

Multi-omics Mendelian randomisation using expression, splicing and protein quantitative trait loci: identification of novel drug targets for gliomagenesis

Authors and Affiliations

Zak A Thornton,^{1 2 3*} Lily J Andrews,^{1 2 3} Huiling Zhao,¹ Jie Zheng,^{1 4 5} Lavinia Paternoster,^{1 2 6} Jamie W Robinson,^{1†} and Kathreena M Kurian^{1 2 3 7 *†}

¹ MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Bristol, BS8 1QU, United Kingdom

² Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

³ Cancer Research Integrative Cancer Epidemiology Programme, University of Bristol, Bristol, UK

⁴ Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

⁵ Shanghai National Clinical Research Center for Metabolic Diseases, Key Laboratory for Endocrine and Metabolic Diseases of the National Health Commission of the PR China, Shanghai Key Laboratory for Endocrine Tumor, State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

⁶ NIHR Bristol Biomedical Research Centre, University Hospitals Bristol and Weston NHS Foundation Trust and University of Bristol

⁷ Brain Tumour Research Centre, Bristol Medical School, University of Bristol, Bristol, UK

† Co-senior author

Contact Information

* Corresponding authors: Mr Zak Thornton – zak.thornton@bristol.ac.uk; Professor Kathreena Kurian – kathreena.kurian@bristol.ac.uk

24 Abstract

25 **Background:** Genetic variants associated with molecular traits that are also associated with liability
26 to glioma can provide causal evidence for the prioritisation of these traits as candidate drug targets.

27 **Methods:** We performed two-sample Mendelian randomisation and genetic colocalisation of a large
28 panel of molecular traits on glioma. Molecular data were taken from studies of expression
29 quantitative trait loci (QTL) [11,985 genes]; splicing QTL [13,285 genes] and protein QTL [7,376
30 proteins] derived from 15 brain tissues. Glioma summary-level data was extracted from a genome-
31 wide association meta-analysis of 12,496 cases and 18,190 controls.

32 **Results:** Our MR analysis showed evidence for a causal effect of 85 molecular traits on glioma – 37
33 were robust according to colocalisation and Steiger filtering. We found causal evidence for 10 genes
34 previously associated with glioma risk. We identified one novel genetic locus with strong causal
35 evidence in the gene expression analysis: *HBEGF* (5q31.3) in all glioma [OR 1.36 (95%CI 1.19 to 1.55);
36 $P = 4.41 \times 10^{-6}$]. We also identified three novel genetic loci with strong causal evidence in the splicing
37 variation analysis: *CEP192* (18p11.21) in glioblastoma [OR 4.40 (95%CI 2.28 to 8.48); $P = 9.78 \times 10^{-4}$];
38 *FAIM* (3q22.3) in all glioma [OR 2.72 to 3.43; $P = 1.03 \times 10^{-5}$ to 1.09×10^{-5}] and *SLC8A1* (2p22.1) in all
39 glioma [OR 0.37 (95%CI 0.24 to 0.56; $P = 5.72 \times 10^{-6}$].

40 **Conclusions:** We provide robust causal evidence for genes previously associated with glioma risk in
41 genome-wide association studies, as well as four novel genes.

42 **Keywords:** glioma, Mendelian randomisation, quantitative trait loci, molecular traits

43 Introduction

44 Glioma is the most common (~80%) primary malignant brain and central nervous system (CNS)
 45 tumour group, with age-standardised incidence rates ranging from 4.67 to 5.73 per 100,000.^{1,2}
 46 Malignant brain tumours are responsible for the greatest years of potential life lost of all cancers
 47 (~20 years), with a comparatively shorter 5-year relative survival rate (34.9%) compared to other
 48 malignancies.^{3,4}

49 In the 2016 World Health Organisation (WHO) guidelines for diagnosis of brain tumours, histological
 50 markers formed a sufficient basis to classify CNS tumours; however, this was expanded in the latest
 51 2021 edition, which suggests molecular profiling as a means of stratifying tumour types.^{5,6} The
 52 reason for this change is to emphasise the shift in how glioma, and brain tumours in general, are
 53 now classified in the clinic. One key example is the use of *MGMT* methylation as a tool to stratify
 54 patients in clinical trials.^{7,8} However, the molecular landscape of glioma remains complex and
 55 heterogeneous, highlighting an area of unmet research need to improve diagnosis, patient
 56 stratification and, hopefully, treatment.

57 One approach to elucidating novel molecular markers comes from the field of genetic epidemiology
 58 in the way of genome-wide association studies (GWAS) and transcriptome-wide association studies
 59 (TWAS), two study designs which leverage large-scale human genotyping of genetic variation to
 60 determine which genes are associated with a phenotype of interest. To date, previous studies have
 61 identified 27 genetic loci associated with glioma risk, implicating a total of approximately 50 genes;
 62 however, further analysis is required to determine both which genes these genetic variants map to,
 63 and whether such variation is causally implicated in gliomagenesis, or simply an observational
 64 association.⁹⁻¹⁴

65 Furthermore, genotyping of large cohorts has been performed to link disease-associated genetic
 66 variants to gene regulatory mechanisms to improve understanding of disease aetiology.¹⁵ Such
 67 genetic variants associated with these molecular traits are known as quantitative trait loci (QTL), and

68 examples of these include measuring relative gene expression levels (eQTLs), splicing variation
 69 (sQTLs) and protein abundance (pQTLs). Integrating multi-dimensional omics data can further our
 70 understanding of the aetiology of complex traits by providing functional and mechanistic insight into
 71 these layered molecular relationships.¹⁶⁻¹⁸

72 Mendelian randomisation (MR) is an established statistical method which utilises genetic predictors
 73 of an exposure to assess the causal relationship between this exposure and an outcome.¹⁹ The main
 74 advantage MR has over traditional observational epidemiological methods is that MR can determine
 75 causality between an exposure and an outcome as it is less liable to common epidemiological biases,
 76 such as confounding and reverse causality. When combined with other sensitivity analyses such as
 77 genetic colocalisation, MR can provide strong evidence of an unbiased causal relationship between
 78 genes and traits of interest.²⁰

79 In this study, we leveraged multi-dimensional QTL data derived in bulk brain tissues in a combined
 80 MR-colocalisation framework, with the aim to identify causal evidence for aetiologically important
 81 genes for gliomagenesis.

82 Methods

83 A diagram of the workflow described in this section can be found in Figure 1.

84 Study Datasets

85 We used QTL summary statistics from genome-wide association studies (GWAS) of gene expression,
86 protein abundance and alternative splicing levels in various brain tissues described below. Overviews
87 of these datasets are given in Supplementary Table S1.

88 Gene expression data were taken from MetaBrain, a meta-analysis of 14 eQTL datasets derived from
89 3,659 samples from 2,683 individuals of European ancestry in five CNS tissues: the basal ganglia
90 (n=208), cerebellum (n=492), cortex (n=2,683), hippocampus (n=168), and spinal cord (n=108)
91 (Supplementary Table S1).²¹

92 The sQTL data were taken from GTEx Portal (v8 release). We used sGenes provided by GTEx,
93 generated using their QC protocol, from a primarily European-American (85.3%) population. sQTL
94 were extracted from 13 regions of the CNS: amygdala (n=129), anterior cingulate cortex (BA24)
95 (n=147), caudate (n=194), cerebellar hemisphere (n=175), cerebellum (n=209), cortex (n=205),
96 frontal cortex (BA9) (n=175), hippocampus (n=165), hypothalamus (n=170), nucleus accumbens
97 (n=202), putamen (n=170), spinal cord (n=126), and substantia nigra (n=114) (Supplementary Table
98 S1).²²

99 Finally, we included pQTL data retrieved from BrainQTL, derived from European ancestry individuals
100 in the religious orders study and memory and ageing projects.^{23,24} pQTL were extracted from the
101 dorsolateral pre-frontal cortex of 330 individuals (Supplementary Table S1).²⁵

102 Glioma summary-level data were derived from a meta-analysis of eight constituent glioma GWAS
103 consisted of 6,191 glioblastoma (GB) cases, 5,819 non-GB cases, 12,496 combined cases and 18,190
104 controls.⁹ We therefore used three outcomes throughout our analyses: the GB-only case load, the
105 non-GB-only case load, and the combined case load (defined as 'all glioma'). Although glioma is

highly heterogeneous, we did this to increase our statistical power by using larger sample sizes. The non-GB cases analysed in this meta-analysis were not explicitly classified but consists mainly of astrocytoma and oligodendroglioma diagnosed patients (Supplementary Table S1).

This study uses previously published summary-level data and thus contains no patient identifiable data. Ethical approval and informed consent from each participant were given and can be found where the dataset was initially described. All procedures performed in studies involving human participants were done in accordance with the ethical standards of the institutional or national research committee and with 1964 Helsinki declaration.

Instrument Selection

We identified *cis*-acting (within 1Mb of the gene coding region) QTLs which met genome-wide significance ($P < 5 \times 10^{-8}$). *Trans*-acting QTLs were excluded from the analysis because of the increased likelihood of horizontal pleiotropy, due to their distant location from the gene whose variation they alter. Instruments were selected to be in linkage disequilibrium ($R^2 < 0.001$) to ensure independence.

Two Sample Mendelian Randomisation

We used a two-sample MR framework to estimate the causal effect of genetically proxied gene expression, protein abundance and alternative splicing on genetic liability to glioma risk. MR estimates were generated using the Wald ratio method for instruments consisting of single SNPs and inverse variance weighted (IVW) method for instruments comprising of multiple SNPs.²⁶ MR estimates were transformed and presented throughout as odds ratios (OR) and were scaled to reflect one standard deviation increase in the respective molecular trait. Following MR analysis, the results had to meet a Bonferroni-corrected P value threshold ($0.05/\text{number of tests performed}$) to adjust for multiple testing.

MR has three assumptions which must hold to produce an unbiased estimate. Firstly, the genetic instrument must associate with the exposure ('relevance'). We tested this by generating the F-

statistic for each instrument, where an F-statistic > 10 is evidence against weak instrument bias, and filtering out instruments which did not surpass this threshold.^{27,28} The second assumption is that the genetic instrument does not share a common cause with the outcome ('independence'). We tested this assumption partly by using genetic colocalisation, which determines whether the molecular trait and glioma share the same causal variant, a necessary (though not sufficient) condition for causality. Using colocalisation in this way has been posited to at least eliminate some unreliable associations when standard follow up sensitivity analyses to evaluate the presence of horizontal pleiotropy (such as MR-Egger) are unavailable.^{27,29} The final assumption is that the genetic instrument affects the outcome only through the effect of the exposure of interest ('exclusion restriction'). This assumption, also known as horizontal pleiotropy, is difficult to assess with single SNP instruments, as is common for 'omic' variables. This was tested by investigating whether there was causal evidence linking the molecular trait of interest with putative risk factors for glioma. The presence of an association between the molecular trait and glioma risk factors may then be evidence of potential horizontal pleiotropy (Figure 2).²⁷

For brevity, we refer to our MR results as causal genes (or causal proteins); however, we recognise that further studies and evidence will be required to validate our results.

