Author:
Robinson, Jamie W

Title:
Identifying new drugs and re-purposing old drugs for the treatment of glioma

Author:
Robinson, Jamie W

Title:
Identifying New Drugs and Re-purposing Old Drugs for the Treatment of Glioma

Identifying New Drugs and Re-purposing Old Drugs for the Treatment of Glioma

By

JAMIE WILLIAM ROBINSON

Bristol Medical School - Translational Health Sciences
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Health Sciences.

FEBRUARY 2021

Word count: 39,355

Abstract

Glioma is the most common type of brain tumour (BT) and results in significant years of life lost. The risk profile of glioma is largely unknown, and there has been little innovation in treatments and therapies available for glioma. These points were expanded upon in Chapter 1.

This thesis conceptualised a pipeline which combined population-level and multi-omic data with causal inference analyses and traditional and pharmaco-epidemiological techniques. Methods and relevant datasets were described in Chapters 2 & 3.

Analyses in Chapter 4 utilised a combined Mendelian randomisation (MR) and colocalisation framework to provide causal evidence for germline genetic variants associated with gene expression levels that affect glioma risk.

Chapter 5 leveraged the same framework with germline genetic variants associated with protein abundance levels. Results were combined with results from the previous Chapter and, with supporting evidence from the literature, inform on putative chemopreventive targets for glioma.

The pharmaco-epidemiological analyses in Chapter 6 provided evidence that exposure to glitazones, a family of anti-type 2 diabetic medications, reduced risk of primary and secondary BT, highlighting these drugs as potential agents for re-purposing.

Chapter 7 concluded that, altogether, analyses presented in this thesis may be combined to form a drug target identification, prioritisation and re-repurposing pipeline to improve patient outcomes in the clinic.

DEDICATION AND ACKNOWLEDGEMENTS

I would first like to extend my heartfelt gratitude to my PhD supervisors for their invaluable support, guidance and knowledge: Kathreena Kurian, Richard Martin, Caroline Relton, Jie Zheng and Kaitlin Wade.

A special thank you and consideration goes to the funders of my PhD, the Bristol Brain Tumour Bank, the North Bristol NHS Trust Charitable Fund and Southmead Hospital Charitable Funds.

With thanks to the individuals and families who participated in the collection of data for the studies used in this thesis, particularly the Gliogene, Glioma International Case Control, MD Anderson, University of California, San Francisco-Mayo, GliomaScan and other glioma genotyping studies. Without their participation in such studies, this thesis would not have been possible.

I give thanks to my collaborators, co-authors and reviewers whose help in those works and my academic career are acknowledged. In particular, a special thank you goes to Melissa Bondy, Martha Elwenspoek, Philip Haycock, Gibran Hemani, Amy Howell, Beatrice Melin, Claire Perks, Mio Ozawa, Yoav Ben-Shlomo and Spiridon Tsavachidis.

I owe a thank you for the enumerable help and guidance provided by members of the Integrative Epidemiology Unit and Integrative Cancer Epidemiology Programme and for allowing me a platform to present my research.

A special thanks to my friends, whose advice, both scientific and otherwise, has been extremely helpful and is always welcome. A special mention belongs to TB, JK, JY, JB, CD, PP, IC, the members of BF8, and, as always, LA, GM, OM and CFP.

Finally, I owe a lot to my family who have supported me throughout my burgeoning academic career.

Part of the Ship, Part of the Crew.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... [redacted] ..... DATE: 02/02/2021

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Expansion
5-ALA	5-aminolevulinic acid
BBB	Blood-brain barrier
BCNU	Carmustine (β-chloro-nitrosurea)
BT	Brain tumour
CCNU	Lomustine
CI	Confidence interval
CNS	Central nervous system
COJO	Conditional and joint analysis
CPRD	Clinical Practice Research Datalink
DAG	Directed acyclic graph
DGIdb	Drug-Genome Interaction database
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptor
EMA	European Medicines Agency
EORTC	European Platform of Cancer Research
eQTL	Expression quantitative trait loci
FDR	False discovery rate
GCTA	Genome-wide Complex Trait Analysis
GBM	Glioblastoma
GP	General practitioner
GTEx	Genotype-Tissue Expression project
GWAS	Genome-wide association study
HbA1c	Glycated haemoglobin
HGG	High-grade glioma
HR	Hazard ratio
IDH	Isocitrate dehydrogenase
IMD	Index of multiple deprivation
IR	Incidence rate
ISM	Intraoperative stimulation mapping
IV	Instrumental variable

IVW ..... Inverse variance weighted
LD ..... Linkage disequilibrium
LGG ..... Low-grade glioma
MGMT ..... O6-alkylguanine DNA-alkytransferase
MR ..... Mendelian randomisation
mRNA ..... Messenger ribonucleic acid
NHS ..... National Health Service
NOS ..... Not otherwise specified
NSC ..... Neural stem cells
ONS ..... Office of National Statistics
OR ..... Odds ratio
OT ..... Open Targets
PCV ..... Procarbazine, lomustine and vincristine
PPAR ..... Peroxisome proliferator-activated receptor
pQTL ..... Protein quantitative trait loci
PWCoCo ..... Pairwise conditional and colocalisation analysis
QTL ..... Quantitative trait loci
RCT ..... Randomised controlled trial
RR ..... Risk ratio
RNA ..... Ribonucleic acid
SD ..... Standard deviation
SE ..... Standard error
SNP ..... Single nucleotide polymorphism
TERT ..... Telomerase reverse transcriptase
TMZ ..... Temozolomide
(T)P53 ..... (Tumour) protein 53
VEGF ..... Vascular endothelial growth factor
WBRT ..... Whole-brain radiotherapy
WHO ..... World Health Organisation
WR ..... Wald ratio

Analyses and results presented in this thesis were also reported in the following pre-published and peer-reviewed articles. For each article relevant to this thesis, an author contribution statement is provided with signatures provided from the first and last authors to confirm these statements.

Peer-reviewed Publications

A. E. Howell, J. W. Robinson, R. E. Wootton, A. McAleenan, S. Tsavachidis, Q. T. Ostrom, M. Bondy, G. Armstrong, C. Relton, P. Haycock, R. M. Martin, J. Zheng, and K. M. Kurian.

Testing for causality between systematically identified risk factors and glioma: a Mendelian randomization study.

BMC Cancer, 20(1):508, 2020. *First author position shared between Howell and Robinson*. (Chapter 1).

My contributions to this publication were as follows. Cleaned and meta-analysed the glioma outcome data. Performed all statistical analyses presented in the paper and interpreted all results thereof. Analyses were replicated between the co-first authors. Co-drafted the manuscript by writing the methods and results sections related to statistical methodoloy, and co-drafted the discussion and conclusions sections. Produced specific figures and tables (figure 2, table 1, additional figures 6 and 7, and additional tables 4 through 7) for publication. Howell performed the literature search and prepared for publication the introduction and other co-drafted sections of the manuscript. Any figures and tables not mentioned were produced by Howell. Supervision was provided by Caroline Relton, Richard Martin, Jie Zheng and Kathreena Kurian. Support was provided from Spiridon Tsavachidis, Quinn Ostrom, Georgina Armstrong and Melissa Bondy for the extraction of the glioma outcome data. Robyn Wootton assisted with collection and analysis of smoking-related data. Alexandra McAleenan supported with the literature search. Philip Haycock provided valuable insight into the MR analyses. All co-authors read and provided comments for the original and subsequent re-drafts of the paper.

FIRST AUTHOR:  ..... LAST AUTHOR 

J. W. Robinson, R. M. Martin, S. Tsavachidis, A. E. Howell, C. L. Relton, G. N. Armstrong, M. Bondy, J. Zheng, and K. M. Kurian.

Transcriptome-wide Mendelian randomization study prioritising novel tissue-dependent genes for glioma susceptibility.

Sci Rep, 11(1):2329, 2021. (Chapter 4).

My contributions to this publication were as follows. Conceptualisation of the project and analysis plan. Identified, obtained and cleaned of all relevant datasets. Conducted and interpreted all analyses. Produced all figures and tables. Wrote the original and subsequent manuscripts. All co-authors read and provided comments for the original draft. Supervision was provided by Richard Martin, Kathreena Kurian and Jie Zheng. Support was provided from Amy Howell, Spiridon Tsavachidis, Georgina Armstrong and Melissa Bondy for the extraction of the glioma outcome data.

FIRST AUTHOR: . [redacted] ..... LAST AUTHOR [redacted]

Pre-print Articles

J. W. Robinson, R. Martin, M. Ozawa, M. M. C. Elwenspoek, M. T. Redaniel, K. Kurian, and Y. Ben-Shlomo.

Role of treatments for diabetes and hyperlipidaemia in risk and mortality of primary and secondary brain tumours.

medRxiv, 2020. (Chapter 6).

My contributions to this pre-print were as follows. Performed data management, and cleaning and linkage of the data. Performed all statistical analyses and interpreted results. Wrote the original version and re-drafts of the manuscript. Produced all figures and tables for eventual publication. All co-authors read and provided comments for the paper. Data were extracted by Martha Elwenspoek with support from Maria Theresa Redaniel. Original analyses were conceptualised by Richard Martin, Mio Ozawa, Maria Theresa Redaniel, Kathreena Kurian and Yoav Ben-Shlomo. Input for further analyses was provided by myself. Mio Ozawa conducted initial exploratory analyses. Writing support was provided by Richard Martin and Yoav Ben-Shlomo, who also provided support for interpretation of results. Supervision was provided by Richard Martin, Kathreena Kurian and Yoav Ben-Shlomo.

FIRST AUTHOR: .. [redacted] ..... LAST AUTHOR [redacted]

Additional Publications

J. Zheng, V. Haberland, D. Baird, V. Walker, P. C. Haycock, M. R. Hurle, A. Gutteridge, P. Erola, Y. Liu, S. Luo, J. Robinson, T. G. Richardson, J. R. Staley, B. Elsworth, S. Burgess, B. B. Sun, J. Danesh, H. Runz, J. C. Maranville, H. M. Martin, J. Yarmolinsky, C. Laurin, M. V. Holmes, J. Z. Liu, K. Estrada, R. Santos, L. McCarthy, D. Waterworth, M. R. Nelson, G. D. Smith, A. S. Butterworth, G. Hemani, R. A. Scott, and T. R. Gaunt.

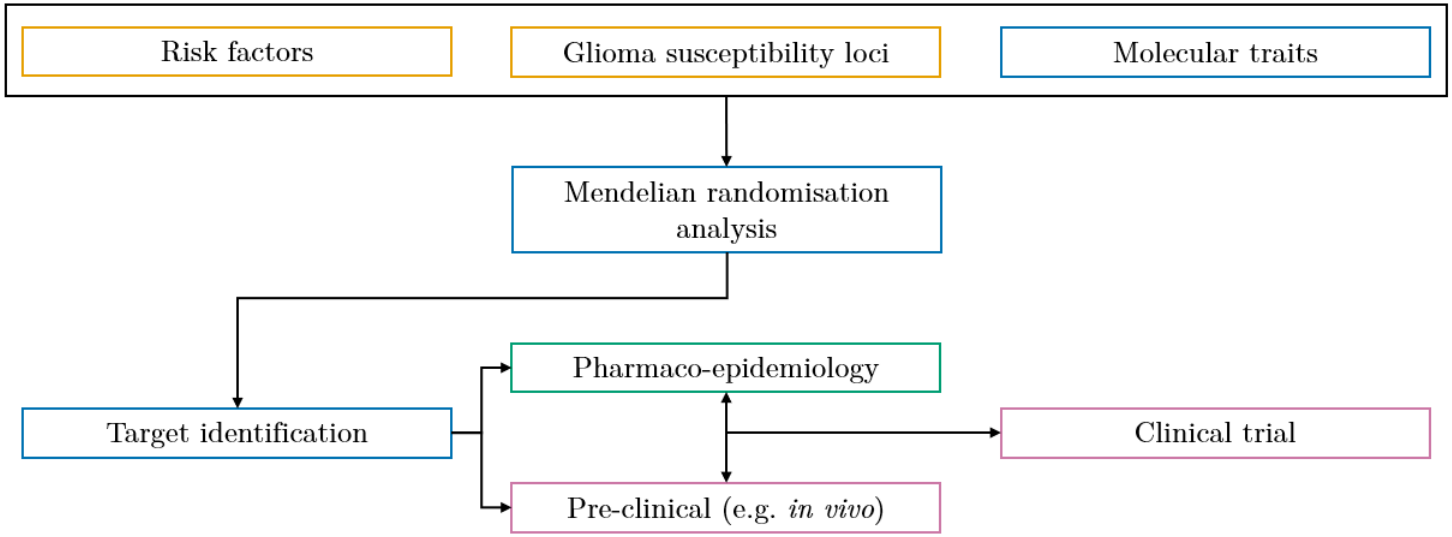
Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases.

Nat Genet, 2020.

J. Yarmolinsky, C. J. Bull, E. E. Vincent, J. Robinson, A. Walther, G. D. Smith, S. J. Lewis, C. L. Relton, and R. M. Martin.

Association between genetically proxied inhibition of HMG-CoA reductase and epithelial ovarian cancer.

JAMA, 323(7):646–655, 2020.



Key

- Inter-connected traits proxied by genetic variants.
- Contextually important, though not the main focus of this thesis (Chapter 1).
- Factors that form the main body of this thesis to inform drug discovery (Chapters 4-5).
- Next step in the drug discovery pipeline to increase evidence for a drug target (Chapter 6).
- Other important steps to provide evidence for drug targets that are beyond the scope of this thesis.

CHAPTER 1

INTRODUCTION

This doctoral thesis used novel approaches to generate insights into the role of gene expression pathways in mediating germline genetic variation on risk of glioma. In addition, analyses levereaged both genetically informed and observational methods to identify novel therapeutic targets for glioma and brain tumour chemoprevention.

1.1 Incidence

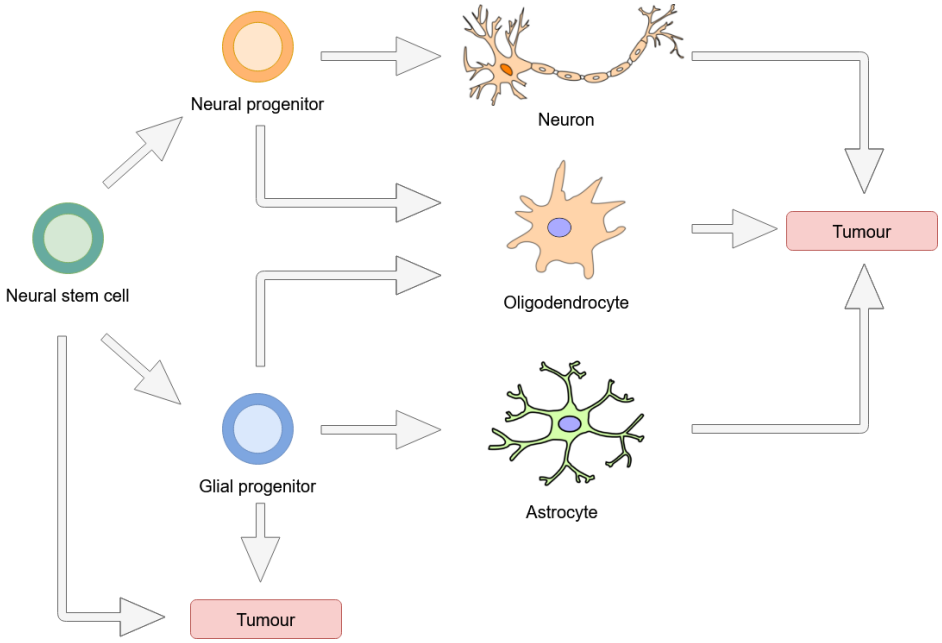
Gliomas are a subset of central nervous system (CNS) tumours and are the most common type thereof, with two thirds of brain, CNS and intracranial tumours in England between 2006 and 2010 diagnosed as gliomas [50]. The overall age-adjusted incidence rate for all brain, CNS and intracranial tumours is roughly 6.4 per 100,000 [327]. Gliomas are the largest group of these with overall age-adjusted incidence rates ranging from 4.67 to 5.73 per 100,000 [247]. It has been observed that gliomas tend to occur more frequently in: the elderly, where median age of diagnosis is 64 [247]; males [50, 209, 247]; and people of white ethnicity [50, 209, 247–249]. However, ethnicity may be subject to confounding as previous studies have shown a relationship between higher socioeconomic status and increased incidence of glioma, which also correlates with white ethnicity [267, 381].

Between 2014 and 2016 in the UK, brain tumours (BT) formed a total of 3% of all total cancer cases [50]; however, there are still many years of life lost and the impact upon quality of life can be significant if the tumour does not prove fatal. According to Cancer Research UK, there has been a 36% increase in BT incidence since the early 1990s [50], which could increase further due to an ageing population.

1.2 Aetiology

Gliomas originate from glial cells and are differentiated based on phenotypic cell characteristics, though these are not always the cell types of origin. The three cell types are ependymal cells (ependymomas), astrocytes (astrocytomas and glioblastoma) and oligodendrocytes (oligodendrogliomas). Mixed gliomas exist and may contain multiple different types of glial cells (Figure 1.1). Further differentiation by tumour grade is based on cell morphology and histological and molecular markers (described later). Generally, gliomas are heterogeneously distributed across the brain though tend to be more commonly found in the cerebral lobes and rarely seen in deeper structures, such as the cerebrum or cerebellum [195].

Figure 1.1: A simplified graph showing how cells may differentiate (arrowheads) within the CNS. Neural stem cells (NSC) produce more NSCs whilst also producing progressively more restricted progenitor cells which eventually differentiate into mature cells types such as neurons, oligodendrocytes and astrocytes. Accumulation of genetic mutations throughout these stages may induce gliomagenesis be it from an NSC, a progenitor cell or a mature cell type.



1.3 WHO Classification

BT are a heterogeneous group of tumours that are generally sub-classified based on the malignancy and histological type of tumour. The World Health Organisation (WHO) grades these tumours from I (benign) to IV (highly malignant), with each grade having histological similarities.

Grade I tumours, which tend to be astrocytomas, are not as proliferative as higher grades. These tumours are commonly seen in children, where more than 80% of astrocytomas are low grade [296], though adult tumours of this type are sometimes seen. Maximal safe resection of the

tumour tends to be the normal treatment regimen, which will generally suffice for treatment due to the low diffusive nature of the tumour [328].

Grade II tumours are more proliferative and tend to reoccur more than grade I tumours due to increased diffusivity. Treatment generally consists of resection of the lesion, though radiotherapy is sometimes considered, the benefits of which are disputed [221, 310, 339]. These tumours can also progress into higher grade tumours, such as diffuse astrocytoma transforming to the higher malignancy grade III anaplastic astrocytoma and glioblastoma [209].

A diagnosis of a grade III tumour depends on clear malignancy. These tumours are diffusive in nature and treatment can be difficult, which includes surgical resection, radiotherapy and/or chemotherapy.

Finally, grade IV tumours are those which progress and infiltrate rapidly throughout brain tissue and the CNS. Glioblastoma (GBM), a grade IV glioma, accounts for more than 60% of all brain, CNS and spinal tumours in adults [294] making it the most common primary brain tumour in adults and has a survival rate of roughly one year to 15 months after diagnosis [335, 343]. Treatment for these tumours involves surgery, radiotherapy and chemotherapy; however, due to the nature of rapid growth, even the post-operative lesion can grow quickly and result in a fatal outcome.

The current WHO guidelines for CNS tumour classification were updated in 2016 to include a range of molecular markers that differentiate brain tumour diagnoses [209]. A revised edition of the guidelines was expected to be released in 2020 with further molecular markers incorporated into the diagnosis procedure, however, the release has been inevitably delayed. A selection of important and relevant clinical molecular biomarkers that are currently used in a clinical setting were described in the following Section 1.6 - *Histological Tumour Markers* based on the 2016 WHO guidelines, and will likely still be included in the future revision.

1.4 Metastatic and Secondary Tumours

Metastatic tumours spread from outside of the brain, CNS and intracranial regions and generally have poorer prognosis than those BT that develop in situ [100]. Incidence of metastatic tumours is not well measured, potentially due to a mixture of issues such as asymptomatic disease or under-reporting in the presence of other metastases, and thus estimates are likely to be conservative; however, incidence is increasing over time due to better outcomes for patients with other systemic cancers [100]. Tumours metastasise most commonly from lung (particularly non-small-cell lung cancer), breast, colorectal, kidney and melanoma by haematogenous dissemination [237].

The term "secondary" may be applied to metastatic tumours or tumours that generally develop from lower-grade primary tumours. A common example is secondary GBM which tends to develop from low-grade or anaplastic gliomas (low-grade diffuse astrocytoma, WHO grade II; or anaplastic astrocytoma, WHO grade III) and has a different histological, epidemiological and clinical outlook

than primary GBM [245].

1.5 Risk Factors

1.5.1 Accepted Risk Factors

There are only two known risk factors for glioma, of which only exposure to ionising radiation is modifiable [318]. Glioma risk has also shown to have a genetic component, as certain Mendelian cancer syndromes appear to predispose those individuals to glioma and other cancers [209]. However, it is important to note that even when exposed to risk factors for glioma, known or unknown, the absolute risk of developing glioma remains low. There is a clear gap in the literature whereby more modifiable risk factors for glioma could be identified in the absence of RCTs by methods that may still provide causal evidence, such as Mendelian randomisation (MR, described in Section 2.3.3 - *Mendelian Randomisation*).

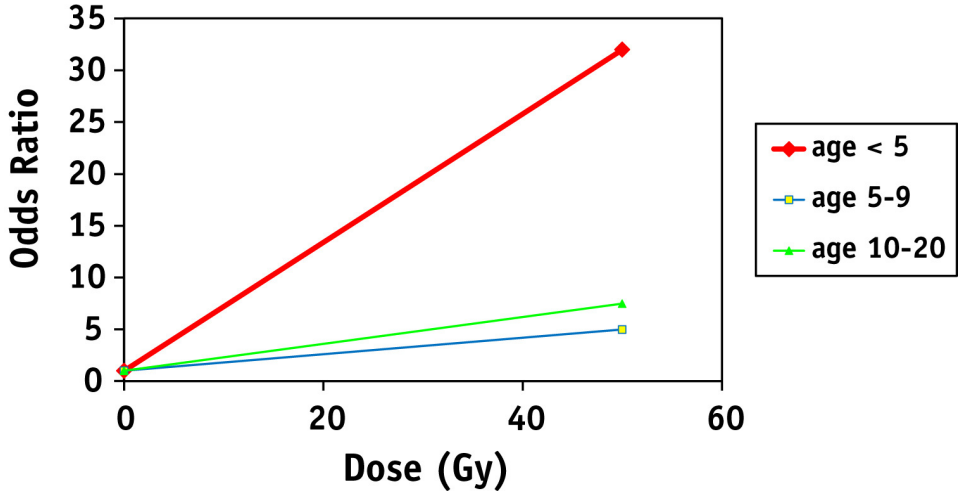
1.5.1.1 Exposure to Ionising Radiation

Previous exposure to ionising radiation is the only established environmental risk factor for glioma [318]. A study of 10,834 individuals in an Israeli cohort for tinea capitis (fungal infection of the skin of the scalp, eyebrows and eyelashes) found an increased risk in both benign meningiomas and malignant BT correlated to larger doses of childhood radiation [299]. Furthermore, another study using a cohort of 14,361 individuals from the Childhood Cancer Survivor Study investigated exposure to radiation therapy and incidence of primary CNS tumours. They found a higher risk of glioma in children irradiated at an early age that was independent of original cancer diagnosis and treatment with chemotherapy [231]. A study by Inskip, *et al.* showed the presence of a dose-response effect where higher radiation doses were associated with higher tumour risk; however, they also showed that age of first cancer also influenced BT development, with younger individuals at higher risk (Figure 1.2) [158]. A retrospective study showed that the use of CT scans in children may triple the risk of brain cancer [262], whilst a systematic review showed that the association between exposure to ionising radiation and glioma was not as strong as that with meningioma but still a risk factor [39].

1.5.1.2 Genetic Susceptibility

Genetic disorders, such as Li-Fraumeni syndrome which can result in *TP53* germline mutations, Turcot syndrome which can cause germline mutations in *APC* and *hMLH1*/*hPSM2*, and neurofibromatosis 1 and 2 have all been shown to increase risk of developing glioma [209, 244]. These disorders are commonly inherited within families, which indicates a genetic susceptibility. This has been shown in a review of genome-wide association studies (GWAS) of glioma [179]. However, these genetic disorders and their risk variants are not sufficient, or even necessary, for

Figure 1.2: Risk of developing glioma after radiation therapy by age of first cancer. A younger age of first cancer correlates strongly with increased risk of developing glioma. While the odds ratio (OR) appears to be very large for those under age five, the relative risk for glioma development remains small. For example, of 12,268 five-year survivors, only 1,003 (8.2%) received a first diagnosis of astrocytoma. Units given in gray (Gy) for radiation. From Inskip, *et al.* [158].



gliomagenesis. Instead, it is additional somatic mutations that allow for tumour development to take place [286].

It is not clear how much influence genetic susceptibility has on "familial" glioma, where glioma is present in two or more close family members. Gliomas that occur in families account for roughly 5% of all glioma cases [250], but family-based studies have shown a high correlation between direct relatives and glioma risk as compared to the general population. The Gliogene Consortium, which genotyped 75 US glioma families, found a moderate-high penetrance susceptibility locus at 17q12-21.32 though more work needs to be done to identify the exact causal gene [166, 325]. Another study using the Gliogene Consortium data also found that families with mutations in *POT1* all had a member diagnosed with oligodendroglioma [14].

1.5.2 Postulated Risk Factors

This Section described several postulated risk factors for glioma that still lack conclusive evidence or are otherwise not yet widely accepted. A selection of papers from 2000 onwards are presented in Table 1.1 to demonstrate the state of the evidence for these risk factors over the last 20 years. Further discussion about each risk factor is given in subsections that follow.

Table 1.1: An overview of postulated risk factors and selected, non-exhaustive papers which provided evidence for or against an association. Papers were selected if they were published after 2000 and to give a broad overview of the state of the literature. For example, there are many studies that have investigated the potential association between atopy and glioma risk; however, papers are selected here to demonstrate how this association has changed over time due to increasing sample sizes and more sophisticated study designs. Ratios are given as presented in the paper: odds ratio (OR), hazard ratio (HR), relative risk (RR) and incidence rate (IR).

Study	**Subtype**	**Sample size**	**Risk factor**	**Ratio (95% CI) [Measure]**	**Notes**
Atopy, allergies and immunological disorders					
Schwartzbaum, *et al.* (2003) [317]	Glioma (also meningioma)	68 gliomas, 51,999 controls	Diagnosis of immune-related condition	0.46 (0.14, 1.48) [HR]	Study suffers from low number of cases.
Linos, *et al.* (2007) [203]	Glioma (also meningioma)	3,450 gliomas	Allergic disease	0.61 (0.55, 0.67) [RR]	Meta-analysis of 12 studies from 1990 to 2006. Authors rule out methodological bias from sensitivity analyses but studies use self-reporting which may introduce bias.
			Asthma	0.68 (0.58, 0.80) [RR]	
			Eczema	0.69 (0.58, 0.82) [RR]	
Lachance, *et al.* (2011) [191]	High-grade glioma (HGG, WHO grade III/IV)	855 HGG, 1,160 controls	History of allergies	0.62 (0.51, 0.76) [OR]	Meta-analysis of HGG cases only that looked at history of allergies in association with certain glioma risk alleles to ascertain potential gliomagenesis mechanisms.

