Multi-omics Mendelian randomisation using expression, protein and splicing quantitative trait loci to identify novel drug targets associated with risk of gliomagenesis

Authors and Affiliations

Zak A. Thornton, [1 2 3] Lily J. Andrews, [1 2 3] Huiling Zhao, [1] Jie Zheng, [1 4 5] Jamie W. Robinson, [1]† Kathreena M Kurian, [1 2 3 6]†

[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, United Kingdom

[3] Cancer Research Integrative Cancer Epidemiology Programme, University of Bristol, Bristol, United Kingdom

[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] Brain Tumour Research Centre, Bristol Medical School, University of Bristol, Bristol, United Kingdom

† Co-senior author

* Corresponding author: Kathreena Kurian (kathreena.kurian@bristol.ac.uk); +44 117 4142405

Abstract

Genetically predicted changes in molecular traits are associated with genetic liability to glioma risk. We sought to identify and provide evidence for the prioritisation of these traits using a combined two-sample Mendelian randomisation and colocalisation pipeline.

We used genetic variants from genome-wide association studies (GWAS) to compare the effects of gene expression, protein abundance and splicing variation on genetic liability to glioma risk.

We performed analysis of quantitative trait loci (QTL) (expression QTL [MetaBrain, n=4,119,012]; protein QTL [BrainQTL, n=786,632] and splicing QTL [GTEx, n=166,721]) derived from 15 different brain tissues and summary-level data from eight glioma GWAS (12,496 cases and 18,190 controls).

We found robust evidence linking QTL for 25 genes in 15 loci with risk of glioma following relevant sensitivity analysis. We identified four novel genes putatively implicated in glioma risk: *BTN3A2* (6p22.2), *FAIM* (3q22.3), *GALNT6* (12q13.13) and *GMPPB* (3p21.31). Identification of these novel genes may improve understanding in pathogenic mechanisms underlying gliomagenesis and should be followed up in further studies, which may inform novel prevention studies following future clinical trials.

Introduction

Glioma makes up the majority (~80%) of malignant brain and central nervous system (CNS) tumours, with age-standardised incidence of glioma ranging from 4.67 to 5.73 per 100,000. [1, 2] Malignant brain tumours have the greatest years of potential life lost (YPLL) of all malignancies at an average of 19.33 in males, second only to other malignant CNS tumours. [3] Currently, gliomas are diagnosed and assessed using brain scans (e.g., MRI and CT), and histological and molecular profiling. [4] Standard treatment uses surgery and chemotherapy (typically temozolomide), adjuvant radiotherapy and pharmaceutical intervention such as *BRAF* inhibitors. [4-5] There is an urgent need for better understanding of gliomagenesis to inform novel therapeutic and preventative strategies because the response and subsequent progression free survival and overall survival of patients is limited.

Glioblastoma (GB, WHO grade 4), is a particularly aggressive form of glioma. [6] Typical hallmarks of GB include an increased capacity to infiltrate surrounding healthy tissue, abnormal vasculature, necrosis and a poor five-year survival rate (~4%). [7-8] According to the latest 2021 WHO classification, the group of lower grade non-glioblastoma (non-GB) gliomas, astrocytoma and oligodendroglioma, tend to be less malignant and are characterised by a Isocitrate Dehydrogenase (IDH) mutation. [6] These subtypes have varied prognosis and response to therapy, but overall tend to have a better survival rate than GB. [9-11]

Genetic variants alone are unable to explain the developing risk of disease. Quantitative trait loci (QTL) are regions of the genome which are associated with a particular quantitative (generally, molecular) trait. Examples of QTL include for gene expression (eQTL), protein abundance (pQTL) and splicing variation (sQTL). Integrating multi-dimensional omics data can allow us to better understand the development of complex traits and drug target-related analyses by providing functional and mechanistic insight into the genome. [12-14]

Previous genetics and MR studies have identified some candidate risk factors and genes associated with glioma risk. [15-18] In this study, we utilised Mendelian Randomisation (MR), an established statistical method which measures variation in genes of known function to assess causal relationship between genetically proxied gene expression, protein abundance and alternative splicing on glioma risk. [19] The main advantage MR has over traditional observational epidemiological methods is that MR can determine causality between an exposure and an outcome. Furthermore, MR is less liable to common epidemiological biases, such as confounding and reverse causality; however, for biases which MR does not account for, relevant sensitivity analyses can establish whether results are robust and not due to these biasing factors. One such method is colocalisation, which is used to identify if a genetic variant is shared by two traits, and is a necessary condition for causality. [20]

We utilised QTL data from different brain tissues in a combined MR-colocalisation framework to investigate how genetically proxied gene expression, protein abundance and alternative splicing may differentially associate with genetic liability to glioma risk. We combined our results with evidence from the literature to identify potential novel drug targets for prevention of glioma and to prioritise molecular traits of interest in further pre-clinical studies.

Methods

Study Datasets

We extracted QTL from GWAS which derived gene expression, protein abundance and alternative splicing levels in various brain tissues (Table 1). [21-24]

We used eQTL data produced by MetaBrain meta-analyses, which was collected from 14 eQTL datasets, composed of 6,518 individuals (Table S1). The data was derived from five CNS tissues (basal ganglia, cerebellum, cortex, hippocampus and spinal cord) from 2,970 individuals. Sample sizes varied from tissue-to-tissue (n = 108 to 2,970). [21]

sQTL data was produced by GTEx (v8 release). We extracted tissue-specific sQTL from 13 independent regions of the central nervous system (amygdala, anterior cingulate cortex (BA24), caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord and substantia nigra) from 209 non-diseased individuals. Sample sizes varied from tissue-to-tissue (n = 114 to 209). [22]

We used pQTL data retrieved from BrainQTL derived from the religious orders study (ROS) and memory and ageing projects (MAP). [25-26] pQTL were extracted from the dorsolateral pre-frontal cortex (DLPFC) tissue from a total of 330 individuals, across both studies. [23] Additionally, we retrieved pQTL data from the National institute on ageing genetics of Alzheimer's disease data study (NIAGADS), which was collected from the parietal lobe tissue of 380 individuals. [24]

The glioma outcome data was derived from eight glioma GWAS (UK-GWAS, French-GWAS, German-GWAS, MDA-GWAS, SFAGS, GliomaScan, GICC and UCSF-Mayo) and consisted of 12,496 cases (6,191 GB and 6,305 non-GB) and 18,190 controls (Table S2). [27] Data was categorised as 'combined glioma', and further sub-categorised into GB and non-GB, depending on subtype. The 'non-GB' group used in this analysis was not defined, and thus we assumed that it was composed of the subtypes we had previously stated.

This MR study uses previously published summary-level data and thus contains no patient identifiable data. Ethical approval and informed consent from each participant was 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

Our instrument selection pipeline is detailed in Figure 1. QTL were categorised into *cis*-acting or *trans*-acting, defined as a SNP which is inside or outside a 1Mb window of the gene regulatory region, respectively. *Trans*-acting SNPs were excluded from the analysis because of an increased likelihood of horizontal pleiotropy due to their distant location within the genome. Instruments were constructed from independent ($r^2 < 0.001$). *Cis*-acting SNPs met genome-wide significance ($P < 5 \times 10^{-8}$). The number of instruments that met this criteria is available in Table S3. Following MR analysis, the results had to meet a Bonferroni-corrected P value threshold which depended on the number of instruments (or tests) performed in that tissue. For example, we imposed a Bonferroni P value threshold of 2.52×10^{-5} (0.05/1,981) on basal ganglia derived eQTL (Table S3).

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, which included all of our instruments. We used the TwoSampleMR R package to implement the MR analyses. [28-31] MR estimates are presented throughout as odds ratios (OR) and were scaled to reflect one standard deviation increase in the molecular trait of interest.

MR has three assumptions which must hold to produce an unbiased estimate, which are that the genetic instrument: 1) must associate with the exposure ("relevance"), 2) does not share a common cause with the outcome ("independence" or "exchangeability"), and 3) affects the outcome only through the effect of the risk factor of interest ("exclusion restriction") (Figure 2).

We performed the following tests to ensure the validity of the MR results. The relevance assumption was tested by generating the F-statistic for each instrument, where an F-statistic > 10 is evidence against weak instrument bias. [32-33] The independence assumption was tested by investigating whether there was causal evidence linking the molecular trait of interest with putative risk factors for glioma. Risk factors we investigated were genetic liability to: allergies, asthma, altered low-density lipoprotein cholesterol (LDLc) levels and alcohol consumption (Table S4). [2, 34] We selected these risk factors based on whether there is evidence linking that risk factor to glioma in the literature, if there are data available for that risk factor in the OpenGWAS database, and whether those data were available for use in an MR analysis. [28, 35-37] Finally, the exclusion restriction assumption, also known as horizontal pleiotropy, is difficult to assess with single SNP instruments, as is common for molecular traits. Therefore, we tested this assumption by using colocalisation, which evaluates whether two traits 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. [38]

Colocalisation

We conducted colocalisation analyses to test for colocalisation between the molecular trait of interest and glioma subtype for each MR result which passed the Bonferroni-corrected P value threshold. We extracted *cis* regions within ±500Kbp around the instrumented SNP for each QTL which we analysed using Pair Wise Conditional Analysis and Colocalisation (PWCoCo), a package which combines Conditional and Joint Analysis (GCTA-COJO) and colocalisation (Coloc). [20, 39, 40] PWCoCo performs a joint analysis to determine independent signals within a region before systematically conditioning and performing colocalisation on each configuration of independent signal. This allows for the single-variant assumption of the "coloc" packages colocalisation analyses to hold. [28] The number of SNPs included in the conditional analysis are given in Table S5a and S6a.