Colocalisation

We tested for colocalisation between the loci of each molecular trait and glioma subtype which had an MR result that passed the Bonferroni-corrected P value threshold. We extracted *cis*-regions consisting of all SNPs within a 1Mb window around the gene coding regions.²⁰ Colocalisation requires the *cis*-region of the molecular trait to be richly genotyped. If there were not enough SNPs to perform colocalisation, we used LDCheck to measure the pairwise LD between the QTL and *cis*-SNPs in the glioma data at a more lenient threshold ($P < 5 \times 10^{-6}$). If the SNPs were in high LD ($R^2 > 0.8$), then we took this as proxy, albeit weak, evidence for colocalisation. This method has been described and used previously when full genome-wide statistics were not available.³⁰

Steiger Filtering

For causal effects which had strong evidence of colocalisation, we applied Steiger filtering to test for the presence of reverse causation, where the instrument explains more of the variance in the outcome than the exposure.²⁹ Results of Steiger filtering are presented as a one of three categorical variables; “pass” if Steiger $P < 0.05$ and the instrument explains more of the variance in the outcome than the exposure, “fail” if Steiger $P < 0.05$ and the instrument explains more of the variance in the exposure than the outcome, and “uncertain” if $P \geq 0.05$.

Evidence thresholds for multi-omic analyses

Our main results comprised of those which passed the Bonferroni correction P value, had strong colocalisation evidence ($H_4 \geq 80\%$) and passed Steiger filtering sensitivity analysis, as they have the most robust evidence. Since Steiger filtering is sensitive to sample sizes, the analysis may be liable to false negatives if one dataset is better powered than the other; therefore, given the relatively small samples in the QTL datasets, we included results which had “uncertain” evidence but with the caveat that these results may be less robust.

Overlap in signals between molecular traits

To determine whether there was any potential overlap in signals between molecular traits, we looked at all MR results for our molecular traits with robust evidence in all tissues, regardless of P value. We then filtered these results using a more lenient Bonferroni correction P value (similarly, by using $0.05/\text{number of tests performed}$) to determine whether any were potentially overlapping signals which failed to pass due to insufficient statistical power.

Specificity Analysis

To annotate the potential causal pathway between the molecular trait and glioma, we used two different approaches: exploring PheWAS data, and risk factors with pre-existing evidence in the literature. To perform a broad search for potential causal pathways, we used the IEU OpenGWAS

project PheWAS tool and PhenoScanner V2 to search for phenotypes which were strongly associated ($P < 5 \times 10^{-8}$) with lead SNPs used to instrument the robust causal effects.³¹⁻³³ We then performed a specificity analysis to determine whether the robust causal effects were being caused by non-specific effects or driven through a putative risk factor for glioma. We investigated risk factors with genetic liability to allergies, asthma, altered low-density lipoprotein cholesterol (LDLc) levels and alcohol consumption. These risk factors were selected based on potential evidence linking that risk factor to glioma in the literature, if there was data available for that risk factor in the IEU OpenGWAS project database, and whether those data were available for use in an MR analysis.³⁴

Differential expression in molecular traits

We explored whether there was differential expression in the genes for which there was robust evidence that expression had a causal effect on at least one glioma subtype, between subtypes of glioma and control patients using the Gliovis resource.³⁵

We investigated the differences in the expression of splicing variants in tissue types in the genes for which there was evidence that gene-splicing had a causal effect on at least one glioma subtype, using the GTEx portal.²²

Annotation of 'druggability'

We annotated robust results which passed sensitivity analyses with evidence of 'druggability', i.e. how likely a gene and its gene product are to be a valid drug target, to build evidence for the prioritisation of genes and proteins as candidate targets. We used the Open Targets platform, a large-scale database that uses genetic and genomic data for systematic drug target identification and prioritisation, to search for gene ontology terms, pathways, and gene interactions.^{36,37} We also used Drug Gene Interaction Database (DGIdb) to search for drug-gene interactions and genome 'druggability'.³⁸

Software

We used the TwoSampleMR R package to implement the MR analyses and Steiger filtering.³⁹⁻⁴¹ We performed genetic colocalisation using Pair Wise Conditional Analysis and Colocalisation (PWCoco), a method which combines Conditional and Joint Analysis (GCTA-COJO) and colocalisation (Coloc R package) to ensure the single causal variant assumption of colocalisation holds.^{20,42,43}

Results

Estimating the causal effects of gene expression on glioma risk

Of the 10,488 genes for which expression could be instrumented, 10,006 (95.4%) had the relevant SNPs available in the glioma GWAS and therefore MR could be performed. We found 29 genes with evidence for a causal effect ($P < 5.00 \times 10^{-6}$) on all glioma, GB, or non-GB in at least one of the tissues tested, resulting in 80 causal effects in total; 40 causal effects in cortex, 30 in cerebellum, 6 in basal ganglia, 3 in hippocampus and 1 in spinal cord (Supplementary Figure S1). In the follow-up sensitivity analyses of the unique 80 expression causal effects, we found that 30 (38%) had strong evidence of colocalisation ($H_4 > 80\%$), 11 (14%) had moderate evidence of colocalisation ($80\% > H_4 > 50\%$) and 39 (49%) had weak evidence of colocalisation ($H_4 < 50\%$). The 30 causal effects (in 10 genes) with strong evidence for colocalisation were then tested for the potential of reverse causality using Steiger filtering; 29 causal effects passed this test, whilst one result was uncertain according to Steiger P-value ($P > 0.05$, Table 1).

Three genes (*GALNT6*, *HEATR3*, *PICK1*) with robust evidence of a causal effect were identified in multiple tissues, with different lead SNPs, but all SNPs were in high LD ($R^2 > 0.7$).³⁰ Two genes (*JAK1* and *PICK1*) with robust evidence had a causal effect in only one subtype (non-GB and GB, respectively) whereas all other genes (*CDKN2B*, *EGFR*, *GALNT6*, *HBEGF*, *HEATR3*, *MDM4*, *RAVER2* and *TERT*) had a causal effect on all glioma, or more than one outcome tested (Table 1).

In summary, we found 10 genes for which there was robust evidence that expression had a causal effect on at least one glioma subtype in at least one brain region (30 causal effects in total, Figure 3). Three genes (*CDKN2B*, *RAVER2*, *TERT*) had strong causal effects ($OR < 0.5$ or > 2). For six genes, increased expression was causally related to increased odds of glioma. All results for all eQTL sensitivity analyses are described in Supplementary Table S1.

Estimating the causal effects of alternative splicing on glioma risk

Of the 6,200 genes for which alternative splicing could be instrumented, 4,496 (72.5%) had the relevant SNPs available in the glioma GWAS are therefore MR could be performed. We found 10 genes with evidence for a causal effect ($P < 1.11 \times 10^{-5}$) on glioma, GB or non-GB in at least one of the thirteen tissues tested, resulting in 71 causal effects in total; five in amygdala, four in anterior cingulate cortex, seven in caudate, six in cerebellar hemisphere, nine in cerebellum, six in cortex, four in frontal cortex, seven in hippocampus, seven in hypothalamus, nine in nucleus accumbens, six in putamen and one in substantia nigra (Supplementary Figure S2). We applied the same sensitivity analysis pipeline. We found that 47 (66%) had strong evidence of colocalisation ($H_4 > 80\%$), five (7%) had moderate evidence of colocalisation ($80\% > H_4 > 50\%$) and 19 (27%) had weak evidence of colocalisation ($H_4 < 50\%$). The 47 causal effects with strong evidence for colocalisation were then tested for the potential of reverse causality using Steiger filtering; 24 (51%) passed this test, however all of these results were uncertain according to Steiger P-value ($P > 0.05$. Table 2).

Three genes (*FAIM*, *PHLDB1*, *RTEL1*) had robust evidence of a causal effect on glioma risk within multiple tissues, with different lead SNPs, but all were in high LD ($R^2 > 0.7$).³⁰ One gene (*CEP192*) with robust evidence of a causal effect on GB, whereas all other genes (*FAIM*, *HEATR3*, *PHLDB1*, *SLC8A1* and *RTEL1*) had a causal effect on all glioma, or more than one outcome tested (Table 2).

In summary, we found six genes for which there was robust evidence that splicing had a causal effect on at least one glioma subtype in at least one brain region (24 causal effects in total, Figure 4). All results for all sQTL sensitivity analyses are described in Supplementary Table S1.

Estimating the causal effects of protein abundance on glioma risk

Of the 618 proteins that could be instrumented, all (100%) had the relevant SNPs available in the glioma GWAS are therefore MR could be performed. We only found one protein with evidence for a causal effect on GB in DLPFC ($P < 8.09 \times 10^{-5}$), SCFD1 (OR 0.17 (95%CI 0.07 to 0.40); $P = 6.67 \times 10^{-5}$), Supplementary Figure S3).

We applied the same sensitivity pipeline but there was not sufficient SNPs in the *cis*-region of SCFD1 to perform colocalisation. We used LDCheck (see methods) to determine weak proxy evidence of colocalisation. However, none of the SCFD1 SNPs in the glioma dataset reached our most lenient threshold of $P < 5 \times 10^{-6}$. Therefore, we could not conclude that this pQTL had sufficient causal evidence to be associated with glioma.

In summary, there were no proteins with robust causal evidence from the MR and sensitivity analyses. All results for causal pQTL sensitivity analysis are described in Supplementary Table S1.

Novel genes

From our multi-omic analysis, we found robust evidence for four genes that not been implicated with glioma risk in previous genetic studies: *CEP192* (18p11.21), *FAIM* (3q22.3), *HBEGF* (5q13.3) and *SLC8A1* (2p22.1).

We found that increased *HBEGF* expression in the cortex is causal for all glioma (OR 1.36 (95%CI 1.19 to 1.55); $P = 4.41 \times 10^{-6}$). We found there was also increased risk of glioma in the GB and non-GB subtypes in the cortex, but they did not meet our Bonferroni-corrected P value threshold ($P = 5.53 \times 10^{-6}$ and 2.01×10^{-2} , respectively) (Table 1, Figure 5).

We found that *FAIM* isoforms in four tissues were causal for all glioma (OR 2.72 to 3.43; $P = 1.03 \times 10^{-5}$ to 1.09×10^{-5}). There was also an increased risk of glioma in the GB and non-GB subtypes in these tissues, but they failed to meet our Bonferroni-corrected P value threshold ($P = 4.36 \times 10^{-5}$ to 9.64×10^{-3}). We found that *CEP192* splicing in the cerebellar hemisphere is causal for GB (OR 4.40

(95%CI 2.28 to 8.48; $P = 9.78 \times 10^{-6}$). There was also an increased risk of glioma in all glioma (but not non-GB), but this failed to meet the Bonferroni-corrected P value ($P = 1.74 \times 10^{-3}$) (Table 2, Figure 6). We found that *SLC8A1* splicing in the substantia nigra was protective for all glioma (OR 0.37 (95%CI 0.24 to 0.56); $P = 5.72 \times 10^{-6}$). There was also a decreased risk of glioma in the GB and non-GB subtypes in this tissue, but they failed to meet our Bonferroni-corrected P value threshold ($P = 5.67 \times 10^{-4}$ and 1.81×10^{-4} , respectively) (Table 2, Figure 6).

Overlap in signals between molecular traits

We analysed all 16 robust molecular traits that we identified in our analysis in each tissue type across all three datasets, regardless of P value. We performed 104 tests, of which 82 molecular traits had passed the Bonferroni corrected P value in the primary analysis. Of the 22 other molecular traits, 12 passed the more lenient Bonferroni-corrected P value ($P = 4.81 \times 10^{-4}$), of which only *FAIM* in the cerebellum in all glioma ($P = 6.93 \times 10^{-5}$) showed potential overlap between molecular traits (gene expression and alternative splicing). Additionally, we found that the effect of *HEATR3* appears to be eQTL-specific ($P = 1.95 \times 10^{-10}$ to 6.74×10^{-12}) despite finding a robust causal effect in the alternative splicing analysis ($P = 1.29 \times 10^{-10}$, Supplementary Table S2). The causal effect is likely driven through expression of the canonical gene isoform, which is incorrectly measured as an alternative splice isoform due to the way in which eQTL and sQTL are measured (see discussion).