			History of allergies and does/does not carry 9p21.3 (*CDKN2B*) risk allele	0.40 (0.28, 0.58) [OR] (w/o risk allele) 0.76 (0.59, 0.97) [OR] (w/ risk allele)	
			History of allergies and does/does not carry 20q13.3 (*RTEL1*) risk allele	0.68 (0.54, 0.86) [OR] (w/o risk allele) 0.44 (0.29, 0.68) [OR] (w/ risk allele)	
Zhao, *et al.* (2014) [399]	Glioma	9,986 cases, 118,950 controls	Allergic condition	0.78 (0.73, 0.83) [OR]	Another meta-analysis of 18 studies from 1990 to 2013. Authors also ruled out publication bias as a potential explanation.
Disney-Hogg, *et al.* (2018) [83]	Glioma	12,488 cases, 18,189 controls	Asthma and hay fever Atopic dermatitis Immunoglobin E (IgE) level Self-reported allergy	0.96 (0.90, 1.03) [OR] 0.96 (0.93, 1.00) [OR] 0.88 (0.69, 1.13) [OR] 1.03 (0.95, 1.11) [OR]	MR study. Authors concluded that this study did not provide strong evidence for an association due to follow-up sensitivity analyses that did not show the same result.
Howell, *et al.* (2020) [152]	Glioma, GBM & non-GBM	5,739 cases, 5,501 controls	Allergic disease	1.29 (1.01, 1.67) [OR]	Another MR study. Unlike other studies, allergies disease was associated with increased risk of GBM.

			IgE level	0.92 (0.57, 1.50)	
Occupational health and hazards					
Schlehofer, *et al.* (2004) [311]	HGG & low-grade glioma (LGG)	1,178 cases, 1,987 controls	Female agricultural workers	0.60 (0.36, 0.99) [OR]	16 occupational categories were tested between men and women. Comparator groups had < 5 years employment history in that occupation. Other results were null or consistent with chance. Results potentially imply a dose-response effect for chemicals used in agricultural and food processing professions.
			Female working in food production or processing	1.95 (1.04, 3.68) [OR]	
Ruder, *et al.* (2012) [297]	Glioma	798 cases, 1,175 controls	Ever farm job (≥ 1 year) vs never	1.36 (0.89, 2.08) [OR]	Interviews used to collect the data may also have been performed via proxy; the authors provided ORs for some of the exposures excluding proxy interviews. The authors also provided results for more than just the two jobs listed here; however, these two results were highlighted by the authors despite small sample sizes.
			Pesticide use in non-farm job vs never	0.77 (0.56, 1.06) [OR]	
			Trade job (vs professional)	0.97 (0.74, 1.27) [OR]	
			Service job (vs professional)	0.94 (0.72, 1.23) [OR]	

		16 cases, 10 controls	Engineers, architects (longest total employment duration)	2.50 (1.12, 5.60) [OR]	
		27 cases, 21 controls	Food processing workers (longest total employment duration)	1.78 (0.99, 3.18) [OR]	
Lacourt, *et al.* (2013) [192]	Glioma	1,800 cases, 5,160 controls	Formaldehyde	0.8 (0.6, 1.1) [OR]	All chemicals were ever exposed vs never exposed, but also investigated potential dose-response effects. The authors concluded none of the agents they investigated were associated with glioma risk due to either results being attenuated or due to lack of a dose-response effect.
			Oil mist	0.8 (0.6, 1.1) [OR]	
			Diesel exhaust emissions	1.0 (0.8, 1.2) [OR]	
			Petrol exhaust emissions	1.0 (0.8, 1.2) [OR]	
			Benzo(a)pyrene	0.8 (0.6, 1.0) [OR]	
			Sulphur dioxide	2.0 (1.0, 3.8) [OR]	
			Asbestos	0.9 (0.8, 1.1) [OR]	
			Wood dust	1.1 (0.8, 1.5) [OR]	
Ruder, *et al.* (2013) [298]	Glioma	798 cases, 1,175 controls	Exposure to any chlorinated solvent	0.86 (0.68, 1.08) [OR]	-
Parent, *et al.* (2017) [258]	Glioma, HGG and GBM	1,800 cases, 5,160 controls	Cadmium	1.1 (0.7, 1.6) [OR]	ORs given for cumulative exposure of the metal. The authors concluded there is no evidence for any effect of the metals (with differing exposure durations) on risk.
			Chromium	0.9 (0.7, 1.1) [OR]	
			Iron	0.9 (0.7, 1.1) [OR]	
			Lead	0.8 (0.7, 1.0) [OR]	
			Nickel	0.9 (0.8, 1.1) [OR]	

			Welding fumes	0.9 (0.7, 1.1) [OR]	
Diet and lifestyle					
Chen, *et al.* (2002) [58]	Glioma	20 cases, 71 controls	Beans	0.4 (0.2, 0.8) [OR]	Study that used food-frequency questionnaires to determine associations between diet and glioma risk. Intake across the participants was split into quartiles; ORs were given for highest quartile vs lowest quartile. The study also included other nutrients such as nitrates, nitrites, vitamin C and saturated fats. Presented here are the main results of the paper.
		47 cases, 113 controls	Dark yellow vegetables	0.6 (0.3, 1.0) [OR]	
		44 cases, 110 to 119 controls	Pro-vitamin A carotenoids	0.5 (0.3, 0.8) [OR]	
		63 cases, 110 to 119 controls	Retinol	1.6 (1.0, 2.8) [OR]	
		47 cases, 110 to 119 controls	Dietary fibre	0.6 (0.3, 0.9) [OR]	
Michaud, *et al.* (2009) [222]	Glioma	335 cases	Processed meat consumption	0.92 (0.48, 1.77) [RR]	Results given for highest quintile vs lowest. Authors also looked at red meat consumption, nitrosamines and other antioxidant measures but there was no evidence.
			Nitrate	1.02 (0.66, 1.58) [RR]	
			Nitrite	1.26 (0.89, 1.79) [RR]	
			Vitamin C	0.88 (0.62, 1.26) [RR]	

Braganza, *et al.* (2014) [40]	Glioma and GBM	65 cases, 477,095 controls	Cigarette smoking (current vs never)	0.83 [HR]	(0.63, 1.09)	This large, prospective study concluded that smoking and alcohol consumption did not increase risk of glioma.
		104 cases, 477,095 controls	> 30 cigarettes per day (vs never)	0.95 [HR]	(0.75, 1.20)	The authors found evidence for protective associations but point out this is contradictory to existing knowledge and concluded may be due to chance or residual confounding (e.g., SES).
		159 cases, 477,095 controls	No alcohol (vs < 1 drink per week)	0.93 [HR]	(0.75, 1.15)	
		545 cases, 477,095 controls	Per drink per day (vs < 1 drink per week)	0.96 [HR]	(0.92, 0.99)	
Takahashi, *et al.* (2018) [341]	Glioma, GBM and non-GBM	12,488 cases, 18,169 controls	Vitamin D levels	1.21 [OR]	(0.90, 1.62)	An MR study wherein the authors concluded their study did not provide evidence for a causal relationship between vitamin D and glioma risk.
Saunders, *et al.* (2020) [309]	Glioma, GBM and non-GBM	12,488 cases, 18,169 controls	Vitamin D levels	0.99 [OR]	(0.86, 1.15)	Another MR study that studied lifestyle and dietary factors, as well as inflammatory and cardiometabolic factors. Despite some of the results showing evidence of an association, these did not hold up in follow-up sensitivity analyses due to bias or confounding.
			Circulating carotenoids	1.09 [OR]	(0.94, 1.26)	
			Mono-unsaturated fatty acids	0.88 [OR]	(0.76, 1.03)	

			Serum calcium	0.84 [OR]	(0.71, 0.98)	
			Serum vitamin B6	1.31 [OR]	(1.06, 1.62)	
			Omega-3 fatty acids	0.83 [OR]	(0.71, 0.98)	
Non-ionising radiation						
Hardell, *et al.* (2003) [132]	BT (inc. metastases)	247 cases, 218 controls	Analogue phone usage (ever vs never)	1.03 [OR]	(1.04, 1.60)	One of the earlier studies Hardell, *et al.* conducted into phone usage and brain tumour risk. Authors concluded analogue cellular phone usage had evidence of an increased risk of brain tumours.
		423 cases, 433 controls	Digital phone usage (ever vs never)	1.04 [OR]	(0.90, 1.30)	
		402 cases, 396 controls	Cordless phone usage (ever vs never)	1.10 [OR]	(0.90, 1.30)	
The INTERPHONE study group (2010) [163]	Glioma (also meningioma)	2,708 cases, 2,972 controls	Regular use in the past ≥ 1 year (yes vs no)	0.81 [OR]	(0.70, 0.94)	No increase in glioma or meningioma risk was associated with mobile phone use. The authors noted that there were suggestions of increased risk at the highest exposure levels but bias and confounding limited causal inference.
			As above, but with tumour in the temporal lobe	0.86 [OR]	(0.66, 1.13)	
			As above, but with tumour in the parietal or frontal lob	0.77 [OR]	(0.62, 0.95)	

			As above, but with tumour in other locations	0.79 [OR]	(0.51, 1.23)	
Frei, *et al.* (2011) [103]	CNS tumours including glioma and meningioma	10,729 cases	≥ 10 years of mobile phone usage, males only (glioma)	1.03 [IR]	(0.83, 1.27)	A large cohort study of mobile phone subscribers in Denmark found that there was no overall increased risk of CNS tumour incidence and mobile phone usage.
			≥ 10 years of mobile phone usage, females only (glioma)	1.04 [IR]	(0.56, 1.95)	

1.5.2.1 Atopy, Allergies and Immunological Disorders

Previous research has shown an inverse association between atopy (allergies) and glioma in a wide range of studies, including cohort [317], case-control [42, 380] and systematic reviews [57, 203, 399]. However, some studies have shown a weaker or null results [28, 117], whilst an MR study showed no association between genetically predicted atopy with glioma risk [83]. A more recent meta-analysis conducted by Amirian, *et al.* sought to provide a consensus on the subject of atopy and glioma by using the Glioma International Case-Control study (GICC) consisting of 4,533 cases and 4,171 controls. They found a history of respiratory allergies resulted in approximately 30% reduction in glioma risk when compared to not having a respiratory allergy, whilst asthma and eczema also had a similar protective effect [5]. As of yet, no mechanisms or pathways have been found that explain this protective effect due to allergies, though some have been proposed. For example, one explanation of how atopy can lead to reduced risk of glioma is through interplay of particularly cytokines (e.g., IL-4) that may lead an immune response to gliomas [203]. This explanation has yet to be shown as causal, however, as the association between glioma and atopy becomes clearer, then research can shift towards that area and a more definitive answer can be found.

1.5.2.2 Occupational Health and Hazards

Occupational hazards have been postulated to increase risk of developing glioma, particularly for those occupations that deal with ionising and non-ionising radiation, chemicals and agriculture [243, 247, 311]. However, the evidence for these hazards affecting glioma risk is contradictory and no relationship has been concretely drawn between any particular occupation and glioma risk [153, 244, 247, 318].

1.5.2.3 Diet and Lifestyle

Diet has received a lot of attention as a potential protective factor for glioma, specifically the ketogenic diet – a low-carbohydrate and high-fat diet that increases ketone bodies in the blood. Originally developed in the 1920s due to its anti-seizure properties, it fell out of use over a decade later due to the development of new anti-convulsants [13]. Ketogenic diets have recently seen renewed interest due to their therapeutic effects for treating refractory epilepsy (drug-resistant epilepsy), particularly in paediatric epileptic patients where adherence to a ketogenic diet has been shown to decrease seizure frequency [102, 143]. It has been postulated that the ketogenic diet can be used to treat other neurological disorders through the mechanism whereby ketone bodies replace glucose in the brain as an energy source, resulting in enhanced mitochondrial function and substrate delivery [19]. In particular, ketogenic diets may treat brain cancers due to the relative metabolic inflexibility that malignant BT display as compared to normal neurons and glia, which readily adapt to a new substrate (ketones instead of glucose) for energy [323].

A study by Zhou, *et al.* demonstrated this effect in mouse models, where mice on a ketogenic diet exhibited anti-tumour and anti-angiogenesis effects as compared to mice on a standard, unrestricted carbohydrate diet [403]. Another study in mice models found that a ketogenic diet also enhanced the anti-tumour effects of radiation, suggested the use of the diet as an adjuvant to standard care of malignant gliomas [1]. However, there exist only a small number of studies examining the efficacy of the ketogenic diet in human BT. A case report for a patient with GBM showed that standard therapy with concomitant ketogenic diet resulted in no re-growth of BT tissue for two months, until the diet was suspended, when tumour recurrence was observed 10 weeks later [406]. A case report of two patients, together with details of a further 19 patients from a trial (NCT01535911), and another 11 from other case reports, showed that the ketogenic diet showed no major side effects and was safe to use in treatment of glioma; however, there was no conclusive evidence that use of the ketogenic diet improved treatment and concluded that further research was necessary [316]. There are currently ongoing RCTs attempting to measure progression-free survival of glioma patients using the ketogenic diet (NCT01754350, NCT03075514, NCT01535911). Some of these trials have finished but have yet to publish their results.

Other aspects of lifestyle, including overall diet and obesity, vitamin intake, alcohol consumption and tobacco usage, have also been implicated in glioma development. Obesity is considered to be a risk factor for meningiomas but there is little evidence of this being true for gliomas as well [84, 322, 379]. Likewise, for other dietary choices, such as intake of nitrates, nitrites, vitamins C and D and cholesterol, there is little-to-no evidence supporting an effect on risk of glioma [58, 87, 222, 243, 341]. Similarly, research into lifestyle factors, in particular alcohol consumption and tobacco usage, have not provided convincing evidence for an increase (or decrease) in risk of glioma [26, 40]. There have been myriad observational studies that focus on self-reported lifestyle choices, though these studies have remained controversial due to many potential explanations of failure to find consistent results, including: small sample sizes for cases; false-positive results relating from sample size; imprecise measurement of exposures due to the self-reported nature of studies, compounded by proxy reporting when glioma patients become unavailable; and disease heterogeneity [244, 318].

1.5.2.4 Exposure to Non-ionising Radiation

In 2013, WHO/International Agency for Research on Cancer (IARC) categorised non-ionising radiation ("radiofrequency electromagnetic fields" in their classification guide) from the likes of mobile phones as a group 2B carcinogen (i.e. possibly carcinogenic to humans) for brain cancer [160]. Due to the increasing prevalence and usage of mobile phones, there has been concern over the potential public health risk this could cause.

IARC created their own study group, called INTERPHONE, to investigate the potential link between mobile phone usage and BT risk. They conducted an interview-based case-control study

of 2,708 glioma and 2,409 meningioma cases from 13 countries, but found no increase in risk of either tumour [163]. They reported a slightly elevated risk for those with the highest exposure levels (1,640 hours or more total recalled call time associated with OR = 1.40 (95% CI: 1.03, 1.89) for glioma, however, given the presence of "improbable data" in this group, it was highly likely this study suffered from some sort of error or bias [163]. Overall, they concluded non-ionising radiation did not increase risk, though the study was limited by biases and errors that did not allow for causal inference.

Another large study from the Danish cohort study linked data from the CANULI study (socioeconomic status), Danish Cancer Registry (cancer identification) and mobile phone subscription records. In this cohort study, 358,403 subscription holders allowed for a total of 3.8 million person years follow-up on binary exposure to non-ionising radiation. They observed no increased risk for BT associated with mobile phone usage and suggested there was little evidence of a causal relationship between the two [103].

A study using a prospective cohort in the UK, the Million Women Study, also investigated the link between mobile phone usage and CNS tumours. Using nearly 800,000 women over a follow-up of seven years, Benson, *et al.* concluded there was no association between mobile phone usage and glioma incidence (ever vs never users of mobile phones, RR = 1.01; 95% CI: 0.90, 1.14; $P = 0.82$) [27].

The driving force for IARC to include radio-frequency electromagnetic fields as a group 2B carcinogen came mostly from Hardell, *et al.* in a series of studies, though there have been smaller studies published which showed a positive association as well [11, 308]. The Hardell studies were a series of case-control studies using linked data from Swedish regional health and population registries that analysed self-reported (or via proxy) use of both cellular and, in later studies, cordless telephones. Hardell's earlier studies showed no association between telephone use and BT risk [133]. However, a follow-up to the 2002 study [131] showed that digital wireless telephone use was associated with increased risk of BT when considering ipsilateral use (i.e., the BT appeared on the same side of the brain as most common usage of the telephone) but not when considering contralateral tumours [132]. Another follow-up analysis sought to ascertain whether geographical location (urban vs rural) should be considered as a covariate which, despite mixed results, Hardell, *et al.* advised should be taken into consideration in future assessments [130]. The rationale for this study and the inclusion of geographical location is based on the idea that a feature of mobile phones called adaptive power control, which is regulated by the distance between base stations, may increase the specific absorption rate of radio frequency due to mobile phones in urban areas, where base stations are closer together. Though the authors conclude that their study provides support that exposure may differ between areas, this may only be relevant to Sweden (where the study was conducted) and may only be relevant for exposure to the Global Systems for Mobile Communications (GSM, also commonly referred to as "2G") standard, which, although still in common use, is outdated due to newer standards. A pooled analysis of the two

1997-2003 case control studies showed a positive risk association across many types of tumour and duration latency; particularly, the result for under 20s was highlighted: OR = 5.0 (95% CI: 1.5, 16) for mobile phone use only [128]. More recently, Hardell, *et al.* performed another pooled analysis of their studies from 1997-2003 and 2007-2009, and found again a positive association between glioma incidence and mobile phone usage across all of their subgroup analyses [129].

Careful interpretation of these analyses is required due to potential sources of bias and error, such as: recall bias, which is likely to affect the results of these studies, especially when considering the outcome of glioma or BT will negatively impact patient's ability to respond; proxy reporting, which can also lead to error in the collection of the data, similar to recall bias; reporting bias, where respondents may selectively withhold information from the study; and finally survival bias, as seen in the earlier studies by Hardell, *et al.* which decided to exclude deceased patients. It is still unclear whether there is an underlying association between non-ionising radiation and brain cancer risk – independent replication of the results from the Hardell, *et al.* studies has yet to be published. Furthermore, causal inference can be difficult due to the ubiquitous nature of mobile phone usage and lack of instruments with which their usage can be proxied in an instrumental variable framework.

1.5.2.5 Genetic Polymorphisms

There has been an explosion of genome-wide association studies (GWAS) within recent years, with the GWAS Catalog containing information from 4,220 publications (as of October 2019) [46]. Since their introduction, GWAS have had many successes in uncovering loci associated with complex disease. In the case of glioma, there have not been as many GWAS conducted when compared to other, more common diseases; however, GWAS still have managed to uncover a number of susceptibility loci associated with glioma risk (Table 1.2) [168, 180, 220, 279, 304, 325, 333, 364, 389].

Whilst GWAS have become more common, a number of issues have been raised about the methodology. One such concern is that GWAS explain only a small portion of the heritability of a complex trait [215]. A classic example is given by Visscher who summarises how human height has been studied within the context of GWAS; a highly heritable trait, estimated to be 80% genetically determined, GWAS of thousands of people have identified around 50 variants which account for only 5% of the phenotypic variation in human height [358]. It has been 12 years since this article was published and, despite technological advances meaning more granular and cheaper genotyping, the heritability of height explained by genetic variants has increased to about 20% [387]. However, in twins studies, for example, the heritability of height has been found to be lower than the aforementioned 80%, and instead it has been posited that environmental and epigenetic effects also play a part in the heritibility of height, and not just genetics [167]. Another issue raised about GWAS is that the variants found may have little biological or clinical utility, rendering the results largely unactionable [359].

Despite the limitations of GWAS, results arising from such studies have become incredibly valuable due to their widespread utility beyond the basic reporting with the phenotype tested. For example, results from GWAS are an important tool in the field of causal inference using MR (described in greater detail in Section 2.3.3 - *Mendelian Randomisation*).

Table 1.2: Results from GWAS for glioma susceptibility loci. OR are in respects to the risk allele, in bold. Table adapted from Kinnersley, *et al.*[179].

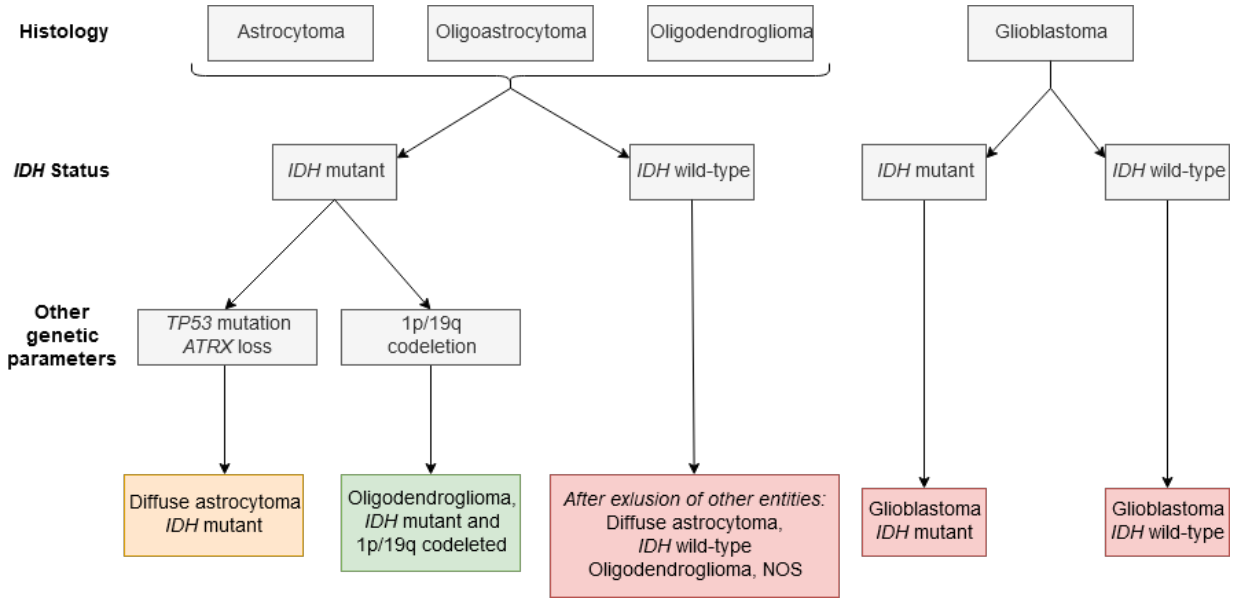
Reference	**Gene**	**Locus**	**SNP**	**Allele**	**MAF**	**OR (95% CI)**	***P***
	TERT	5p15.33	rs2736100	T/**G**	0.49	1.27 (1.19, 1.37)	1.50×10^{-17}
	CCDC26	8q24.21	rs4295627	**G**/T	0.17	1.36 (1.28, 1.43)	2.34×10^{-18}
Shete, *et al.*	*CCDC26*	8q24.21	rs891835	**G**/T	0.22	1.24 (1.17, 1.30)	7.54×10^{-11}
2009 [325]	*CDKN2A/B*	9p21.3	rs4977756	A/**G**	0.38	1.24 (1.19, 1.30)	7.24×10^{-15}
	PHLDB1	11q23.3	rs498872	C/**T**	0.31	1.18 (1.13, 1.24)	1.07×10^{-8}
	RTEL1	20q13.33	rs6010620	**G**/A	0.24	1.28 (1.21, 1.35)	2.52×10^{-12}
Wrensch, *et*	*CDKN2A/B*	9p21.3	rs1412829	**C**/T	0.38	1.42 (1.27, 1.58)	1.85×10^{-10}
al. 2009	*RTEL1*	20q13.33	rs6010620	**G**/A	0.24	1.52 (1.32, 1.75)	3.40×10^{-9}
[389]	*RTEL1*	20q13.33	rs4809324	**C**/T	0.11	1.60 (1.37, 1.87)	1.70×10^{-9}
Sanson, *et*	*EGFR*	7p11.2	rs11979158	**A**/G	0.16	1.23 (1.15, 1.35)	7.72×10^{-8}
al. 2011	*EGFR*	7p11.2	rs2252586	**T**/C	0.29	1.18 (1.11, 1.25)	2.09×10^{-8}
[304]							
Stacey, *et*	*TP53*	17p13.1	rs78378222	T/**G**	0.01	2.35 (1.61, 3.44)	1.0×10^{-5}
al. 2011							
[333]							
Jenkins,	*CCDC26*	8q24.21	rs55705857	A/**G**	0.06	6.3 (4.6, 8.8)	2.2×10^{-28}
et al.							
2012							
[168]							
Walsh, *et*	Near	3q26.2	rs1920116	**G**/A	0.29	1.30 (1.19, 1.42)	8.3×10^{-9}
al. 2014	*TERC*						
[364]							
	VTI1A	10q25.2	rs11196067	**A**/T	0.40	1.19 (1.12, 1.27)	4.32×10^{-8}
Kinnersley,	*ZBTB16*	11q23.2	rs648044	C/**T**	0.42	1.25 (1.17, 1.34)	6.26×10^{-11}
et al. 2015	Intergenic	12q21.2	rs12230172	**G**/A	0.47	1.23 (1.16, 1.32)	7.53×10^{-11}
[180]	*POLR3B*	12q23.3	rs3851634	**T**/C	0.29	1.23 (1.15, 1.32)	3.02×10^{-9}
	ETFA	15q24.2	rs1801591	G/**A**	0.09	1.36 (1.23, 1.51)	5.71×10^{-9}
Melin, *et al.*	*JAK1*	1p31.3	rs12752552	**T**/C	0.13	1.22 (1.15, 1.31)	2.04×10^{-9}
2017 [220]	*MDM4*	1q32.1	rs4252707	G/**A**	0.19	1.19 (1.12, 1.26)	3.34×10^{-9}
	AKT3	1q44	rs12076373	**G**/C	0.19	1.23 (1.16, 1.32)	2.63×10^{-10}

Near *IDH1*	2q33.3	rs7572263	**A**/G	0.24	1.20 (1.13, 1.26)	2.18×10^{-10}
LRIG1	3p14.1	rs11706832	A/**C**	0.45	1.15 (1.09, 1.20)	7.66×10^{-9}
OBFC1	10q24.33	rs11598018	**C**/A	0.33	1.14 (1.09, 1.20)	3.39×10^{-8}
Intergenic	11q14.1	rs11233250	**C**/T	0.12	1.24 (1.16, 1.33)	9.95×10^{-10}
MAML2	11q21	rs7107785	**T**/C	0.50	1.16 (1.11, 1.33)	3.87×10^{-10}
AKAP6	14q12	rs10131032	**G**/A	0.09	1.33 (1.22, 1.44)	5.07×10^{-11}
Near *MPG*	16p13.3	rs2562152	A/**T**	0.14	1.21 (1.13, 1.29)	1.93×10^{-8}
LMF1	16p13.3	rs3751667	C/**T**	0.23	1.18 (1.12, 1.25)	2.61×10^{-9}
HEATR3	16q12.1	rs10852606	T/**C**	0.30	1.18 (1.13, 1.24)	1.29×10^{-11}
SLC16A8	22q13.1	rs2235573	**G**/A	0.47	1.15 (1.10, 1.20)	1.76×10^{-10}
Near *TERC*	3q26.2	rs3772190	**G**/A	0.15	1.11 (1.06, 1.15)	2.25×10^{-6}
TERT	5p15.33	rs10069790	C/**T**	0.01	1.61 (1.53, 1.69)	8.33×10^{-74}
EGFR	7p11.2	rs75061358	T/**G**	0.03	1.63 (1.50, 1.76)	4.94×10^{-34}
EGFR	7p11.2	rs723527	**A**/G	0.43	1.25 (1.20, 1.31)	4.79×10^{-23}
CCDC26	8q24.21	rs55705857	A/**G**	0.05	3.39 (3.09, 3.71)	7.28×10^{-149}
CDKN2A/B	9p21.3	rs634537	T/**G**	0.41	1.37 (1.31, 1.43)	7.23×10^{-45}
VTI1A	10q25.2	rs11599775	**G**/A	0.36	1.16 (1.10, 1.22)	3.44×10^{-9}
ZBTB16	11q23.2	rs648044	**A**/G	0.42	1.19 (1.13, 1.25)	4.66×10^{-12}
PHLDB1	11q23.3	rs12803321	**G**/C	0.14	1.42 (1.35, 1.49)	6.33×10^{-43}
Intergenic	12q21.2	rs1275600	**T**/A	0.42	1.16 (1.10, 1.21)	3.72×10^{-9}
RFX4	12q23.33	rs12227783	**A**/T	0.10	1.16 (1.08, 1.24)	1.60×10^{-5}
ETFA	15q24.2	rs77633900	G/**C**	0.08	1.35 (1.25, 1.46)	1.60×10^{-13}
TP53	17p13.1	rs78378222	T/**G**	0.01	2.53 (2.19, 2.91)	8.64×10^{-38}
RTEL1	20q13.3	rs2297440	T/**C**	0.22	1.48 (1.40, 1.56)	3.66×10^{-46}

1.6 Histological Tumour Markers

Histological biomarkers are used in the delineation of what could otherwise appear as similar tumours and come in two types: prognostic and predictive. Prognostic biomarkers can inform the outcome of the disease regardless of treatment whilst predictive biomarkers can help clinicians decide on a course of therapy specific for each individual's tumour [16]. The 2016 WHO classification guide specifically includes histological markers for use of diagnosis and prognosis for use in practice [209] with isocitrate dehydrogenase (*IDH*) mutations, 1p/19q co-deletion and O6-alkylguanine DNA alkyltransferase (*MGMT*) promoter methylation routinely assessed in a clinical setting due to their diagnostic and prognostic properties (Figure 1.3), though others are

Figure 1.3: How histological markers guide diagnosis of diffuse tumours. Colour of diagnosis represents the patient outlook: green is favourable, orange is intermediate, and red is poor. Adapted from Louis, *et al.* from *WHO Classification of Tumours of the Central Nervous System* [209].



also used (as described in the following Sections).