As tissues differed between the eQTL and sQTL datasets, we used the RNA tissue grouping described in de Klein et al. to match the tissues. [21] Cerebellar hemisphere was matched with cerebellum, putamen was matched with basal ganglia and frontal cortex and anterior cingulate cortex were matched with cortex. Nucleus accumbens was initially matched with the hypothalamus in the eQTL dataset, but due to small sample size (<100), *cis*-eQTL were not calculated and removed from analysis. For the purpose of our analysis, we matched nucleus accumbens with basal ganglia; the nucleus accumbens forms the majority part of the ventral striatum, which is a principal component of the basal ganglia. We excluded the *FAIM* sQTL in the amygdala from this analysis, as we could not match this with a tissue in the eQTL dataset. Amygdala was similarly excluded from the MetaBrain analysis due to small sample size (<100). We performed this between-trait analysis to determine whether the eQTL and sQTL are driven by distinct causal pathways, as both QTL are measured using

quantities of mRNA. This can result in a 'double counting' effect, as they would be measuring the same thing.

Steiger Filtering

Finally, we applied Steiger filtering methods to test for the presence of reverse causation, where the outcome has an influence on the exposure. [38] Results of Steiger filtering are presented as a one of three categorical variables; true if the direction of the effect is from exposure to outcome and $P<0.05$, false if the direction of the effect is from outcome to exposure and $P<0.05$, and uncertain if $P\geq0.05$, regardless of the direction of effect (Table S5a and S6a).

Evidence thresholds for multi-omic analyses

Our main results comprised of those which passed the Bonferroni correction P value, had strong colocalisation evidence ($H_4\geq80\%$) and passed Steiger filtering sensitivity analysis with certainty. These results have the most robust evidence with a reduced likelihood of bias e.g., by LD structure or horizontal pleiotropy, due to strong evidence of colocalisation and passing Steiger filtering sensitivity analysis.

Enrichment Analysis

We performed an enrichment analysis to build evidence for the prioritisation of proteins, whose coding genes had robust MR and coloc evidence, as candidate drug 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 (molecular functions, biological pathways and cellular components), pathways and gene interactors. [41-42] We also used Drug Gene Interaction Database (DGIdb) to search for drug-gene interactions and genome 'druggability'. [43]

Data Availability

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 NIAGADS pQTL data can be found at https://www.niagads.org/datasets/ng00102. The glioma data may be accessed under the European Genome-phenome Archive accession number EGAD00010001657 (https://www.ebi.ac.uk/ega/datasets/EGAD00010001657).

Results

Estimating the causal effects of gene expression on glioma risk

The pipeline used for our analyses can be found in Figure 1. In total, our MR analyses of brain eQTL identified 62 associations that met the tissue-dependent Bonferroni corrected *P* value threshold ($P < 4.34 \times 10^{-5}$ to 2.06×10^{-6}, depending on tissue; Table S3) for 23 genes associated with risk of combined glioma, GB or non-GB (Table S5a).

The 62 associations consisted of the tissue in which the QTL was derived and the glioma subtype for which the association was measured. We refer to these associations as tissue-subtype associations. For example, we found 38 eQTL in the cortex, 10 in the cerebellum, six in the basal ganglia, five in the hippocampus and three in the spinal cord. Six genes had associations in multiple tissues - *GALNT6, GMEB2, GMPPB, HEATR3, PICK1* and *POLR3K;* all other genes had results in only one tissue (Table S5d & S5e). Figures 3-5 shows Manhattan plots for MR results for eQTL from the cortex.

Of the 62 tissue-subtype associations, we found that 51 had strong evidence of colocalisation (H_4>80%), four had moderate evidence of colocalisation (80% > H_4 > 50%) and seven had weak evidence of colocalisation (H_4 < 50%). All eQTL were tested for the potential of reverse causality using Steiger filtering; 60 tissue-subtype associations showed the correct direction of effect from exposure to outcome, whilst two were uncertain due to P-value (P > 0.05).

In summary, 51 tissue-subtype MR associations for 20 genes showed robust causal evidence from the MR and sensitivity analyses. Full details of these associations can be found in Table 2. All results for eQTL sensitivity analysis are described in Table S5.

We found four genes in the eQTL analysis (*GALNT6, GMPPB, HEATR3, POLR3K*) which had an association with glioma risk within different brain tissues. In all of the genes identified in the eQTL analysis, the association in each tissue was instrumented by a different lead SNP. Three of the four genes showed both an increase and decrease in the odds ratio (OR) of glioma risk, dependent on SNP. However, there was no significant differences in the risk between subtypes instrumented by the same SNP, although only the combined glioma dataset, and non-GB subtypes were implicated (Table 2).

Figure 6 shows the robust MR effect estimates for the 51 associations in 20 genes (in eQTL) in the combined glioma dataset, and GB and non-GB subtypes. The direction of the estimated causal effects broadly agreed across subtypes where more than one was implicated.

Estimating the causal effects of alternative splicing on glioma risk

We identified 72 tissue-subtype associations that the met tissue-dependent Bonferroni corrected P-value threshold ($P < 2.42 \times 10^{-5}$ to 8.79×10^{-6}, depending on tissue; Table S3) for nine genes associated with the combined glioma dataset, or GB or non-GB (Table S6). The breakdown of sQTL by tissue-subtype and by gene can be found in Table S6d and 6e, respectively. In total, nine tissue-subtype associations had strong evidence of colocalisation (H_4 > 80%). Steiger filtering was likely underpowered for this dataset, as none of the results showed the correct directionality of effect (32 tissue-subtype associations) or did not reach nominal significance (40 tissue-subtype associations). However, seven associations passed the Steiger filtering directionality test and had strong colocalisation results; we kept these associations as part of our main results. Full details of these associations can be found in Table 3. All results for the sQTL sensitivity analysis are described in Table S6.

Figure 7 shows the MR effect estimates for the nine associations in the three genes (in sQTL) in the combined glioma dataset, and non-GB subtypes.

Estimating the causal effects of protein abundance on glioma risk

Using the brain tissue-derived pQTL from the BrainQTL and NIAGADS datasets, we found that none of the tissue-subtype MR associations met the Bonferroni corrected P value threshold ($P < 2.70 \times 10^{-5}$). We performed a *post-hoc* analysis using a more lenient P value threshold for selection of pQTL ($P < 5 \times 10^{-5}$), with the limitation that these results may be biased due to, for example, weak instrument bias. Despite including instruments for 618 proteins, none of these results met the Bonferroni corrected P value threshold (Table S7).

Determining potential overlap of eQTL and sQTL

We tested for colocalisation between the molecular traits to establish whether the results are driven by the same lead SNP(s). Of the six regions tested, five were located in *IFT46* and one was located in *RTEL1*, and all associations were present in the non-GB subtype. All associations showed weak evidence of colocalisation (Table S8).

Enrichment Analysis

We identified drugs that currently target the genes identified in our analyses. Of the genes with robust tissue-subtype associations, seven genes (*AKT3, CDKN2A, CDKN2B, EGFR, JAK1, MDM4, TERT*) currently had drugs which were clinically available or in phase 1+ trials (Table S9a). [43]

Furthermore, we then analysed genes which had been categorised as part of the 'druggable genome'. Similarly, seven genes were annotated as part of the druggable genome, including one novel gene in the context of the glioma (*BTN3A2)* (Table S9b and S10a).

Sensitivity Analysis

For those molecular traits which had strong MR, colocalisation and Steiger filtering evidence, we performed a follow up sensitivity analysis to determine whether those molecular traits also had causal evidence with putative glioma risk factors. We followed the STROBE-MR guidelines to adhere to best practice. [44] We found that 58 molecular traits had strong MR evidence with evidence with the risk factors for glioma (Table 2 & 3).

Discussion

In this study, we used a combined MR-colocalisation framework to estimate the effect of genetically proxied gene expression, protein abundance and alternative splicing levels on genetic liability of glioma subtype risk. We found robust evidence that causally implicated 21 genes with genetic liability to glioma risk – four of which were not previously reported in the literature.

Novel Genes

We compared our results with previously identified GWAS associations with glioma risk (noted in a review by Kinnersley et al. [18]). We found that *BTN3A2* (6p22.2), *FAIM* (3q22.3), *GALNT6* (12q13.13) and *GMPPB* (3p21.31) are putative novel genes implicated in glioma risk that are also not located on a known glioma risk locus. *BTN3A2* (Butyrophilin subfamily 3 member A2) is a protein-coding gene which has roles in the T-cell responses of adaptive immune system. It also has higher levels of RNA expression in other immune-related cells such as NK-cells and dendritic cells. [45] It resides in the juxta-telomeric region of the major histocompatibility class 1 locus, which is renowned for marked genetic variation between individuals. [46] Previous MR has found evidence of a variant in *BTN3A2* which increases the risk of schizophrenia, highlighting a potential role in the development of neurological disease. [47] *FAIM* (Fas apoptotic inhibitory molecule) is a protein-coding gene that has a role in mediating Fas resistance produced by surface immunoglobulin engagement in B cells. It is involved in pathways associated with inhibiting cell death processes. [48] According to the Human Protein Atlas, there is low tissue specificity in the brain, but markedly high RNA expression in cells of the retina, especially rod photoreceptor cells. [49] *GALNT6* (Polypeptide N-acetylgalactosaminyltransferase-6) is a protein-coding gene which encodes an enzyme in the GalNAc-T family. This family of enzymes is responsible for the transfer of an N-acetyl-D-galactosamine residue to threonine or serine residues on protein receptors. [50] Song et al. found that *GALNT6* promotes invasion and metastasis of lung adenocarcinoma through O-glycosylating chaperone protein GRP78, which is found to be highly upregulated in the cell membrane of glioma cells. [51, 52] *GMPPB* (GDP-mannose pyrophosphorylase B) is a protein-coding gene that encodes an enzyme which catalyses the conversion of mannose-1-phosphate to GTP and GDP-mannose, and is involved in the production of N-linked oligosaccharides. It has low tissue and cell specificity, but has higher expression in plasma B cells. It has previously been associated with other pathologies such as muscular dystrophy and Walker-Warburg syndrome. [53] Further research such as protein function studies into these genes is needed to explain potential pathways by which these genes affect glioma risk.