Specificity analysis

To annotate the potential causal pathways between the robust molecular traits and glioma, we used PheWAS data, and known glioma risk factors with pre-existing evidence. *HBEGF* is commonly linked with other cancer types so we wanted to explore this putative oncogene further.⁴⁴ Similarly, *RTEL1* is linked with telomere length which has been previously implicated in observational and MR studies to have an association with glioma risk. Additionally, the considerable number of tissues in which we found a causal effect warranted further analyses.⁴⁵⁻⁴⁷

We used the IEU OpenGWAS project PheWAS tool and PhenoScanner v2 to conduct a broad search for phenotypes which were strongly associated with the lead SNP(s) for *HBEGF* (rs4150197) and *RTEL1* (rs6062302, rs2297440, rs2315009, rs3208007, rs2777941) at genome-wide significance ($P < 5 \times 10^{-8}$).^{31,33} We found that rs4150197 was associated with anthropometric traits such as hip circumference and weight, and various image-derived phenotypes from diffuse MRI (Supplementary Table S2). We found that rs6062302 and rs2297440 were already associated with glioblastoma ($P = 1.00 \times 10^{-13}$ and $P = 4.00 \times 10^{-46}$, respectively) and rs2297440 was associated with glioma ($P = 2.00 \times 10^{-42}$). We also found significant associations with phenotypes associated with the immune system such as dermatitis, eczema, Crohn's disease, inflammatory bowel disease and allergic rhinitis (Supplementary Table S2).

For the 16 molecular traits that had strong evidence for a causal effect on glioma, we performed a follow-up ad hoc analysis to determine whether these molecular traits also had causal evidence with putative glioma risk factors. We found that of the molecular traits with robust causal evidence, only *HBEGF* showed no evidence for non-specific effects with any of the traits we identified. The trait that was associated with the most potential glioma genes was LDLc, which was associated ($P < 0.05$) with seven genes (*CDKN2B*, *FAIM*, *GALNT6*, *HEATR3*, *PHLDB1*, *PICK1*, *RTEL1*). (Supplementary Table S2).

Differential expression in molecular traits

We searched for differential expression of *HBEGF* using the Rembrandt and TCGA studies, found in the GlioVis tool. We found no significant differences ($P < 0.05$) between non-tumour samples and any of the glioma subtypes but did find differential expression between GB and non-GB subtypes ($P = 3.10 \times 10^{-5}$ to 1.90×10^{-10}) (Supplementary Figure S4A&B).⁴⁸

Additionally, we searched for differences in isoform expression of *RTEL1* using the GTEx portal tool.²² We found that isoform expression was significantly lower between wild-type (CC) and the homozygous variant (TT) in our lead SNPs ($P = 6.64 \times 10^{-18}$ to 4.23×10^{-75}). We found that the cerebellum had the highest isoform (ENST00000425905.5) expression (3.17 transcripts per million

(TPM)), one of the tissues in which we found our results, followed by the cerebellar hemisphere (2.96 TPM) and testes (1.31 TPM).²²

Annotation of 'druggability'

We used OpenTargets and DGIdb to gather evidence of 'druggability' for molecular traits with robust evidence.³⁶⁻³⁸ We found that 4/10 (40%) of the molecular traits identified in the expression analysis, and none of the molecular traits identified in the gene-splicing analysis were categorised as being part of the 'druggable genome'. One novel result, *HBEGF*, had three drug-gene interactions (cetuximab, panitumumab and KHK-2866), the latter being an anti-HBEGF antibody which was trialled for use in advanced solid tumours and ovarian cancer. However, the study was discontinued after 22 months due to patients experiencing reversible neuropsychiatric toxicity, although the aetiology was not understood (Supplementary Table S2).⁴⁹

Discussion

In this study, we used a combined MR-colocalisation framework to estimate the causal effect of genetically proxied gene expression, alternative splicing and protein abundance levels on genetic liability of glioma risk. We found robust evidence that causally implicated 16 molecular traits with glioma risk: ten gene expression evidence levels and six alternative splicing events (one gene shared both expression and splicing causal effects on glioma). Assessing different ‘omics’ data sources can be beneficial to understand the development of complex traits and for drug target-related analyses; each QTL provides an insight into different molecular processes.^{16,17}

For 11 of the genes that we found robust evidence for here, the loci have been previously implicated in GWAS/TWAS of glioma risk.^{9,10} However, these studies did not establish causality of all these genes, as we have. Additionally, we found four genes that not been implicated in previous genetic studies: *CEP192* (18p11.21), *FAIM* (3q22.3), *HBEGF* (5q13.3) and *SLC8A1* (2p22.1). For brevity, we discuss *RTEL1* and a novel result, *HBEGF*.

RTEL1 (Regulator of telomere elongation helicase 1) is a gene which encodes for a DNA helicase responsible for the elongation of telomeres, and is known to interact with the telosome complex, a group of proteins which protect the telomere caps of DNA.⁵⁰ *RTEL1* has previously been identified as a risk factor for glioma, and studies have tried to identify whether telomere length has an established causative effect on the risk of glioma, although no definitive conclusion has been reached.^{12,14, 45-47} *RTEL1*-associated sQTLs were associated with an exon-skipping event (GTEx intron ID: 63689132:63689750:clu_27064) present only in *RTEL1* transcript ENST00000425905 ($P = 4.23 \times 10^{-75}$ to 6.64×10^{-18}). Therefore, it may be the case that previous conflicting evidence on the directionality of the effect of *RTEL1* expression on glioma risk may be confounded due to non-canonical transcript variants; though, further studies would be required to ascertain if this is the case. Furthermore, we observed that *RTEL1* sQTLs rs6063202 and rs2297440, were found to be strongly associated to glioblastoma ($P = 1.00 \times 10^{-13}$ and 4.00×10^{-46} , respectively) and with allergy-

related traits ($P = 4.78 \times 10^{-8}$ to 7.84×10^{-19} , Supplementary Table S2). These results beget further study and could help to understand the links between allergies, glioma risk and *RTEL1*.

HBEGF (Heparin-binding epidermal growth factor) is a member of the EGF family of growth factors which is involved in the MAPK, STAT and PI3K/AKT pathways.^{44,51} These molecular pathways have roles in regulating biological processes, such as positive regulation of cell growth, migration, and proliferation.⁵² The role of *HBEGF* in the CNS is well characterised: it is highly expressed throughout; has roles in both the developing and adult brain, with studies indicating *HBEGF* contributes to glial and stem cell proliferation and is a ligand for EGFR.^{53,54} Furthermore, expression of *HBEGF* has been shown to be increased significantly in many human cancer types, and *HBEGF* mRNA expression has been shown to be increased two- to five-fold in glioblastoma cell lines compared to control brain tissue.⁵⁵ We found no significant differences between control and any glioma subtypes but did find differential expression between GB and non-GB subtypes in two independent datasets (Supplementary Figure S4A&B).^{48,56} We also found that the instrumented *HBEGF* eQTL (rs4150197) had no associations with cancer risk or any putative risk factors for glioma (Supplementary Table S2). Unlike *RTEL1*, we observed no evidence for an effect through non-canonical transcripts or proteins. Altogether, our causal evidence implicates the expression of *HBEGF* in the brain as a potentially novel biomarker, though more research would be required to ascertain its role. Given that there are drugs which target *HBEGF*, glioma and *HBEGF* may make for an interesting target-indication pair for future studies.

Strengths

We aimed to identify molecular traits with the most robust evidence for causal roles in glioma risk, by combining strong MR evidence with colocalisation and Steiger filtering. To test if the causal effects were driven via putative glioma risk factors, we tested for potential pleiotropic effects in known risk factors for glioma: allergies, asthma, altered low-density lipoprotein cholesterol (LDLc) levels and alcohol consumption. This may provide evidence of horizontal pleiotropy, which may

make it harder to draw decisive conclusions of causality^{27,29} However, this would require follow up studies to investigate which could lead to potentially elucidating how these putative risk factors and molecular traits are linked to glioma.

Limitations

Despite using relatively large datasets, our analyses are still likely to suffer from limited statistical power due to restricted sample sizes, particularly in the sQTL dataset (n = 114 to 209). This might lead to some important causal effects not being identified. All three QTL datasets were derived from a mixture of case and control samples. However, the cases were derived from individuals whose disease (e.g. Alzheimer's disease in BrainQTL) have no known link to glioma risk.

Additionally, as we used the GTEx sGenes, only the top splice event per gene was measured, and therefore there were many splice events which were not tested in our MR analysis.

Most of the MR analyses used a single SNP instrument, which restricts the type of sensitivity analyses that could be performed; however, this is a common phenomenon observed when conducting MR with molecular traits.²⁹

SNPs can act via multiple molecular QTL pathways and are not mutually exclusive; SNPs affecting gene expression can also be associated with alternative gene splicing of the same genes. As eQTL and sQTL are both measured by quantifying mRNA levels, this was not unexpected.⁵⁷ These SNPs may be legitimately affecting both gene expression and alternative splicing, however if a particular splice variant alters the ability of a gene to be efficiently measured, it appears that gene expression is altered, and the effect is driven by the splicing event.⁵⁸ Furthermore, some probes for mRNA will detect commonly splice variants as canonical transcripts, which are included in 'bulk tissue' eQTL analysis. This can lead to the false assumption that both expression and splicing events are driving the causal effect, when this is not the case.⁵⁹

407 Our analysis is limited to individuals with European ancestry; therefore, it will be important to
 408 extend these analyses to individuals of alternative ancestries as such data becomes available.⁶⁰
 409 Previous GWAS and observational studies in non-European cohorts have found novel associations
 410 which we did not find in our study e.g., *GSTP1*, which may illustrate that the loci appear to be
 411 population specific, or may be observational, and not causal.^{61,62}
 412 The nature of glioma can result in variation in inter-patient, intra-tumoural, subtype and
 413 spatiotemporal heterogeneity, which can lead to variation in tissue recovery at biopsy.^{63,64} These
 414 factors can result in deviation in the profile of disease, which can make our results more
 415 generalisable and less specific to one subtype of glioma. To make our analysis more precise, we
 416 would require more granular tissue and cell-specific data.

Conclusion

We conducted a robust multi-omic causal analysis for gene and protein molecular traits in brain tissue on glioma risk. We used tissue-specific molecular data and glioma subtype information to explore the nature of the causal relationships identified. Here, we combined MR with colocalisation and Steiger filtering to ensure the robustness of our results.

We provide robust multi-omic causal evidence for 11 previously implicated genes which affect the risk of glioma. We also present novel evidence for a causal effect of increased *HBEGF* expression in cortex on increased risk of all glioma, and present novel evidence for causal effects of *CEP192*, *FAIM* and *SLC8A1* splicing in multiple tissues on variable risk of all glioma and GB. Additionally, we show evidence for the causal effect of increased *HEATR3* expression and alternative splicing on increased risk of all glioma, although we propose that this effect is eQTL-specific.