1.6.1 *IDH* Mutation

Mutations in *IDH1* and *IDH2*, referred to simply as *IDH* mutation, occur early in the development of diffuse gliomas [368]. This histological biomarker is commonly seen in stage II and III astrocytomas (roughly 70%) and in about 10% of GBMs, mostly in secondary GBMs [123]. It has been postulated that *IDH* mutations are not simply a consequence of the development of tumours but are actually a driver in oncogenesis in potentially two ways: one such way is that *IDH* mutations are causally associated with syndromes such as Ollier's disease (also referred to as enchrondromatosis, a disorder whereby benign growths of cartilage, called enchondromas, develop between bones due to *IDH* mutations) and Maffucci syndrome (similar to Ollier's disease but characterised by additional growths of tangles of abnormal blood vessels called haemangiomas), which carry an increased risk for glioma [256, 281]; another such way arises from increased tumour cell proliferation and decreased ability to differentiate as observed in experiments [184, 211]. Typically, *IDH* mutations occur in much younger patients than those with *IDH* wild-type tumours (i.e. those tumours without *IDH* mutations) [136]. Most grade II and III gliomas, as well as secondary GBM (GBM that has started as a lower grade astrocytoma, for example, and form roughly 10% of all adult GBM cases) have *IDH* mutations and generally have a much better prognosis than those that do not [371]. Thus, presence of *IDH* mutations is

prognostically considered favourable.

1.6.2 1p/19q Co-deletion

1p/19q co-deletion, defined by the complete deletion of the short arm of chromosome 1 and the long arm of chromosome 19, occurs early in the pathogenesis of oligodendrogliomas, is mostly associated with those only [49, 209], and seen in about 80% of cases [288]. The biological role behind 1p/19q codeletion is currently unknown, but two candidate genes with observed mutations have been highlighted as avenues of research: *FUBP1*, located on 1p, which is mutated in a small subset of oligodendrogliomas; and *CIC*, located on 19q, which is mutated in most tumours [29, 169, 301]. As a predictive biomarker, three randomised trials have shown presence of 1p/19q co-deletion indicates that a patient will respond well to both radiotherapy and chemotherapy bettering their outcome and chances of survival [159, 350, 375].

1.6.3 *MGMT* Promoter Methylation

MGMT is a DNA repair enzyme, expressed ubiquitously throughout human tissue, that increases resistance to alkylating agents [263]. *MGMT* methylation is seen in about 40% of GBMs and 80% of *IDH*-mutated, low grade gliomas [209]. Methylation of the promoter which controls *MGMT* activity silences the gene and ceases production of MGMT in cells, decreasing resistance to alkylating agents. *MGMT* levels vary greatly between tumours, even of the same type, with some 30% of gliomas lacking any sort of MGMT production [329, 330]; those tumours that lacked MGMT production were shown to have increased chemosensitivity and could predict response to alkylating agents [93, 94, 138]. Although *MGMT* methylation status cannot be used as a diagnostic biomarker, it can be used to predict response to alkylating agent chemotherapy and, as Weller, *et al.* proposed, should be used to influence clinical decision-making [372]. Such is the power of *MGMT* promoter methylation as a predictive biomarker, Wick, *et al.* advised in their paper that *MGMT* status should determine whether a patient is given temozolomide (TMZ) therapy – going as far to advise withholding TMZ therapy from elderly GBM patients whose tumours do not have *MGMT* promoter methylation [376]. However, as of yet, WHO do not include *MGMT* status in their guidelines, despite being the most widely performed and clinically acted-upon biomarker test.

1.6.4 *TP53* Mutation

Tumour protein 53 (*TP53*) on 17p encodes for a protein of the same name, although the protein is sometimes referred to as p53. This protein is a tumour suppressor which keeps in check cell proliferation and in cases where cell damage is present, can determine whether the damaged cells undergoes DNA repair or apoptosis. It is through this mechanism of stopping aberrant cell division that p53 helps to prevent tumourigenesis. Within glioma, *TP53* mutations has long

been associated with an increased risk of glioma and is present in approximately 30-40% of astrocytomas, both anaplastic and not, and GBMs [326, 367].

1.6.5 *TERT* Promoter Mutation

Telomerase reverse transcriptase (*TERT*) is a protein coding gene that forms part of the telomerase complex which lengthens telomeres on DNA strands. Through this mechanism, *TERT* may endow cells with immortality and self-renewal properties akin to stem cells, however differentiated cells do not express *TERT* themselves. Over- or re-expression of *TERT* can therefore cause the formation of telomerase in differentiated cells. This feature of *TERT* has been implicated in human tumourigenesis, though the genetic bias of this is not well understood. Two studies showed that germline and somatic mutations of the *TERT* promoter were observed in melanomas and other cancer cell lines [151, 156], whilst another looked at a range of cancers including CNS tumours. This study found that mutations in the *TERT* promoter were found in 43% of all cases, of which GBM had the highest frequency at 62% [357]. Analysis of *TERT* promoter mutations and GBM has shown an association between diagnosis at an older age [190], and poor prognosis either alone (e.g. *IDH* wild-type tumours) or with *EGFR* amplification [7, 174, 190]. These findings have provided evidence for the now-routinely tested for *TERT* promoter mutations in GBM tumours for patient prognosis.

1.6.6 *EGFR* Amplification

Epidermal growth factor receptor (*EGFR*) encodes for a transmembrane protein of the same name that acts as a receptor for the epidermal growth factor (EGF) type of extracellular protein ligands. Over-expression of *EGFR* has been observed and linked to many different cancers, including GBM (40-98% of cases [209]) where a specific mutation of *EGFR*, called *EGFRvIII* (27-33% of cases [209]), is often seen. By itself, *EGFR* amplification status is not a prognostic factor for outcome in GBM patients, except for cases with other defining characteristics, for example with *TERT* status; however, it has been shown that *EGFRvIII* expression may form a negative prognostic factor after one year of survival [139]. Moreover, overexpression of *EGFR* has also been linked to resistance of treatment, in particular radiotherapy [21, 77, 313].

1.6.7 Expression of *VEGF*

Vascular endothelial growth factor (VEGF) is a protein that stimulates proliferation and migration of vascular endothelial cells and plays a key role in angiogenesis. Expression of *VEGF* has been associated with a poor clinical outcome in patients with GBM [229]; however, more importantly is that VEGF represents an appealing target for therapeutics in GBM tumours as the key-angiogenic driving factor [89]. It is for this reason, whether or not to provide anti-angiogenic treatment to patients, that VEGF testing has become important within the clinic.

1.7 Treatment Regimen

Treatment for gliomas generally fall into three categories: surgery, radiotherapy and chemotherapy. Low-grade gliomas (LGG, WHO grade I/II), as the least invasive and angiogenic, tend (though not always) to be treated with surgery with follow up adjuvant therapy where applicable, whereas high-grade gliomas (HGG, WHO grade III/IV) are the most aggressive and thus require multi-modal approaches. Advice exists to guide clinicians on how to treat their patients though can be contentious due to conflicting evidence.

1.7.1 Surgical Treatment

The distinction between LGG and HGG is important as it generally categorises the diffusivity of the glioma allowing the surgeon to make an informed decision on the resection of the tumour; however, there exist no guidelines on the extent of resection, which can vary both on an intra- and intertumoural basis. There is evidence available that indicates the patient's age, histology and *IDH* status are as predictive of the patient's outcome as extent of resection [208, 236, 239, 386]. However, this does not mean maximal resection should not be strived for. A 2008 review of all major clinical publications since 1990 on the role of extent of resection in glioma outcome found evidence that extensive resection resulted in longer life expectancy for both LGG and HGG [303]. A study by Jakola, *et al.* examined whether early resection was favourable versus biopsy and "watchful waiting", finding that the early intervention was better (estimated 5-year survival of 74%, 95% CI: 64%, 84%) when compared to the watchful group (estimated 5-year survival of 60%, 95% CI: 48%, 72%) in LGG [165]. Evidence is mounting that the goal of surgery should be maximal resection of the tumour, allowing increased survival for the patient; however, this must be balanced with the preservation of the functionality of the pathways within the brain. Therefore, surgeons must ensure maximal, safe resection whilst considering the quality of life for the patient.

New methods are being developed that can aid surgeons perform more aggressive surgery without harming the patient. For example, a meta-analysis of the impact of intraoperative stimulation mapping (ISM), a technique used to monitor brain function through direct electrical stimulation (often under anaesthesia), found 3.4% (95% CI: 2.3%, 4.8%) of cases had severe neurological deficits after resections aided by ISM compared to 8.2% (95% CI: 5.7%, 11.4%) for those without ISM aided resections [73]. New technologies such as neuro-navigation and intraoperative magnetic resonance imaging (iMRI) are providing surgeons new tools to increase safety of surgery, and recently, fluorescence guided surgeries have shown promise in aiding peri-operative diagnosis of a glioma as LGG or HGG. One such study aimed to use 5-aminolevulinic acid (5-ALA), designed to fluoresce in HGG and not LGG, in this manner and found that out of 88 patients, 81 had visible fluorescence (1 LGG, 78 HGG (99% concordance with HGG classification, 99% CI: 91%, 99.9%) and 2 NOS), while 7 patients had no fluorescence (6 LGG and 1 HGG) [369].

Whilst the use of 5-ALA in this way does not inherently increase the safety of surgery for the patient, it does allow the surgeon to identify and thus resect more of the high-grade disease, potentially resulting in better surgical outcomes.

1.7.2 Radiotherapy

Similar to surgery, the guidelines for radiotherapy depends on the grade and diagnosis of the tumour.

The benefits of radiotherapy use for the treatment of LGG is still contentious. A phase 2/3 study (NCT00003375) sought to determine how radiotherapy with and without combination chemotherapy (in this case, procarbazine lomustine and vincristine) affected patients with LGG (WHO grade II). The study enrolled 251 patients and found that patients who received the combination therapy had longer median overall survival compared with those who received only radiotherapy (13.3 vs 7.8 years, HR = 0.59, 95% CI: 0.42, 0.83; P = 0.003). However, there was a cohort of patients who did not benefit from the combination therapy, with the authors unable to determine why this was. Furthermore, toxicity and side effects were higher in patients who received the combination therapy compared to the monotherapy. Regardless of this, Buckner, *et al.* conclude their results showed a clear benefit for the majority of patients and recommended combined therapy for LGG patients [45]. Another randomised, open-label, phase 3 study (European Platform of Cancer Research (EORTC)-22033-26033) randomly assigned 477 patients with an LGG diagnosis (astrocytoma, oligoastrocytoma or oligodendroglioma) to receive either radiotherapy or TMZ chemotherapy. They found after a median follow-up of 48 months, median progression-free survival in the TMZ group was 39 months (95% CI: 35, 44 months) and 46 months (95% CI: 40 to 56 months) in the radiotherapy group, and concluded there was no observable difference in progression-free survival between the two groups [22]. The same study also conducted an exploratory analysis in patients whose tumours were molecularly defined and found patients with *IDH* mutations and 1p/19q co-deletion tumours had a longer progression-free survival when treated radiotherapy (HR = 1.86; 95% CI: 1.21, 2.87; P = 0.0043) than with TMZ; this same effect was not seen in patients with *IDH* mutations but without 1p/19q co-deletion or in *IDH* wildtype [22]. A similar effect was seen in an earlier EORTC trial (EORTC 22845) comparing long-term efficacy of radiotherapy versus controls (delayed treatment) for LGG and found radiation after surgery increased period without progression but had no effect on overall survival [348]. However, another study found an increase in 5-year overall and progression-free survival: 57% for patients given post-operative radiotherapy compared to 47% for those who did not receive radiotherapy [395]. A recent systematic review on the management of LGG found that early radiation was associated only with better progression-free survival after 5 years compared to patients who had delayed or no radiation [44]. The literature surrounding the survival benefits of radiotherapy are contradicting for LGG and no clear consensus has yet to be drawn. Yet there are studies that have examined the effects of radiotherapy on quality of life and have found an

adverse effect on mid- to long-term cognitive function [43, 86, 182, 339], however this could be due to too-high radiation dosages or the effects could be managed by, for example, anti-epileptic drugs [20, 182].

Despite this mix of evidence, the aforementioned Buckner, *et al.* study has become practice defining for determining treatment for LGG; however, in the case of HGG, radiotherapy is frequently included in the treatment regimen [45]. A systematic review by Laperriere, *et al.* advised a variable dose of external beam radiotherapy as the standard treatment for malignant gliomas, depending on the patient's age [194]. Another study found that each delayed week until the start of radiotherapy increased risk of death by 8.9% (95% CI: 2.0%, 16.1%), which contradicts the "watch and wait" approach sometimes taken with LGG [164].

1.7.3 Chemotherapy

Temozolomide is the mainstay of chemotherapy for the treatment of GBM. The National Institute of Health and Care Excellence (NICE) granted authorisation for the use of TMZ as a second-line therapy for metastatic glioma patients in 2001 and then allowed it as a front-line drug for newly diagnosed cases in 2007. Dosage amounts depend on sub-diagnosis, which differs between non-GBM and GBM cases, and whether there is presence of *MGMT* promoter methylation. TMZ as an oral alkylating agent induces DNA methylation and tumour cytotoxicity and depletes the DNA-repair enzyme MGMT [138, 335]. Furthermore, TMZ is readily bioavailable, does not require hepatic metabolism for activation and easily passes the blood-brain barrier (BBB) [96]. Overall, the efficacy and safety of TMZ in either combined or adjuvant therapy is well documented in the literature [38, 135, 335, 375, 396].

For LGG, chemotherapy normally consists of a combination of chemotherapeutic drugs: procarbazine, lomustine (abbreviated as CCNU) and vincristine, and together are commonly referred to as PCV. The efficacy of PCV over carmustine (described below) was shown as early as 1990 in a study by Levin, *et al.*, where post-radiotherapy PCV was found to increase time to progression and survival, substantially so for patients with anaplastic glioma compared with carmustine alone [200]. Similar improvements to patient outcomes were seen in a practice-defining trial (NCT00003375) and in a post-hoc analysis by Bell, *et al.*, which included patients with grade II glioma sub-classified based on *IDH* and 1p/19q co-deletion status. In particular, patients with *IDH*-mutant and 1p/19q non-codeleted/co-deleted tumours had longer progression-free survival (HR = 0.32; P = 0.03; HR = 0.13; P = 0.001) and overall survival (HR = 0.38; P = 0.01; HR = 0.21; P = 0.03), respectively [24].

Carmustine, known also by its trade name, BCNU (β-chloro-nitrosourea), is an alkylating agent that is used mostly as an interstitial therapy for malignant glioma. This involves soaking biodegradable "wafers" in BCNU (known by the brand name, Gliadel wafers) and applying them to the tumour site post-operatively as a local treatment. BCNU treatment is normally reserved for malignant and metastatic brain tumours, such as GBM, but benefits have also been observed

in medulloblastoma, astrocytoma and tumours of the brain stem [373, 374]. Within randomised phase III clinical trials, BCNU wafers were found to have increased median survival versus placebo (13.8 months versus 11.6 months; HR = 0.73; 95% CI: 0.56, 0.95; $P = 0.018$) [374], but these studies also showed several severe complications and adverse events, including seizures, cerebral oedema, hydrocephalus and cerebral spinal fluid leaks [373, 374]. After the introduction of systematic TMZ as the mainstay of chemotherapy for glioma, studies sought to show how concomitant BCNU wafer and TMZ therapy is well tolerated and presents with no increase in morbidity or adverse advents compared to simple BCNU exposure [217, 253].

Recently, bevacizumab (brand name, Avastin), an anti-angiogenic therapy targeting VEGF, is becoming a popular choice of treatment for clinicians to treat highly vascularised, recurrent gliomas (mostly GBM). Whilst it is thought that bevacizumab does not cross the BBB [53], it has been observed that this is not necessary due to its effect to neutralise VEGF in the capillaries of the brain regardless [370]. A major phase II trial examining the effects of bevacizumab and irinotecan (a cytotoxic agent) in recurrent malignant glioma of 32 patients found that median progression-free survival was 23 weeks for all patients (95% CI: 15, 30 weeks; 20 weeks for grade IV patients and 30 for grade III), with 6-month progression-free survival probability of 38% and 6-month overall survival probability of 72% [362]. Another seminal study found bevacizumab was both safe and efficacious for patients as a combination therapy with chemotherapy (TMZ) [240]. A number of studies corroborate these findings, recommending bevacizumab as a therapy option for recurrent GBM and showing the potential as a first-line treatment as well [124, 186, 269, 363]. This evidence led the United States Food and Drug Agency (FDA) to grant accelerated approval for bevacizumab to be used in the treatment of recurrent GBM in 2009; however, the European Medicines Agency (EMA) decided to reject approval for bevacizumab as a second line treatment for GBM, though it is still administered off-label in Europe. A correspondence piece written by Wick, *et al.* published in the Journal of Clinical Oncology in 2010 provides a good summary of the European opinion of bevacizumab [377]. In summary, the EMA concluded that there was not enough evidence from studies to show that bevacizumab was efficacious in treatment of brain tumours, with questions still to be answered regarding optimal dosage, best timing of treatment and whether suboptimal study designs were used (for example, uncontrolled trials were used as evidence to support authorisation of the drug by the FDA) [377]. Further, a Cochrane review in 2014 (updated 2018) investigating anti-angiogenic therapy for HGG found no improvement on overall survival with its addition (pooled HR= 0.95; 95% CI: 0.88, 1.02; $P = 0.16$; 11 studies, 3,743 participants) and, with a number of follow-up analyses, concluded anti-angiogenic therapy should not be used in the treatment of newly diagnosed GBM [4]. The same study also concluded that combination therapy with chemotherapy, as opposed to chemotherapy alone, may result in a small improvement in overall survival in those with recurrent GBM [4]. However, this study did not consider adverse events which may negatively impact a patient's quality of life. A phase III trial (NCT00943826) found a slight improvement in progression-free survival for newly diagnosed

GBMs with bevacizumab as a combination therapy but with an increased rate of adverse events [59].

Another potential agent of interest in treating glioma is erlotinib (brand name Tarceva), which is a tyrosine kinase inhibitor that is primarily used to treat non-small-cell lung cancer and pancreatic cancer by inhibiting *EGFR*. Erlotinib is a relatively new, potential addition to the treatment regimen of glioma and as such, research is still on-going to ascertain its efficacy and safety, both as a sole and combination therapy. There is evidence to support the hypothesis that erlotinib passes the BBB which identifies it as a drug of interest for BT treatment [270]. An initial phase I trial found that GBM with high levels of *EGFR* expression and low levels of phosphorylated PKB/Akt responded better (defined as 50% decrease in tumour area measured by cross-sectional diameter, and with no increase in steroid doses) to erlotinib treatment (no response in 22 tumours with high levels of phosphorylated PKB/Akt levels compared to response in eight of 18 with low levels) [125]. A phase II trial (NCT00671970) sought to investigate the effect of combined bevacizumab and erlotinib therapy in patients with recurrent malignant glioma and found a progress-free survival of 6 months and median overall survival of 28% and 42 weeks for GBM patients and 44% and 71 weeks for anaplastic glioma patients; the authors conclude their results are similar to, or worse than, previous studies that investigate sole-bevacizumab therapy [307]. Another phase II trial (EORTC-26034-16031) by van den Bent, *et al.* investigated erlotinib versus TMZ or BCNU treatment in recurrent GBM and found, although well tolerated, 6-month progression-free survival was 11.4% (95% CI: 4.6%, 21.5%) in the erlotinib arm and 24% in the control arm [349]. Overall, there is currently very little evidence for the use of erlotinib in the treatment regimen for glioma.

However, there have been some successes in utilising targeted therapies for glioma treatment. One such example is combination therapy with kinase inhibitors, dabrafenib (brand name, Tafinlar) and trametinib (brand name, Mekinst), which are used to treat melanoma and non-small cell lung cancer by inhibiting BRAF and MEK respectively. Recent case studies have shown regression in *BRAF V600E*-mutant (a specific type of *BRAF* mutation) epithelioid gliomas and pleomorphic xanthoastrocytoma when given this treatment regime, both in adult and in paediatric cases [171, 223, 290]. There are trials underway to determine the efficacy of this therapy in glioma patients with *BRAF* mutations (NCT02684058, NCT04201457). More recently, there has also been limited evidence that a TRK inhibitor, larotrectinib (brand name, Vitrakvi), may be used to treat tumours which harbour TRK mutations, a key oncogenic driver. Research published in 2018 presented a case study detailing near-total resolution of lesions for a paediatric patient with TRK-mutated HGG [405]. Evidence is limited for treatment, but there is a pilot trial currently undergoing recruitment for use of larotrectinib in children with newly diagnosed HGG with NTRK fusion (NCT04655404) which hopes to provide more evidence for this drug as an efficacious treatment option.

1.7.4 Immunotherapy

Immunotherapy, defined as either activating or supressing the immune system to fight disease, is a new potential route of treatment being explored in multiple diseases and cancers. However, despite being recognised as an alternative treatment option for brain cancer, limited knowledge of the immune system has held back potential therapeutic benefits and applications. Although there have been advances in immunotherapy for other cancers, there are unique difficulties for CNS tumours involving immunity, specifically that these tumours grow in what is known as sanctuary sites [268]. However, studies are starting to show the brain is not completely immune privileged, for example, in the case of combatting John Cunningham virus (*human polyomavirus 2*, a common, frequently benign, virus that has been found to infect the brain) [252]. Future observation and experiments to identify possible candidate antigens are required before testing or trials can begin.

1.7.5 Treatment for Metastatic Tumours

Whole-brain radiotherapy (WBRT) is the main treatment option for metastatic tumours although its use has been waning due to the deleterious effects this can have on a patient's quality of life and other modalities of treatment that are equally as efficacious [300]. WBRT is also used as a preventative measure against metastases for patients with primary cancers that commonly migrate to the brain, in particular non-small-cell lung cancer [33]. Modern approaches to metastatic tumours now also involve stereotactic radiosurgery – a more localised and precisely-targeted radiotherapy – as well as medications to better manage symptoms (e.g. anti-convulsants) and chemotherapeutic and surgical therapies. However metastatic tumours are highly malignant and have a poor prognosis, highlighting the need for further research to improve the therapeutic options available. A promising avenue is personalising treatment to each patient's specific tumour, for example by prescribing *BRAF* inhibitors to patients whose tumours have *BRAF* mutations [213].

1.7.6 Summary

Overall, there remains many avenues of exploration that may improve patient outcomes and quality of life, whether this be through new operative techniques, technologies to aid surgical procedures, or development or re-purposing of drugs for treatment. As our knowledge of glioma and brain tumours improves, so too will our ability to better determine potential treatment modalities for clinical use.

1.8 New Approaches to Understanding Glioma Aetiology and Treatment

Diagnosis and stratification for many diseases has steadily improved over the last few decades due to novel methods and approaches that have expanded our knowledge of the aetiology of such diseases. Similar advancements have also improved treatment options for many diseases. For example, cancers with historically poor prognoses have seen steadily increasing survival rates, e.g. lung cancer [398], breast cancer [246] and prostate cancer [204]. Patient outcomes may be improved due to personalised medicine, whereby better per-patient diagnoses may be used to stratify risk and treatments, or simply due to better and more wide-ranging therapeutical options. These advancements are being applied to glioma, albeit slowly, as evidenced by the 2018 report from the UK's Department of Health and Social Care task and finish working group on BT research which has led to increased funding in the area [79].

1.8.1 Genetic heterogeneity between Tumours

Whilst genetic mutations are used to diagnose different tumour types, research has recently sought to understand the heterogeneity of inter-tumoural genetic mutations by use of spatio-temporal mapping. One such study examined the genetic profile of recurrent GBM in 38 patients, and found that the more distal the tumour from the primary site, the larger the divergence in driver mutations was from that initial tumour [175]. Another recent study found that the genetic and histological profile of metastatic GBM differed between primary site and in lung metastases, highlighting the need for and importance of genetic testing for diagnosis and effective treatment [111]. Similar heterogeneity was observed in diffuse intrinsic pontine glioma, an aggressive paediatric brain tumour, where whole exome sequencing showed evolutionary and spatial heterogeneity in non-primary sites [238].

1.8.2 Harnessing Germline Genetic Variation

Section 1.5 - *Risk Factors* described how there is a genetic component to glioma risk and that previous GWAS have found genetic variations at various loci that associate with glioma risk [220]. Many glioma GWAS have focussed on uncovering variants associated with risk, concentrating on increasingly larger sample sizes which can be beneficial for a rare disease like glioma. In the wider field of cancer epidemiology, GWAS and their results have been expanded to also detail variants which may also be associated with prognosis (as with breast cancer in 2002 [115]), treatment response (in this 2009 study examining drug responsiveness for paediatric acute lymphoblastic leukaemia [392]) and susceptibility across ethnicities [216]. Recently, results from GWAS have been increasingly used to leverage a statistical method called Mendelian randomisation (MR).