Subtype Analysis

When estimating variations in risk across glioma subtype, our MR results varied in the direction of effect for risk of the combined glioma dataset, and GB and non-GB. In the eQTL analysis, there was a large variation in the estimated effect OR (0.21 to 7.37). In the sQTL analysis, the results only found a reduction in risk (OR: 0.08 to 0.36), but there were no MR associations in the GB subtype. Of the 21 genes, there were five which were exclusive to GB (*GMPPB, PICK1, POLR2F, POLR3K, SLC16A8*) and three which were exclusive to non-GB (*AKT3, BTN3A2, IFT46*). There was no marked differences in the magnitude of the OR between subtypes likely due to the similar sample sizes between GB and non-GB cases (n=6,191 and n=6,390, respectively). However, the differences in the statistical power could be due to the heterogeneity of brain tumours rather than small sample sizes. Overall, there was limited evidence to suggest whether there is a substantial difference between the gene expression risk profile of GB and non-GB gliomas. Further analysis with larger datasets is required to validate differences by enhancing statistical power of the analysis.

Sensitivity Analysis

We performed follow-up sensitivity analysis for molecular traits which had strong MR, colocalisation and Steiger filtering evidence. To test if the molecular traits also had causal evidence with putative glioma risk factors, we tested the second MR assumption – that the genetic instrument does not share a common cause with the outcome ("independence" or "exchangeability"). However, it should be noted that this methodology is not perfect as known glioma risk factors are not instrumentable. Therefore, we looked at putative risk factors identified in the literature – allergies, asthma, altered low-density lipoprotein cholesterol (LDLc) levels and alcohol consumption, to determine whether these associations could be driven by horizontal pleiotropy and therefore violate the second assumption of MR. [28, 35-37] However, if the putative risk factor is part of the molecular pathway by which the molecular trait affects the risk of glioma, this is simply annotating the functional pathway – this is vertical pleiotropy (Figure S1). This would require follow up studies to investigate and could lead to potentially elucidating how these putative risk factors and molecular traits are linked to glioma.

Link Between expression and splicing QTL associations

In this study, we integrated eQTL, sQTL and pQTL to establish a multi-omic genetic profile of glioma risk. Integrating different omics data can be beneficial to understand the development of complex traits and for drug target-related analyses as they allow for better functional annotation of the genome; each QTL provides an insight into different molecular processes. Whilst eQTL and sQTL have independently shown to be beneficial in studying complex disease and drug targets, integration of multi-omics analysis has proven to be extremely useful in different complex diseases. [12-13, 54]

eQTL are measured using abundance of protein expression, whereas sQTL are measured using RNA expression. However, it is still not fully understood whether eQTL and sQTL are capturing the same effects, and what the extent of the overlap between these are. Eales et al. (2021) found that in a study on hypertension and the kidney, carrying out multi-omic analysis using eQTL, sQTL and methylation QTL (mQTL), approximately 15% of blood pressure GWAS signals may be driven by a kidney gene at a different level of molecular regulation. [13] In our analyses, we do not have sufficient levels of evidence to draw similar conclusions, but further analysis using larger datasets could provide significant evidence to this. However, we did perform colocalisation of genes with robust causal evidence identified in both analyses to determine whether these expression and splicing QTL were driven by the same lead SNP(s). All of these associations showed poor weak evidence of colocalisation. This highlights that the QTL are driven by two independent SNPs rather than sharing a common causal SNP, which shows that our main results are unable to currently support the idea that expression and splicing QTL affect glioma risk through shared biological molecular regulation.

Limitations

Despite using relatively large datasets, our analyses are still likely to suffer from limited statistical power due to limited sample sizes, particularly in the sQTL dataset. All our main results were instrumented by a single SNP (using Wald ratio), which restricts the type of sensitivity analyses we could perform. Colocalisation can be used to remove some false associations which are driven by horizontal pleiotropy, due to the fact that a shared causal variant for two traits is necessary, but not sufficient for them to be causally related. [35] Another limitation of using MR is that effect estimates are for lifetime exposure to gene expression, and does not take into account that gene expression fluctuates in tissues both temporally and spatially. For example, cells at the leading edge of glioma appear to have a different genetic expression than those at the core. [55] Whilst our main results appeared to pass sensitivity analyses, there is still the possibility that bias remains through confounding and horizontal pleiotropy.

A common theme of GWAS is that the analysis focuses on European-derived ancestry and does not take into consideration other ethnic groups such as African, South Asian and Han Chinese populations. [56] This is due to the availability of large-scale datasets, which are much more readily available in European cohorts compared to other ethnicities. Previous studies in non-European cohorts have found novel targets which we did not find in our study, which illustrates that the loci appear to be population agnostic. [57, 58] Our results therefore may only be applicable to those with European-ancestry; we would need to validate these results in GWAS datasets from other ethnic groups in order to confirm their applicability in these groups.

Adult gliomas occur most commonly in the anterior fossa cortical regions with the frontal lobe typically having the greatest incidence (~40%), whereas paediatric tumours occur in the posterior fossa, cerebellum and brainstem. Gliomagenesis also occurs more rarely in deeper regions of the brain in stem-like subventricular zones. [59-61] The majority of our associations tended to be in the cortex, particularly in the eQTL analysis, this could be due to differences in other tissues' sample sizes. The cortex had a much larger sample size (n=2,970) compared to any of the other sample sizes in the eQTL analysis, which could lend to why there were more significant associations within this tissue. This is apparent in the sQTL analysis as the sample sizes in all tissues were extremely small (n=114 to 209) (Table S3). Ultimately, further analysis with larger datasets would be required to establish if there is a relationship between tissue and the number of MR associations.

Strengths

Our analysis had numerous strengths including the use of genetic variants that proxy for gene expression levels. This should reduce the effect of bias and confounding through reverse causation. These genetic variants were obtained from large meta-analyses, which increases statistical power and allows for more precise estimates. By using tissue-specific data separated into glioma subtypes, we can derive a much more accurate profile of the risk of glioma.

Mendelian Randomisation also contributes to the strength of our analyses. MR is less liable to sources of bias and confounding and provides evidence of causal relationships between genetic expression splicing, alternative splicing and glioma risk. This evidence is further supported when MR is combined with colocalisation by determining if the two traits are shared by a true, causal variant, or rather through two independent SNPs which are in LD with one another, which MR alone cannot determine. This colocalisation method assumes that there is only a single causal variant within the genome being tested. We overcome this by integrating conditional analysis prior to the colocalisation analysis which will condition upon marginal associations in the region of the lead SNP(s) and allow the single causal assumption to hold. [20] Despite evidence of causality, MR does not establish definitive causality and further studies are needed to do so.

Conclusion

We used MR to identify putative causal genes for glioma risk. We found causal evidence for four novel genes (*BTN3A2, FAIM, GALNT6, and GMPPB*) associated with differential glioma risk in eQTL and sQTL analysis. Our enrichment analysis provides rationale to whether these genes are viable drug targets for prevention of glioma. However, we acknowledge that limited sample size and population diversity restricts the power of these results and further studies in larger datasets are needed to validate our results.

Acknowledgements

ZAT, LJA, JWR and KMK 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 is funded by Cancer Research UK (grant number C30758/A2979). JWR receives funding from Biogen.

Declaration of Interests

JWR receives funding from Biogen for unrelated research.