We did not identify any causal proteins for glioma risk, likely due to the limited sample size of the currently available brain pQTL data.

We focussed on the causal effects of increased *HBEGF* expression and alternative splicing of *RTEL1* and use multiple investigate methods to provide evidence to understand the biological mechanisms that these genes may play in the risk of glioma.

434 Abbreviation List

435 **CNS** – Central nervous system

436 **DGIdb** – Drug gene interaction database

437 **eQTL** – Expression quantitative trait loci

438 **GB** – Glioblastoma

439 **GWAS** – Genome-wide association study

440 **LD** – Linkage disequilibrium

441 **LDLc** – Low-density lipoprotein cholesterol

442 **MR** – Mendelian randomisation

443 **OR** – Odds ratio

444 **pQTL** – Protein quantitative trait loci

445 **QTL** – Quantitative trait loci

446 **sQTL** – Splicing quantitative trait loci

447 **TPM** – Transcripts per Million

448 **WHO** – World Health Organisation

Declarations

Ethics approval and consent to participate

This study uses previously published summary-level data and thus contains no patient identifiable data. Ethical approval and informed consent from each participant were given and can be found where the dataset was initially described. All procedures performed in studies involving human participants were done in accordance with the ethical standards of the institutional or national research committee and with 1964 Helsinki declaration.

Consent for publication

Not applicable

Availability of data and materials

The MetaBrain meta-analyses eQTL data can be accessed at <https://www.metabrain.nl/>. The GTEx sQTL v8 data can be accessed at <https://gtexportal.org/home/datasets>. The BrainQTL pQTL data can be found at <https://www.synapse.org/#!Synapse:syn24172458>. The glioma data may be accessed under the European Genome-phenome Archive accession number EGAD00010001657 (<https://www.ebi.ac.uk/ega/datasets/EGAD00010001657>).

Competing interests

The authors declare they have no competing interests.

Funding

ZAT, LJA, LP and JWR receive support from the UK Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/4, MC_UU_00011/1). ZAT receives funding from Southmead Hospital Charitable Funds: Brain tumour bank and research fund 8036. LJA and KMK are funded by Cancer Research UK (grant number C30758/A29791). JWR received funding from Biogen for unrelated projects. KMK is funded by Innovate (grant number 10027624). This study was supported by the National Institute for Health and Care Research Bristol Biomedical Research

473 Centre (NIHR203315) and Cancer Research UK (grant number C18281/A29019 and C18281/A30905).
 474 The views expressed are those of the author(s) and not necessarily those of the NIHR or the
 475 Department of Health and Social Care.

476 Authors' contributions

477 ZAT wrote the manuscript, performed all statistical analysis and interpreted results; LJA and HZ
 478 revised the manuscript; JZ advised on statistical methodology; LP substantially revised the
 479 manuscript and assisted with statistical analysis; JWR substantially revised the manuscript, assisted
 480 with statistical analysis and interpreted results and designed the study; KMK designed the study. All
 481 authors have approved the submitted version of the manuscript.

482 Acknowledgements

483 We would like to acknowledge all patients who provided data that made this work possible.

References

1. Goodenberger, M.L., and Jenkins, R.B. (2012). Genetics of adult glioma. *Cancer Genet.* 205, 613-621. 10.1016/j.cancergen.2012.10.009.
2. Ostrom, Q.T., Bauchet, L., Davis, F.G., Deltour, I., Fisher, J.L., Langer, C.E., Pekmezci, M., Schwartzbaum, J.A., Turner, M.C., Walsh, K.M., et al. (2014). The epidemiology of glioma in adults: a "state of the science" review. *Neuro-Oncology* 16, 896-913. 10.1093/neuonc/nou087.
3. Rouse, C., Gittleman, H., Ostrom, Q.T., Kruchko, C., and Barnholtz-Sloan, J.S. (2016). Years of potential life lost for brain and CNS tumors relative to other cancers in adults in the United States, 2010. *Neuro-Oncology* 18, 70-77. 10.1093/neuonc/nov249.
4. Ostrom, Q.T., Gittleman, H., Liao, P., Vecchione-Koval, T., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J.S. (2017). CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro-Oncology* 19, V1-V88. 10.1093/neuonc/nox158.
5. Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* 131, 803-820. 10.1007/s00401-016-1545-1.
6. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology* 23, 1231-1251. 10.1093/neuonc/noab106.
7. Butler, M., Pongor, L., Su, Y.T., Xi, L.Q., Raffeld, M., Quezado, M., Trepel, J., Aldape, K., Pommier, Y., and Wu, J. (2020). *MGMT* Status as a Clinical Biomarker in Glioblastoma. *Trends Cancer* 6, 380-391. 10.1016/j.trecan.2020.02.010.

- 509 8. Hau, P., Stupp, R., and Hegi, M.E. (2007). MGMT methylation status: The advent of stratified
510 therapy in glioblastoma? *Dis. Markers* 23, 97-104. 10.1155/2007/159242.
- 511 9. Melin, B.S., Barnholtz-Sloan, J.S., Wrensch, M.R., Johansen, C., Il'yasova, D., Kinnersley, B.,
512 Ostrom, Q.T., Labreche, K., Chen, Y.W., Armstrong, G., et al. (2017). Genome-wide
513 association study of glioma subtypes identifies specific differences in genetic susceptibility to
514 glioblastoma and non-glioblastoma tumors. *Nature Genetics* 49, 789-+. 10.1038/ng.3823.
- 515 10. Atkins, I., Kinnersley, B., Ostrom, Q.T., Labreche, K., Il'yasova, D., Armstrong, G.N., Eckel-
516 Passow, J.E., Schoemaker, M.J., Nothen, M.M., Barnholtz-Sloan, J.S., et al. (2019).
517 Transcriptome-Wide Association Study Identifies New Candidate Susceptibility Genes for
518 Glioma. *Cancer Research* 79, 2065-2071. 10.1158/0008-5472.can-18-2888.
- 519 11. Schwartzbaum, J.A., Fisher, J.L., Aldape, K.D., and Wrensch, M. (2006). Epidemiology and
520 molecular pathology of glioma. *Nat. Clin. Pract. Neurol.* 2, 494-503. 10.1038/ncpneuro0289.
- 521 12. Saunders, C.N., Kinnersley, B., Culliford, R., Cornish, A.J., Law, P.J., and Houlston, R.S. (2022).
522 Relationship between genetically determined telomere length and glioma risk. *Neuro-*
523 *Oncology* 24, 171-181. 10.1093/neuonc/noab208.
- 524 13. Kinnersley, B., Houlston, R.S., and Bondy, M.L. (2018). Genome-Wide Association Studies in
525 Glioma. *Cancer Epidemiol. Biomarkers Prev.* 27, 418-428. 10.1158/1055-9965.epi-17-1080.
- 526 14. Haycock, P.C., Burgess, S., Nounu, A., Zheng, J., Okoli, G.N., Bowden, J., Wade, K.H., Timpson,
527 N.J., Evans, D.M., Willeit, P., et al. (2017). Association Between Telomere Length and Risk of
528 Cancer and Non-Neoplastic Diseases A Mendelian Randomization Study. *JAMA Oncol.* 3, 636-
529 651. 10.1001/jamaoncol.2016.5945.
- 530 15. Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and
531 disease. *Nat. Rev. Genet.* 16, 197-212. 10.1038/nrg3891.
- 532 16. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., De La Torre-Ubieta, L.,
533 Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic Control of Expression and

534 Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* 179, 750-+.
535 10.1016/j.cell.2019.09.021.

536 17. Eales, J.M., Jiang, X., Xu, X.G., Saluja, S., Akbarov, A., Cano-Gamez, E., McNulty, M.T., Finan,
537 C., Guo, H., Wystrychowski, W., et al. (2021). Uncovering genetic mechanisms of
538 hypertension through multi-omic analysis of the kidney. *Nature Genetics* 53, 630-+.
539 10.1038/s41588-021-00835-w.

540 18. Sun, Y.V., and Hu, Y.J. (2016). Integrative Analysis of Multi-omics Data for Discovery and
541 Functional Studies of Complex Human Diseases. In *Advances in Genetics*, Vol 93, T.
542 Friedmann, J.C. Dunlap, and S.F. Goodwin, eds. (Elsevier Academic Press Inc), pp. 147-190.
543 10.1016/bs.adgen.2015.11.004.

544 19. Smith, G.D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology
545 contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1-
546 22. 10.1093/ije/dyg070.

547 20. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and
548 Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association
549 Studies Using Summary Statistics. *PLoS Genet.* 10, 15, e1004383.
550 10.1371/journal.pgen.1004383.

551 21. de Klein, N., Tsai, E.A., Vochteloo, M., Baird, D., Huang, Y.F., Chen, C.Y., van Dam, S., Oelen,
552 R., Deelen, P., Bakker, O.B., et al. (2023). Brain expression quantitative trait locus and
553 network analyses reveal downstream effects and putative drivers for brain-related diseases.
554 *Nature Genetics*, 24. 10.1038/s41588-023-01300-6.

555 22. Aguet, F., Barbeira, A.N., Bonazzola, R., Brown, A., Castel, S.E., Jo, B., Kasela, S., Kim-
556 Hellmuth, S., Liang, Y.Y., Parsana, P., et al. (2020). The GTEx Consortium atlas of genetic
557 regulatory effects across human tissues. *Science* 369, 1318-1330. 10.1126/science.aaz1776.

558 23. Bennett, D.A., Schneider, J.A., Arvanitakis, Z., and Wilson, R.S. (2012). Overview and Findings
559 from the Religious Orders Study. *Curr. Alzheimer Res.* 9, 628-645.

- 560 24. Bennett, D.A., Schneider, J.A., Buchman, A.S., Barnes, L.L., Boyle, P.A., and Wilson, R.S.
561 (2012). Overview and Findings from the Rush Memory and Aging Project. *Curr. Alzheimer*
562 *Res.* 9, 646-663.
- 563 25. Robins, C., Liu, Y., Fan, W., Duong, D.M., Meigs, J., Harerimana, N.V., Gerasimov, E.S.,
564 Dammer, E.B., Cutler, D.J., Beach, T.G., et al. (2021). Genetic control of the human brain
565 proteome. *Am. J. Hum. Genet.* 108, 400-410. 10.1016/j.ajhg.2021.01.012.
- 566 26. Burgess, S., Small, D.S., and Thompson, S.G. (2017). A review of instrumental variable
567 estimators for Mendelian randomization. *Stat. Methods Med. Res.* 26, 2333-2355.
568 10.1177/0962280215597579.
- 569 27. Lawlor, D.A., Harbord, R.M., Sterne, J.A.C., Timpson, N., and Smith, G.D. (2008). Mendelian
570 randomization: Using genes as instruments for making causal inferences in epidemiology.
571 *Stat. Med.* 27, 1133-1163. 10.1002/sim.3034.
- 572 28. Staiger, D., and Stock, J.H. (1997). Instrumental variables regression with weak instruments.
573 *Econometrica* 65, 557-586. 10.2307/2171753.
- 574 29. Hemani, G., Bowden, J., and Smith, G.D. (2018). Evaluating the potential role of pleiotropy in
575 Mendelian randomization studies. *Hum. Mol. Genet.* 27, R195-R208. 10.1093/hmg/ddy163.
- 576 30. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring
577 population-specific haplotype structure and linking correlated alleles of possible functional
578 variants. *Bioinformatics* 31, 3555-3557. 10.1093/bioinformatics/btv402.
- 579 31. Elsworth, B.L., Lyon, M.S., Alexander, T., Liu, Y., Matthews, P., Hallett, J., Bates, P., Palmer, T.,
580 Haberland, V., Smith, G.D., et al. (2020). The MRC IEU OpenGWAS data infrastructure.
581 *bioRxiv*. 10.1101/2020.08.10.244293.
- 582 32. Mitchell, R., Elsworth, B.L., Mitchell, R., Raistrick, C.A., Paternoster, L., Hemani, G., and
583 Gaunt, T. (2019). MRC IEU UK Biobank GWAS pipeline version 2.
- 584 33. Kamat, M.A., Blackshaw, J.A., Young, R., Surendran, P., Burgess, S., Danesh, J., Butterworth,
585 A.S., and Staley, J.R. (2019). PhenoScanner V2: an expanded tool for searching human

586 genotype-phenotype associations. *Bioinformatics* 35, 4851-4853.