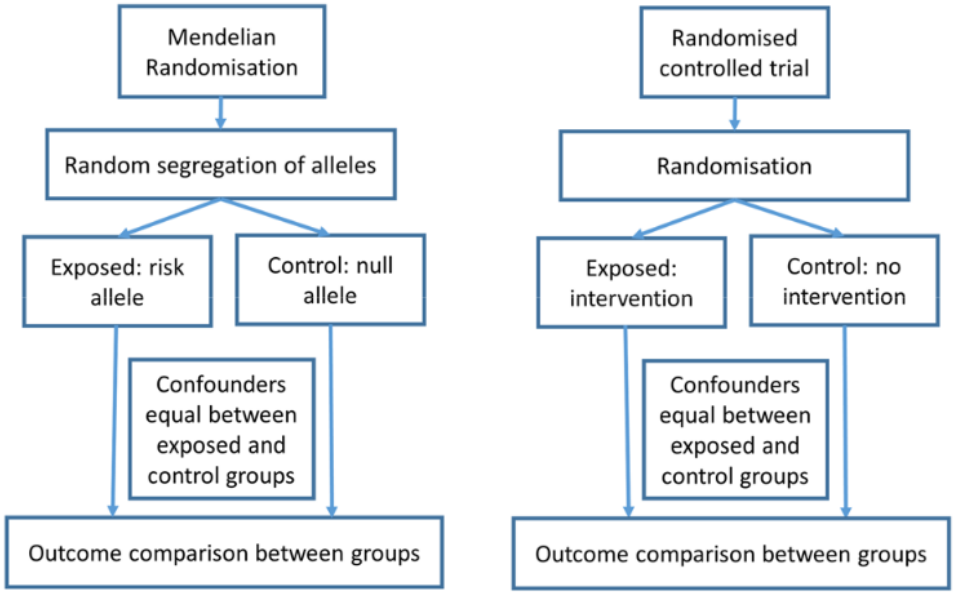
1.8.3 Epigenetics and Methylation in Glioma

Epigenetics and DNA methylation are becoming increasingly important in understanding many diseases, including cancer. It has been shown that many epigenetic mechanisms are aberrant in cancer, for example, DNA methylation, histone modifications and non-coding RNAs [324]. In the context of CNS tumours, DNA methylation has been successfully used to differentiate between subtypes and shown to be clinically and pathologically relevant [51]. Epigenetics in cancer research is an evolving field. Recently, datasets are being published from large-scale consortia, such as the Genetics of DNA Methylation Consortium (GoDMC), which capture the genetic effects on DNA methylation [224]. These data could further elucidate the role of epigenetics in disease aetiology and may be used in the aforementioned MR methodology.

1.8.4 Mendelian Randomisation to Strengthen Causal Inference

MR is a statistical method that utilises germline genetic variants to estimate causal relationships between exposures and outcomes [196] in a way that is analogous to the design of an RCT (Figure 1.4). Generally, MR has been applied to estimate the relationship between a modifiable risk factor and disease that can then inform public health advice, but usage has expanded to include applications such as the identification and validation of drug targets [312]. Investigators are increasingly using MR due to its ability to estimate causality and overcome unmeasured confounding that is common in other observational study designs [67]. This Section described a high-level overview of MR and its applications to glioma research, with the details of the methodology explained in Section 2.3.3 - *Mendelian Randomisation*.

Figure 1.4: Analogy of RCT and MR study designs. Randomly inherited germline genetic variants mimic the randomisation process of an RCT. From Pingault, *et al.* [266].



1.8.5 Strengthening Causal Inference for Putative Glioma Risk Factors using Mendelian Randomisation

I was the co-first author of a recent publication (Howell, Robinson, et al.) that aimed to identify novel risk factors for glioma, GBM and non-GBM risk [152]. My contributions to this paper were as follows. I cleaned and meta-analysed the outcome (glioma) data and performed all statistical (MR with sensitivity analyses and polygenic risk score) analyses and interpretation thereof. All analyses were independently repeated by both first authors. I co-drafted the manuscript, producing table 1, figure 2, additional tables 4 through 7 and additional figures 6 and 7 for publication, as well as writing parts of the methods and results sections related to statistical methodology. I also co-drafted the discussion and conclusions sections. Howell conducted the literature search for which this paper was based on, produced the first draft of the paper with any other figures or tables not otherwise stated.

There have been some MR studies that have sought to identify putative risk factors for glioma [83, 84, 341], including one paper for which I was co-first author [152]. In this paper, Howell and I used MR to investigate an assortment of risk factors and glioma risk, where risk factors were found by way of a systematic search of the literature and included if the risk factor could be instrumented by genetic variants in an MR analysis (i.e., that there were publicly available genetic variants associated with the risk factor). There was evidence that four traits increased risk of glioma, GBM or non-GBM. These were: longer leukocyte telomere length (odds ratio for non-GBM subtype ($OR_{\text{non-GBM}}$) = 4.05, $OR_{\text{all glioma}}$ = 4.09), liability to allergic disease

(OR_{GBM} = 1.29), increased alcohol consumption (OR_{GBM} = 8.37, $OR_{all\ glioma}$ = 4.42) and liability to childhood extreme obesity (defined as > 3 standard deviations away from the mean and excluded participants with known monogenic causes of obesity, $OR_{all\ glioma}$ = 1.11, OR_{GBM} = 1.12). Two traits also decreased the risk of non-GBM gliomas: increased levels of low-density lipoprotein cholesterol (LDLc, $OR_{non\text{-}GBM}$ = 0.79) and triglyceride levels (TG, $OR_{non\text{-}GBM}$ = 0.77). Results are summarised in Table 1.3. This study provided key causal evidence implicating these modifiable traits and glioma risk leading the way for future research to investigate underlying mechanisms driving these associations; however, given the less stringent P value threshold, these results should be interpreted with caution. Triangulation with other results form studies with orthogonal sources of bias would be required for these results to be informative for prevention of glioma.

Table 1.3: Summary of the MR results presented in Howell, Robinson, *et al.* [152] to identify risk factors associated with glioma subtypes. OR for disease risk are per unit increase in the exposure. A Bonferroni-corrected P value threshold was defined at $P < 1.00\times10^{-3}$ and a suggestive threshold at $1.00\times10^{-3} < P < 0.05$. Results for other subtypes which did not meet the P value threshold are presented in Appendix A.

Risk Factor	**Subtype**	**OR (95% CI)**	***P***
Alcohol consumption	All glioma	4.42 (1.07, 18.30)	0.041
	GBM	8.37 (1.69, 41.54)	0.0094
Allergic disease	GBM	1.29 (1.01, 1.67)	0.048
LDLc	Non-GBM	0.79 (0.63, 0.99)	0.04
Obesity (childhood extreme)	All glioma	1.11 (1.02, 1.21)	0.016
	GBM	1.12 (1.02, 1.22)	0.021
Telomere length	All glioma	4.09 (1.13, 14.86)	0.032
Triglycerides	Non-GBM	0.77 (0.59, 1.00)	0.049

This work was a key contribution to the literature regarding the identification of novel risk factors for glioma. However, it became apparent during this study that many other MR studies focussed on identifying modifiable risk factors for glioma, and only one paper (at the time of publication) utilised molecular traits in such a context [10]. Therefore, this formed part of the motivation for the work conducted and presented in this thesis.

1.8.6 Leveraging Mendelian Randomisation to Identify and Prioritise Drug Targets and other Interventions

Published GWAS of molecular traits can be integrated into an MR study design to garner novel insights into how such traits are causally implicated in disease [361, 404]. For example, GWAS have identified germline genetic variants associated with gene expression and protein abundance

levels [92, 99, 121, 336, 338, 393], including brain tissue-specific data [121, 289], which could provide novel insights into glioma risk.

As proteins are overwhelming the target of almost all pharmaceutical drugs today [193], an MR study that establishes a putative causal link between variation in protein abundance levels and glioma risk could then inform on a drug target for intervention to reduce risk. A drug target identified in this way could then be prioritised in further studies, e.g. *in vivo* or pharmaco-epidemiological studies, before a trial to determine efficacy and safety. MR could therefore form the first step in a drug discovery and prioritisation pipeline.

It is important to note that throughout the work presented in this thesis, the MR analyses utilised glioma data relating to risk. Therefore, putative drug targets identified from such analyses may be clinically relevant for primary prevention; it does not necessarily follow that the effect of such a drug would be seen in secondary or tertiary prevention. These conclusions would only be permissible if data from a progression GWAS were used, though utilising MR in this way is still in its infancy and comes with its own set of limitations [259].

1.8.7 Further Approaches to gain Insights into Novel Therapeutics in Glioma Treatment

There are other methods to gain insights for novel therapeutics in glioma treatment that also use large, population-level health datasets. Pharmaco-epidemiological techniques can be used in this way with large datasets to garner such insights, as described in Section 2.3.4 - *Pharmaco-Epidemiology*. An example of such a study comes from 2016, where Seliger, *et al.* sought to determine the relationship between diabetes, use of anti-diabetic drugs and glioma risk using a large longitudinal dataset consisting of digitised primary care medical records [320]. This dataset is called the Clinical Practice Datalink (CPRD, described in Section 3.3 - *Clinical Practice Research Datalink*) and may be used to provide observational evidence for how a drug affects both glioma risk and survival.

1.9 Aims of Thesis

The aims of this thesis were to:

1. Leverage germline genetic variation associated with glioma to elucidate mediating molecular pathways, via gene expression, on gliomagenesis.

2. In a similar way, leverage germline genetic variation associated glioma to elucidate mediating molecular pathways, via protein abundance levels, on gliomagenesis.

3. Utilise multi-omic data from the previous two aims, i.e. for gene expression and protein abundance levels, to identify novel targets for intervention.

4. Determine whether anti-hyperlipidaemia and anti-diabetes drugs can be repurposed to reduce incidence or improve prognosis of brain tumours.

CHAPTER 2

METHODOLOGICAL BACKGROUND

2.1 Introduction

The research objectives for this thesis are primarily to use germline genetic variation to strengthen causal evidence about molecular pathways linked to glioma risk. As such, epidemiological methods are well-suited to achieve these objectives. In particular, an important application of epidemiology is to inform public health strategies and interventions that can reduce the incidence and burden of disease which, for glioma, has been increasing as populations have become older (described in Section 1.1 - *Incidence*).

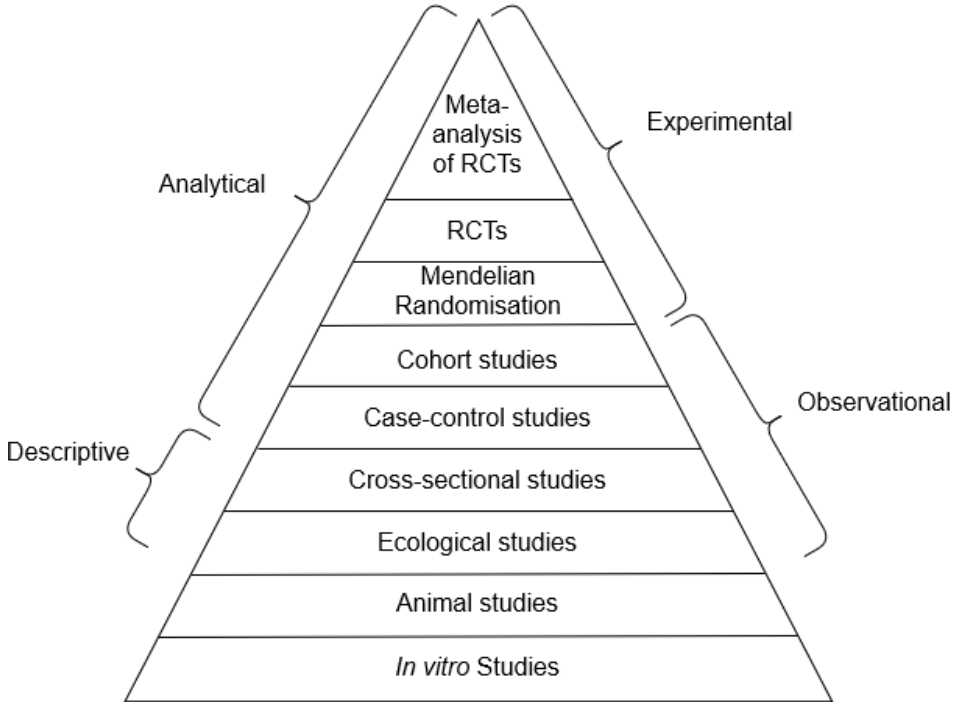
This Chapter provides a brief introduction to epidemiological research methodology and how this is relevant to the research presented in this thesis.

2.2 Causal Inference in Epidemiology

Causality and prediction are two interlinked and important concepts of epidemiology that both inform on public health. Prediction describes a relationship or phenomenon whereby an event, X, can predict an outcome, Y. Causality describes a relationship where if X causes Y then changing X will lead to a change in Y. When free of confounding or bias, causality is a subset of prediction as the occurrence of X will inherently predict Y; however, in practice, the presence of known or unknown confounding factors can make these relationships difficult to ascertain. Many valid associations in epidemiology are not necessarily causal – causality is oftentimes difficult to prove – but such predictive observations are still informative for guiding interventions. Therefore, many techniques in epidemiology seek to strengthen causal inference rather than prove causality (Figure 2.1), each of which is limited by various biases and complications that make causal inference difficult. It is common to supplement evidence from different studies with orthogonal

sources of bias to increase confidence in the association or result (sometimes reffered to as triangulation [197]).

Figure 2.1: Pyramid of study designs for causal inference. As the level of the pyramid increases, so too does the quality of evidence provided. Even at the highest level, a meta-analysis of RCTs provides evidence of, and does not singularly prove, causality meaning it is still important to pool evidence from many different sources with orthogonal sources of bias. Figure adapted from Lucas, *et al.* [212].



The research presented in this thesis is analytical and observational in nature, and made use of an instrumental variables approach called Mendelian randomisation (MR). MR has been adapted from the econometrics literature to help inform the process of making causal inference using observational data [196]. Section 1.8 - *New Approaches to Understanding Glioma Aetiology and Treatment* described how MR can be applied to glioma to provide insight into how variation in molecular traits affect glioma risk.

2.2.1 Observational Epidemiology

Observational epidemiology consists of studies wherein the investigators do not intervene on the exposure of interest, unlike in an experimental study. These studies may be descriptive, by describing a disease or condition of interest and generating hypotheses, or analytical, by testing hypotheses by investigating determinants, risk factors or causes of diseases in a more complex and rigorous framework. These study designs have different strengths and limitations,

Table 2.1: Summary of advantages and disadvantages of the case-control and cohort study designs. Adapted from Song, *et al.* [332].

Advantages	**Disadvantages**
Case-control study design	
May be used to study rare diseases.	If retrospective in design, the study can suffer from recall bias.
Generally, less expensive and time-consuming to conduct.	Cannot be used for incidence or diagnostic data.
Multiple exposures can be examined simultaneously.	Selection of controls can be difficult.
Provides initial evidence of associations for further research.	Selection bias can be introduced if not considered.
Cohort study design	
Temporal sequence is observed, allowing for assessment of causality.	A large cohort may be required for follow-up over a long period of time.
Multiple outcomes may be examined for a single exposure.	Selection bias can be introduced if not considered.
May be used to study rare exposures.	*Prospective cohort studies:*
Incidence, relative risk and risk difference can all be calculated from a cohort study.	Can be expensive and time-consuming.
	Loss to follow-up may induce bias.
	Retrospective cohort studies:
	Susceptible to recall bias.
	Data may be of poor quality due to less control over variables.

as summarised in Table 2.1 for the study designs used in the research presented in this thesis (case-control and cohort).

2.2.2 Statistical Methods used in Observational Epidemiology

The research presented in Chapter 6 used regression models to analyse data from the nested case-control and cohort studies to determine the effects of exposure to two drugs on brain tumour (BT) risk and survival. It is important to note that all BT were included in this analysis, and not just gliomas, to increase sample sizes (described in Section 6.2.1 - *Participants*). Regression analyses are statistical techniques that allow for the examination of the relationship between two, or more, variables of interest. Naively, regression analyses are a set of statistical methodologies that fit a line of best fit between the variables of interest that may then be interpreted to draw conclusions about the relationship of the variables being examined. One such type of regression analysis is linear regression, which is used to investigate continuous dependent variables and

fits a straight line to the data:

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{2.1}$$

where $(x_i, y_i), i = 1, .., n$ describes n datapoints for the independent and dependent variables, respectively, and ϵ denotes the error term. Equation (2.1) is of the same form of the equation for a straight line ($y = \alpha + \beta x$) and thus the relationship of the coefficients α and β with respect to the dataset characterises the linear regression model. Alternatively, logistic regression fits a logistic function to a dichotomous (binary) dependent variable. Given that the dependent variable is binary, let

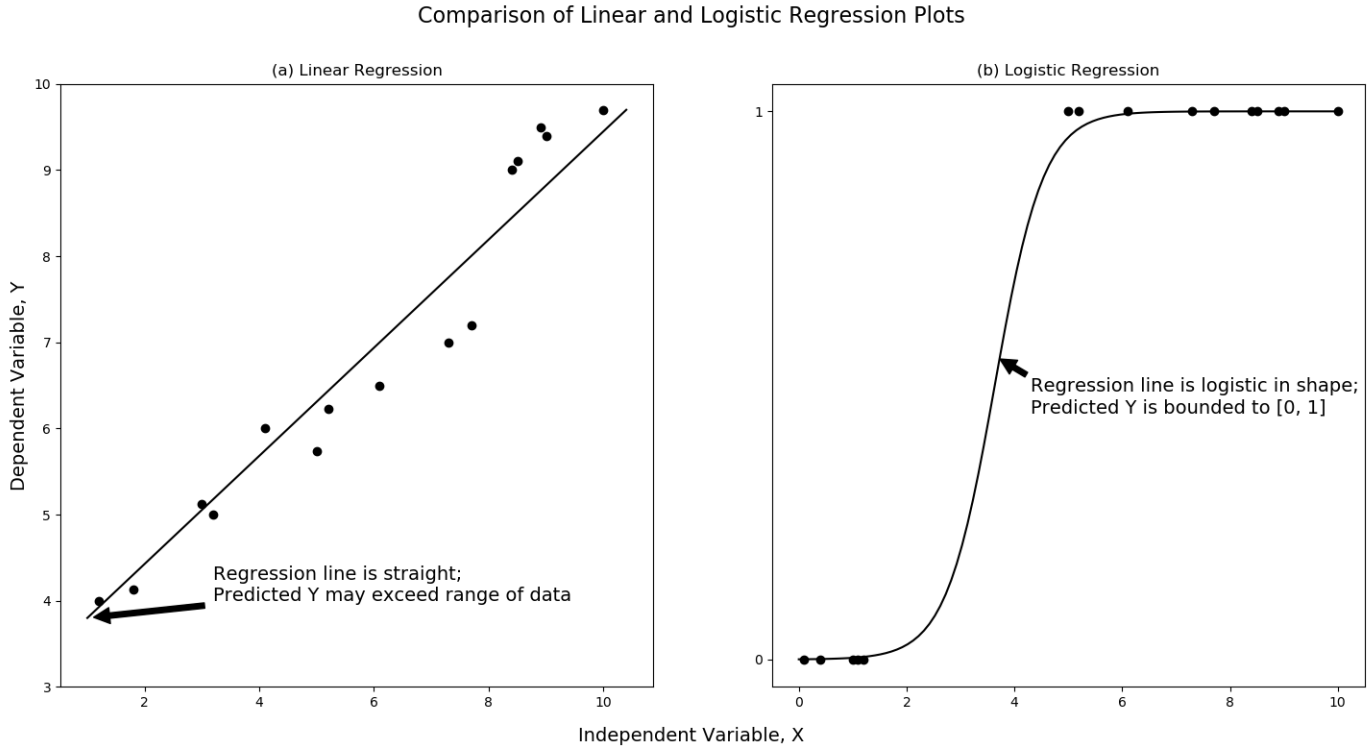
$$p = P(Y = 1) \tag{2.2}$$

such that p contains the probability that the dependent variable is true or another arbitrary dichotomous state. For $i = 1, ..., n$ datapoints, the logistic regression model may be characterised using Equation (2.2):

$$\log_b \frac{p}{1-p} = \beta_0 + \beta_i x_i \tag{2.3}$$

for some logarithm base, b. This will give the log-odds and may be exponentiated to find the odds of the dependent variable. These equations are plotted and compared in Figure 2.2.

Figure 2.2: A comparison of the shapes of linear and logistic regressions. Linear regressions (a) fit straight lines to the data and predicted values of the dependent variable may exceed that of the range of the data. A linear regression can also accept a categorical or binary dependent (outcome) variable. A logistic regression fits a logistic function (b) and accepts only dichotomous dependent variable and, as such, predicted values will be bounded to this range. For these graphs, data was randomly generated, and lines plotted to demonstrate the fitting shapes of the models.



2.2.3 Experimental Epidemiology

Randomised controlled trials (RCTs), if conducted appropriately, are considered the gold-standard for providing causal evidence for the effects of new drugs, treatments, or interventions due to the reduction of many biases inherent to other epidemiological study designs. However, they are generally prohibitively expensive, and many are not feasible to conduct due to ethical concerns (for example, by withholding efficacious treatment from the control arm). Furthermore, whilst an RCT will deliver a less biased estimate of the effect of the particular intervention on the outcome of interest in the patient group examined, this does not necessarily translate directly to the benefit of all patient groups. Whilst one should always conduct an RCT if it is possible, many times it is not, so evidence must be supplemented from other study designs such as those described above.

2.3 Integrative and Molecular Epidemiology

Integrative and molecular epidemiology is a large field that encompasses many different methodologies. Therefore, this Section focussed on genetic and pharmaco-epidemiology as these were pertinent to the research presented in this thesis.

2.3.1 Genetic Epidemiology

Genetic epidemiology is the study of the effects of genetic determinants on health and disease. One of the two accepted risk factors for BT is genetic susceptibility, highlighting the potential use of genetic epidemiology in understanding the aetiology and progression of the disease.

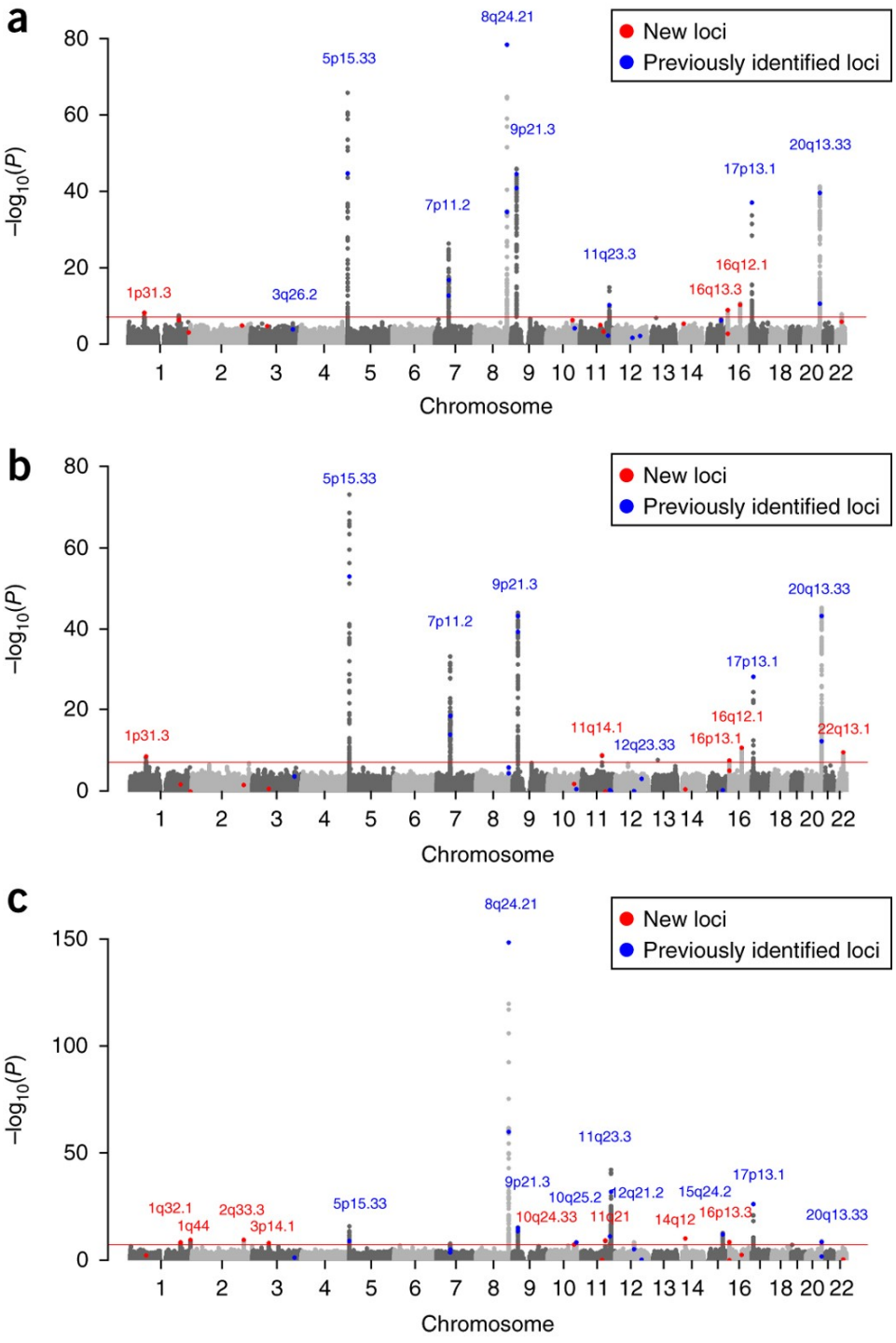
Genome-wide association studies (GWAS) are a common genetic epidemiological, hypothesis-free tool used to find genetic variants (usually single nucleotide polymorphisms, SNPs) associated with a phenotype of interest. They generally have little measurement error and thus produce robust and reliable results. An important concept for GWAS is linkage disequilibrium (LD), defined as the non-random association of alleles. Two alleles which are commonly inherited together, and are therefore correlated, are said to be in strong or high LD. GWAS leverage tag SNPs, i.e. a SNP which has strong LD with a haplotype or group of SNPs, to identify genetic variation in phenotypes without necessarily genotyping every SNP within a given region.

Despite repeated success in highlighting important loci for many diseases and traits, frequently variants identified by GWAS are found to not be causal of variation in the phenotype but may correlate or tag with the actual key gene. This can arise if the tag-SNP used by the microarray is not the causal variant but is still associated with the trait due to being in high LD with that variant [107]. Furthermore, many GWAS tag-SNPs are located in intergenic or intronic regions, meaning it is more likely that a variant will affect gene regulation through other mechanisms besides protein expression, e.g. transcriptional or epigenetic regulation [91].

There are methods to analyse GWAS results to identify potentially causal variants, such as fine-mapping, colocalisation and other functional studies [109].

The importance of GWAS cannot be understated. Section 1.5.2.5 - *Genetic Polymorphisms* described how GWAS have identified new variants associated with glioma risk in Section 1.5 - *Risk Factors*. Results from the latest glioma GWAS, are presented below as a Manhattan plot (Figure 2.3).

Figure 2.3: Manhattan plot of the GWAS results highlighting loci that have been implicated in (a) glioma, (b) glioblastoma (GBM) and (c) non-GBM glioma risk. This plot visualises the locations of previously (in blue) and newly (in red) identified loci which are associated with glioma risk. From Melin, *et al.* [220].



2.3.2 Statistical Colocalisation

GWAS frequently highlight associated loci that are not necessarily causal for the trait of interest. Statistical colocalisation (sometimes called genetic colocalisation) may be used as a follow-up functional analysis to provide evidence that, given two independent association signals at the same locus in two GWAS, a single, shared causal variant is that which is consistent with both signals. Typically, colocalisation has been used in the context of comparing GWAS of a trait and a GWAS derived from RNA sequencing detailing messenger RNA (mRNA) levels, referred to as expression quantitative trait loci (eQTL) studies, though in practice, colocalisation may be used for any two phenotypes.