References

1. Goodenberger, M.L., and Jenkins, R.B. (2012). Genetics of adult glioma. Cancer Genet 205, 613-621.
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.
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.
4. Weller, M., van den Bent, M., Preusser, M., Le Rhun, E., Tonn, J.C., Minniti, G., Bendszus, M., Balana, C., Chinot, O., Dirven, L., et al. (2021). EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. Nat Rev Clin Oncol 18, 170-186.
5. Andrews, L.J., Thornton, Z.A., Saincher, S.S., Yao, I.Y., Dawson, S., McGuinness, L.A., Jones, H.E., Jefferies, S., Short, S.C., Cheng, H.Y., et al. (2021). Prevalence of BRAFV600 in glioma and use of BRAF Inhibitors in patients with BRAFV600 mutation-positive glioma: systematic review. Neuro-Oncology 24, 13.
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.
7. D'Alessio, A., Proietti, G., Sica, G., and Scicchitano, B.M. (2019). Pathological and Molecular Features of Glioblastoma and Its Peritumoral Tissue. Cancers 11, 19.
8. Poon, M.T.C., Sudlow, C.L.M., Figueroa, J.D., and Brennan, P.M. (2020). Longer-term (>= 2 years) survival in patients with glioblastoma in population-based studies pre- and post-2005: a systematic review and meta-analysis. Sci Rep 10, 10.
9. Tunthanathip, T., Ratanalert, S., Sae-heng, S., Oearsakul, T., Sakaruncchai, I., Kaewborisutsakul, A., Chotsampancharoen, T., Intusoma, U., Kitkhuandee, A., and Vaniyapong, T. (2020). Prognostic Factors and Nomogram Predicting Survival in Diffuse Astrocytoma. J Neurosci Rural Pract 11, 135-143.
10. Scheie, D., Meling, T.R., Cvancarova, M., Skullerud, K., Mork, S., Lote, K., Eide, T.J., Helseth, E., and Beiske, K. (2011). Prognostic variables in oligodendroglial tumors: a single-institution study of 95 cases. Neuro-Oncology 13, 1225-1233.
11. Wahner, H.C.W., Trager, M., Bender, K., Schweizer, L., Onken, J., Senger, C., Ehret, F., Budach, V., and Kaul, D. (2020). Predicting survival in anaplastic astrocytoma patients in a single-center cohort of 108 patients. Radiat Oncol 15, 9.
12. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., De La Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. Cell 179, 750-+.
13. Eales, J.M., Jiang, X., Xu, X.G., Saluja, S., Akbarov, A., Cano-Gamez, E., McNulty, M.T., Finan, C., Guo, H., Wystrychowski, W., et al. (2021). Uncovering genetic mechanisms of hypertension through multi-omic analysis of the kidney. Nature Genetics 53, 630-+.
14. Sun, Y.V., and Hu, Y.J. (2016). Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. In Advances in Genetics, Vol 93, T. Friedmann, J.C. Dunlap, andS.F. Goodwin, eds. (San Diego, Elsevier Academic Press Inc), pp 147-190.
15. Schwartzbaum, J.A., Fisher, J.L., Aldape, K.D., and Wrensch, M. (2006). Epidemiology and molecular pathology of glioma. Nat Clin Pract Neurol 2, 494-503.

16. Haycock, P.C., Burgess, S., Nounu, A., Zheng, J., Okoli, G.N., Bowden, J., Wade, K.H., Timpson, N.J., Evans, D.M., Willeit, P., et al. (2017). Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases A Mendelian Randomization Study. JAMA Oncol 3, 636-651.
17. Saunders, C.N., Kinnersley, B., Culliford, R., Cornish, A.J., Law, P.J., and Houlston, R.S. (2022). Relationship between genetically determined telomere length and glioma risk. Neuro-Oncology 24, 171-181.
18. Kinnersley, B., Houlston, R.S., and Bondy, M.L. (2018). Genome-Wide Association Studies in Glioma. Cancer Epidemiol Biomarkers Prev 27, 418-428.
19. Smith, G.D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology 32, 1-22.
20. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. Plos Genetics 10, 15.
21. De Klein, N., Tsai, E.A., Vochteloo, M., Baird, D., Huang, Y., Chen, C.Y., van Dam, S., Deelen, P., Bakker, O.B., El Garwany, O., et al. (2021). Brain expression quantitative trait locus and network analysis reveals downstream effects and putative drivers for brain-related diseases. bioRxiv. Preprint. https://doi.org/10.1101/2021.03.01.433439
22. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nature Genetics 45, 580-585.
23. Robins, C., Liu, Y., Fan, W., Duong, D.M., Meigs, J., Harerimana, N.V., Gerasimov, E.S., Dammer, E.B., Cutler, D.J., Beach, T.G., et al. (2021). Genetic control of the human brain proteome. American Journal of Human Genetics 108, 400-410.
24. Yang, C.R., Farias, F.H.G., Ibanez, L., Suhy, A., Sadler, B., Fernandez, M.V., Wang, F.X., Bradley, J.L., Eiffert, B., Bahena, J.A., et al. (2021). Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. Nat Neurosci 24, 1302-1312.
25. Bennett, D.A., Schneider, J.A., Arvanitakis, Z., and Wilson, R.S. (2012). Overview and Findings from the Religious Orders Study. Curr Alzheimer Res 9, 628-645.
26. Bennett, D.A., Schneider, J.A., Buchman, A.S., Barnes, L.L., Boyle, P.A., and Wilson, R.S. (2012). Overview and Findings from the Rush Memory and Aging Project. Curr Alzheimer Res 9, 646-663.
27. Melin, B.S., Barnholtz-Sloan, J.S., Wrensch, M.R., Johansen, C., Il'yasova, D., Kinnersley, B., Ostrom, Q.T., Labreche, K., Chen, Y.W., Armstrong, G., et al. (2017). Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. Nature Genetics 49, 789-+.
28. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47, D1005-D1012.
29. Burgess, S., Small, D.S., and Thompson, S.G. (2017). A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res 26, 2333-2355.
30. R Core Team. R: A Language and Environment for Statistical Computing. https://www.R-project.org/. (2021)
31. RStudio Team. RStudio: Integrated Development for R. http://www.rstudio.com/. (2020)

32. Staiger, D., and Stock, J.H. (1997). Instrumental variables regression with weak instruments. Econometrica 65, 557-586.
33. Lawlor, D.A., Harbord, R.M., Sterne, J.A.C., Timpson, N., and Smith, G.D. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine 27, 1133-1163.
34. Howell, A.E., Robinson, J.W., Wootton, R.E., McAleenan, A., Tsavachidis, S., Ostrom, Q.T., Bondy, M., Armstrong, G., Relton, C., Haycock, P., et al. (2020). Testing for causality between systematically identified risk factors and glioma: a Mendelian randomization study. Bmc Cancer 20, 11.
35. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. eLife 7, 29.
36. Elsworth, B., Lyon, M., Alexander T., Liu, Y., Matthews, P., Hallett, J., Bates, P., Palmer, T., Haberland, V., Smith, G.D., et al. (2020). The MRC IEU OpenGWAS data infrastructure. bioRxiv. Pre-print. https://doi.org/10.1101/2020.08.10.244293.
37. Mitchell, R., Elsworth, B.L., Mitchell, R., Raistrick, C.A., Paternoster, L., Hemani, G., Gaunt, T.R. (2019). MRC IEU UK Biobank GWAS pipeline version 2. bioRxiv. Pre-print. https://doi.org/10.5523/bris.pnoat8cxo0u52p6ynfaekeigi.
38. Hemani, G., Bowden, J., and Smith, G.D. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. Human Molecular Genetics 27, R195-R208.
39. Zheng, J., Haberland, V., Baird, D., Walker, V., Haycock, P.C., Hurle, M.R., Gutteridge, A., Erola, P., Liu, Y., Luo, S., et al. (2020). Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nature Genetics 52, 1122-+.
40. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nature Genetics 44, 369-U170.
41. Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Uriarte, A., Malangone, C., Miranda, A., Fumis, L., Carvalho-Silva, D., Spitzer, M., et al. (2021). Open Targets Platform: supporting systematic drug-target identification and prioritisation. Nucleic Acids Res 49, D1302-D1310.
42. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., et al. (2014). The MIntAct project-IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 42, D358-D363.
43. Freshour, S.L., Kiwala, S., Cotto, K.C., Coffman, A.C., McMichael, J.F., Song, J.J., Griffith, M., Griffith, O.L., and Wagner, A.H. (2021). Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res 49, D1144-D1151.
44. Skrivankova, V.W., Richmond, R.C., Woolf, B.A.R., Yarmolinsky, J., Davies, N.M., Swanson, S.A., VanderWeele, T.J., Higgins, J.P.T., Timpson, N.J., Dimou, N., et al. (2021). Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA-J Am Med Assoc 326, 1614-1621.
45. Cai, P.A., Lu, Z.H., Wu, J.J., Qin, X., Wang, Z.T., Zhang, Z., Zheng, L., and Zhao, J.M. (2020). BTN3A2 serves as a prognostic marker and favors immune infiltration in triple-negative breast cancer. J Cell Biochem 121, 2643-2654.
46. Wieczorek, M., Abualrous, E.T., Sticht, J., Alvaro-Benito, M., Stolzenberg, S., Noe, F., and Freund, C. (2017). Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. Front Immunol 8, 16.