587 10.1093/bioinformatics/btz469.

588 34. Howell, A.E., Robinson, J.W., Wootton, R.E., McAleenan, A., Tsavachidis, S., Ostrom, Q.T.,

589 Bondy, M., Armstrong, G., Relton, C., Haycock, P., et al. (2020). Testing for causality between

590 systematically identified risk factors and glioma: a Mendelian randomization study. *Bmc*

591 *Cancer* 20, 11. 10.1186/s12885-020-06967-2.

592 35. Bowman, R.L., Wang, Q.H., Carro, A., Verhaak, R.G.W., and Squatrito, M. (2017). GlioVis data

593 portal for visualization and analysis of brain tumor expression datasets. *Neuro-Oncology* 19,

594 139-141. 10.1093/neuonc/now247.

595 36. Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Uriarte, A., Malangone, C.,

596 Miranda, A., Fumis, L., Carvalho-Silva, D., Spitzer, M., et al. (2021). Open Targets Platform:

597 supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* 49,

598 D1302-D1310. 10.1093/nar/gkaa1027.

599 37. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell,

600 N.H., Chavali, G., Chen, C., del-Toro, N., et al. (2014). The MIntAct project-IntAct as a

601 common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42,

602 D358-D363. 10.1093/nar/gkt1115.

603 38. Freshour, S.L., Kiwala, S., Cotto, K.C., Coffman, A.C., McMichael, J.F., Song, J.J., Griffith, M.,

604 Griffith, O.L., and Wagner, A.H. (2021). Integration of the Drug-Gene Interaction Database

605 (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res.* 49, D1144-D1151.

606 10.1093/nar/gkaa1084.

607 39. Hemani, G., Zhengn, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess,

608 S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal

609 inference across the human phenome. *eLife* 7, 29, e34408. 10.7554/eLife.34408.

610 40. RCoreTeam (2021). R: A Language and Environment for Statistical Computing.

611 <https://www.R-project.org/>.

612 41. RStudioTeam (2020). RStudio: Integrated Development for R. <http://www.rstudio.com/>.

613 42. Robinson, J.W., Hemani, G., Babaei, M.S., Huang, Y., Baird, D.A., Tsai, E.A., Chen, C.-Y.,

614 Gaunt, T.R., and Zheng, J. (2022). An efficient and robust tool for colocalisation: Pair-wise

615 Conditional and Colocalisation (PWCoCo). BioRxiv.

616 <https://doi.org/10.1101/2022.08.08.503158>

617 43. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G.,

618 Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-

619 SNP analysis of GWAS summary statistics identifies additional variants influencing complex

620 traits. *Nature Genetics* 44, 369-U170. 10.1038/ng.2213.

621 44. Medicine, N.C.f.B.I.N.L.o. HBEGF. <https://www.ncbi.nlm.nih.gov/gene/1839>.

622 45. Walsh, K.M., Codd, V., Smirnov, I.V., Rice, T., Decker, P.A., Hansen, H.M., Kollmeyer, T.,

623 Kosel, M.L., Molinaro, A.M., McCoy, L.S., et al. (2014). Variants near TERT and TERC

624 influencing telomere length are associated with high-grade glioma risk. *Nature Genetics* 46,

625 731-735. 10.1038/ng.3004.

626 46. Walsh, K.M., Wiencke, J.K., Lachance, D.H., Wiemels, J.L., Molinaro, A.M., Eckel-Passow, J.E.,

627 Jenkins, R.B., and Wrensch, M.R. (2015). Telomere maintenance and the etiology of adult

628 glioma. *Neuro-Oncology* 17, 1445-1452. 10.1093/neuonc/nov082.

629 47. Chen, Y.B., Wu, Y.S., Huang, X.J., Qu, P., Li, G., Jin, T.B., Xing, J.L., and He, S.M. (2015).

630 Leukocyte telomere length: a novel biomarker to predict the prognosis of glioma patients. *J.*

631 *Cancer Res. Clin. Oncol.* 141, 1739-1747. 10.1007/s00432-015-1938-x.

632 48. Gusev, Y., Bhuvaneshwar, K., Song, L., Zenklusen, J.C., Fine, H., and Madhavan, S. (2018). The

633 REMBRANDT study, a large collection of genomic data from brain cancer patients. *Sci. Data*

634 5, 9, 180158. 10.1038/sdata.2018.158.

635 49. Sarantopoulos, J., Mita, M.M., Birrer, M.J., Cranmer, L.D., Campos, L.T., Zhang, X.P., Bristow,

636 P., Kaito, H., Strout, V., and Camacho, L.H. (2016). Phase 1 Study of Monotherapy with

637 KHK2866, an Anti-Heparin-Binding Epidermal Growth Factor-Like Growth Factor Monoclonal

638 Antibody, in Patients with Advanced Cancer. *Target. Oncol.* *11*, 317-327. 10.1007/s11523-
639 015-0394-5.

640 50. Medicine, N.C.f.B.I.N.L.o. RTE1. <https://www.ncbi.nlm.nih.gov/gene/51750>.

641 51. Jin, K.L., Mao, X.O., Guerra, G.D., Jin, L., and Greenberg, D.A. (2005). Heparin-binding
642 epidermal growth factor-like growth factor stimulates cell proliferation in cerebral cortical
643 cultures through phosphatidylinositol 3'-kinase and mitogen-activated protein kinase. *J.*
644 *Neurosci. Res.* *81*, 497-505. 10.1002/jnr.20510.

645 52. Hatanpaa, K.J., Burma, S., Zhao, D.W., and Habib, A.A. (2010). Epidermal Growth Factor
646 Receptor in Glioma: Signal Transduction, Neuropathology, Imaging, and Radioresistance.
647 *Neoplasia* *12*, 675-684. 10.1593/neo.10688.

648 53. Wan, J., Ramachandran, R., and Goldman, D. (2012). HB-EGF Is Necessary and Sufficient for
649 Muller Glia Dedifferentiation and Retina Regeneration. *Dev. Cell* *22*, 334-347.
650 10.1016/j.devcel.2011.11.020.

651 54. Lemmon, M.A., and Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. *Cell*
652 *141*, 1117-1134. 10.1016/j.cell.2010.06.011.

653 55. Mishima, K., Higashiyama, S., Asai, A., Yamaoka, K., Nagashima, Y., Taniguchi, N., Kitanaka,
654 C., Kirino, T., and Kuchino, Y. (1998). Heparin-binding epidermal growth factor-like growth
655 factor stimulates mitogenic signaling and is highly expressed in human malignant gliomas.
656 *Acta Neuropathol.* *96*, 322-328. 10.1007/s004010050901.

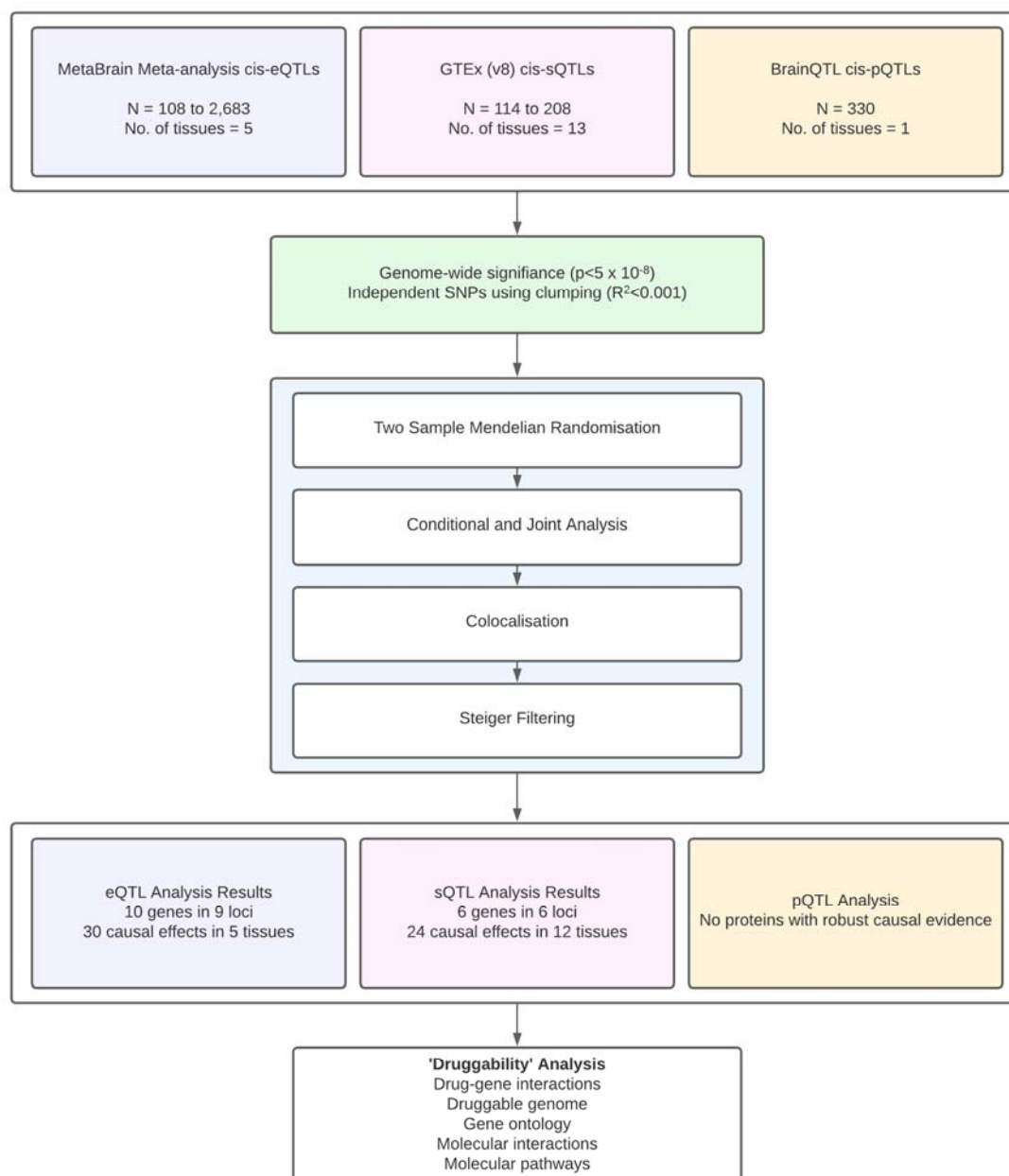
657 56. Network, T.C.G.A.T.R. The Cancer Genome Atlas (TCGA) Research Network.
658 <https://www.cancer.gov/tcga>.

