11216930	11	117,993,992	*PHLDB1*	1.22	1.62×10^{-9}	0.29	1.40×10^{-9}
rs573905	11	118,077,477	*PHLDB1*	1.17	3.59×10^{-8}	7.00	1.34×10^{-9}
rs11216943	11	118,061,608	*PHLDB1*	0.84	3.91×10^{-8}	0.09	1.99×10^{-8}
rs10892258	11	118,085,075	*PHLDB1*	0.84	1.73×10^{-7}	0.82	6.29×10^{-8}
rs6010620	20	61,780,283	*RTEL1*	1.24	1.89×10^{-9}	4.54	3.89×10^{-11}
rs2736100	5	1,339,516	*TERT*	1.25	1.10×10^{-14}	4.69	1.03×10^{-12}
rs2853676	5	1,341,547	*TERT*	0.79	1.39×10^{-14}	4.32	2.52×10^{-13}

Table S1: SNPs not on 7p11.2 found at a significance level of 5×10^{-7} or better in the fixed effects meta-analysis. All SNPs are located in previously found regions: 5p.15.33 (*TERT*), 8q24.21 (*CCDC26*), 9p21.3 (*CDKN2A/CDKN2B*), 11q23.3 (*PHLDB1*) and 20q13.33 (*RTEL1*). Q is the between study heterogeneity statistic.
* ORs per copy of the A/T allele

Number of risk alleles	Controls (%)	Cases (%)	OR (95% CI)
0-3	225 (3.0)	53 (1.3)	0.39 (0.29 - 0.53)
4	589 (7.9)	174 (4.2)	0.49 (0.41 - 0.59)
5	1114 (15.0)	412 (9.9)	0.62 (0.54 - 0.71)
6	1655 (22.3)	738 (17.8)	0.74 (0.66 - 0.83)
7	1637 (22.0)	984 (23.7)	1.00 (0.89 - 1.12)
8	1255 (16.9)	872 (21.0)	1.16 (1.03 - 1.30)
9	626 (8.4)	554 (13.4)	1.47 (1.28 -1.69)
10	259 (3.5)	262 (6.3)	1.68 (1.39 - 2.03)
11+	98 (1.0)	98 (2.4)	2.17 (1.59 - 2.97)
Total	7,435	4,147	1.24 (1.21 - 1.27) $P_{trend} = 2.89 \times 10^{-72}$

Table S2: Odds ratio corresponding to increasing numbers of risk alleles in rs2252586, rs11979158 and the previously found five loci (rs2736100, rs4295627, rs498872, rs4977756 and rs6010620). The median number of risk alleles, seven, is used as the reference group for the odds ratios

		7p11.2 SNPs	
		rs2252586	**rs1197586**
Previously found loci	**rs2736100**	0.662 (11 559)	0.253 (11 558)
	rs4295627	0.055 (11 572)	0.386 (11 572)
	rs4977756	0.321 (11 573)	0.234 (11 573)
	rs498872	0.380(11 574)	0.008 (11 574)
	rs6010620	0.309 (11 575)	0.530 (11 576)
7p11.2 SNPs	**rs2252586**	-	0.191 (11 574)
	rs11979586	-	-

Table S3: Interactions between 7p11.2 loci (rs11979186 and rs2252586) and five previously validated loci (rs2736100, rs4295627, rs4977756, rs498872 and rs6010620). *P* values are given for pairwise interaction analyses with the number of samples each test is based on in parentheses.

	Germany				France				UK				US			
rs2252586	CC	TC	TT	MAF	CC	TC	TT	MAF	CC	TC	TT	MAF	CC	TC	TT	MAF
Grade 1	46	43	8	0.304	5	1	1	0.214	6	4	1	0.273	1	3	1	0.500
Grade 2	45	27	18	0.350	231	247	71	0.354	64	55	11	0.296	119	125	13	0.294
Grade 3	76	103	16	0.346	189	191	55	0.346	64	61	16	0.330	144	121	32	0.311
Grade 4	184	186	61	0.357	175	197	58	0.364	112	131	28	0.345	301	287	67	0.321
P value				0.758				0.266				0.171				0.251
rs11979158	AA	AG	GG	MAF	AA	AG	GG	MAF	AA	AG	GG	MAF	AA	AG	GG	MAF
Grade 1	69	24	4	0.165	5	2	0	0.143	10	1	0	0.045	4	1	0	0.100
Grade 2	65	22	3	0.156	415	125	9	0.130	90	37	3	0.165	183	70	4	0.152
Grade 3	135	55	4	0.162	346	82	8	0.112	108	28	5	0.135	212	73	12	0.163
Grade 4	340	85	6	0.113	325	100	5	0.128	206	58	7	0.133	489	155	11	0.135
P value				0.206				0.486				0.975				0.228

Table S4: Correlation between 7p11.2 glioma susceptibility SNPs and histopathology. Analysis was done on a subset of data for which classification by WHO grade was available. P_{trend} values were calculated excluding grade 1 tumors due to the small numbers in this group.

SNP		*EGFR*		*p16INK4a*		*IDH1*	
		Amplified	**Normal**	**Deleted**	**Normal**	**Altered**	**Normal**
rs11979158 7p11.2 *EGFR*	AA	90	477	134	421	171	268
	AG	29	158	38	146	71	86
	GG	0	4	1	7	3	5
	P value	0.999*		0.585*		0.381*	
rs2252586 7p11.2 *EGFR*	TT	18	79	20	73	26	44
	TC	58	297	85	267	110	166
	CC	42	267	68	233	109	149
	P value	0.419		0.823		0.704	
rs2736100 5p15.33 *TERT*	AA	16	121	27	107	57	51
	AC	68	315	96	277	120	184
	CC	35	207	50	190	68	124
	P value	0.203		0.246		**0.011**	
rs4295627 8q24.21 *CCDC26*	AA	82	376	118	332	130	233
	AC	33	229	48	209	100	108
	CC	4	38	7	33	15	18
	P value	0.099*		**0.05**		**0.014**	
rs4977756 9p21.3 *CDKN2A/B*	AA	46	217	62	200	89	125
	AG	53	320	78	284	124	173
	GG	20	106	33	90	32	61
	P value	0.531		0.473		0.421	
rs6010620 20q13.33 *RTEL1*	AA	2	97	4	14	6	8
	AG	30	287	48	172	76	103
	GG	87	259	121	388	163	248
	P value	0.481*		0.889*		0.804	
rs498872 11q23.3 *PHLDB1*	AA	8	16	15	88	43	37
	AG	59	194	84	254	120	165
	GG	52	433	74	232	82	157
	P value	0.052		0.082		0.007	

Table S5: Associations between glioma risk SNPs and EGFR amplification, p16INK4a deletion and IDH1 mutation status. Results are from a subset of French cases for whom information was available. Bold *P* values are < 0.05 and * denotes *P* values calculated using Fisher's exact test.

SNP & genotypes		Germany			France		
		Grade II	Grade III	Grade IV	Grade II	Grade III	Grade IV
rs2252586	TC	0.39	0.61	0.07	0.97	0.74	0.78
	TT	0.30	0.48	0.21	0.90	0.81	0.59
rs11979158	AG	0.52	0.11	0.95	0.53	0.52	0.34
	AA	0.99	NA	0.79	0.40	0.60	0.29

Table S6: Results from survival analysis for the subset of French and German patients for whom survival data was available. *P* values are given for each histology Grade after adjusting for age at diagnosis, sex, preoperative KPI, degree of resection, chemotherapy and radiotherapy.