47. Wu, Y., Bi, R., Zeng, C.H., Ma, C.G., Sun, C.L., Li, J.Z., Xiao, X., Li, M., Zhang, D.F., Zheng, P., et al. (2019). Identification of the primate-specific gene BTN3A2 as an additional schizophrenia risk gene in the MHC loci. EBioMedicine 44, 530-541.
48. Huo, J.X., Xu, S.L., and Lam, K.P. (2019). FAIM: An Antagonist of Fas-Killing and Beyond. Cells 8, 12.
49. The Human Genome Atlas Project. FAIM. https://www.proteinatlas.org/ENSG00000158234-FAIM. (2022).
50. The Human Genome Atlas Project. GALNT6. https://www.proteinatlas.org/ENSG00000139629-GALNT6. (2022).
51. Song, J., Liu, W.W., Wang, J.Z., Hao, J.X., Wang, Y.Y., You, X., Du, X.H., Zhou, Y., Ben, J., Zhang, X.R., et al. (2020). GALNT6 promotes invasion and metastasis of human lung adenocarcinoma cells through O-glycosylating chaperone protein GRP78. Cell Death Dis 11, 14.
52. Kang, B.R., Yang, S.H., Chung, B.R., Kim, W., and Kim, Y. (2016). Cell surface GRP78 as a biomarker and target for suppressing glioma cells. Sci Rep 6, 7.
53. Astrea, G., Romano, A., Angelini, C., Antozzi, C.G., Barresi, R., Battini, R., Battisti, C., Bertini, E., Bruno, C., Cassandrini, D., et al. (2018). Broad phenotypic spectrum and genotype-phenotype correlations in GMPPB-related dystroglycanopathies: an Italian cross-sectional study. Orphanet J Rare Dis 13, 9.
54. Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. Science 352, 600-604.
55. Hoelzinger, D.B., Mariani, L., Weis, J., Woyke, T., Berens, T.J., McDonough, W.S., Sloan, A., Coons, S.W., and Berens, M.E. (2005). Gene expression profile of glioblastoma multiforme invasive phenotype points to new therapeutic targets. Neoplasia 7, 7-16.
56. Medina-Gomez, C., Felix, J.F., Estrada, K., Peters, M.J., Herrera, L., Kruithof, C.J., Duijts, L., Hofman, A., van Duijn, C.M., Uitterlinden, A.G., et al. (2015). Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. Eur J Epidemiol 30, 317-330.
57. 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 al. (2021). Genetic variants of CYP4F12 gene are associated with glioma susceptibility. Int J Cancer 149, 1910-1915.
58. 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, G.D. (2012). Selected polymorphisms of GSTP1 and TERT were associated with glioma risk in Han Chinese. Cancer Epidemiol 36, 525-527.
59. Larjavaara, S., Mantyla, R., Salminen, T., Haapasalo, H., Raitanen, J., Jaaskelainen, J., and Auvinen, A. (2007). Incidence of gliomas by anatomic location. Neuro-Oncology 9, 319-325.
60. Smith, M.A., Freidlin, B., Ries, L.A.G., and Simon, R. (1998). Trends in reported incidence of primary malignant brain tumors in children in the United States. J Natl Cancer Inst 90, 1269-1277.
61. Duffau, H., and Capelle, L. (2004). Preferential brain locations of low-grade gliomas. Cancer 100, 2622-2626.

Figures

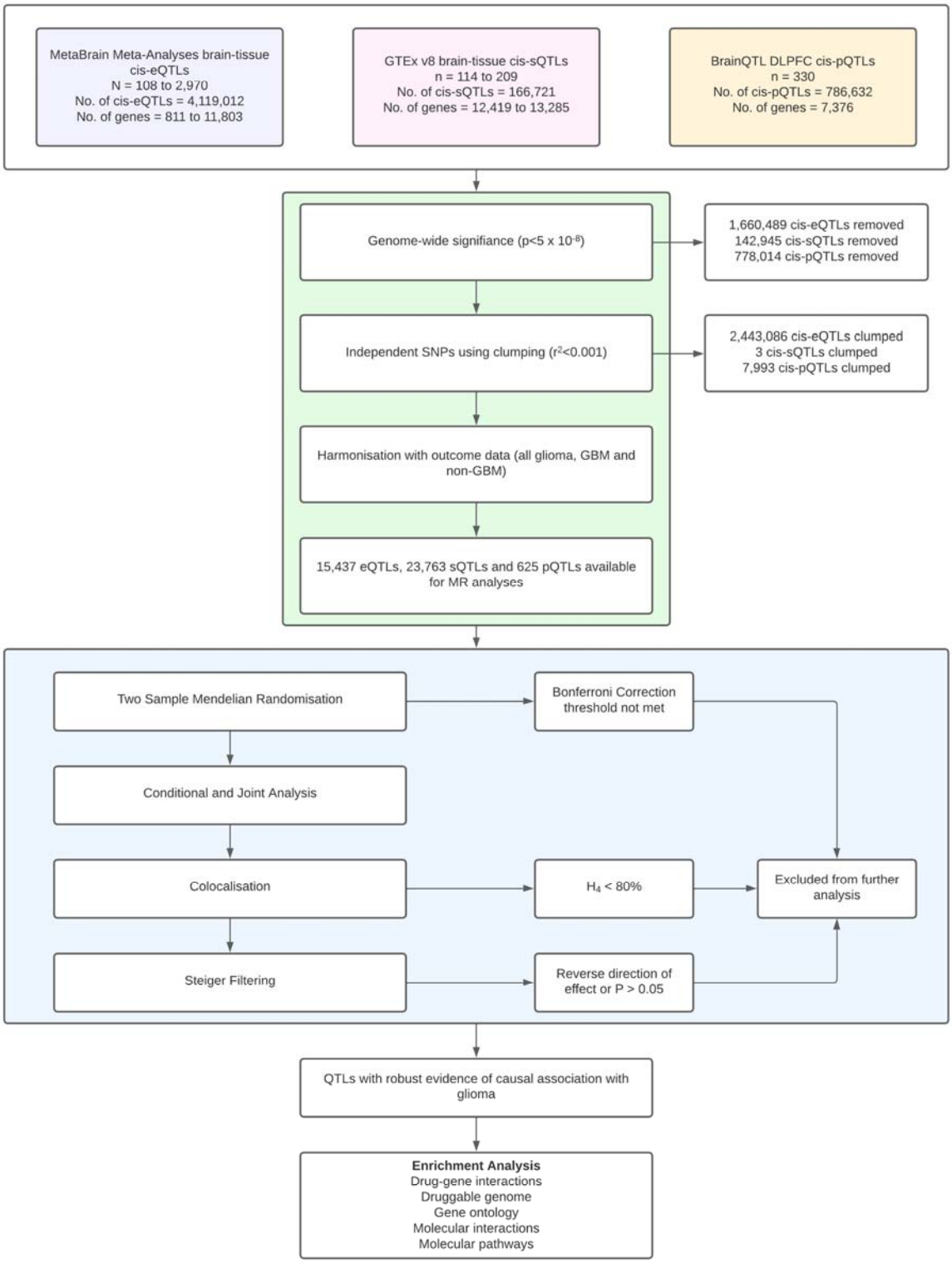
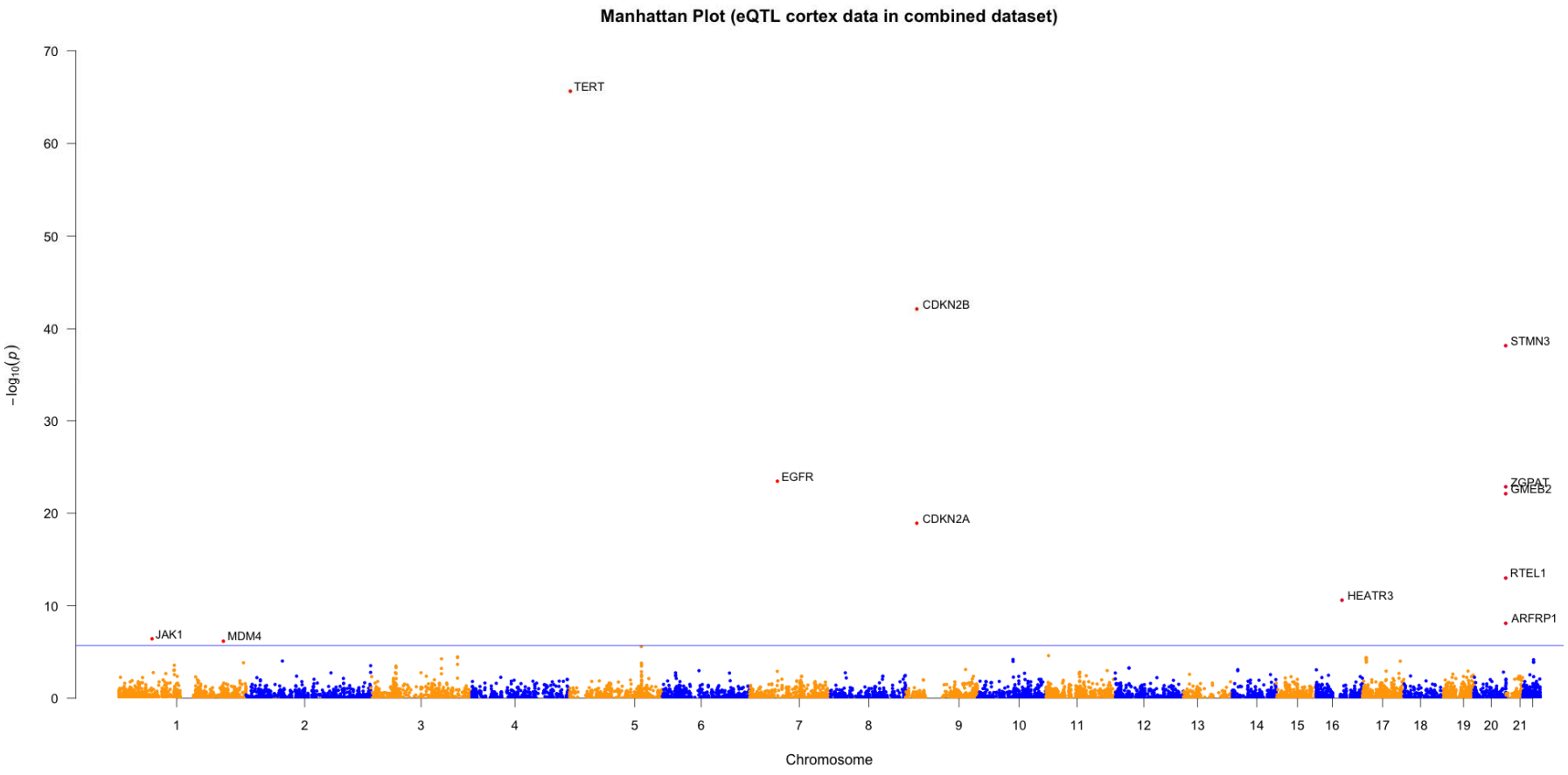
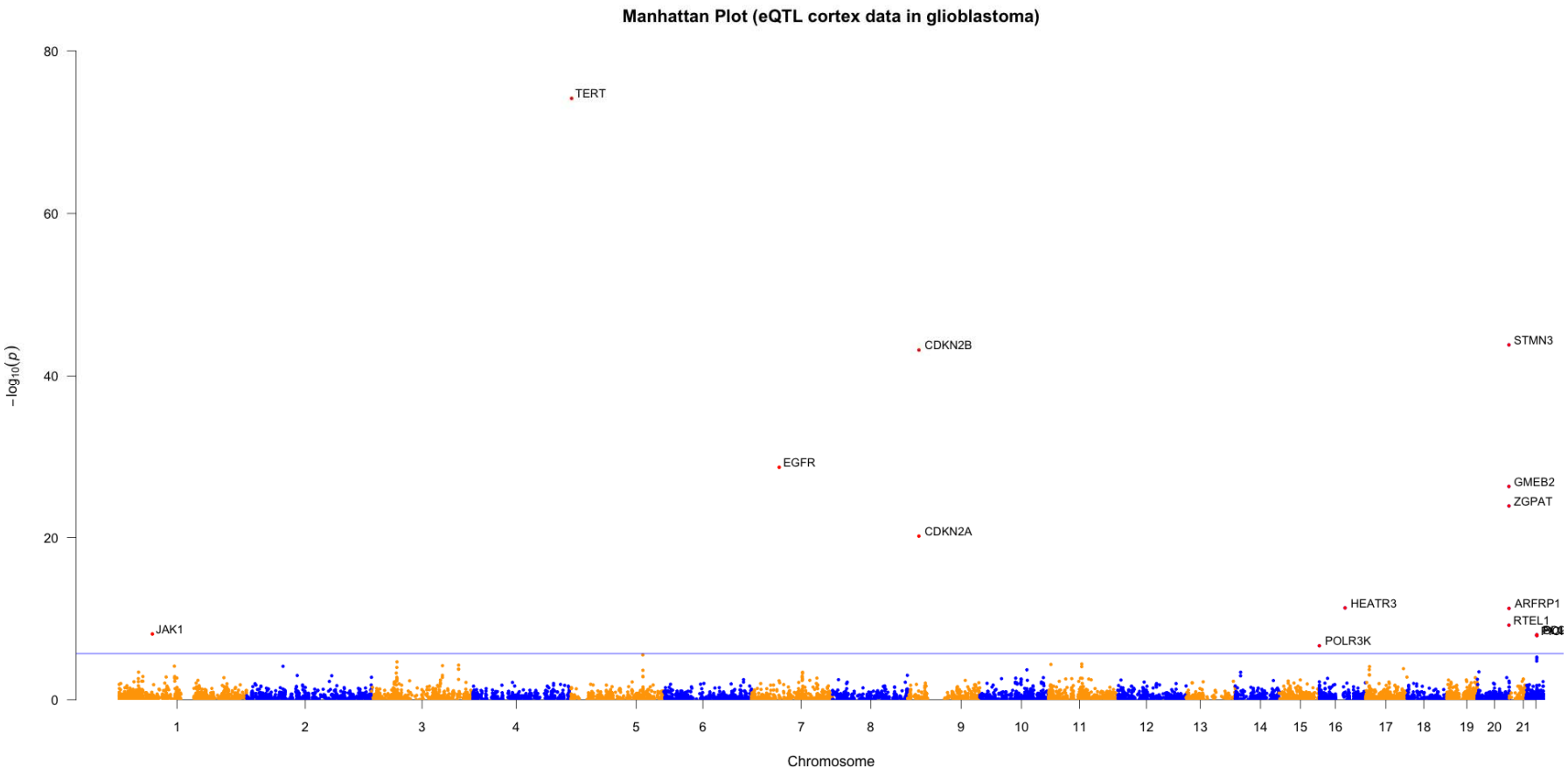