659 57. Yamaguchi, K., Ishigaki, K., Suzuki, A., Tsuchida, Y., Tsuchiya, H., Sumitomo, S., Nagafuchi, Y.,
660 Miya, F., Tsunoda, T., Shoda, H., et al. (2022). Splicing QTL analysis focusing on coding
661 sequences reveals mechanisms for disease susceptibility loci. *Nat. Commun.* *13*, 13, 4659.
662 10.1038/s41467-022-32358-1.

- 663 58. Mucaki, E.J., Shirley, B., and Rogan, P.K. (2020). Expression Changes Confirm Genomic
664 Variants Predicted to Result in Allele-Specific, Alternative mRNA Splicing. *Front. Genet. 11*,
665 16, 109. 10.3389/fgene.2020.00109.
- 666 59. Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A.,
667 Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X.G., and Mortazavi, A. (2016). A survey of
668 best practices for RNA-seq data analysis. *Genome Biol. 17*, 19, 13. 10.1186/s13059-016-
669 0881-8.
- 670 60. Medina-Gomez, C., Felix, J.F., Estrada, K., Peters, M.J., Herrera, L., Kruithof, C.J., Duijts, L.,
671 Hofman, A., van Duijn, C.M., Uitterlinden, A.G., et al. (2015). Challenges in conducting
672 genome-wide association studies in highly admixed multi-ethnic populations: the Generation
673 R Study. *Eur. J. Epidemiol. 30*, 317-330. 10.1007/s10654-015-9998-4.
- 674 61. Li, G., Jin, T.B., Wei, X.B., He, S.M., Liang, H.J., Yang, H.X., Cui, Y., Chen, C., Cai, L.B., and Gao,
675 G.D. (2012). Selected polymorphisms of GSTP1 and TERT were associated with glioma risk in
676 Han Chinese. *Cancer Epidemiol. 36*, 525-527. 10.1016/j.canep.2012.06.008.
- 677 62. Li, N., Shi, H.Y., Hou, P.F., Gao, L., Shi, Y.Q., Mi, W.Y., Zhang, G., Wang, N., Dai, W., Wei, L., et
678 al. (2021). Genetic variants of CYP4F12 gene are associated with glioma susceptibility. *Int. J.*
679 *Cancer 149*, 1910-1915. 10.1002/ijc.33755.
- 680 63. Comba, A., Faisal, S.M., Varela, M.L., Hollon, T., Al-Holou, W.N., Umemura, Y., Nunez, F.J.,
681 Motsch, S., Castro, M.G., and Lowenstein, P.R. (2021). Uncovering Spatiotemporal
682 Heterogeneity of High-Grade Gliomas: From Disease Biology to Therapeutic Implications.
683 *Frontiers in Oncology 11*, 16, 703764. 10.3389/fonc.2021.703764.
- 684 64. Eisenbarth, D., and Wang, Y.A. (2023). Glioblastoma heterogeneity at single cell resolution.
685 *Oncogene 42*, 2155-2165. 10.1038/s41388-023-02738-y.

686

Figures



688

689 *Figure 1. Flowchart showing the pipeline used in the multi-omic analysis.*

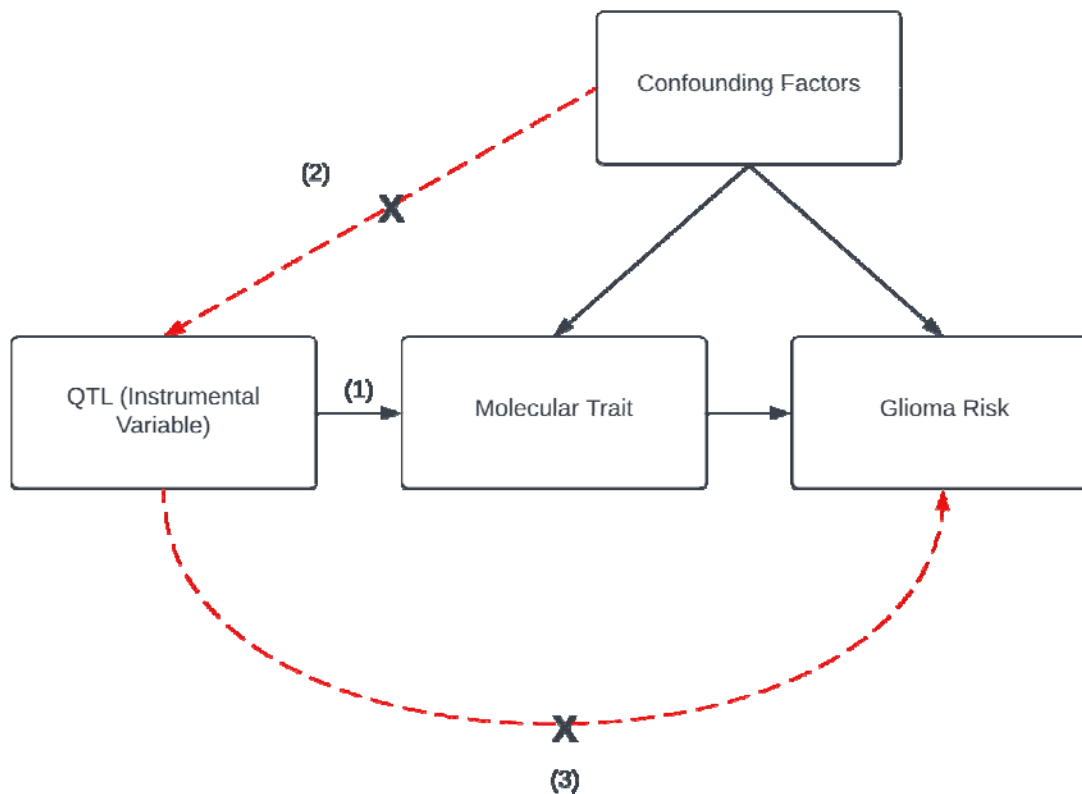
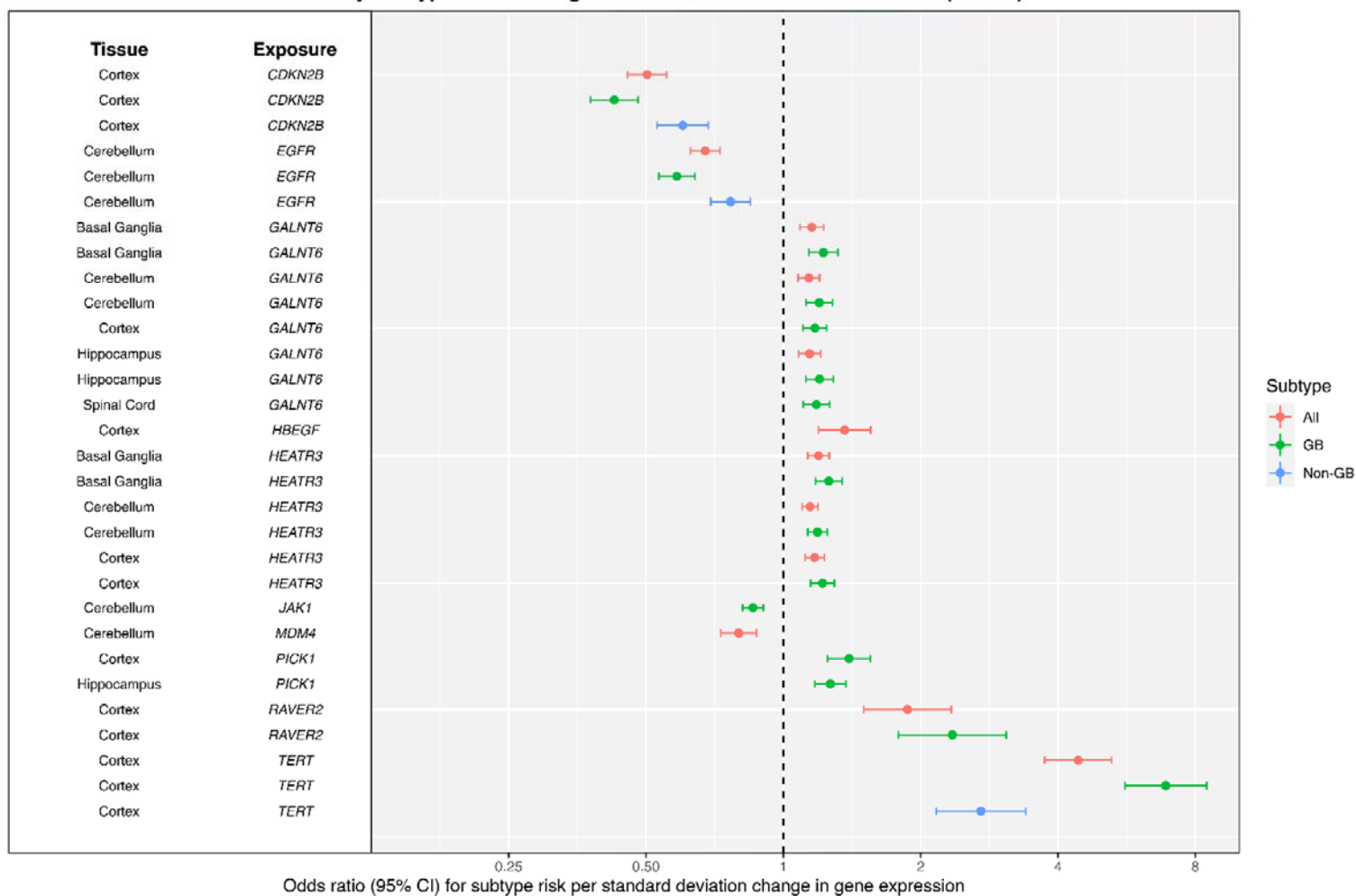


Figure 2. Directed Acyclic Graph (DAG) showing a visual representation of potential causal relationships. This DAG also shows the three assumptions of Mendelian randomisation. (1) The genetic variant(s) being used as an instrument is associated robustly with the exposure. (2) The instrument is independent of measured and unmeasured confounding factors of the association between exposure and outcome. (3) There must be no independent pathway between the instrument and outcome other than through the exposure.

Mendelian randomisation results by subtype for the ten genes of interest in five brain tissues (eQTLs)



696

697 Figure 3. Forest plot showing Mendelian randomisation results by subtype (all glioma, GB and non-GB) for the ten genes of interest in five brain tissues in the
698 gene expression analysis. Results have strong evidence of colocalization ($H_4 \geq 80\%$) and correct Steiger directionality.