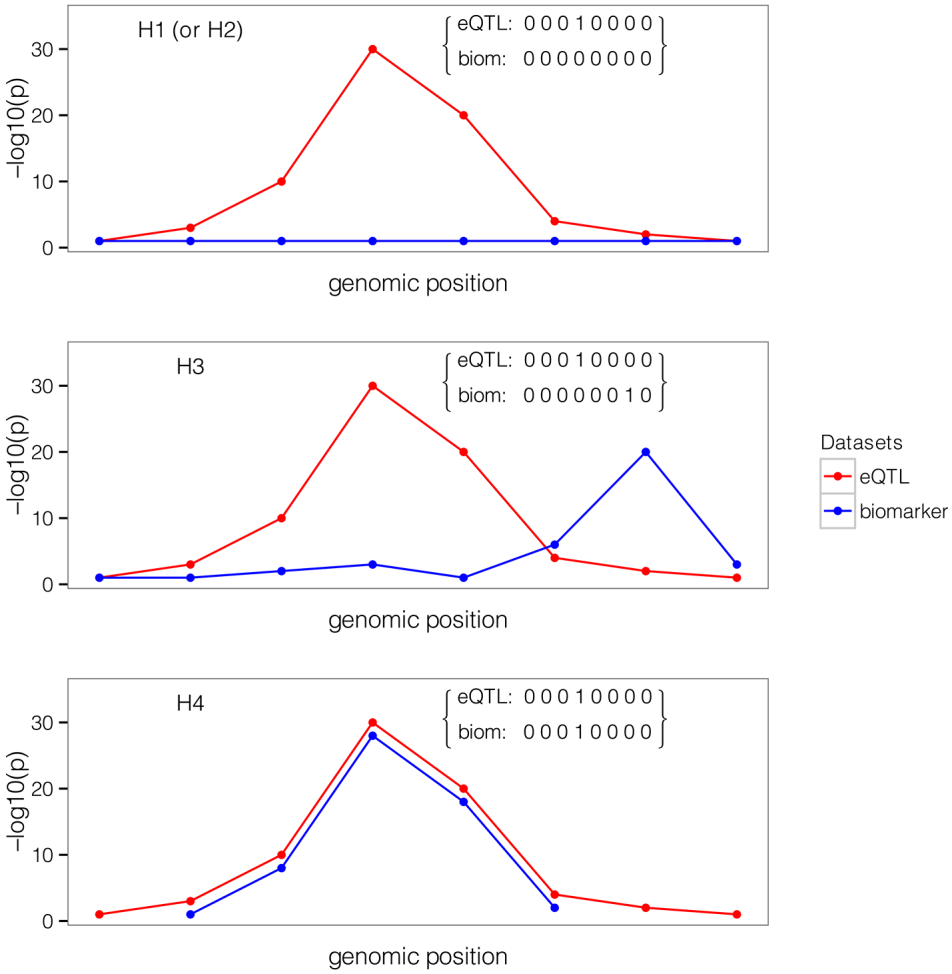
Colocalisation may be explained graphically using two GWAS, one that investigates mRNA levels (an eQTL GWAS) and another which investigates some biomarker (Figure 2.4). This simplified example shows the configurations of how an association signal can arise in both of these datasets. Whilst one can conduct a naive comparison using graphical tools (such as LocusZoom [274]) these lack any statistical rigour. Methods such as the Bayes test for colocalisation (coloc) [113] or eQTL and GWAS causal variant identification in associated regions (eCAVIAR) [150] use the summary statistics from each study to provide a statistical probability that a given variant is shared between both datasets.

There are many different tools, libraries and packages that can implement colocalisation analyses, for example: coloc [113], an R package that integrates a Bayesian statistical test for colocalisation; eCAVIAR [150] and FINEMAP [25] which are akin to fine-mapping methods; the heterogeneity in dependent instruments (HEIDI) method [404] which uses a heterogeneity test to distinguish between pleiotropy and linkage; and the joint likelihood mapping (JLIM) method [60] which uses individual-level data for one of the phenotypes of interest to provide evidence of colocalisation at the locus level. Each methodology has its own strengths and limitations and no one method is preferred in the literature over others; however, coloc, HEIDI and eCAVIAR tend to be the most popular because they only require summary-level statistics (a major limitation of the JLIM method). Due to requiring only summary-level data, the research in this thesis used the "coloc" R package written by Giambartolomei, *et al.* [113].

However, the coloc method has an inherent limitation called the single causal SNP assumption, whereby the method assumes that there will be only one single causal SNP in the entire region of interest that may be shared by both phenotypes. This is an obvious limitation given how complex the structure of linkage disequilibrium can be at any given region of the genome. Both coloc and HEIDI, as well as others, are subject to this limitation. Certain methods attempt to overcome this limitation, for example, eCAVIAR allows the user to specify how many shared causal SNPs they suspect are in the region (which opens the analysis to human error) with the authors of the method advising no more than six or seven so that the program remains computationally viable [150].

Therefore, it is advisable – and fast becoming standard practice – to combine colocalisation

Figure 2.4: Graphical explanation of colocalisation. Each subplot shows the genomic region for four hypotheses (H). H_1/H_2 shows an association signal in one dataset but not the other indicating these datasets do not colocalise. H_3 shows a signal in both datasets that is not shared, implying two separate causal variants that may (or may not) have been identified. Finally, H_4 shows when both datasets colocalise due to a single, shared causal variant in both regions. Adapted from Giambartolomei, *et al.* [113].



analyses with a step-wise conditional analysis (e.g., by use of the conditional and joint analysis (COJO) tool [391]) to identify conditionally independent signals within each phenotypic region of interest before applying a colocalisation analysis. This provides an effective work-around for the single causal SNP assumption by ensuring the colocalisation analyses are conducted on only conditionally independent association signals and not the marginal association statistics which can increase the type II error rate (or "false-negative", and in this context would mean weak evidence of colocalisation). This concept is displayed in Figure 2.5. In their paper detailing their colocalisation pipeline, Zheng, *et al.* treated datasets independently from one another such that conditional analyses are performed to systematically condition upon independent signals within the region of interest for both datasets separately [400]. Therefore, the same pipeline was used

for analyses presented in this thesis.

I am leading on a project which will present this pipeline of conditional and colocalisation analyses, called Pair-Wise Conditional and Colocalisation analysis (PWCoCo) as a stand-alone tool. This project is still ongoing but will seek to resolve questions about the methodology, for example, whether datasets should be treated dependently or independently (i.e., both datasets are systematically conditioned upon all association signals within the region in both datasets, regardless if that signal exists in that dataset or not). However, considerations need to be made for the increased computational burden this will put on the tool by increasing the amount of conditional analyses that will be run. These extra analyses may even be superfluous; consider two datasets, the first has a signal driven by SNP 1 and the second a signal driven by SNP 2. If SNP 1 and SNP 2 are the same, the tool will inherently condition upon that SNP, or signal, in both datasets. If SNP 1 and SNP 2 are different, then no conditional analysis on SNP 1 in dataset 2 should be necessary, given that that SNP, or signal, was not found to be either conditionally independent from other signals in that dataset, or was not strong enough to constitute a signal in the first place. Regardless, these hypotheses will be tested in that project; however, for the analyses in this project, the prevailing method of treating datasets independently for the conditional analyses was used.

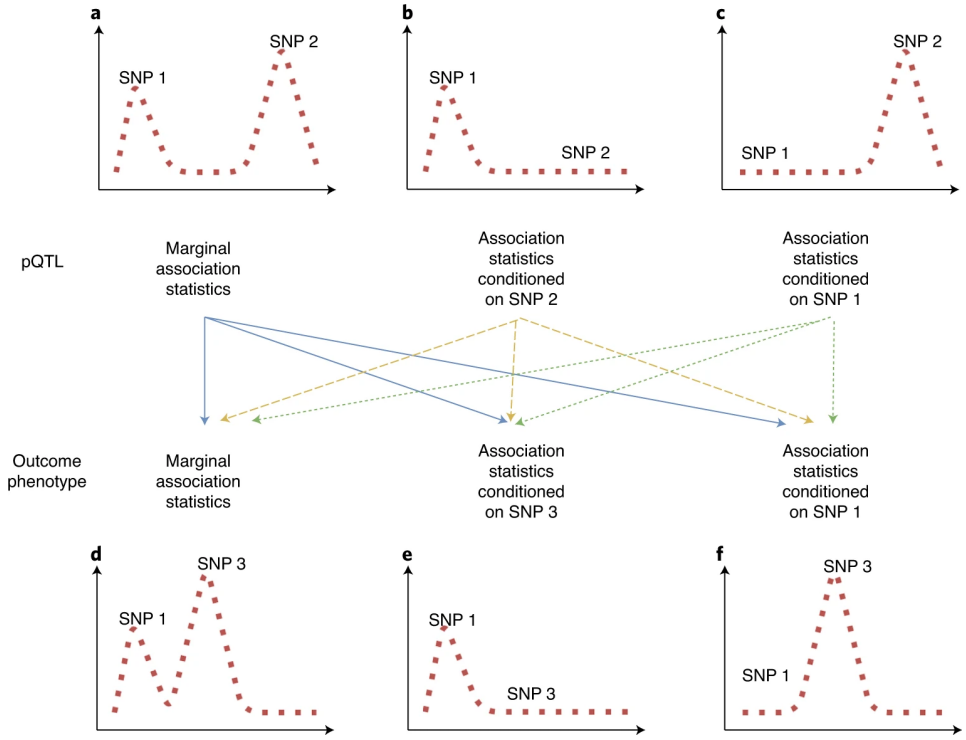
2.3.3 Mendelian Randomisation

MR is another observational method that is relevant to this thesis and uses germline genetic variants as instrumental variables (IVs) [196]. It is important to note that these genetic variants are germline, i.e. the mutation occurs in the gamete and is inherited from either parent, as opposed to somatic, whereby the mutation occurs in somatic cells meaning that tissues derived from that cell are affected, and are not inherited. IVs are variables that are associated with the exposure of interest and are used to estimate or infer causal relationships between the exposure and an outcome of interest. A variable requires three core assumptions to hold true to be used as an IV; therefore, these assumptions broadly underpin the MR methodology:

- The relevance assumption. This states that the genetic variants must associate with the exposure of interest.
- The independence assumption. This states there are no measured or unmeasured confounders between the genetic variants and outcome.
- The exclusion restriction. This states that the effect of the genetic variants on the outcome is only through the exposure of interest (also known as horizontal pleiotropy).

These assumptions can be portrayed graphically using a directed acyclic graph (DAG), a widely used epidemiological tool (Figure 2.6).

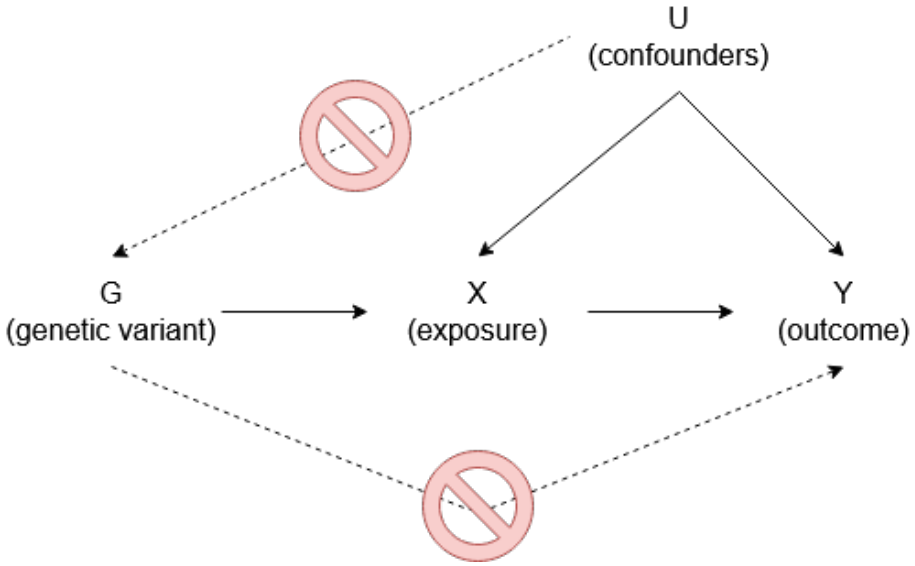
Figure 2.5: Schematic of how a step-wise conditional analysis before conducting a colocalisation analysis can reduce the rate of type II error in genomic regions with more than one association signal. The paper from which this figure was adapted (Zheng, *et al.* [400]) examined the effects of protein quantitative trait loci (pQTL) on a host of phenotypes. If there were two independent association signals in the exposure and outcome datasets, a naive colocalisation analysis using the marginal association statistics ((a) and (d)) would show little evidence of colocalisation due to the lead SNPs 2 and 3 not colocalising. However, performing a step-wise conditional analysis to identify independent association signals, and conducting colocalisation on SNP 1 – the lead SNP of an independent association signal in both phenotypes – then this would show strong evidence of colocalisation. Adapted from Zheng, *et al.* [400].



Instruments used in MR typically come from GWAS and may be used in either a single sample framework, whereby the causal exposure-outcome effect is estimated using individual-level data in one sample, or a two-sample framework, whereby two datasets of summary-level data are used to estimate the variant-exposure and variant-outcome associations which are combined to give an estimate of the exposure-outcome effect. Two-sample MR has become common practice over single sample MR due to the large amount of GWAS available, the fact that it is not always possible for the exposure and outcome to be measured in the same sample, and statistical power concerns. All of the MR analyses conducted in this thesis used two independent GWAS datasets in a two-sample framework.

Results from MR analyses are given as beta coefficients, which represent the causal estimate in the units of the outcome or log odds of the outcome per unit increase in the exposure. Commonly, these are presented as odds ratios (OR) to aid comprehension. In the case of a binary exposure

Figure 2.6: DAG of underlying assumptions of MR. Arrowheads show the direction of effect. Bias may be introduced when one, or more, of these assumptions are broken: when there is an association between the genetic variant or IV, G, and any measured or unmeasured confounder, U; furthermore, if G is not independent of the outcome, Y, then this association can also break the assumptions.



when estimates are base 2, this represents the OR per doubling of the genetic liability to the exposure. On the other hand, if the exposure is continuous then the OR represents the unit change in the outcome per unit increase in the exposure.

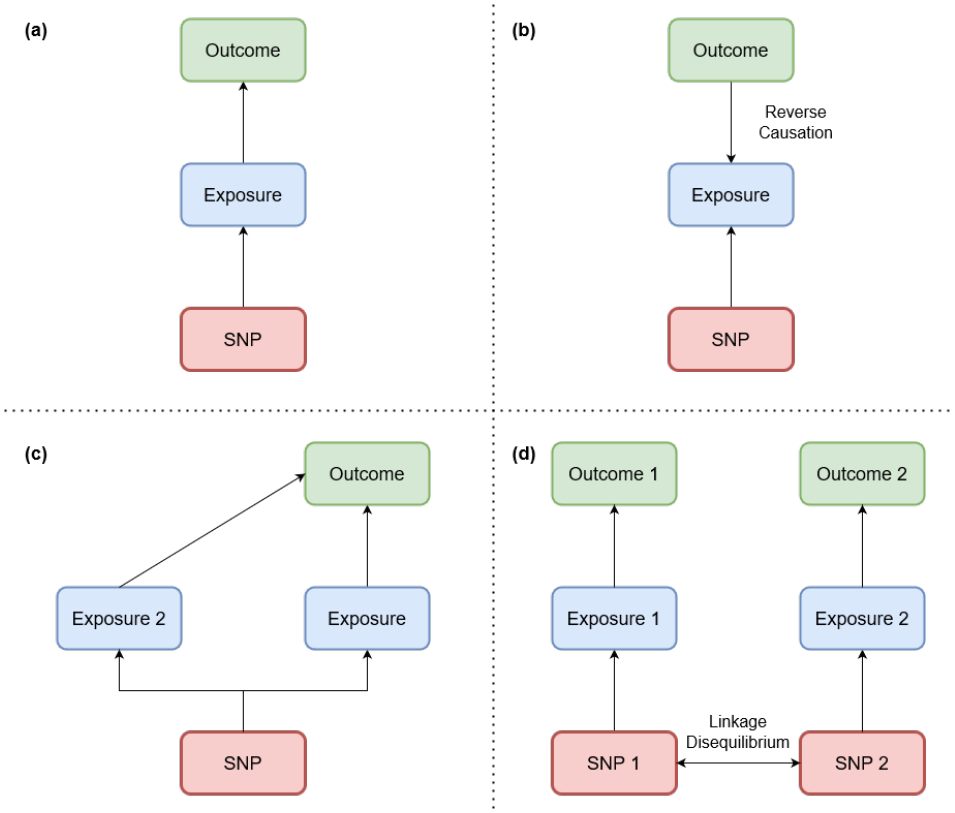
It is imperative for an investigator conducting an MR study to address invalidations or violations of the IV assumptions that underpin the methodology. This is normally achieved through the use of sensitivity analyses. Figure 2.7 depicts a simplified version of three common types of invalidations that are also relevant to the research presented in this thesis. Generally, sensitivity analyses used to assess these violations include methods such as colocalisation, MR-Egger [34], weighted median [35] and MR pleiotropy residual sum and order (MR-PRESSO) [355] that may assess the presence of horizontal pleiotropy or Steiger filtering to probe direction of effect (reverse causation) [141]. Not all of these methods are relevant to the research presented in this thesis; analyses used in this thesis are explained in depth in the corresponding results Chapters.

The main MR method used in this thesis is called the Wald ratio (WR). Obtaining the WR is relatively simple and is defined as the SNP-outcome effect divided by the SNP-exposure effect:

$$\beta_{WR} = \frac{\beta_Y}{\beta_X} \tag{2.4}$$

where X represents the exposure and Y represents the outcome. The WR is generally used for single SNP instruments, which is common in 'omics analyses (proteomics, transcriptomics,

Figure 2.7: Selected common layouts for a simple MR analysis examining the effect of an exposure instrumented by a SNP on some outcome. (a) Depicts a standard, albeit simple, MR study that will provide an unbiased, undistorted estimate of the effect of the exposure on the outcome. (b) Reverse causation can occur MR studies when the outcome affects the exposure of interest. (c) Horizontal pleiotropy can arise if a SNP is used to instrument for an exposure of interest and is also associated with the outcome through another, independent pathway, either directly or indirectly. If this SNP is used to estimate the relationship between the exposure and the outcome then this estimate will be distorted by the presence of horizontal pleiotropy. (d) If SNP 1 is used to estimate the effect of Exposure 1 on Outcome 2, the estimate will show a pleiotropic association because SNP 1 is in linkage disequilibrium with SNP 2 [140].



etc.). Care must be taken to ensure the SNPs chosen to instrument the exposure adhere to the IV assumptions (laid out above in Section 2.3.3 - *Mendelian Randomisation*), otherwise this may bias the causal estimate. As instruments from these types of analyses will be generally composed of a single SNP, follow up sensitivity analyses are limited due to the requirement of many SNPs.

In the case of an instrument with more than one SNP, there is a continuously growing number of methods that have been developed in the context of two-sample methods using summary-level data. Of these methods, the research in this thesis used the fixed effects inverse-variance weighted (IVW) method described by Bowden, *et al.*[35]. This method takes into consideration each of the SNP's effect on the exposure and outcome:

$$X|G_i = \gamma_0 + \gamma_i G_i + \epsilon x_i \tag{2.5}$$

$$Y|G_i = \Gamma_0 + \Gamma_i G_i + \epsilon y_i \tag{2.6}$$

Where G is the instrument consisting of i SNPs. γ and Γ are the SNP effect for the exposure and outcome, respectively. Similar to the WR, the causal effect of the exposure and outcome is estimated:

$$\hat{\beta}_i = \frac{\hat{\Gamma}_i}{\hat{\gamma}_i} \tag{2.7}$$

From Equation (2.7), using a fixed-effects meta-analysis technique, the overall estimate for the IVW method can be calculated:

$$\hat{\beta}_{IVW} = \frac{\sum_i \hat{\gamma}_i^2 \sigma_{\bar{Y}_i^2} \hat{\beta}_i}{\sum_i \hat{\gamma}_i^2 \sigma_{\bar{Y}_i^2}} \tag{2.8}$$

where $\sigma_{\bar{Y}_i^2}$ is the standard error of the gene-outcome association estimate for variant i. Naively, this can be considered as a meta-analysis of the WR for each SNP. Furthermore, two assumptions are made by the weightings in the meta-analysis: that the SNP-exposure and SNP-outcome effects are not correlated; and that there is no measurement error (the NOME assumption) in the SNP-exposure association. These assumptions can be examined using common meta-analysis sensitivity techniques such as Cochran's Q statistic or the F statistic [140].

2.3.4 Pharmaco-Epidemiology

Pharmaco-epidemiology is the study of drugs and their effects in large populations. Whilst pharmaco-epidemiology can be and is used to study the efficacy and safety of new drugs, it is also used to research whether existing drugs can be re-purposed for use in other diseases. This approach has many benefits over development of novel therapeutic substances. To illustrate this point, DiMasi, *et al.* estimated in 2016 that the average cost of bringing a new compound to market cost USD $2558 million [82]. Furthermore, it is a real possibility for new drugs to fail in early stage I or II safety trials. In a 2010 paper, DiMasi, *et al.* estimated the success rates of novel compounds can range from as low as 8% to as high as 30% depending on product type and pharmacodynamics [81]. It is relevant to the research in this thesis that they specifically found that the lowest of rate of success, 8%, belonged to drugs targeting the CNS. Finally, it is common for the entire process to last for at least a decade, from research and development to marketable product [81]. Altogether, these three points illustrate why the pharmaceutical industry is keen to investigate how existing drugs may be repurposed to treat off-target diseases.

However, whilst pharmaco-epidemiology studies may appear easier to conduct when compared to an RCT, there are many sources of bias and confounding that need to be considered to ensure an accurate result. The most prominent of these is confounding by indication which can arise due to some indication (reason to prescribe a drug) being associated with both the exposure and the outcome, and thus distorting that association. An example from the Catalogue of Bias [8] comes

from a study which suggested that paracetamol use in childhood is associated with increased risk of asthma, rhinoconjunctivitis and eczema [23]. This result may have been confounded by an indication, such as fever, which can cause these outcomes and also prompt the use of paracetamol in the first place [8]. Careful study design can limit the influence of this type of confounding. The research presented in Chapter 6 investigated the effects of fibrate and glitazone prescription on BT incidence and survival. In that study, there was risk of confounding by indication which necessitated careful study design to limit the risk of this confounding distorting the observed effects.

2.4 Summary

Described within this Chapter were methodologies and study designs that were relevant to and used throughout the research presented in this thesis, such as MR and pharmaco-epidemiology. Overviews provided for these methods included their strengths and limitations and important considerations when applying these methods which were relevant to the studies conducted in this thesis.

CHAPTER 3

DATA SOURCES

3.1 Introduction

The work presented in this doctoral thesis utilised a range of datasets from different organisations and research groups (Table 3.1). This Chapter provides a broad overview and discusses strengths and limitations for each of those datasets.

3.2 Glioma GWAS

3.2.1 Background

The glioma GWAS meta-analysis reported by Melin, *et al.* [220] is the largest of its type and contains fully genotyped data for 12,496 cases and 18,190 controls. Participants in the meta-analysis were non-overlapping adults (> 18 years old) of European ancestry. Data were adjusted for age, sex, and the first two principle components.

Due to data sharing limitations, only a subset of these data was used for the gene expression project conducted in Chapter 4, however the full meta-analysis as detailed here was used for the protein levels project in Chapter 5.

3.2.2 Materials and Methods

The constituent GWAS (Table 3.2) that were included in the authors' original meta-analysis were all assayed using the Illumina OncoArray-500K array [6]. Studies that participated in the Glioma International Case-Control study (GICC) (UK-GWAS, French-GWAS, German-GWAS, MD Anderson (MDA)-GWAS, the San Francisco Adult Glioma Study (SFAGS) and GICC) all underwent the same quality control protocol: participants with > 80% estimated European

Table 3.1: Overview of the datasets used in throughout the research in this doctoral thesis.

Study	**Summary**	**How is this data used?**
Glioma GWAS from Melin, *et al.* [220]	Glioma case-control GWAS	Summary statistics are used as outcome data in MR analyses. Used in Chapters 4 and 5.
Clinical Practice Research Datalink (CPRD)	Primary care database	Individual-level data including medical records, personal details, prescriptions, diagnoses, etc. are used to determine if fibrates and glitazones affect risk and prognosis of brain tumours. Used in Chapter 6.
Genotype-Tissue Expression (GTEx) project [121]	Genotypic effects on gene expression in a range of tissues	Summary statistics are used to investigate tissue-specific gene expression on glioma risk. Used in Chapter 4.
eQTLGen Consortium [361]	Genotypic effects on gene expression in whole blood	Summary statistics are used to determine the effect of gene expression in whole blood on glioma risk. Used in Chapter 4.
Brain tissue expression quantitative trait loci (eQTL) meta-analysis [276]	Meta-analysis of genotypic effects on gene expression in brain tissues	Summary statistics are used to examine the effects of eQTLs derived from brain tissues on glioma risk. Used in Chapter 4.
Whole blood protein quantitative trait loci (pQTL) dataset [400]	Pooled analysis of genotypic effects on protein levels in whole blood	Summary statistics are used to determine how genetically predicted protein levels affect glioma risk. Used in Chapter 5
Brain tissue pQTL (BrainQTL) study [289]	Genotypic effects on protein levels in brain tissue	Summary statistics are used to determine how genetically predicted protein levels affect glioma risk. Used in Chapter 5.

Table 3.2: Number of cases, controls and imputed SNPs included from each glioma GWAS. Adapted from Supplementary Table 2 published by Melin, *et al.* [220]. Number of imputed SNPs is for all glioma; 6,790,270 SNPs for GBM and 6,769,856 SNPs for non-GBM total.

Study	**Cases (GBM/non-GBM)**	**Controls**	**Imputed SNPs**
UK-GWAS	631 (270/361)	2,699	8,954,681
French-GWAS	1,423 (430/993)	1,190	9,113,681
German-GWAS	846 (431/415)	1,310	9,012,806
MDA-GWAS	1,175 (652/523)	2,236	9,043,003
SFAGS	677 (511/166)	3,940	10,679,291
GliomaScan	1,653 (903/472)	2,725	9,161,499
GICC	4,572 (2,468/1,897)	3,285	10,783,269
UCSF-Mayo	1,519 (526/992)	804	6,915,238
Total	12,496 (6,191/6,305)	18,190	6,887,412

ancestry was determined using FastPop [202] (with HapMap version 2 [162]) and reference populations of Utah Residents with Northern and Western European Ancestry (CEU), Japanese in Tokyo (JPT)/Han Chinese in Beijing (CHB) and Yoruba in Ibadan, Nigeria (YRI)). GWAS data was imputed to > 10 million SNPs using IMPUTE2 (v2.3) [154] and reference panels 1000 Genomes Project (phase 1 integrated version 3 release) [110] and UK10K [347]. SNPs that were poorly imputed (information rate < 0.40 with IMPUTE2) or exhibited large deviation from Hardy-Weinberg equilibrium ($P < 1{\times}10^{-8}$) were excluded in controls. Case-control matching was tested for adequacy was evaluated using Q-Q plots of test statistics (more details are given in the Supplementary Figure 1 in the original publication [220]). The UCSF-Mayo study used STRUCTURE [273], with 1000 Genomes [110] as a reference panel, to assess population admixture, with non-Caucasians excluded. SNPs were imputed using the Michigan Imputation Server [65] with HapMap version 1 [162] as the reference. The meta-analysis was performed using the fixed-effects inverse-variance method using META (v1.6) [205]. The data were adjusted for age, sex and the first two principal components, generated for the GICC, GliomaScan, MDA-GWAS and SFAGS using PLINK [275].