Figure 1. Flowchart showing the pipeline used in the multi-omic analysis.

(A)


Manhattan Plot (eQTL cortex data in combined dataset)
−log10(p)
70
60
50
40
30
20
10
0
TERT
CDKN2B
STMN3
EGFR
ZGPAT
GMEB2
CDKN2A
RTEL1
HEATR3
ARFRP1
JAK1
MDM4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Chromosome

(B)


Manhattan Plot (eQTL cortex data in glioblastoma)
$-\log_{10}(p)$
80
60
40
20
0
TERT
CDKN2B
STMN3
EGFR
GMEB2
ZGPAT
CDKN2A
HEATR3
ARFRP1
RTEL1
JAK1
POLR3K
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Chromosome

(C)

Figure 2. Series of Manhattan plots showing the results of the MR analysis investigating the effects of genetically proxied gene expression levels on genetic liability to glioma subtype in cortex tissue in (A) combined dataset (B) glioblastoma and (C) non-glioblastoma.

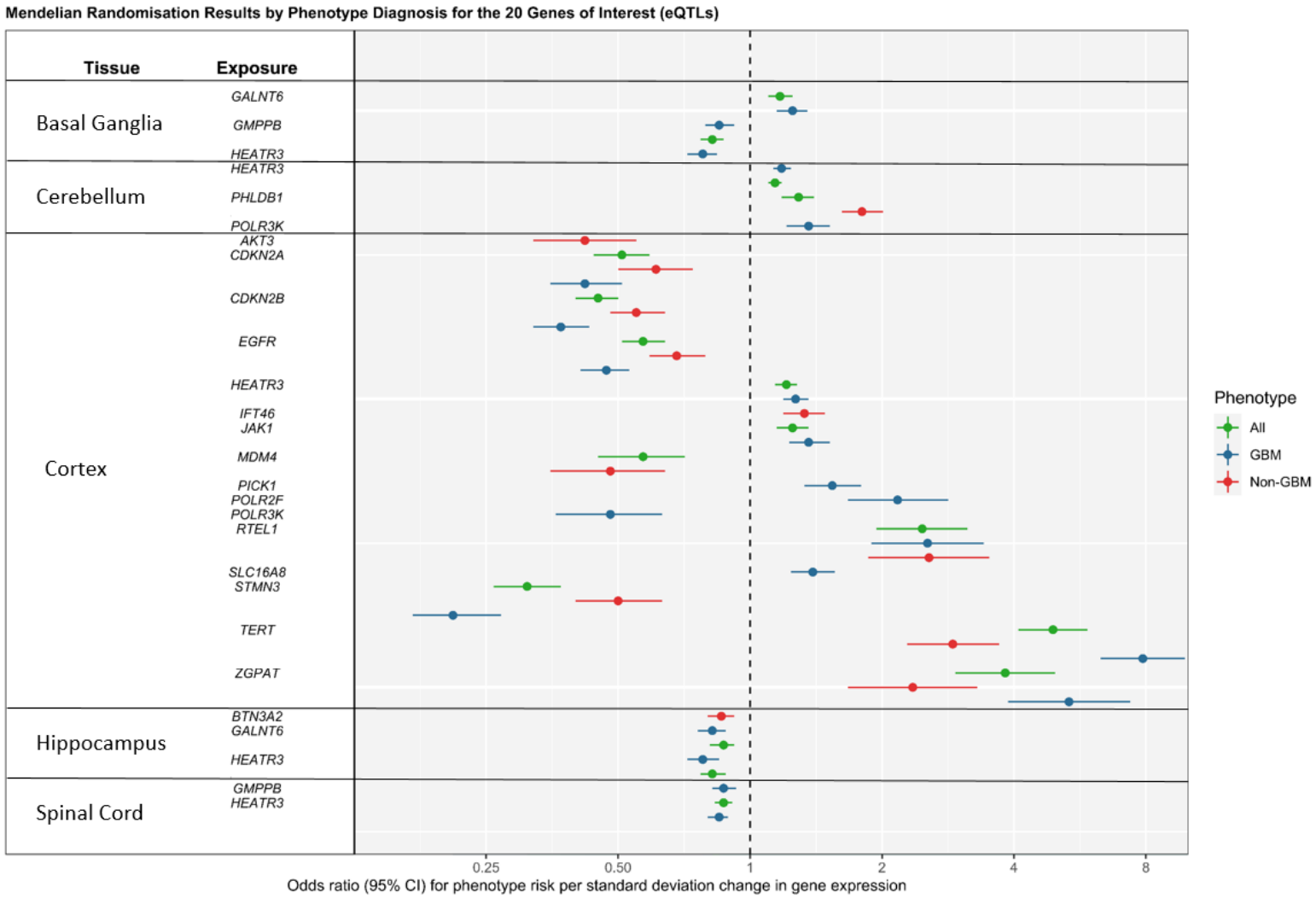


Figure 3. Forest plot showing Mendelian randomisation results by subtype diagnosis (combined dataset (all), GB and non-GB) for the 20 genes of interest in the MetaBrain eQTL dataset.