Mendelian randomisation results by phenotype for the six genes of interest in 12 brain tissues (sQTLs)

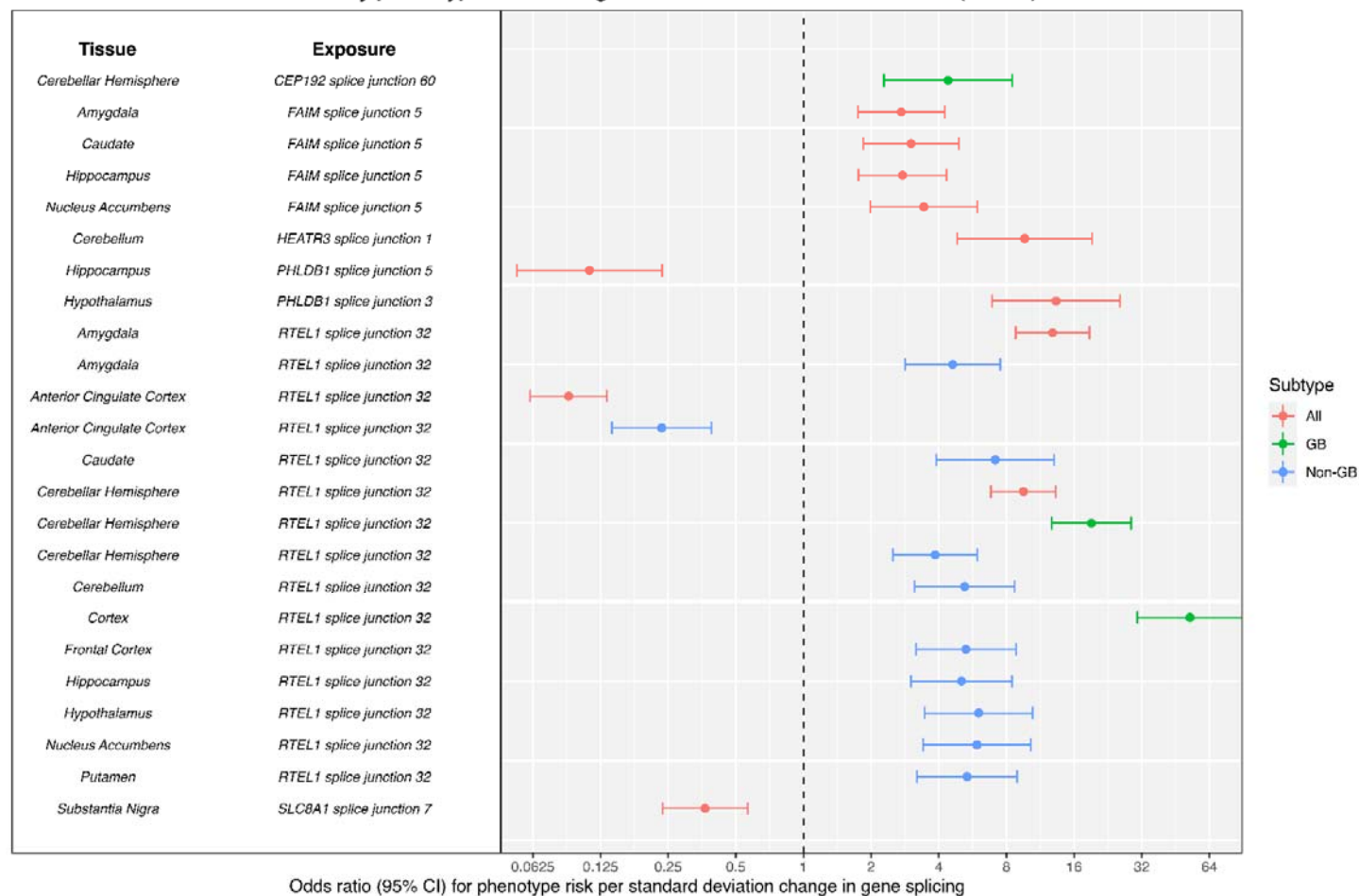


Figure 4. Forest plot showing Mendelian randomisation results by subtype (all glioma, GB and non-GB) for the six genes of interest in 12 brain tissues in the gene-splicing analysis. Results have strong evidence of colocalization ($H_4 \geq 80\%$) and correct Steiger directionality.

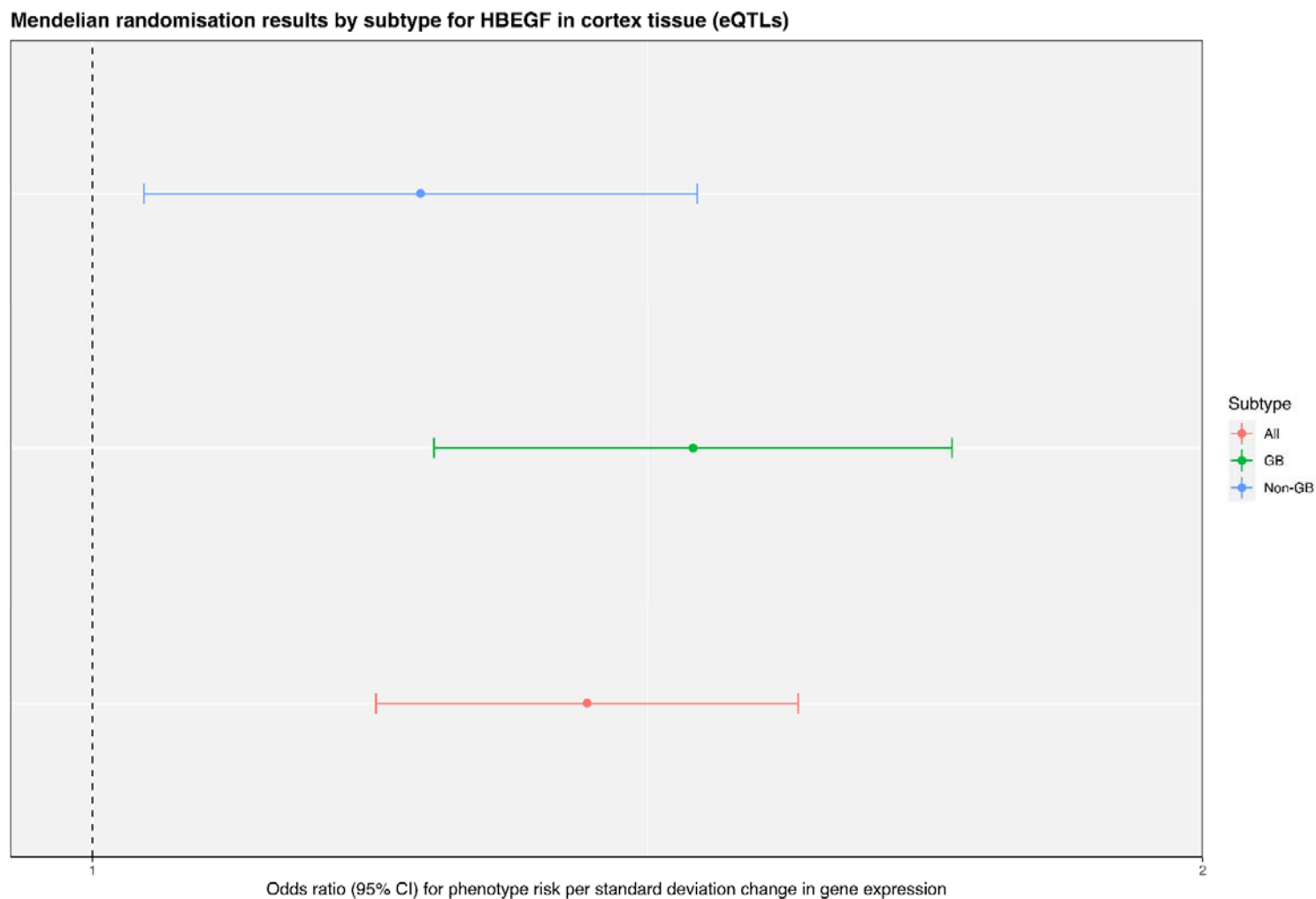


Figure 5. Forest plot showing Mendelian randomisation results by subtype (all glioma, GB and non-GB) for HBEGF in cortex tissue in the gene expression analysis. All subtypes show an increased risk of developing glioma, but only all glioma surpassed our Bonferroni-corrected P value threshold ($P = 5.00 \times 10^{-6}$).

Mendelian randomisation results by phenotype for the three genes of interest in six brain tissues (sQTLs)

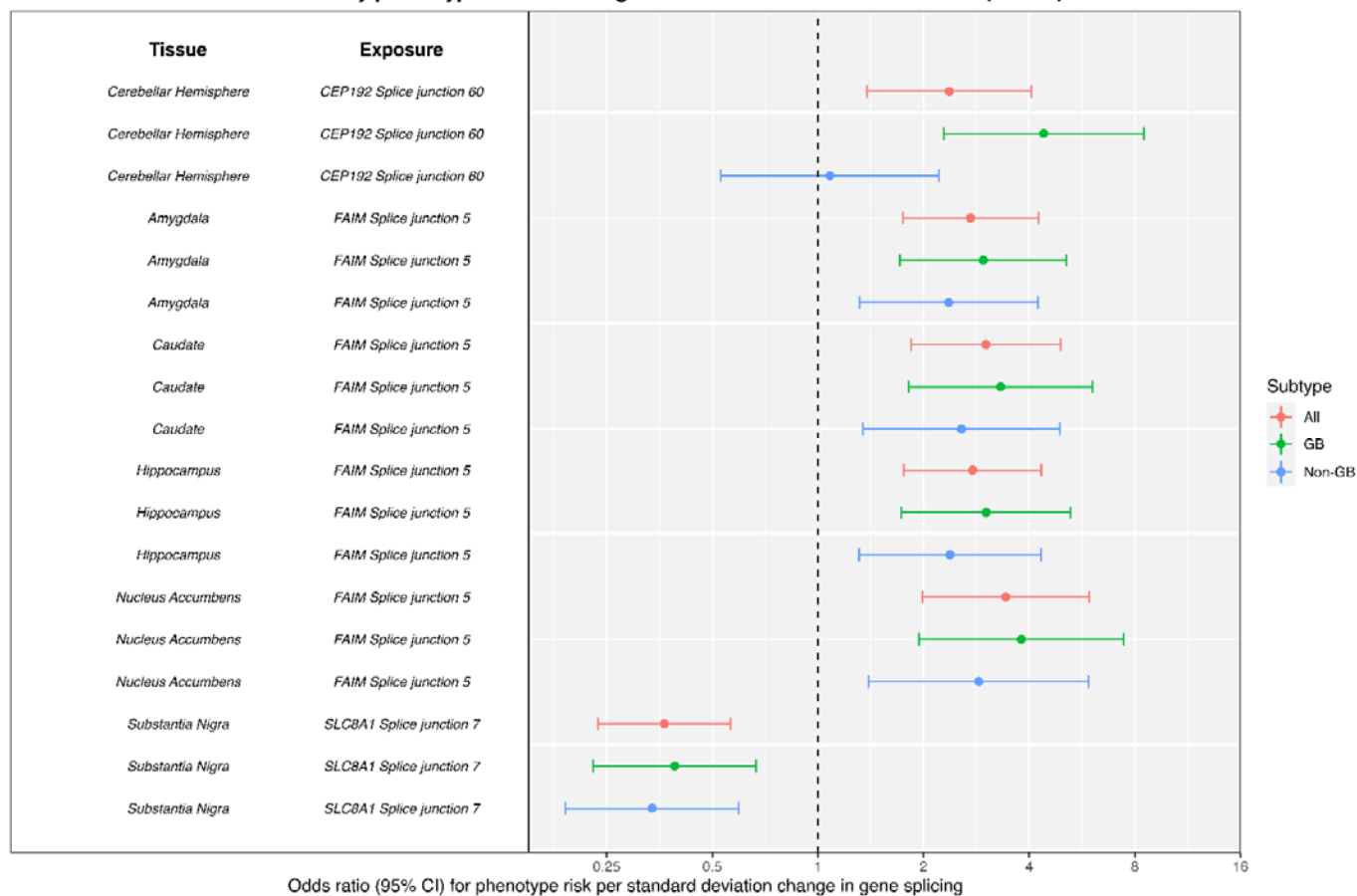


Figure 6. Forest plot showing Mendelian randomisation results by subtype (all, GB and non-GB) for CEP192, FAIM and SLC8A1 in six tissues in the gene-splicing analysis. All subtypes showed the same direction of effect on glioma risk within tissues, but not all passed our Bonferroni-corrected P value threshold ($P = 1.11 \times 10^{-5}$)

Tables

Table 1. Results of the MR analysis investigating the effects of genetically proxied gene expression levels on genetic liability to glioma subtype.