3.2.3 Data Cleaning and Analysis

Due to limitations in data sharing, only three of the constituent GWAS were used in the eQTL analyses conducted in Chapter 4 which were the GICC, GliomaScan and MDA datasets. Therefore, to increase statistical power, these GWAS were meta-analysed in METAL [384] to obtain beta coefficients and standard errors for SNP-glioma associations and resulted in a sample size of 7,400 cases (3,112 GBM, 2,411 non-GBM) and 8,257 controls. SNPs that had incomplete or missing data were removed as these could not be used for MR.

3.2.4 Strengths and Limitations

This dataset meta-analysed all currently available glioma GWAS conducted in a European ancestry sample, which resulted in the largest single source of genotype data for glioma patients. One limitation of using this dataset is that to increase the number of participants, all diagnosed glioma tumours were included. Furthermore, the study investigators did not record detailed subtype diagnoses of participants; tumours are encoded as either non-GBM gliomas or GBM. Whilst this will increase statistical power when using this dataset, there is large somatic genetic heterogeneity between tumour types meaning MR results using germline genetic variants may not be completely applicable when considering specific subtypes [390]. Therefore, the MR analyses presented in Chapters 4 and 5 analysed the all glioma, GBM and non-GBM datasets separately to investigate how associations compared across subtype diagnoses.

Another potential limitation is the possibility of selection bias in the original selection of participants by the investigators. This could then induce collider bias within the MR analysis. Collider bias is induced when controlling for a variable which is associated with both the exposure and the outcome [198]. However, considering the only two known risk factors for glioma are genetic liability (which this study seeks to investigate) and previous exposure to ionising radiation, unless there is an association between the latter and selection into this study, the risk of collider bias is likely minimal [198].

3.3 Clinical Practice Research Datalink

3.3.1 Background

The Clinical Practice Research Datalink (CPRD) [145] is a primary care database that consists of data collected from participating general practices around the UK. The CPRD forms a uniquely large resource (coverage of 15.5 million patients from 674 practices in the UK at the time of writing) with detailed information on each patient's interaction with their primary care provider through their unique NHS identifier, including: prescription histories, diagnostic and prognostic tests, consultations, immunisation records and referrals. A portion of participating GPs (75% of English practices, or 58% of all UK practices [145]) have also allowed for linkage to third-party databases, such as the Official of National Statistics (ONS) for mortality data and the Index of Multiple Deprivation (IMD) and Townsend scores for socioeconomic data, creating an ideal resource for epidemiological research.

Two nested case-control studies and one cohort study were constructed using participants in the CPRD. These were used to evaluate how exposure to anti-hyperlipidaemic fibrates and anti-type 2 diabetic glitazones affected risk and prognosis of BT, compared to other indications for the treatment of the same diseases. These drugs target two proteins, peroxisome proliferator-activated receptors (PPAR)-α and -γ, which have been implicated in BT biology previously as targets of interest which may reduce risk and improve prognosis [137, 257].

3.3.2 Patient Counts

Of the 15,538,338 patients present in the CPRD GOLD database in the August 2018 snapshot, there were a total of 14,022 primary or secondary BT patients between the study dates of January 2000 and December 2016. It is important to note that for the definition of cases, a diagnosis of BT was needed, and not solely glioma, to increase the number of cases eligible for inclusion into the study.

3.3.2.1 Case-control Study of Anti-Hyperlipidaemic Medication

Chapter 6 detailed the case-control study of how fibrates, an anti-hyperlipidaemia medication, affect BT risk compared to other medications for the treatment of hyperlipidaemia. Of the 14,022 BT patients in the CPRD snapshot, 3,755 of these were treated with any kind of anti-hyperlipidaemia medication. After ensuring prescription preceded diagnosis (n = 3,251) and cases were eligible for ONS linkage (n = 2,238), the final amount of cases included after data cleaning and linkage was 1,916. 7,757 controls (no diagnosis of BT at any time) were identified and matched to these cases. Figure 3.1 shows participant selection for this study, with more details given in Chapter 6.

3.3.2.2 Case-control Study of Anti-Diabetic Medication

Similarly, a case-control design was used to assess the association between glitazones, a type 2 diabetes medication, and BT risk compared to other medications for the treatment of type 2 diabetes. There were 1,338 BT cases eligible for inclusion, of which 920 were available for ONS linkage and 791 were prescribed a drug of interest before diagnosis. This amounted to 445 cases after data cleaning and linkage, which were matched to 1,885 controls. Figure 3.2 shows participant selection for this study, with more details given in Chapter 6.

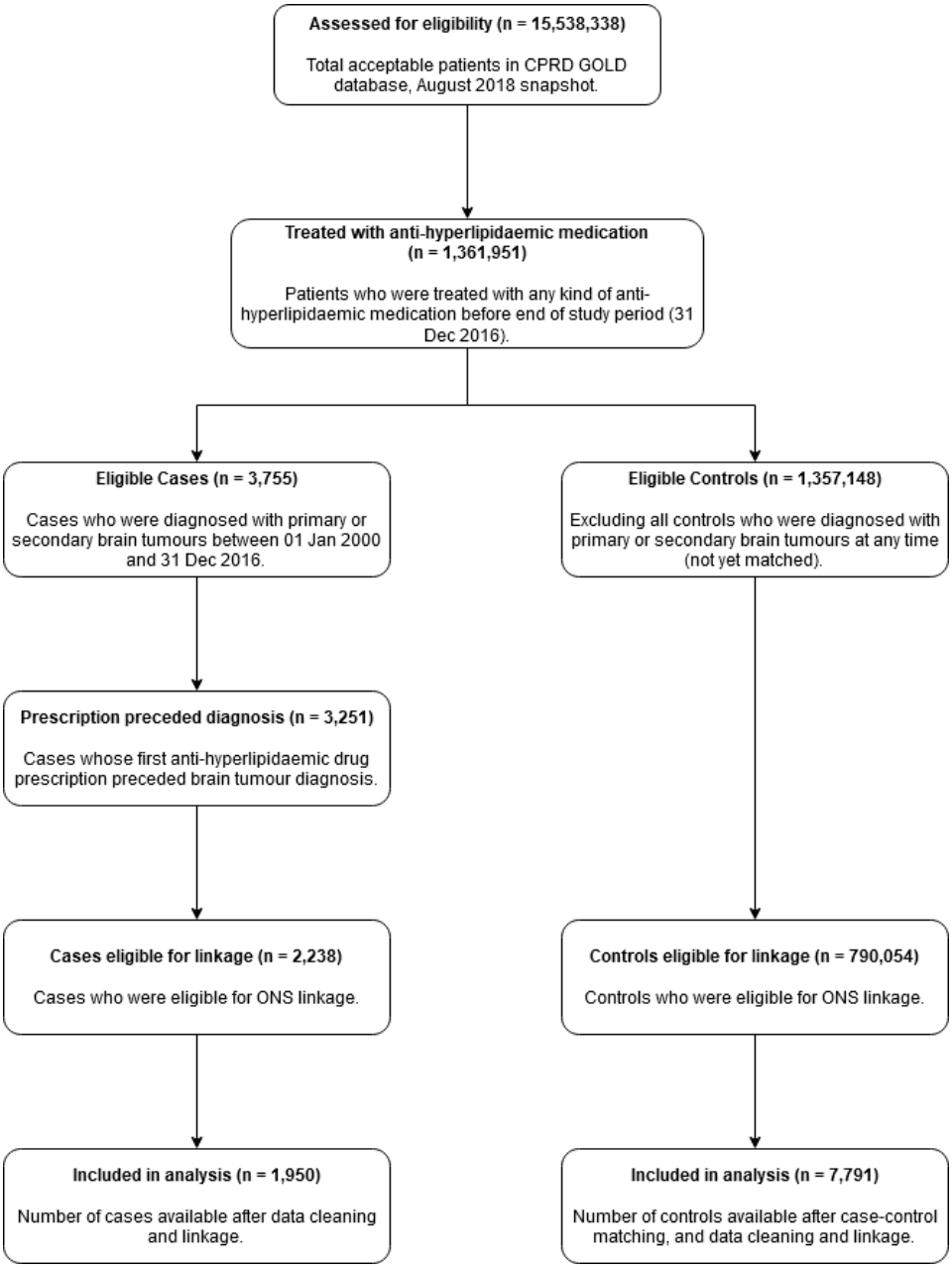
3.3.2.3 Cohort Study of Anti-Hyperlipidaemic and Anti-Diabetic Medications

All 14,022 cases within the CPRD snapshot were included in the cohort study to assess the effect of these medications on BT mortality. 7,449 cases were available after data cleaning and linkage. Figure 3.3 shows participant selection for this study, with more details given in Chapter 6.

3.3.3 Strengths and Limitations

The CPRD has wide and deep coverage of a large number of patients over their entire interaction with the National Health Service (NHS) via a participating GP. A unique strength of the CPRD is that it is broadly representative of the UK population in age, sex, ethnicity and body mass index (BMI) by comparison with the UK census of 2011 [145]. Furthermore, the breadth of information, such as each patient's consultation, diagnostic and procedure history, is a huge benefit for health-related research. The CPRD also contains information regarding a patient's prescription history

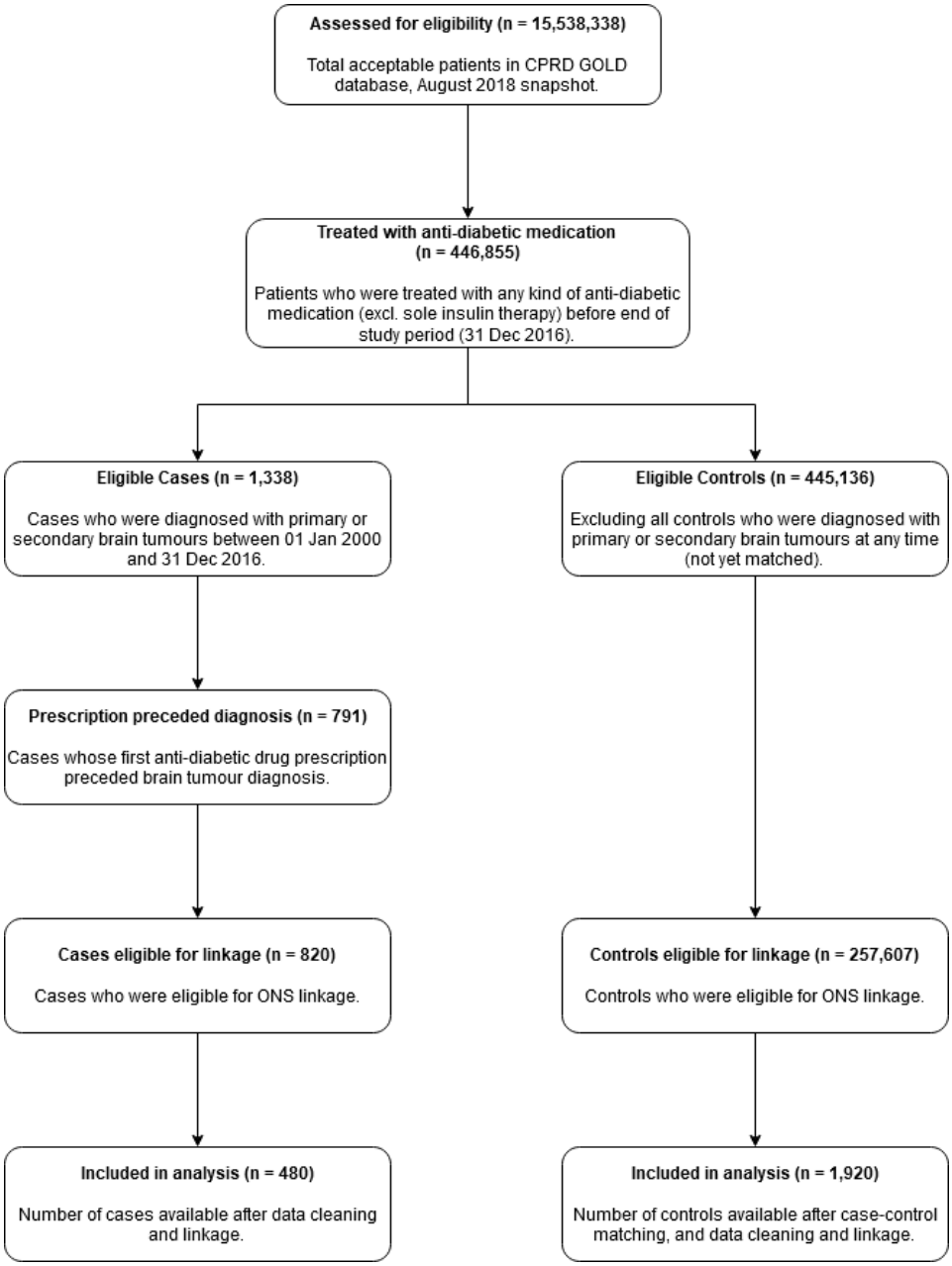
Figure 3.1: Participant selection for the anti-hyperlipidaemic medications case-control study.



though this will not include records of use for over-the-counter medications which could hinder some studies. Patients within the CPRD can also be linked to other statistics, such as those collected by the ONS for mortality data and area-based deprivation data as measured by IMD scores. Previous studies have validated data extracted from the CPRD [146, 170].

Missing or incomplete data within the CPRD is common due to its nature as a retrospectively digitised database. There is also little standardisation of codes and measurements. However, if care is taken by the investigator of a CPRD study, then these limitations can be minimised

Figure 3.2: Participant selection for the anti-diabetic medications case-control study.



or removed completely with methods such as imputation or careful and informed derivation of variables of interest. This was observed in the glitazones nested case-control study where mean glycated haemoglobin (haemoglobin A1c, HbA1c) levels were used as a covariate to proxy for severity of diabetes. Nearly 55% of participants did not have HbA1c readings due to missing data, therefore it was assumed that their GP did not deem it necessary for the test to be run and so these participants were categorised in the lowest category (signifying well-controlled diabetes). In the univariate models, there was weak evidence that worse-controlled diabetes

was negatively associated with BT risk; however, the assumption likely induced bias and upon testing this further with imputation and dropping missing values, there was no evidence of an association between HbA1c levels and BT risk. Therefore, limitations around missing data can limit inferences in CPRD, however can be minimised if addressed by the investigator. This was discussed in Chapter 6 in further detail.

3.4 Exposure Datasets

The following datasets are used as exposure datasets in MR analyses to ascertain the effect of genetically predicted gene expression and protein levels on glioma risk. These datasets are also discussed in more detail in the relevant Chapters in which they appear.

3.4.1 The Genotype-Tissue Expression Project

3.4.1.1 Background

The GTEx project [121] is a public data resource and tissue bank created with the aim to study tissue-specific gene expression and explore the genetic basis for complex human disease. The database contains different types of molecular data, chiefly eQTLs which are genetic variants that are associated with gene expression. These data have been generated from 54 tissues from 1000s samples included in GTEX version 8 (v8) [121] using genotyping and RNA sequencing. In version 8 of the GTEx release, there are genotyped samples from 13 brain tissues with a sample size ranging from 114 to 209 (Table 3.3).

3.4.1.2 Materials and Methods

There were 838 participants of the latest GTEx v8 release with RNA sequenced and genotyped data, the majority of which are European Americans (85.3%), but also included 12.3% African Americans and 1.4% Asian Americans. In total, 15,253 samples were collected from these participants and 15,201 samples were used to discover QTLs. Whole genome sequencing detected roughly 43 million SNPs after quality control. mRNA levels for each tissue sample were also sampled to construct eQTLs, which were sub-categorised as either *cis* or *trans*, defined as local or distal SNPs relative to the associated gene location, respectively. eQTLs were subject to a false discovery rate (FDR) of 5%. Further details about how these data were collected and analyses are given in the GTEx v8 publication [3].

3.4.1.3 Strengths and Limitations

The wide range of tissues which are genotyped and from which eQTLs are derived is a unique strength of the GTEx data. Indeed, other eQTL studies rarely collect data from as many tissues as GTEx. This is especially true for brain tissues, where it is common that brain-derived eQTLs

Table 3.3: Breakdown of brain tissues present in GTEx v8, with number of genes tested with an associated eQTL that passed quality control (sometimes referred to as eGenes) and (genotyped) sample size for each brain tissue type.

Tissue	**No. of Genes**	**Sample Size**
Amygdala	3,726	129
Anterior cingulate cortex (BA24)	5,630	147
Caudate (basal ganglia)	8,362	194
Cerebellar hemisphere	10,027	175
Cerebellum	11,240	209
Cortex	9,082	205
Frontal cortex (BA9)	7,335	175
Hippocampus	5,517	165
Hypothalamus	5,499	170
Nucleus accumbens (basal ganglia)	8,198	202
Putamen (basal ganglia)	6,902	170
Spinal cord (cervical c-1)	4,483	126
Substantia nigra	3,301	114

will be collected from a few tissues, like the cortex or cerebellum, and analysed together [276]. Data from the GTEx project therefore allowed for a nuanced and granular exploration of how gene expression across the entire brain differs according to tissue type. This does, however, lead to limitation of the GTEx data, in that only a small sample size is provided for most tissues which in turn limits statistical power.

Roughly 13% of the samples used to construct eQTLs in the GTEx v8 release are from participants with non-European ethnicity. Furthermore, there is no option provided to filter eQTLs based on participant ethnicity despite analyses conducted by the GTEx original authors based on ethnicity. This could induce bias in studies for whose outcomes vary depending on ethnicity. However, the amount of non-European ethnic samples is relatively small given the total sample size. Furthermore, as described in Section 1.1 - *Incidence*, there is no definitive evidence for an association between glioma and ethnicity and so this may not prove an issue for the MR analyses conducted in Chapter 4.

3.4.2 eQTLGen Consortium

3.4.2.1 Background

The eQTLGen Consortium (eQTLGen) [361] is a publicly available meta-analysis of 37 datasets which have derived whole blood eQTLs. In total, the eQTLGen dataset has studied 19,942 genes in 31,684 samples, making it the largest whole blood eQTL study currently available. Similar to the GTEx project, eQTLGen aims to find associations between gene expression observed in blood and complex traits by performing *cis*- and *trans*-eQTL meta-analyses [361].

3.4.2.2 Materials and Methods

The 37 constituent GWAS used a variety of different expression profiling platforms. Samples were derived from mostly European ethnic participants, though some studies included non-European ethnic participants, e.g. 1,404 Bangladeshi, 175 Arabs and Amazighs and 115 Chinese samples were included. The authors of the eQTLGen meta-analysis conducted pre-processing and quality control analyses for each study separately, as detailed in the publication [361]. In total there were four profiling methods used: Illumina (55% of included studies), Affymetrix U291 (8.7%), Affymetrix HuEx v1.0 ST (16%) expression arrays and RNA-seq (20.3%). 1,000 Genomes phase 1 integrated version 3 release was used by all but one study as the imputation reference panel for genotyped data, which were derived differently on a per-study basis [110]. *Cis*- and *trans*-eQTLs were identified in each dataset separately (defined in this study as SNPs inside and outside 500Kb window of the associated gene, respectively) and subjected to a FDR of 5%. QTLs were then replicated amongst studies and meta-analysed separately.

3.4.2.3 Strengths and Limitations

This dataset overcomes the limitations inherent to small sample sizes by performing a meta-analysis to increase power. This highlights the strength of using the eQTLGen dataset over other whole blood eQTL datasets, in that a larger sample size means that my statistical analyses were more likely to be better powered to detect the effects of gene expression on glioma risk. Similarly with the GTEx v8 dataset, some of the samples are not from ethnically European samples; however, relatively only few samples are included in this way and may not be a problem in the analyses conducted in Chapter 4.

3.4.3 Brain Tissue eQTL Meta-Analysis

3.4.3.1 Background

For their 2018 paper investigating gene expression in brain tissue and blood, Qi, *et al.* performed a meta-analysis of brain-derived eQTLs to overcome limitations related small sample sizes for their MR analyses [276]. In total, their estimated effective sample was 1,194, massively increasing sample sizes compared to the constituent studies (Table 3.4), which included data from GTEx v6 [121], the CommonMind Consortium (CMC) [106] and the Religious Orders Study (ROS) and Memory and Ageing Project (MAP) (ROSMAP) [234].

3.4.3.2 Materials and Methods

All samples from the constituent GWAS were of European ethnicity and passed quality control measures as detailed in the original papers [106, 121, 234]. All genotyped data were imputed using the 1000 Genomes phase 1 integrated version 3 release [110]. Gene expression data in all GWAS were measured using RNA-Seq and annotated using GENCODE [134]. The authors

Table 3.4: Details of the constituent GWAS that were used in the Qi, *et al.* meta-analysis [276]. Note that the reason the number of genes provided for GTEx v6 in this table are larger than those for GTEx v8 in Table 3.3 is because the numbers here are also given for number of probes used and may not necessarily result in a significant eQTL being found for a gene. Adapted from Qi, *et al.* [276].

Dataset	**Tissue**	**No. of Genes/Probes**	**Sample Size**
GTEx v6	Anterior cingulate cortex (BA24)	23,509	72
	Caudate (basal ganglia)	24,621	100
	Cerebellar hemisphere	24,065	89
	Cerebellum	24,762	103
	Cortex	24,366	96
	Frontal cortex (BA9)	24,120	92
	Hippocampus	24,880	81
	Hypothalamus	24,654	81
	Nucleus accumbens (basal ganglia)	24,542	93
	Putamen (basal ganglia)	23,362	82
CMC	Dorsolateral prefrontal cortex	14,366	467
ROSMAP	Cortex	12,979	494

derived a meta-analysis technique specifically for use of these data due to correlated samples, an important consideration as GTEx v6 data were collected from brain tissue samples derived from the same subjects. This method, called meta-analysis *cis*-eQTL data in correlated samples (MeCS), was integrated in their summary-data-based Mendelian randomisation (SMR) package [404]. MeCS uses summary-level statistics to estimate the correlation between estimates for eQTLs (the original authors used *cis*-eQTLs which reached a P value threshold of 5×10^{-8}), corrected for estimation errors which other methods (such as naive Spearman or Pearson correlations) do not account for. In brief, the authors used these correlations and so-called "null" SNPs (defined as SNPs with $P > 0.01$) within the *cis* region to quantify sampling correlation between the estimated SNP effects of the brain-eQTL datasets, thereby accounting for potential sample overlap. Further details of MeCS are given in the original publication [276]. Effect estimates across the studies were standardised and re-scaled based on the expression level per gene in standard devitation (SD) units [276].

3.4.3.3 Strengths and Limitations

Strengths of this dataset are similar for the eQTLGen dataset, in that by conducting a meta-analysis, Qi *et al.* have produced a dataset with a larger sample size than other comparative studies and hence will increase statistical power for methods such as MR. Limitations include the loss of nuance and granularity allowed by analysis of discrete tissues, though this can be addressed by performing follow-up analyses in other datasets, e.g. GTEx v8, as described in

Table 3.5: The five pQTL studies that formed part of the Zheng, *et al.* blood pQTL pooled analysis [400].

Study	**No. of Proteins (with pQTLs)**	**Sample Size**	**Array**
Emilsson, *et al.* [92]	776	3,200	SOMAScan
Folkersen, *et al.* [99]	58	3,394	Olink
Suhre, *et al.* [336]	284	1,000	SOMAScan
Sun, *et al.* [338]	1,478	3,301	SOMAScan
Yao, *et al.* [393]	70	6,861	xMAP

Chapter 4.

3.4.4 Whole Blood pQTL Pooled Analysis

3.4.4.1 Background

A recent dataset made available by Zheng, *et al.*, consists of a pooled analysis of five different whole blood pQTL studies [400]. The five constituent studies are detailed in Table 3.5.

3.4.4.2 Materials and Methods

All participants were genotyped as part of other studies which are detailed in the relevant constituent GWAS papers. In brief, all participants were of European ethnicity and provided blood samples from which genotyping and protein measurements were conducted. Genotyping was conducted using Illumina in the Sun, *et al.* (Omni 2.5 array version 8), Folkersen, *et al.* (CardioMetabochip and Immunochip arrays) and Emilsson, *et al.* (Hu370CNV array) studies and Affymetrix arrays in the Suhre, *et al.* and Yao, *et al.* studies. All studies underwent similar quality control processes and were imputed to the 1000 Genomes phase1 integrated version 3 release reference panel [110]. Three arrays were used to measure protein levels (Table 3.5). To account for differences due to this, Zheng, *et al.* mapped platform IDs for each protein to Uniprot IDs and conducted validation, heterogeneity and consistency tests across the studies.

3.4.4.3 Strengths and Limitations

This dataset benefits from being a pooled analysis of constituent GWAS resulting in more pQTLs and proteins to be tested in the MR analyses. A further strength of this dataset is the lengths to which the authors went to clean and prepare the dataset; they also intended to use the data for MR and so selected and categorised instruments into tiers depending on their suitability and usability for such analyses. The first step of their instrument selection process involved checking for specificity, i.e. if a pQTL associated with five or more proteins, then that pQTL was non-specific and may be difficult to analyse using MR due to horizontal pleiotropy (described in

Figure 2.7). Although the choice to define a cut-off at five or more proteins was arbitrary, the authors justified this as removing potentially pleiotropy instruments which would contravene the core assumptions of MR. This step was important as there were no analyses that could have been conducted to assess the presence of horizontal pleiotropy due to many instruments consisting of single SNPs. Furthermore, the authors also used the EpiGraphDB platform [207] to examine the pathway of pQTLs with between two and five associated proteins to determine whether those non-specific instruments were likely exhibiting horizontal or vertical pleiotropy. The second step of this process checked for consistency between studies which was necessitated as the studies were conducted using three different array types. Consistency was tested using colocalisation and heterogeneity analyses. These two validation steps informed on the three tiers, which are as follows:

1. *Tier 1* - These instruments passed both validation steps and were eligible for inclusion into MR analyses.

2. *Tier 2* - These instruments passed the specificity tests but failed the consistency tests. These instruments were eligible for inclusion in the MR study but required further sensitivity analyses.

3. *Tier 3* - These instruments failed the specificity tests and were removed from the analysis.

For analyses in Chapter 5, only pQTLs in tiers 1 and 2 were included in the MR analyses, as with in the original publication.

3.4.5 Brain pQTL Dataset

3.4.5.1 Background

The brain pQTL dataset is a recently derived dataset of pQTLs from dorsolateral prefrontal cortex brain tissue [289]. In total, the authors identified 864 protiens with associated pQTLs from 144 participants from the ROS and MAP studies [234]. The authors also provided data for similarly derived eQTLs that they used to compare gene expression and protein abundance levels as predicted by e- and pQTLs; analyses described in this study did not use the eQTLs they provided, chiefly because these would have been included in the Qi, *et al.* meta-analysis that also included eQTLs derived from the ROSMAP study.

3.4.5.2 Materials and Methods

Brain tissue samples were obtained from ethnically European participants in the CMC and ROSMAP studies [106, 234]. Genotyped data were from whole genome sequencing obtained using Illumina HiSeq X, imputed to 1000 Genomes phase 1 version 3 [110] and underwent quality control as detailed in the original paper by De Jager, *et al.* [70]. Protein abundance levels were

measured using tandem mass tag isobaric labelling mass-spectrometry. Details of this are given in the original publication [289].