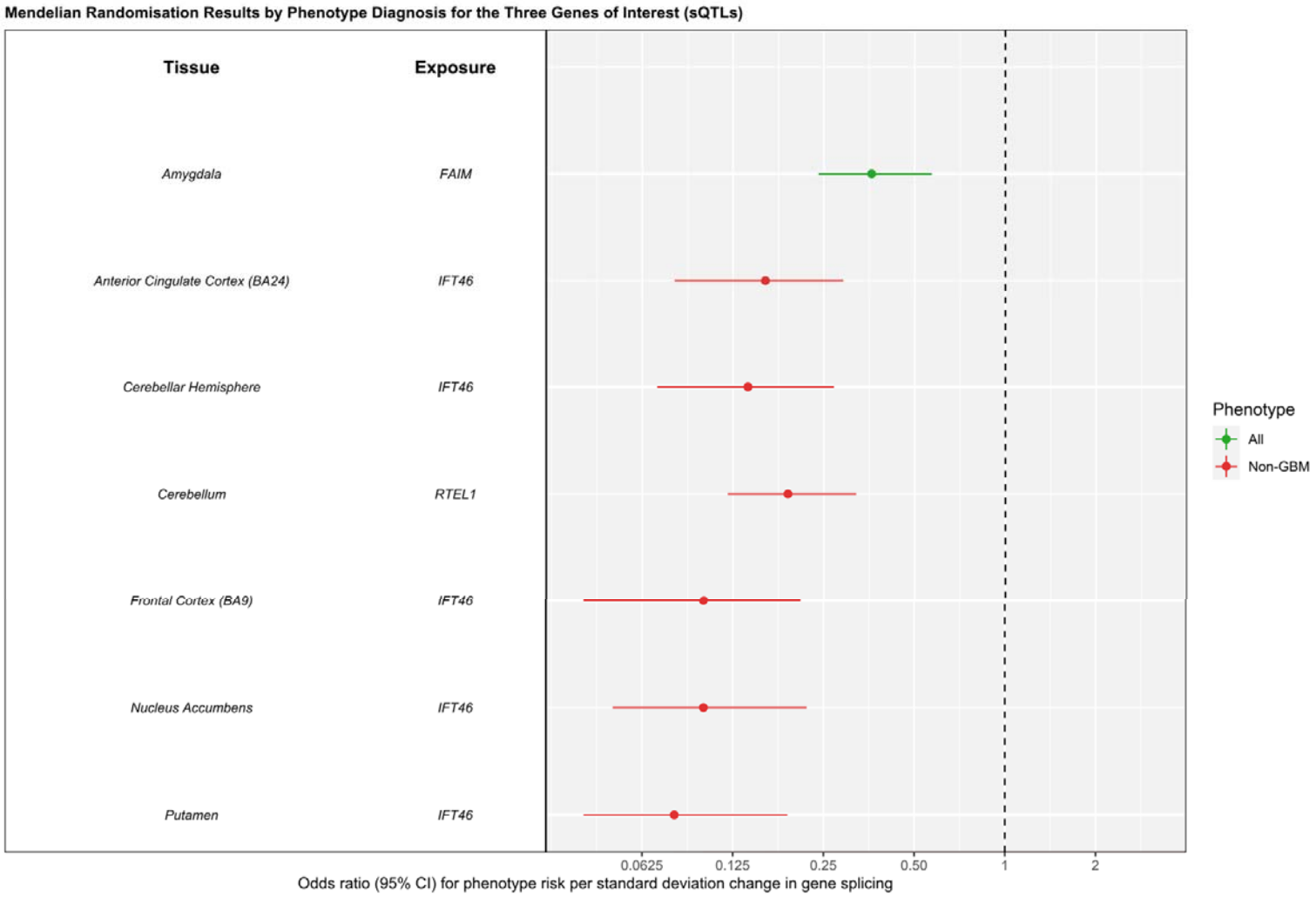


Figure 4. Forest plot showing Mendelian randomisation results by subtype diagnosis (combined dataset (all) glioma and non-GB) for the seven genes of interest in the GTEx sQTL dataset. There were no robust associations for the GB subtype.

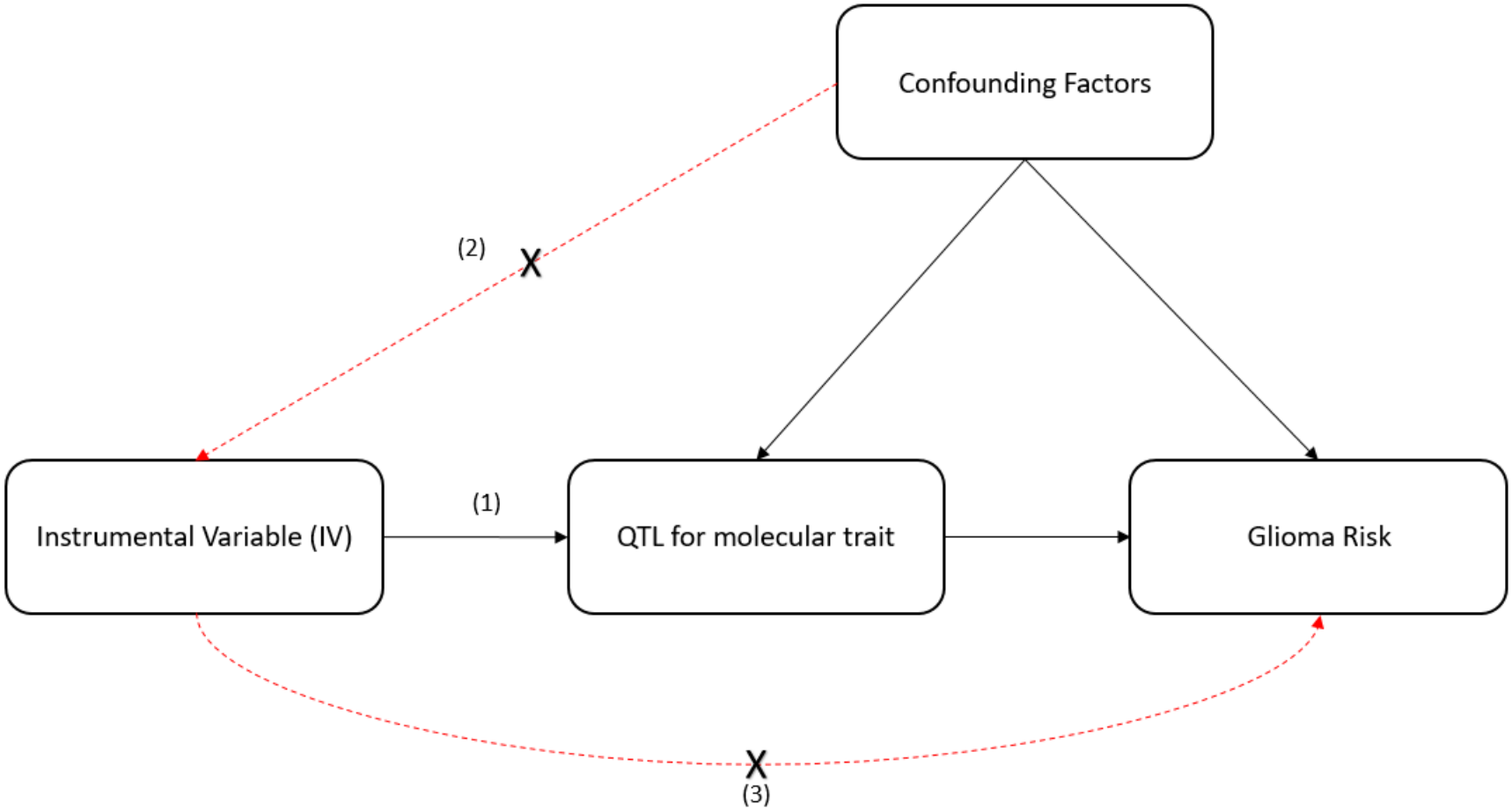


Figure 5. Directed Acyclic Graph (DAG) showing a visual representation of potential causal relationships. This DAG also shows the three assumptions of Mendelian Randomisation. (1) The genetic variant(s) being used as an instrument is associated robustly with the exposure. (2) The instrument is independent 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.

Tables

Table 1. The number of SNPs, genes and sample size in the different datasets separated by brain tissue. BA = Brodmann's Area, DLPFC = Dorsolateral Prefrontal Cortex

Dataset	Tissue	No. of SNPs	No. of Genes	Sample Size
MetaBrain	Basal Ganglia	77,084	1,317	208
	Cerebellum	597,690	6,865	492
	Cortex	1,530,207	11,803	2,970
	Hippocampus	53,384	990	168
	Spinal Cord	42,019	811	108
GTEx v8	Amygdala	12,583	12,575	129
	Anterior Cingulate Cortex (BA24)	12,975	12,967	147
	Caudate	13,062	13,053	194
	Cerebellar Hemisphere	12,647	12,641	175
	Cerebellum	12,548	12,542	209
	Cortex	13,013	13,005	205
	Frontal Cortex (BA9)	13,079	13,071	175
	Hippocampus	12,671	12,662	165
	Hypothalamus	13,293	13,285	170
	Nucleus Accumbens	13,177	13,168	202
	Putamen	12,427	12,419	170
	Spinal Cord (Cervical C1)	12,772	12,762	126
	Substantia Nigra	12,474	12,464	114
BrainQTL	DLPFC	786,632	7,376	330
NIAGADS	Cortex (Parietal Lobe)	3,706,355	713	343

Table 2. Results of the MR analysis investigating the effects of genetically proxied gene expression levels on genetic liability to glioma subtype. Results presented here are those which were deemed the most robust, i.e., met the Bonferroni-correct P value threshold for the MR analysis, had strong evidence of colocalisation (H4 > 80%) and correctly orientated Steiger filtering direction.