Tissue	Gene	Subtype	Method	No. of SNPs	P Value	Odds Ratio (95% CI)	H4 (%)	Steiger Direction	Steiger P Value	Steiger Flag
Cortex	<i>CDKN2B</i>	All	Wald ratio	1	7.17E-43	1.36 (1.19 - 1.55)	93	TRUE	2.32E-19	TRUE
Cortex	<i>CDKN2B</i>	GB	Wald ratio	1	6.22E-44	1.14 (1.08 - 1.21)	97	TRUE	3.19E-15	TRUE
Cortex	<i>CDKN2B</i>	Non-GB	Wald ratio	1	1.59E-14	1.22 (1.15 - 1.29)	92	TRUE	2.91E-21	TRUE
Cerebellum	<i>EGFR</i>	All	Wald ratio	1	1.07E-25	1.22 (1.14 - 1.32)	96	TRUE	2.16E-10	TRUE
Cerebellum	<i>EGFR</i>	GB	Wald ratio	1	2.89E-31	0.67 (0.63 - 0.73)	97	TRUE	4.29E-09	TRUE
Cerebellum	<i>EGFR</i>	Non-GB	Wald ratio	1	1.60E-07	0.58 (0.53 - 0.64)	98	TRUE	2.65E-11	TRUE
Basal Ganglia	<i>GALNT6</i>	All	Wald ratio	1	3.14E-06	0.5 (0.46 - 0.55)	100	TRUE	4.78E-09	TRUE
Basal Ganglia	<i>GALNT6</i>	GB	Wald ratio	1	9.08E-08	0.43 (0.38 - 0.48)	100	TRUE	1.23E-08	TRUE
Cerebellum	<i>GALNT6</i>	All	Wald ratio	1	3.14E-06	1.2 (1.12 - 1.28)	100	TRUE	7.94E-25	TRUE
Cerebellum	<i>GALNT6</i>	GB	Wald ratio	1	9.08E-08	1.2 (1.12 - 1.29)	100	TRUE	1.76E-23	TRUE
Cortex	<i>GALNT6</i>	GB	Inverse variance weighted	2	1.53E-07	1.19 (1.13 - 1.25)	100	TRUE	6.87E-130	TRUE
Hippocampus	<i>GALNT6</i>	All	Wald ratio	1	4.17E-06	6.89 (5.60 - 8.47)	99	TRUE	5.29E-08	TRUE
Hippocampus	<i>GALNT6</i>	GB	Wald ratio	1	2.15E-07	1.87 (1.50 - 2.34)	100	TRUE	1.09E-07	TRUE
Spinal Cord	<i>GALNT6</i>	GB	Wald ratio	1	8.40E-07	2.71 (2.16 - 3.40)	99	TRUE	6.88E-07	TRUE
Cortex	<i>HBEGF</i>	All	Wald ratio	1	4.41E-06	1.17 (1.12 - 1.23)	99	TRUE	8.18E-17	TRUE
Basal Ganglia	<i>HEATR3</i>	All	Wald ratio	1	1.50E-10	0.77 (0.69 - 0.85)	93	TRUE	7.61E-11	TRUE
Basal Ganglia	<i>HEATR3</i>	GB	Wald ratio	1	1.55E-11	0.60 (0.53 - 0.68)	95	TRUE	2.58E-11	TRUE
Cerebellum	<i>HEATR3</i>	All	Wald ratio	1	3.93E-11	1.15 (1.09 - 1.23)	96	TRUE	6.45E-54	TRUE
Cerebellum	<i>HEATR3</i>	GB	Wald ratio	1	6.74E-12	1.17 (1.11 - 1.24)	97	TRUE	1.25E-51	TRUE
Cortex	<i>HEATR3</i>	All	Wald ratio	1	1.95E-10	1.26 (1.18 - 1.34)	97	TRUE	1.15E-139	TRUE

Cortex	<i>HEATR3</i>	GB	Wald ratio	1	8.20E-11	1.14 (1.1 - 1.19)	97	TRUE	3.76E-125	TRUE
Cerebellum	<i>JAK1</i>	GB	Wald ratio	1	9.28E-09	1.14 (1.08 - 1.20)	89	TRUE	2.68E-38	TRUE
Cerebellum	<i>MDM4</i>	All	Wald ratio	1	9.54E-07	1.18 (1.11 - 1.26)	91	TRUE	2.10E-08	TRUE
Cortex	<i>PICK1</i>	GB	Wald ratio	1	1.92E-09	1.19 (1.13 - 1.26)	91	TRUE	2.14E-31	TRUE
Hippocampus	<i>PICK1</i>	GB	Wald ratio	1	3.95E-09	4.43 (3.74 - 5.25)	81	TRUE	1.65E-06	TRUE
Cortex	<i>RAVER2</i>	All	Wald ratio	1	3.19E-08	0.80 (0.73 - 0.87)	94	TRUE	8.11E-06	TRUE
Cortex	<i>RAVER2</i>	GB	Wald ratio	1	7.92E-10	0.86 (0.81 - 0.90)	98	TRUE	1.42E-04	TRUE
Cortex	<i>TERT</i>	All	Wald ratio	1	2.32E-66	1.27 (1.17 - 1.37)	100	TRUE	3.55E-03	TRUE
Cortex	<i>TERT</i>	GB	Wald ratio	1	6.92E-75	1.39 (1.25 - 1.55)	100	TRUE	1.18E-01	UNCERTAIN
Cortex	<i>TERT</i>	Non-GB	Wald ratio	1	6.69E-18	2.35 (1.79 - 3.08)	100	TRUE	3.26E-05	TRUE

712

713 *Results presented are those which were deemed the most robust, i.e., met the Bonferroni-correct P value threshold ($P = 5.00 \times 10^{-6}$) for the MR analysis, had*
714 *strong evidence of colocalisation ($H4 > 80\%$) and correctly orientated Steiger filtering direction.*

715 Table 2. Results of the MR analysis investigating the effects of genetically proxied gene splicing levels on genetic liability to glioma subtype.

Tissue	Gene (Splice Junction)	Subtype	Method	No. of SNPs	P Value	Odds Ratio (95% CI)	H4 (%)	Steiger Direction	Steiger P Value	Steiger Flag
Amygdala	<i>FAIM Splice Junction 5</i>	All	Wald ratio	1	1.03E-05	2.72 (1.74 - 4.25)	97	TRUE	6.91E-01	UNCERTAIN
Amygdala	<i>RTEL1 Splice Junction 32</i>	All	Wald ratio	1	4.38E-40	12.83 (8.80 - 18.72)	100	TRUE	9.68E-01	UNCERTAIN
Amygdala	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	8.36E-10	4.61 (2.83 - 7.51)	98	TRUE	7.34E-01	UNCERTAIN
Anterior Cingulate Cortex	<i>RTEL1 Splice Junction 32</i>	All	Wald ratio	1	3.89E-33	0.09 (0.06 - 0.13)	94	TRUE	9.46E-01	UNCERTAIN
Anterior Cingulate Cortex	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	2.39E-08	0.23 (0.14 - 0.39)	92	TRUE	7.24E-01	UNCERTAIN
Caudate	<i>FAIM Splice Junction 5</i>	All	Wald ratio	1	1.09E-05	3.01 (1.84 - 4.92)	97	TRUE	6.83E-01	UNCERTAIN
Caudate	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	1.56E-10	7.13 (3.91 - 13.02)	100	TRUE	8.68E-01	UNCERTAIN
Cerebellar Hemisphere	<i>CEP192 Splice Junction 60</i>	GB	Wald ratio	1	9.78E-06	4.40 (2.28 - 8.48)	97	TRUE	7.79E-01	UNCERTAIN
Cerebellar Hemisphere	<i>RTEL1 Splice Junction 32</i>	GB	Wald ratio	1	1.58E-45	19.1 (12.7 - 28.72)	100	TRUE	7.79E-01	UNCERTAIN
Cerebellar Hemisphere	<i>RTEL1 Splice Junction 32</i>	All	Wald ratio	1	4.38E-40	9.52 (6.82 - 13.28)	100	TRUE	8.87E-01	UNCERTAIN
Cerebellar Hemisphere	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	8.36E-10	3.85 (2.51 - 5.93)	95	TRUE	6.23E-01	UNCERTAIN
Cerebellum	<i>HEATR3 Splice Junction 1</i>	All	Wald ratio	1	1.29E-10	9.64 (4.83 - 19.24)	91	TRUE	9.38E-01	UNCERTAIN

Cerebellum	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	2.32E-10	5.21 (3.13 - 8.68)	99	TRUE	7.43E-01	UNCERTAIN
Cortex	<i>RTEL1 Splice Junction 32</i>	GB	Wald ratio	1	4.42E-47	52.42 (30.59 - 89.81)	100	TRUE	6.97E-01	UNCERTAIN
Frontal Cortex	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	2.00E-10	5.28 (3.16 - 8.82)	99	TRUE	7.94E-01	UNCERTAIN
Hippocampus	<i>FAIM Splice Junction 5</i>	All	Wald ratio	1	1.09E-05	2.76 (1.76 - 4.34)	97	TRUE	6.64E-01	UNCERTAIN
Hippocampus	<i>PHLDB1 Splice Junction 5</i>	All	Wald ratio	1	7.83E-09	0.11 (0.05 - 0.24)	87	TRUE	9.29E-01	UNCERTAIN
Hippocampus	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	8.36E-10	5.05 (3.01 - 8.46)	96	TRUE	8.08E-01	UNCERTAIN
Hypothalamus	<i>PHLDB1 Splice Junction 3</i>	All	Wald ratio	1	8.40E-15	13.29 (6.92 - 25.55)	99	TRUE	9.94E-01	UNCERTAIN
Hypothalamus	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	2.00E-10	6.02 (3.46 - 10.47)	99	TRUE	8.12E-01	UNCERTAIN
Nucleus Accumbens	<i>FAIM Splice Junction 5</i>	All	Wald ratio	1	1.03E-05	3.43 (1.98 - 5.93)	97	TRUE	7.29E-01	UNCERTAIN
Nucleus Accumbens	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	2.32E-10	5.92 (3.41 - 10.25)	99	TRUE	8.16E-01	UNCERTAIN
Putamen	<i>RTEL1 Splice Junction 32</i>	Non-GB	Wald ratio	1	1.56E-10	5.35 (3.2 - 8.93)	100	TRUE	7.66E-01	UNCERTAIN
Substantia Nigra	<i>SLC8A1 Splice Junction 7</i>	All	Wald ratio	1	5.72E-06	0.37 (0.24 - 0.56)	98	TRUE	7.01E-01	UNCERTAIN

716

717 Results presented here are those which were deemed the most robust, i.e., met the Bonferroni-correct P value threshold (1.11×10^{-5}) for the MR analysis, had
718 strong evidence of colocalisation ($H4 > 80\%$) and correctly orientated Steiger filtering direction.