3.4.5.3 Strengths and Limitations

Unlike the few other published and publicly available pQTL studies, the unique strength of this dataset is the ability to explore protein abundance levels in brain tissue. The ability to specifically determine whether a protein expressed in brain tissue is causally implicated in glioma risk is much more meaningful clinically than if that protein were instead observed to be in blood, chiefly due to the blood-brain barrier. However, the data were derived from a small sample, which inevitably limits statistical power, and the full dataset was not publicly available. This limited colocalisation analyses which required data from whole regions of the genome to conduct.

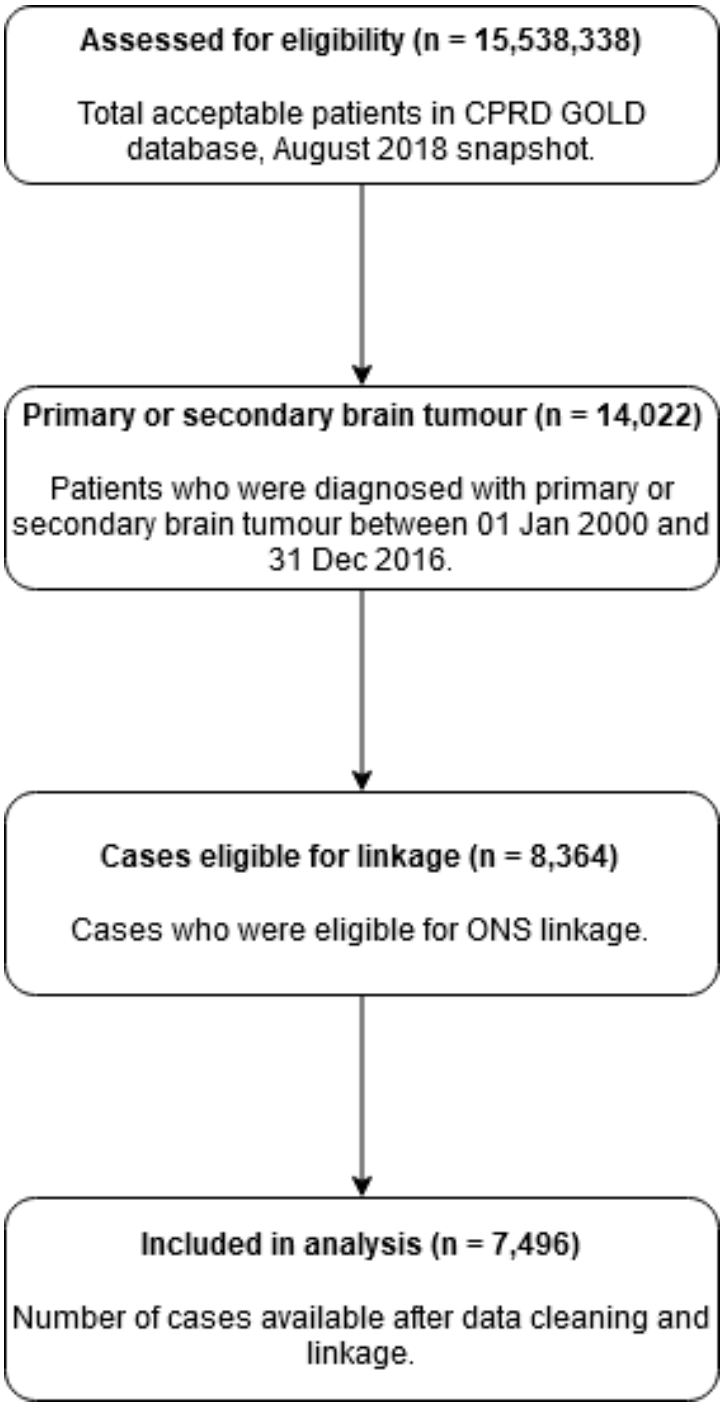
The dataset also did not include allelic information which is useful for MR analyses. Another group had already published using this data within the context of using MR to uncover potentially causal proteins for neurological phenotypes [173]. The authors made available their cleaned version of the data which was MR-ready and included allelic information derived from the original ROSMAP studies and was linked to the summary statistics, allowing for potentially useful allelic information in the MR analyses. The data download they provided had instruments selected based on their selection process, which was similar to those in the analyses detailed in Chapter 5. However, these selection steps were repeated as a form of independent validation. These steps are discussed further in Section 5.2.2 - *Instrument Selection*.

Finally, another consideration for this dataset is that the authors did not clarify the scale on which they measured the protein abundance levels. Therefore, results from this dataset and from the whole blood pQTL pooled analysis are not directly comparable. To enable comparisons between the two sets of data, MR results were scaled from analyses that use these datasets by one SD of the MR betas in each dataset. This is further explained in Section 5.2.3 - *Identifying the Influence of the Plasma and Brain Proteome on Subtype Risk*.

3.5 Summary

This Chapter described the datasets used in the research presented throughout this thesis. Much of the data used was derived from GWAS and, as the collection and analysing of these data are beyond the scope of this thesis, only summaries are presented here for those datasets. This Chapter also discussed some of the processes used to clean and derive the data for analyses presented later in this thesis. These steps are described in more detail in the relevant results Chapters (Chapter 4, 5 & 6).

Figure 3.3: Participant selection for the cohort study.



CHAPTER 4

TRANSCRIPTOME-WIDE MENDELIAN RANDOMISATION STUDY IDENTIFYING BRAIN-SPECIFIC GENES INFLUENCING GLIOMA DEVELOPMENT

The work presented in this Chapter was undertaken as part of a published piece of work [292]. The original idea for the project was devised by myself and my supervisors, who are joint first and joint last authors on the paper, namely Richard Martin, Kathreena Kurian and Jie Zheng. Glioma data were extracted with help from Amy Howell, Spiridon Tsavachidis, Georgina Armstrong and Melissa Bondy, who are named authors on the paper. My contributions to this paper are as follows. I identified, obtained and cleaned all datasets used throughout the analyses presented in this Chapter. Similarly I conducted all analyses produced all figures and tables and interpreted all the results with supervision from Jie Zheng, Richard Martin and Kathreena Kurian. I wrote the original and subsequent draft copies of the manuscript, with supervision from Richard Martin, Kathreena Kurian and Jie Zheng.

4.1 Introduction

There are only two accepted risk factors for glioma: exposure to ionising radiation and inheritable genetic disorders (as described in Section 1.5.1 - *Accepted Risk Factors*). Research into other causes of glioma has shown promise by highlighting the aetiological and diagnostic importance of somatic genetic, epigenetic and molecular markers, such as *TP53* and *IDH1* mutations, 1p/19q co-deletion and *MGMT* promoter methylation (see Section 1.6 - *Histological Tumour Markers* for further explanations of these). Genome-wide association studes (GWAS) haved documented strong statistical associations of germline genetic variants at thousands of loci with complex human traits and diseases; however, it is difficult to identify

directly causal genes using GWAS results, as the majority of associated genetic variants (90%) are found in non-coding regions of the genome and thus their influence is likely to occur via gene regulation, rather than a direct influence on protein structure or function [147]. Similarly, trait-associated SNPs have been shown to be three-times more likely to be associated with gene expression (known as expression quantitative trait loci, eQTLs) than non-trait associated SNPs [144, 235] and so understanding SNP-trait associations means investigating the potential role of these SNPs in the regulation of gene expression. The latest glioma GWAS identified 27 loci that are associated with glioma risk, but estimated that we have uncovered only about a third of the risk posed by familial or inheritable factors (27% for glioblastoma (GBM) and 37% for non-GBM) [220], indicating a large portion of genetic glioma risk is still to be uncovered.

Investigating and understanding how genes are differentially expressed in tumour subtypes has led to a better understanding of gliomagenesis through potentially related mechanisms and pathways and has also improved clinical outcomes for patients due to differential treatments. Previous studies have shown that genes are differentially expressed in glioma dependent on subtype [55, 118, 155, 242, 302, 319]. Furthermore, gene expression profiling was proposed as a better method of diagnosis over the previous clinical practice of histological grading because classification based on gene expression seemed to better predict for survival [104, 118, 242]. In 2016, the World Health Organisation (WHO) classification for central nervous system tumours updated their diagnostic rubric to include analysis of the tumour genome, albeit limited to, for example, IDH and 1p19q status (described previously in Section 1.6) [209]. It is likely that in the latest WHO guidelines, which at the time of writing are undergoing consultation, will still include genetic diagnostic and prognostic biomarkers to inform glioma classification and outcome [210]. Whilst including genetic factors into the classification criteria has seen measurable benefits for patients, functional studies have been limited and it is not known if certain mutations are merely correlated with gliomagenesis and subtype differentiation or play a causal role in risk. Furthermore, previous studies showed that germline genetic variants correlate according to somatic molecular mutations present in gliomas [90, 189, 390]. How germline genetic markers, which are common to all tissue and cell types, associated with tissue-specific expression differ by subtype diagnosis is therefore important both to elucidating mechanisms of glioma risk and development, and to further improve clinical outcomes for patients.

This study utilised Mendelian randomisation (MR) – an established instrumental variable method described in Section 2.3.3 - *Mendelian Randomisation* – to assess the causal relationship between genetically predicted gene expression on glioma subtype risk. To recap, MR suffers less from biases, such as reverse causation, and confounding, that invariably limit causal inference in traditional epidemiological studies [66, 67]. Colocalisation is a statistical method that can identify whether a putative causal genetic variant is shared by two traits – a necessary condition for causality[113]. Use of a combined MR and colocalisation pipeline can strengthen causal inferences by discounting MR results which arise due to confounding through linkage disequilibrium (LD)

– which can arise when another genetic variant in high LD with the variant of interest is also associated with the outcome conditional on the exposure and which MR struggles to differentiate [352, 400]. Integrating MR analyses with expression data from brain tissues provided insight into how tissue-specific gene expression may differentially alter glioma risk across the brain. These data were linked to eQTLs derived from blood to determine how the risk profile for glioma differs between brain tissue and whole blood.

4.2 Methods

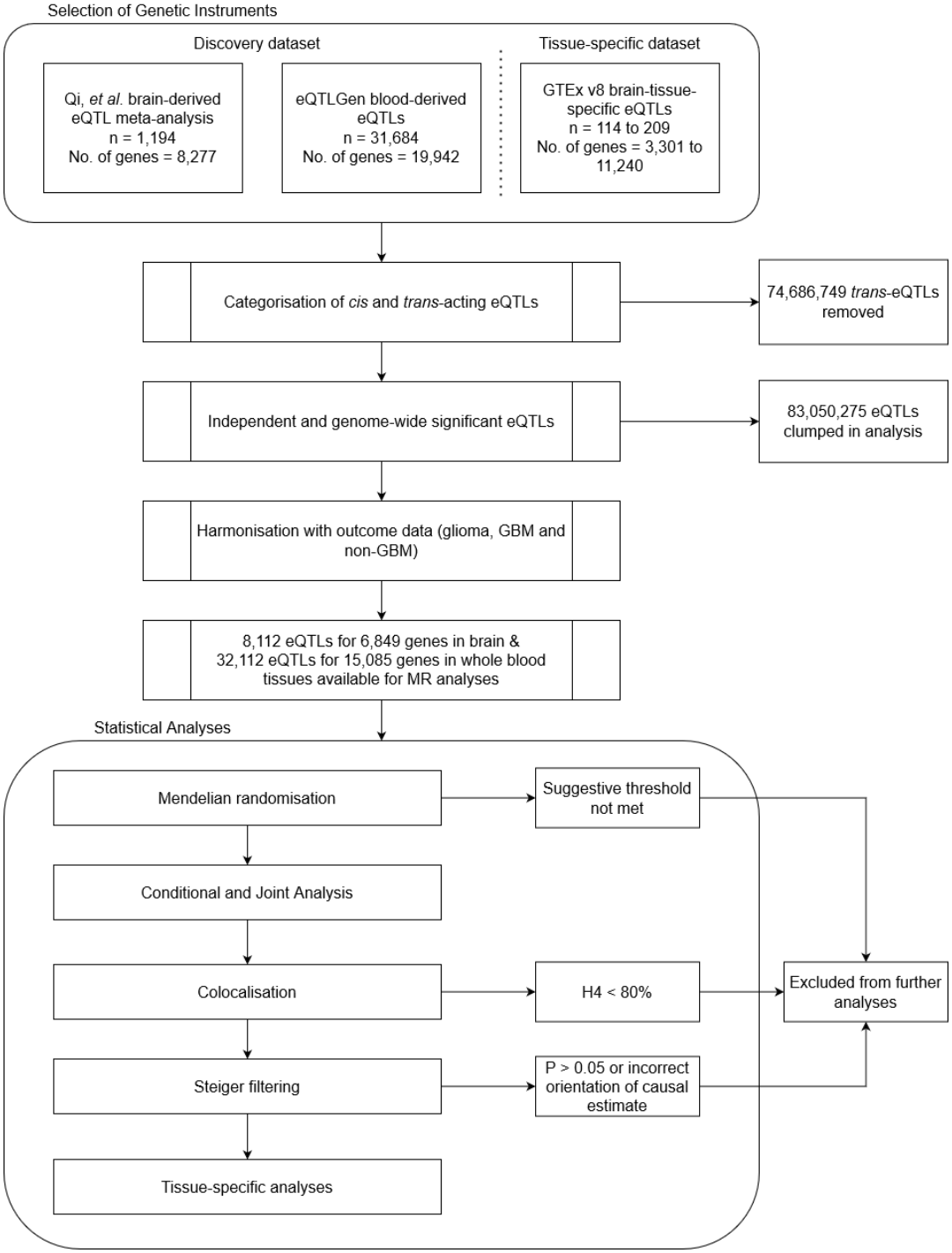
4.2.1 Data

Summary-level data were used from different GWAS to compare eQTLs from brain tissue (estimated effective $n = 1,194$, [276]) and from whole blood ($n = 31,684$, [361]). The analysis pipeline for this Chapter involved a two-sample MR framework, whereby the exposure and outcome data comprised independent populations, to estimate the causal effect of gene expression variation on glioma risk (based on subtype diagnoses of all glioma, GBM and non-GBM). In follow-up sensitivity analyses, eQTLs from Genotype-Tissue Expression version 8 (GTEx v8, $n = 114$ to 209, [121]) were used to examine tissue-specific effects of gene expression. All datasets used in this Chapter were described in Section 3.2 - *Glioma GWAS* for the outcome data and Section 3.4 - *Exposure Datasets* for the eQTL datasets.

4.2.2 Instrument Selection

The instrument selection process and analysis pipeline for this Chapter is summarised in Figure 4.1. eQTLs were categorised based on whether they were *cis*–acting or *trans*–acting, defined as SNPs inside and outside a 1Mbp window of the gene regulatory region (±500Kb), respectively. For analyses in this Chapter, only *cis*–acting eQTLs were included. The exclusion of *trans*–acting eQTLs can be preferable as they are more prone to horizontal pleiotropy due to their distal nature on the genome. This pleiotropy makes results for *trans*-acting eQTLs difficult to interpret in the context of providing evidence for drug targets because of potential off-target or unwanted effects. Furthermore, including *trans*–acting eQTLs, and thus more genes, would increase the burden on correction for multiple testing and may result in more reliable *cis* results being rejected. Instruments were constructed from each dataset using independent ($r^2 < 0.01$) SNPs that met genome-wide significance ($P < 5.00 \times 10^{-8}$). The amount of eQTLs available for the MR analysis in the brain-derived eQTL meta-analysis was 6,849. Therefore, it was pre-specified that results must meet a strict Bonferroni-corrected P value threshold of 7.30×10^{-6} ($0.05/6849$) or a suggestive P value threshold of 9.49×10^{-5} ($0.05/6849 \times 13$, multiplied by 13 for each brain tissue type). Results were considered for sensitivity analyses if they passed the suggestive P value threshold.

Figure 4.1: Schematic of the analysis and instrument selection pipeline employed for this study.



4.2.3 Identifying the Causal Effects of Genetically Predicted Gene Expression on Glioma Risk

Two-sample MR methods were applied to estimate the causal relationship between eQTLs and glioma using the MR-Base R package [142]. 86% of tests consisted of the Wald ratio (WR) MR method as many eQTLs were instrumented by a single SNP; eQTLs that were instrumented by multiple SNPs were analysed using the inverse variance weighted (IVW) method – both the WR and IVW methods were described in Section 2.3.3 - *Mendelian Randomisation*. MR results associated with all glioma risk were obtained first and then using the subset of GBM and non-GBM cases to obtain results associated with subtype risk.

To test for colocalisation between gene expression and glioma, regions of SNPs were extracted within ±500Kbp (forming a 1Mbp window) around the instrumented SNP(s) for each eQTL with an MR result that met at least the suggestive P value threshold ($P < 9.49 \times 10^{-5}$). These regions were subjected to a conditional analysis on the marginal associations (i.e. secondary/tertiary/etc.) in the region of the lead SNP(s) using Conditional and Joint analysis (GCTA-COJO) [391] and colocalisation was tested using the coloc R package [113]. Colocalisation analyses were conducted with the "coloc.abf" function using each SNPs' SNP-trait effect (β coefficient) and corresponding standard error, minor allele frequency, P value and sample size. The combined conditional and colocalisation pipeline was performed systematically and automated using a tool called Pair-wise conditional and colocalisation (PWCoCo) analysis, described previously in Section 2.3.2. Default parameters were also used. Throughout this Chapter, colocalisation estimates are presented with regard to the H_4 hypothesis though full results are available in the appendices. Further details about colocalisation and the rationale for using conditional analyses were described in Section 2.3.2 - *Statistical Colocalisation*.

Steiger filtering analyses were conducted to ensure results were not distorted due to the presence of reverse causation [140]. Steiger filtering calculates the variance explained (r^2, not to be confused with the same shorthand for LD) for each SNP and is used to identify and remove those SNPs which explain more of the variance in the outcome than the exposure. In this study, Steiger filtering was utilised as a post-MR sensitivity analysis to ensure that SNPs which are highlighted by the MR analyses were also not reverse-causal, and was conducted using default parameters. For each SNP, Steiger filtering returned a direction of effect (dependent on which dataset has more variance explained by the SNP) and a P value. To pass the Steiger filtering sensitivity analysis, the causal direction must have been correctly orientated from the exposure to the outcome (i.e. the SNP explains more variance in the exposure than the outcome) and $P < 0.05$. The Steiger P value was not corrected for multiple testing because the Steiger filtering analysis was included to provide supporting evidence that the direction of effect was from the exposure to the outcome after passing a corrected threshold for the MR results and standard threshold of 80% for colocalisation. Further thresholding on the Steiger filtering results may therefore exclude some results as false negatives, especially in the datasets with smaller sample

sizes which Steiger filtering can be sensitive to.

The main results for this Chapter consisted of those results which passed at least the suggestive MR P value threshold ($P < 9.49\times10^{-5}$), had strong evidence of colocalisation ($H_4 \geq 80\%$) and passed the Steiger filtering sensitivity analysis (causal direction correctly orientated from exposure to outcome and Steiger $P < 0.05$). These results were the most reliable and least likely to be biased, e.g. by LD structure or horizontal pleiotropy, due to the strong evidence of colocalisation and by passing the Steiger filtering sensitivity analysis. It is important to note that horizontal pleiotropy remains an alternative explanation for these results however.

4.2.4 Examining Tissue-Specific Effects of Gene Expression on Glioma Risk

Each MR association which passed the suggestive P value threshold ($P < 9.49\times10^{-5}$) was included in a follow-up sensitivity analysis to examine tissue-specific effects of gene expression on all glioma, GBM and non-GBM risk. SNPs were systematically mapped to relative genes across 13 brain tissues from GTEx v8 based on Ensembl IDs (ENSG) to form eQTLs. Only *cis*–acting eQTLs were included to avoid potentially pleiotropic *trans*–acting eQTLs. Instruments were constructed from independent ($r^2 < 0.01$) SNPs that met an arbitrary, lenient threshold ($P < 5.00\times10^{-4}$) to ensure a greater chance that there will be an eQTL for each tissue type. This threshold should be viewed as enabling a heuristic approach to the tissue-specific analyses by allowing for more genes to be instrumented in different tissues [334]. In particular, eQTLs may have been constructed from the larger tissues (e.g. cerebellum, n = 209) but, due to the relatively large differences in sample sizes between brain tissues in GTEx v8, the same eQTLs may not reach genome-wide significance in the smaller tissues (e.g. substantia nigra, n = 114). For a gene to be included in this analysis, its MR association must have passed the suggestive P value threshold already, thereby signifying this gene as a gene of interest, justifying a more lenient threshold to construct instruments in other brain tissues as a comparator. Constructing instruments in this way exposes the MR analysis to potentially pleiotropic instruments and so results were also tested for heterogeneity, which is a way to gauge potentially pleiotropic SNPs in multi-SNP instruments [140] and may highlight the presence of pleiotropic results. These data were analysed similar to the main analyses by using the MR-Base R package [142]. The magnitude and direction of the effect estimate across tissue types and subtype diagnosis were compared both graphically and systematically using the R "cor" command.

To ascertain whether tissue-specific gene expression differentially altered risk of glioma, GBM and non-GBM, heterogeneity analyses were conducted using the Q statistic and derived using the MR-Base R package [142]. The tau-score was also calculated by using SNP-trait effects (β coefficient) as described by Kryuchkova-Mostacci and Robinson-Rechavi [187]. The tau-score is a quantitative measure of tissue-specificity, and can be naively summarised as summing the weighting of a gene's expression in a single tissue against the maximum expression over all tissues. The tau-score ranges between 0 and 1, where 0 means the gene is broadly expressed and

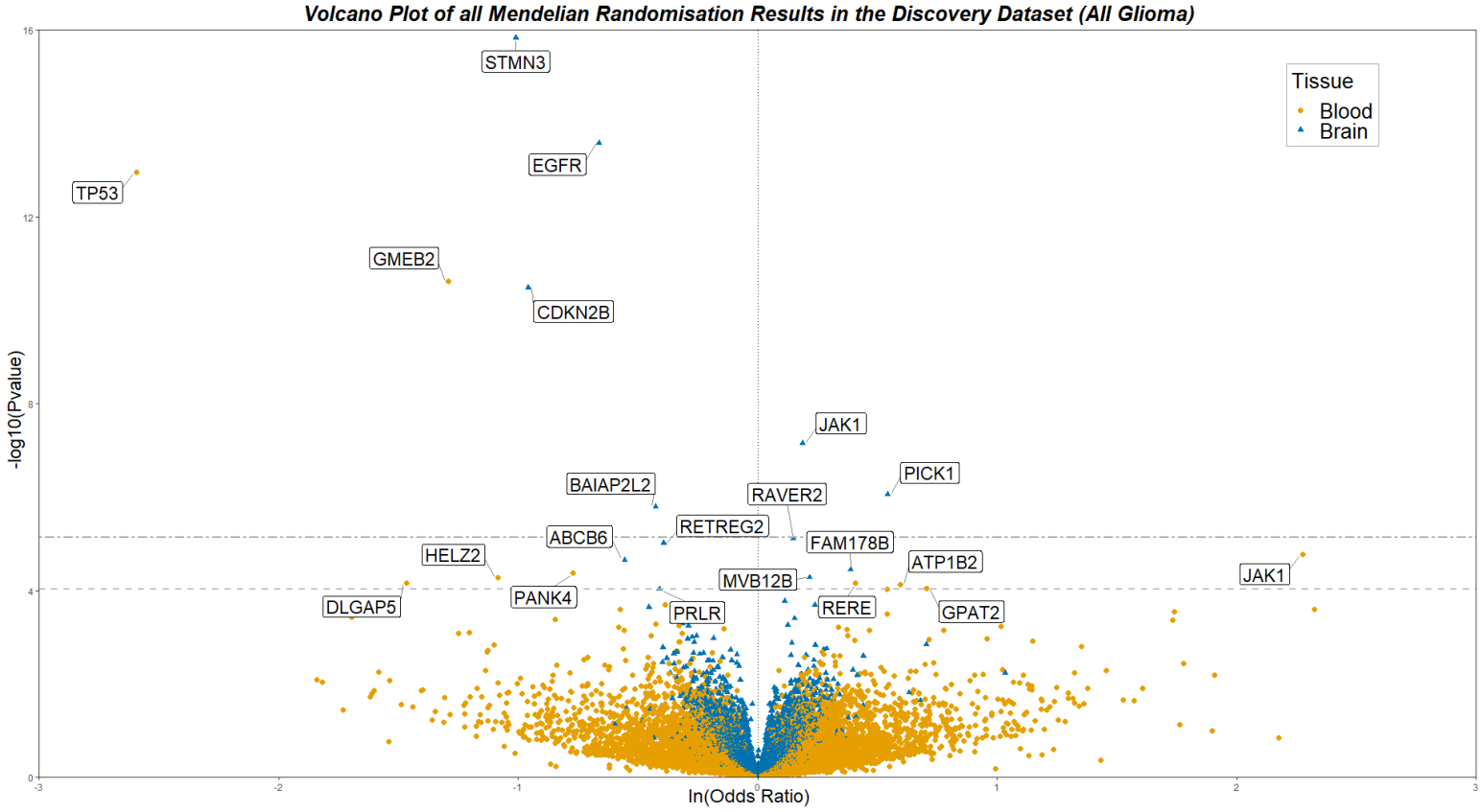
1 means specific expression. In their paper, Kryuchkova-Mostacci and Robinson-Rechavi defined a threshold cut-off for specific expression at 0.8 which was used for these analyses [187].

To test whether MR results using brain- and blood-derived eQTLs correlated, causal estimates were systematically linked for eQTLs that appeared in both brain tissue and blood and compared to determine: i) whether blood-based eQTLs were correlated with brain-based eQTLs, and ii) whether easier-to-gather blood data could proxy sufficiently for brain data in both magnitude and direction of effect. If blood-derived eQTLs correlated with brain-derived eQTLs, then higher powered blood eQTLs could be used in analyses such as this without fear of loss of applicability to risk of glioma. This analysis was performed using the R "cor" command and ggplot2 [378].

4.3 Results

In total, the MR analysis of brain and blood eQTLs identified 34 associations that met at least the suggestive P value threshold ($P < 9.49 \times 10^{-5}$) for 17 genes associated with risk of glioma, GBM or non-GBM (Figure 4.2). Altogether, six genes were instrumented by eQTLs in blood and 12 genes had eQTLs from brain tissue – one gene, *JAK1*, had an associated eQTL in both brain and blood. 20 associations had strong evidence of colocalisation ($H_4 \geq 80\%$), 4 associations had moderate evidence ($80\% > H_4 \geq 50\%$) and 10 associations had weak evidence ($H_4 < 50\%$). Steiger filtering revealed the direction of the causal estimate was correctly orientated from gene expression to subtype diagnosis in 29 associations; the remaining five showed an uncertain result due to the P value for Steiger filtering not reaching 0.05. Overall, 17 tissue-subtype associations for 12 genes showed robust causal evidence from the MR and colocalisation analyses and passed the Steiger filtering analysis. These 17 associations and 12 genes formed the main results for this Chapter, are presented in Table 4.1, and were subjected to the tissue-specific analyses. Appendix B contains MR results for genes in other subtypes for comparison and Appendix C contains the unabridged colocalisation results. All results are given in the Online Appendix, Tables S1-4 (see Appendix K).

Figure 4.2: Volcano plot of all results from the main MR analysis of brain and blood eQTLs and all glioma. The horizontal dot-dashed line represents the Bonferroni-corrected P value threshold ($P < 7.30\times10^{-6}$) and the horizontal dashed line is the suggestive P value threshold ($P < 9.49\times10^{-5}$). Genes are labelled if they pass at least the suggestive threshold.