Gene	Location	SNP	Subtype	Tissue	OR (95% CI)	P Value	H4 (%)	Steiger P Value	Steiger Flag
AKT3	1q43-44	rs3006933	Non-GBM	Cortex	0.42 (0.32, 0.55)	3.15E-10	100	5.79E-05	True
BTN3A2	6p22.2	rs68072215	Non-GBM	Hippocampus	0.86 (0.80, 0.92)	1.83E-05	97	3.12E-12	True
CDKN2A	9p21.3	rs3731239	All	Cortex	0.51 (0.44, 0.59)	1.22E-19	100	7.41E-09	True
CDKN2A	9p21.3	rs3731239	Non-GBM	Cortex	0.61 (0.50, 0.74)	5.30E-07	100	2.04E-10	True
CDKN2A	9p21.3	rs3731239	GBM	Cortex	0.42 (0.35, 0.51)	6.79E-21	100	1.82E-07	True
CDKN2B	9p21.3	rs6475604	All	Cortex	0.45 (0.40, 0.50)	1.59E-19	100	9.23E-14	True
CDKN2B	9p21.3	rs6475604	Non-GBM	Cortex	0.55 (0.48, 0.64)	7.17E-43	100	7.73E-17	True
CDKN2B	9p21.3	rs6475604	GBM	Cortex	0.37 (0.32, 0.43)	1.59E-44	100	2.88E-11	True
EGFR	7p11.2	rs77126132	All	Cortex	0.57 (0.51, 0.64)	3.52E-24	100	3.71E-16	True
EGFR	7p11.2	rs77126132	Non-GBM	Cortex	0.68 (0.59, 0.79)	3.82E-07	100	3.07E-19	True
EGFR	7p11.2	rs77126132	GBM	Cortex	0.47 (0.41, 0.53)	2.19E-29	100	2.00E-13	True
GALNT6	12q13.13	rs3782473	All	Basal Ganglia	1.17 (1.10, 1.25)	3.14E-06	100	1.73E-07	True
GALNT6	12q13.13	rs3782473	GBM	Basal Ganglia	1.25 (1.15, 1.35)	9.08E-08	100	3.60E-07	True
GALNT6	12q13.13	rs4785382	GBM	Hippocampus	0.82 (0.76, 0.88)	9.08E-08	100	9.29E-07	True
GALNT6	12q13.13	rs4785382	All	Hippocampus	0.87 (0.82, 0.93)	3.14E-06	100	4.91E-07	True
GMPPB	3p21.31	rs4855885	GBM	Basal Ganglia	0.85 (0.79, 0.92)	1.50E-05	92	5.62E-10	True
GMPPB	3p21.31	rs6226673	GBM	Spinal Cord	0.87 (0.82, 0.93)	2.00E-05	96	8.65E-09	True
HEATR3	16q12.1	rs2058815	All	Basal Ganglia	0.82 (0.77, 0.87)	2.45E-10	92	9.45E-09	True
HEATR3	16q12.1	rs2058815	GBM	Basal Ganglia	0.78 (0.72, 0.84)	1.52E-10	94	1.83E-08	True
HEATR3	16q12.1	rs2287197	GBM	Cerebellum	1.18 (1.13, 1.24)	6.74E-12	97	1.86E-57	True
HEATR3	16q12.1	rs2287197	All	Cerebellum	1.14 (1.10, 1.18)	3.93E-11	96	6.99E-59	True
HEATR3	16q12.1	rs12933325	All	Cortex	1.21 (1.14, 1.28)	2.54E-11	100	2.35E-104	True
HEATR3	16q12.1	rs12933325	GBM	Cortex	1.27 (1.19, 1.36)	4.78E-12	100	3.15E-100	True
HEATR3	16q12.1	rs3782473	GBM	Hippocampus	0.78 (0.72, 0.85)	5.14E-09	100	1.06E-06	True
HEATR3	16q12.1	rs3782473	All	Hippocampus	0.82 (0.77, 0.88)	6.87E-09	100	6.64E-07	True
HEATR3	16q12.1	rs4785382	All	Spinal Cord	0.87 (0.83, 0.91)	6.87E-09	100	1.40E-10	True
HEATR3	16q12.1	rs4785382	GBM	Spinal Cord	0.85 (0.80, 0.89)	5.14E-09	100	2.25E-11	True

IFT46	11q23.3	rs11552421	Non-GBM	Cortex	1.33 (1.19, 1.48)	4.06E-07	100	1.07E-45	True
JAK1	1p31.3	rs310215	All	Cortex	1.25 (1.15, 1.36)	3.79E-07	100	3.39E-45	True
JAK1	1p31.3	rs310215	GBM	Cortex	1.36 (1.23, 1.52)	7.74E-09	100	3.83E-42	True
MDM4	1q32.1	rs10793765	All	Cortex	0.57 (0.45, 0.71)	6.85E-07	100	8.35E-06	True
MDM4	1q32.1	rs10793765	Non-GBM	Cortex	0.48 (0.35, 0.64)	7.41E-07	100	4.66E-05	True
PHLDB1	11q23.3	rs11603023	All	Cerebellum	1.29 (1.18, 1.40)	2.48E-26	100	1.19E-09	True
PHLDB1	11q23.3	rs11603023	Non-GBM	Cerebellum	1.80 (1.62, 2.01)	3.51E-09	100	1.78E-07	True
PICK1	22q13.1	rs2076369	GBM	Cortex	1.54 (1.33, 1.79)	1.22E-08	100	7.00E-18	True
POLR2F	22q13.1	rs1004764	GBM	Cortex	2.17 (1.67, 2.83)	9.58E-09	100	4.34E-05	True
POLR3K	16p13.3	rs11866537	GBM	Cerebellum	1.36 (1.21, 1.52)	1.74E-07	100	5.86E-08	True
POLR3K	16p13.3	rs4145232	GBM	Cortex	0.48 (0.36, 0.63)	2.21E-07	100	5.15E-05	True
RTEL1	20q13.33	rs6011014	All	Cortex	2.47 (1.94, 3.13)	1.04E-13	85	1.06E-03	True
RTEL1	20q13.33	rs6011014	GBM	Cortex	2.54 (1.89, 3.41)	6.30E-10	85	1.19E-03	True
RTEL1	20q13.33	rs6011014	Non-GBM	Cortex	2.56 (1.86, 3.51)	6.79E-09	100	1.08E-03	True
SLC16A8	22q13.1	rs4821735	GBM	Cortex	1.39 (1.24, 1.56)	1.14E-08	100	1.80E-33	True
STMN3	20q13.33	rs6011016	All	Cortex	0.31 (0.26, 0.37)	6.42E-39	100	2.77E-05	True
STMN3	20q13.33	rs6011016	Non-GBM	Cortex	0.50 (0.40, 0.63)	2.07E-09	100	3.60E-08	True
STMN3	20q13.33	rs6011016	GBM	Cortex	0.21 (0.17, 0.27)	1.42E-44	100	1.83E-03	True
TERT	5p15.33	rs10069690	All	Cortex	4.91 (4.09, 5.88)	2.32E-66	100	1.13E-02	True
TERT	5p15.33	rs10069690	Non-GBM	Cortex	2.90 (2.28, 3.70)	6.69E-18	100	5.99E-05	True
TERT	5p15.33	rs10069690	GBM	Cortex	7.87 (6.31, 9.81)	6.92E-75	100	1.96E-01[a]	Uncertain
ZGPAT	20q13.33	rs2256814	All	Cortex	3.82 (2.94, 4.96)	1.42E-23	99	1.04E-02	True
ZGPAT	20q13.33	rs2256814	Non-GBM	Cortex	2.35 (1.67, 3.30)	8.92E-07	99	4.90E-04	True
ZGPAT	20q13.33	rs2256814	GBM	Cortex	5.34 (3.88, 7.36)	1.29E-24	99	5.91E-02[a]	Uncertain

[a] *Steiger filtering indicated true directionality (exposure associated with outcome), but failed to pass $P < 0.05$.*

Table 3. Results of the MR analysis investigating the effects of genetically proxied gene splicing levels on genetic liability to glioma subtype. Results presented here are those which were deemed the most robust, i.e., met the Bonferroni-correct P value threshold for the MR analysis, had strong evidence of colocalisation (H4 > 80%) and correctly orientated Steiger filtering direction.

Gene	Location	SNP	Subtype	Tissue	OR (95% CI)	P Value	H4 (%)	Steiger P Value	Steiger Flag
FAIM	3q22.3	rs361073	All	Amygdala	0.36 (0.24, 0.57)	1.03E-05	99	6.91E-01[a]	Uncertain
IFT46	11q23.3	rs3132828	Non-GBM	Anterior Cingulate Cortex (BA24)	0.16 (0.08, 0.29)	4.87E-09	100	8.53E-01[a]	Uncertain
IFT46	11q23.3	rs3132828	Non-GBM	Cerebellar Hemisphere	0.14 (0.07, 0.27)	4.87E-09	100	8.69E-01[a]	Uncertain
IFT46	11q23.3	rs3132828	Non-GBM	Frontal Cortex (BA9)	0.10 (0.04, 0.21)	4.87E-09	93	9.61E-01[a]	Uncertain
IFT46	11q23.3	rs3132828	Non-GBM	Nucleus Accumbens	0.10 (0.05, 0.22)	4.87E-09	100	9.49E-01[a]	Uncertain
IFT46	11q23.3	rs3132828	Non-GBM	Putamen	0.08 (0.04, 0.19)	4.87E-09	100	9.89E-01[a]	Uncertain
RTEL1	20q13.33	rs6062302	Non-GBM	Cerebellum	0.19 (0.12, 0.32)	2.32E-10	100	7.45E-01[a]	Uncertain

[a] *Steiger filtering indicated true directionality (exposure associated with outcome), but failed to pass P < 0.05.*