Comparing these results with previously identified GWAS associations (listed previously in Table 1.2 from a review conducted by Kinnersley, *et al.* [179]) revealed that *RETREG2* (also known as *FAM134A*), *FAM178B* and *MVB12B* (also known as *FAM125B*) are putative novel genes implicated in glioma risk that are not also located on a known glioma risk locus and formed part of the main results. The remaining results have been previously implicated in glioma risk through GWAS associations or are located on a known susceptibility locus.

Figure 4.3 shows the MR effect estimates for each of the 12 genes and all glioma, GBM and non-GBM subtypes. The direction and magnitude of the estimated causal effect broadly agreed across all genes and subtypes. However, the non-GBM results were attenuated, for example, in the case of *STMN3*, possibly due to the heterogeneity of cases captured by such a broad categorisation.

Table 4.1: Main associations which showed robust evidence from the MR, colocalisation and Steiger filtering analyses. The colocalisation result for a single, shared causal variant between the gene and all glioma is provided (H_4, given as %). Steiger filtering showed the correct orientation for the direction of effect between gene expression and subtype risk for all results in this table. Unit of exposure is per SD increase in gene expression [276].

Gene	**SNP(s)**	**Tissue**	**Subtype**	**OR (95% CI)**	**P value**	H_4	**Steiger *P***
ABCB6	rs75450661	Brain	All glioma	0.57 (0.44, 0.74)	2.20×10^{-5}	97	7.41×10^{-6}
BAIAP2L2	rs1004764	Brain	All glioma	0.65 (0.55, 0.78)	1.62×10^{-6}	96	2.36×10^{-10}
			GBM	0.60 (0.49, 0.73)	2.85×10^{-7}	81	1.25×10^{-9}
EGFR	rs6979446, rs759170	Brain	GBM	0.45 (0.38, 0.53)	9.99×10^{-20}	81	3.53×10^{-6}
FAM178B	rs13407036	Brain	All glioma	1.47 (1.23, 1.77)	3.59×10^{-5}	94	1.97×10^{-16}
JAK1	rs2780902	Brain	All glioma	1.21 (1.13, 1.29)	6.95×10^{-8}	81	6.89×10^{-139}
			GBM	1.27 (1.17, 1.37)	1.56×10^{-9}	95	1.84×10^{-134}
MVB12B	rs4837096	Brain	All glioma	1.24 (1.12, 1.38)	5.27×10^{-5}	97	2.53×10^{-23}
PANK4	rs2985862	Blood	All glioma	0.46 (0.32, 0.67)	4.30×10^{-5}	97	4.62×10^{-10}
PICK1	rs5756894	Brain	All glioma	1.72 (1.39, 2.14)	8.82×10^{-7}	97	4.34×10^{-7}
			GBM	1.96 (1.54, 2.51)	6.60×10^{-8}	92	2.13×10^{-6}
PRLR	rs67975005	Brain	All glioma	0.66 (0.54, 0.82)	9.33×10^{-5}	91	1.13×10^{-7}
RETREG2	rs1996719	Brain	All glioma	0.68 (0.57, 0.80)	9.54×10^{-6}	98	7.90×10^{-11}
			GBM	0.67 (0.55, 0.81)	6.13×10^{-5}	95	9.91×10^{-11}
STMN3	rs6011016	Brain	All glioma	0.36 (0.29, 0.46)	1.44×10^{-16}	96	1.44×10^{-16}
			GBM	0.29 (0.22, 0.38)	4.55×10^{-19}	97	7.88×10^{-3}
TP53	rs35850753	Blood	Non-GBM	0.17 (0.09, 0.32)	9.61×10^{-8}	98	3.35×10^{-2}

Figure 4.3: Forest plot of MR results for the 12 genes which had robust causal evidence and also passed sensitivity analyses. Note that the effect estimates for the all glioma dataset are not the summed average of both subtypes due to some cases not having accurate subtype diagnoses, meaning those cases were excluded from the GBM and non-GBM subtype datasets and thus analyses.

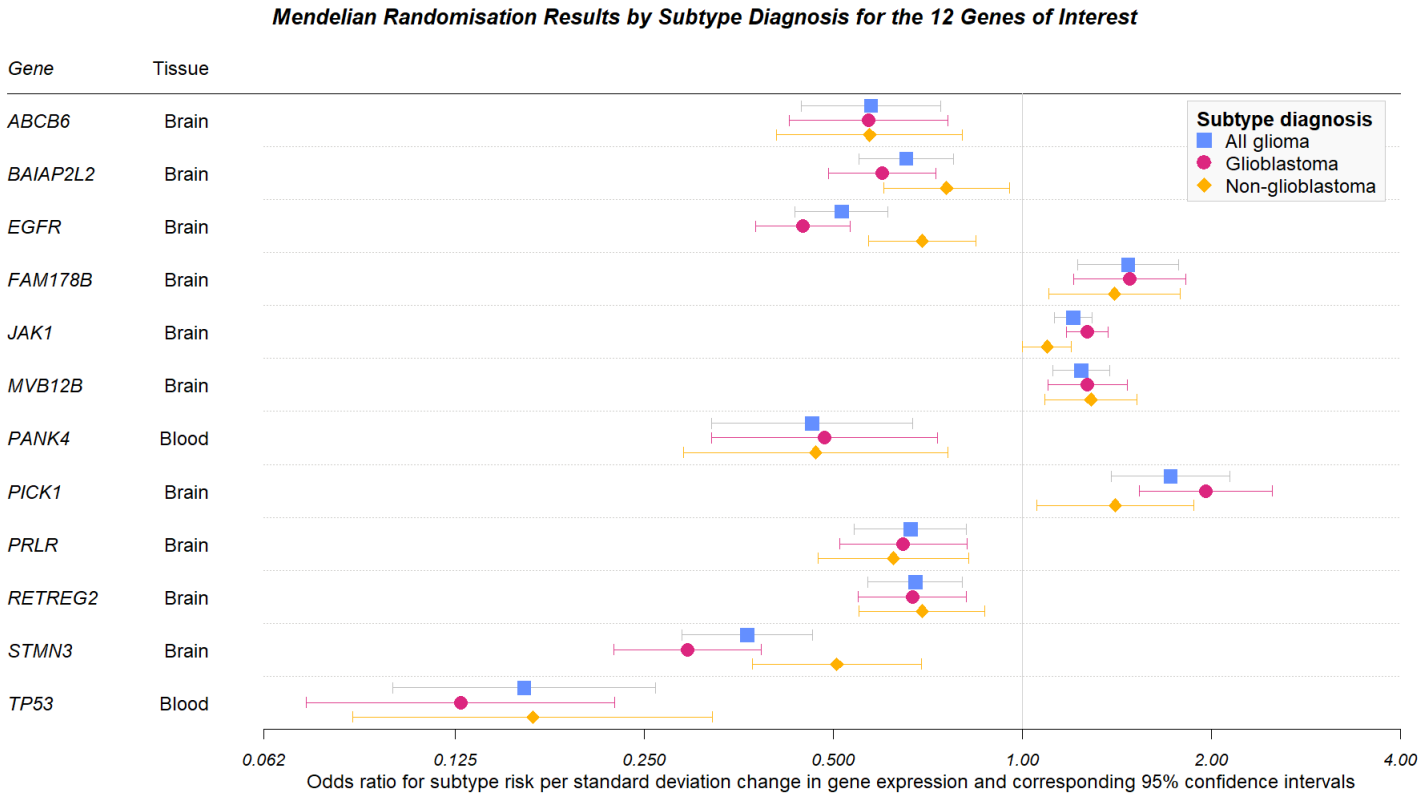


Table 4.2: Amount (denoted by #) and relative percentage (denoted by %) of MR results for highlighted genes in the tissue-specific analysis which had MR evidence ($P < 7.30\times10^{-6}$) and colocalisation evidence ($H_4 \geq 80\%$) and passed the Steiger filtering sensitivity analysis. Breakdown per tissue.

Tissue	**MR evidence**		**& Coloc**		**& Steiger**	
	#	**%**	**#**	**%**	**#**	**%**
Amygdala	1	2%	0	0%	0	0%
Anterior cingulate cortex (BA24)	5	8%	2	5%	0	0%
Caudate (basal ganglia)	7	11%	6	14%	0	0%
Cerebellar Hemisphere	6	9%	5	12%	2	13%
Cerebellum	7	11%	5	12%	3	19%
Cortex	7	11%	7	17%	3	19%
Frontal Cortex (BA9)	7	11%	3	7%	2	13%
Hippocampus	4	6%	1	2%	1	6%
Hypothalamus	2	3%	2	5%	2	13%
Nucleus accumbens (basal ganglia)	7	11%	4	10%	0	0%
Putamen (basal ganglia)	8	12%	5	12%	1	6%
Spinal cord (cervical c-1)	2	3%	1	2%	1	6%
Substantia nigra	2	3%	1	2%	1	6%

To examine how tissue-specific gene expression affected glioma risk, the 12 genes which formed the main results were systematically linked to eQTLs across 13 brain tissues using GTEx v8. The effects of tissue-specific expression of most genes were assessed using MR in 8 to 13 tissues (mean = 10 tissues) except for *ABCB6*, which had data in four tissues. The tissue-specific results broadly agreed with the main MR analysis, though results had generally wider CI due to the smaller sample sizes present in the GTEx v8 dataset. Applying the same threshold for the discovery MR analysis ($P < 7.30\times10^{-6}$) revealed that 56% of the results arise in five tissues: putamen (basal ganglia) (12%), cortex (11%), cerebellum (11%), caudate (basal ganglia) (11%) and nucleus accumbens (basal ganglia) (11%) (Table 4.2). Furthermore, 100% of these results arose due to four genes: *JAK1*, *STMN3*, *PICK1* and *EGFR* (Table 4.3). The tissue-specific analyses also provided evidence of replication for these same four genes as they had strong evidence of both MR and colocalisation and also passed the Steiger filtering sensitivity analysis in at least one tissue.

Two tissue-specific results showed evidence of high heterogeneity. These were *EGFR* for all glioma (Q = 155.96, $P = 1.72\times10^{-28}$) and GBM (Q = 162.38, $P = 3.49\times10^{-27}$) subtype analyses. Only *PICK1* showed potential tissue-specific expression with a tau-score of 0.78. All Q statistics and tau-scores are given in Table 4.4.

Finally, effect estimates from the MR analysis were compared for genes expressed in brain and blood tissues. Four P value thresholds ($P < 0.1, 0.05, 0.01, 0.005$) were applied to examine whether the strength of the MR association influences the correlation between estimates in brain and blood tissues. Overall, a low correlation was observed between brain and blood eQTLs

Table 4.3: Amount (denoted by #) and relative percentage (denoted by %) of MR results in the tissue-specific analysis which had MR evidence ($P < 7.30 \times 10^{-6}$) and colocalisation evidence ($H_4 \geq 80\%$) and passed the Steiger filtering sensitivity analysis. Breakdown per gene.

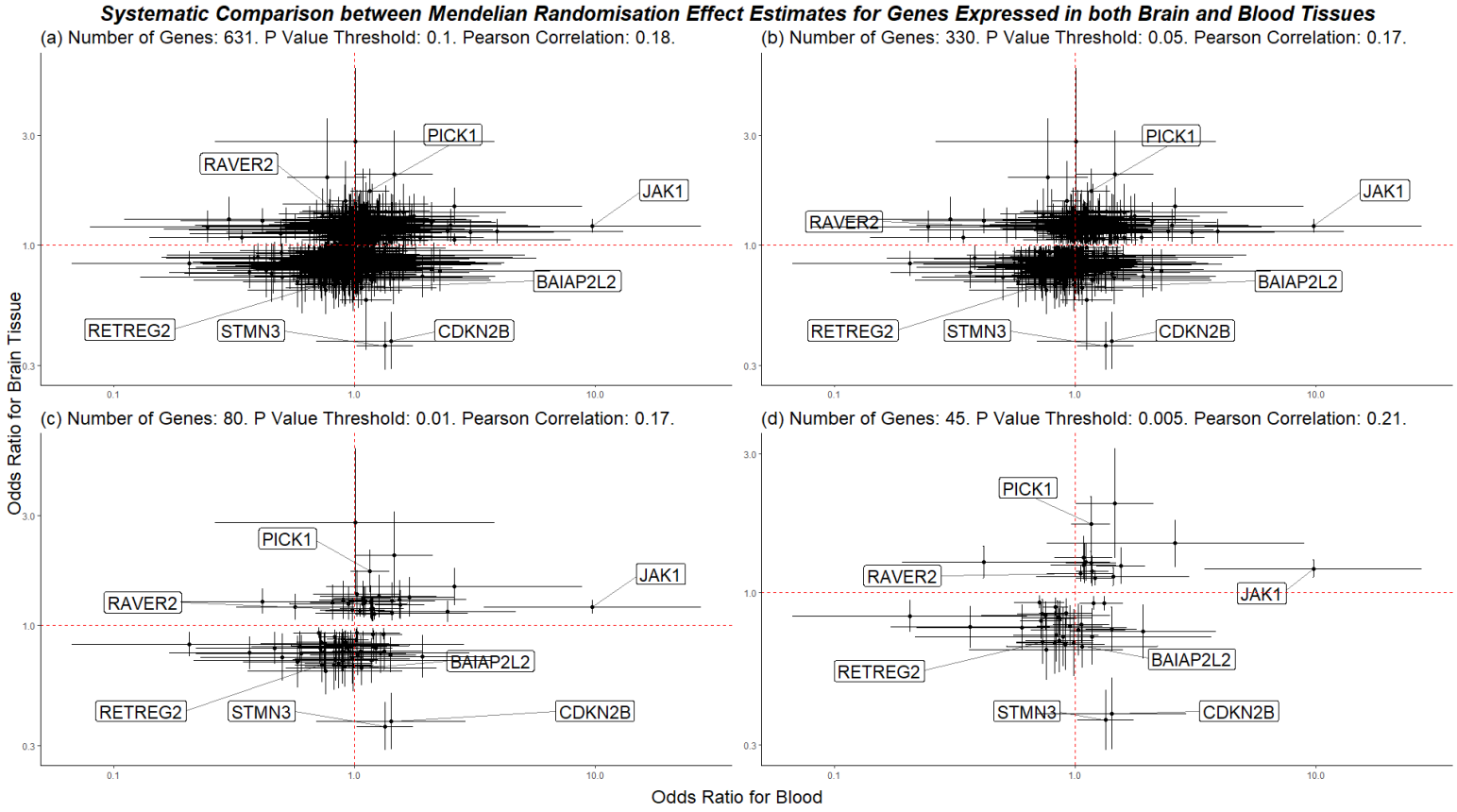
Gene	**MR evidence**		**& Coloc**		**& Steiger**	
	#	*%*	**#**	*%*	**#**	*%*
ABCB6	0	0%	0	0%	0	0%
BAIAP2L2	0	0%	0	0%	0	0%
EGFR	17	26%	6	14%	1	6%
FAM178B	0	0%	0	0%	0	0%
JAK1	12	18%	10	24%	7	44%
MVB12B	0	0%	0	0%	0	0%
PANK4	0	0%	0	0%	0	0%
PICK1	15	23%	9	21%	6	38%
PRLR	0	0%	0	0%	0	0%
RETREG2	0	0%	0	0%	0	0%
STMN3	21	32%	17	40%	2	13%
TP53	0	0%	0	0%	0	0%

(Pearson correlation = 0.18, number of genes = 632) at the highest P value threshold ($P < 0.1$). After applying the more stringent threshold ($P < 0.005$), the correlation increased but remained low (Pearson correlation = 0.21, number of genes = 45). These correlations broadly agree with what has been found in the literature previously. A paper by McKenzie, *et al.* showed a mean overlap between brain and blood eQTLs of roughly 19% [218]. Comparisons between MR derived effect estimates for eQTLs in brain and blood appears to be novel, however. These results are shown in Figure 4.4.

Table 4.4: Results from the tissue-specific sensitivity analyses: the Q statistic and Tau-score. Q is the Q statistic. Q (df) is the degrees of freedom equivalent to the amount of tissues minus one. Q (P) is is the corresponding P value for the Q test. Tau is the tau-score for tissue specificity.

Gene	**Subtype**	**Q**	**Q (df)**	**Q (P)**	**Tau**
ABCB6	All	22.339	3	5.54×10^{-5}	0.49
ABCB6	GBM	17.513	3	5.54×10^{-4}	0.49
ABCB6	Non-GBM	5.582	3	1.34×10^{-1}	0.49
BAIAP2L2	All	12.177	8	1.43×10^{-1}	0.70
BAIAP2L2	GBM	24.007	8	2.29×10^{-3}	0.70
BAIAP2L2	Non-GBM	2.057	8	9.79×10^{-1}	0.70
EGFR	All	155.960	12	3.49×10^{-27}	0.38
EGFR	GBM	162.379	12	1.72×10^{-28}	0.38
EGFR	Non-GBM	32.155	12	1.31×10^{-3}	0.38
FAM178B	All	17.177	8	2.83×10^{-2}	0.49
FAM178B	GBM	9.876	8	2.74×10^{-1}	0.49
FAM178B	Non-GBM	12.699	8	1.23×10^{-1}	0.49
JAK1	All	32.609	8	7.24×10^{-5}	0.56
JAK1	GBM	45.677	8	2.74×10^{-7}	0.56
JAK1	Non-GBM	5.033	8	7.54×10^{-1}	0.56
MVB12B	All	25.309	12	1.34×10^{-2}	0.61
MVB12B	GBM	22.797	12	2.95×10^{-2}	0.61
MVB12B	Non-GBM	23.595	11	1.46×10^{-2}	0.58
PANK4	All	17.007	11	1.08×10^{-1}	0.68
PANK4	GBM	13.254	11	2.77×10^{-1}	0.68
PANK4	Non-GBM	10.914	11	4.50×10^{-1}	0.68
PICK1	All	31.628	12	1.58×10^{-3}	0.78
PICK1	GBM	39.272	12	9.49×10^{-5}	0.78
PICK1	Non-GBM	10.422	12	5.79×10^{-1}	0.78
PRLR	All	8.822	7	2.66×10^{-1}	0.47
PRLR	GBM	14.880	7	3.76×10^{-2}	0.47
PRLR	Non-GBM	4.370	7	7.36×10^{-1}	0.47
RETREG2	All	17.260	9	4.48×10^{-2}	0.48
RETREG2	GBM	20.921	9	1.30×10^{-2}	0.48
RETREG2	Non-GBM	8.382	9	4.96×10^{-1}	0.48
STMN3	All	83.032	10	1.27×10^{-13}	0.43
STMN3	GBM	89.269	10	7.48×10^{-15}	0.43
STMN3	Non-GBM	51.587	10	1.36×10^{-7}	0.43
TP53	All	7.024	10	7.23×10^{-1}	0.61
TP53	GBM	6.303	10	7.89×10^{-1}	0.61
TP53	Non-GBM	10.399	10	4.06×10^{-1}	0.61

Figure 4.4: Systematic comparison between the MR results from brain tissues and blood. Any eQTL that appeared in both brain and blood datasets was included in this analysis. ORs for blood against brain are plotted. Results were included based on a P value threshold: (a) $P < 0.1$, (b) $P < 0.05$, (c) $P < 0.001$ and (d) $P < 0.005$. Labels are provided for genes which had an association that passed at least the suggestive P value threshold ($P < 9.49 \times 10^{-5}$).



4.4 Discussion

For this Chapter, MR and colocalisation analyses were combined to estimate the genetically predicted gene expression on glioma risk which provided causal evidence for 12 genes. Three of these genes are novel in the context of glioma risk. Overall, these results were robust to sensitivity analyses, including Steiger filtering and tissue-specific analyses.

RETREG2 (or *FAM134A*), *FAM178B* and *MVB12B* (or *FAM125B*) appear to be novel findings related to glioma risk. Reticulophagy Regulator Family Member 2 (*RETREG2*, *FAM134A*) is a protein-coding gene whose function is largely unknown. Examination of the Human Protein Atlas reveals that expression of *RETREG2* RNA and protein is primarily located in the brain and testes, though tissue specificity is low [346]. Family with Sequence Similarity 178 Member B (*FAM178B*) is another protein-coding gene with an undocumented function. It is an important paralogue of gene *SLF2* whose protein plays a role in the DNA damage response. Finally, Multivesicular Body Subunit 12B (*MVB12B*, *FAM125B*) is a regulator of vesicle trafficking and has been implicated in lipid and ubiquitin binding. Overexpression of this gene and its protein inhibits HIV-1 infectivity by regulating ESCRT (endosomal sorting complex required for transport-I)-mediated virus budding [226]. A 2014 study created a nine-gene-signature panel, which included *MVB12B*, that accurately predicted prognosis for glioma patients, further implicating the gene's role in glioma biology [17]. Further research into these genes is warranted to provide replication and to elucidate potential pathways by which these genes affect glioma risk.

When considering differential risk across subtype diagnosis, the MR results showed agreement in the direction of effect for risk of all glioma, GBM and non-GBM. However, associations with non-GBM risk tended to be weaker in magnitude than associations with the other two subtypes of glioma and were generally attenuated. Case numbers were broadly similar – 3,112 GBM cases versus 2,411 non-GBM cases – but may still be underpowered in the non-GBM analysis. This was evidenced by the consistently larger P values for the MR results of the non-GBM analysis when compared to the GBM analysis. Whether this attenuation was due to the heterogeneous nature of brain tumours – which will be particularly more pronounced in the non-GBM subtype analyses due to inherently capturing many different tumour types – or due to, for example, the lack of power in the subtype analysis, requires follow-up analyses in larger datasets. Overall, however, there was little evidence to conclude that there exists a large difference in the risk profile driven by variance in gene expression of GBM and non-GBM tumours. Therefore, future studies in this area may seek to consolidate data independent of subtype diagnosis so that larger statistical power may be achieved.

Gliomas may develop across the entirety of the central nervous system but are generally found in the cerebrum, particularly the frontal and temporal lobe, and less commonly in the cerebellum depending on the age of the patient [195]. Analyses sought to determine how genetically predicted gene expression in the 13 brain tissue types collected by GTEx v8 [3] compared to the anatomical regions within which tumours are found. The 12 genes that formed the main results were matched

to, on average, an eQTL in 10 brain tissues allowing for a broad investigation on how glioma risk is affected by gene expression in disparate tissues. Applying MR and a similar P value threshold ($P < 7.30 \times 10^{-6}$) revealed that 56% of the results that met that threshold arose in five of the 13 tissues. Two of these are common/uncommon tissues for glioma (cortex (11%), cerebellum (11%), respectively). The other three tissues were from the deep brain and are considered rarer locations for glioma (putamen (basal ganglia) (12%), caudate (basal ganglia) (11%) and nucleus accumbens (basal ganglia) (11%)). These analyses provided evidence that gene expression in these five tissues potentially drives glioma risk. These five tissues are generally locations where glioma is not found – this could be explained by the diffusive nature of the gliomas, though further studies would need to investigate this.

Furthermore, 100% of the tissue-specific results arose due to four genes: *JAK1*, *STMN3*, *PICK1* and *EGFR* providing evidence of validation and highlighting these genes of high importance for follow-up studies. *EGFR* also showed high heterogeneity for risk of all glioma and GBM subtype, indicating gene expression in certain tissues may affect risk differently – for *EGFR* these were the hippocampus, hypothalamus and substantia nigra. However, follow-up analyses showed little evidence of tissue-specific gene expression, with one gene showing suggestive evidence of tissue specificity (*PICK1*, tau-score = 0.78). Larger sample sizes will allow greater clarity as to how exactly gene expression across multiple tissues differentially affects glioma risk. Broadly, these analyses showed that germline variants associated with gene expression across the entire brain, agnostic of tissue site, drives glioma risk as opposed to specific to certain tissues, though the same gene expressed in different tissues may differentially affect risk.

Analyses also investigated whether blood eQTLs, for which there are datasets of large sample sizes available, can proxy for relatively low powered brain tissue eQTLs. MR effect estimates were compared for eQTLs that were systematically linked between datasets based on their associated gene. Applying an increasingly stringent P value threshold showed little correlation between the MR associations for brain and blood in the context of glioma risk. These results showed that genetically predicted gene expression in brain which was associated with glioma risk did not necessarily mean the same gene expression in blood could be assumed to also affect risk similarly; in some cases, the risk profile of genes expressed in brain and blood differed in direction of effect, e.g. variants associated with *STMN3* expression seemed to increase risk in blood and decrease risk in brain tissues.

It is not clear from these analyses as to why some genes have discordant directions of effects across brain and blood tissues. One explanation could include the presence of the blood-brain barrier, where some tissue cells could cross and appear elsewhere thus indicating expression of that gene in a disparate tissue. Another, similar, explanation could be that the blood-brain barrier can break down and weaken due to the presence of particularly aggressive glioma, and therefore brain tissue cells may start to appear in the blood, which would explain the somewhat contradictory findings that some genes appear to have opposite directions of effect when found in

blood and brain tissues. However, while either of these explanations could explain results seen in a study which directly measured gene expression levels, this study used germline genetic variants associated with gene expression measured from actual blood and brain tissues. Therefore, the effect estimates from the MR analyses were with respects to the effects of genetically predicted gene expression in blood and brain tissues on glioma risk. It could be that for protein-coding genes, the resultant protein may be able to cross the blood-brain barrier from blood into brain and affect gliomagenesis in that way; however, this is speculation and would require further downstream and functional analyses to ascertain whether this is the mechanism of effect the MR estimates are capturing.

A potential explanation for the discordant effects observed comes from a paper from Mizuno, *et al.* where they investigated so-called "opposite eQTL effects". These were defined as eQTLs with differing directions of effect observed in different tissues [225]. The authors found that these opposite eQTL effects may be a common phenomenon that does not necessarily mean that results with opposite directions of effect contradict each other and, instead, might provide further insight into the physiological development of diseases. This could mean that, in the case of *STMN3*, for example, having a different direction of effect in brain and blood tissues might point to differential effects when expressed in those tissues. Therefore, these results may not be contradictory, highlighting this gene and its protein product as targets of interest to investigate further. It is still unknown, at this time, whether these opposite eQTL effects will have relevant clinical interpretations. For the time being, these results remain difficult to interpret and remains an open question in drug target and multi-omics/-tissue MR analyses; therefore, an interesting potential avenue of future research would be to therefore determine why some genes, like *STMN3*, differentially affect risk depending if they are expressed in brain or blood.

Thresholds were applied to results in this Chapter to correct for multiple testing based on the number of MR tests conducted. This was reasonable as analyses presented herein were hypothesis free. However, some important genes could potentially have been missed due to not reaching that threshold, which was why a lenient, or suggestive, threshold was also implemented. Despite this, genes such as *HEATR3*, a known risk loci for glioma, still did not reach even the lenient threshold. This could have been due to power, which remains low in molecular trait GWAS when compared to other traits. Integration of data with differing sources of bias, in a process known as triangulation of evidence [197], may allow for an even more lenient threshold should that evidence agree with and support the result in question. This was used in Chapter 5 where pQTL datasets with lower power than the eQTL datasets were used. However, the analyses in this Chapter revealed genes *RETREG2*, *FAM178B* and *MVB12B* were implicated in glioma risk, which would not have had supporting evidence as they appear to be novel genes. Furthermore, the genes found to affect glioma risk in this Chapter were the most reliable results and so should be prioritised over other genes; however, because of the (sometimes) stringent thresholds applied to these results, other genes should not necessarily be de-prioritised as a result of these analyses.

Whilst conducting analyses to provide evidence for de-prioritisation of targets is valuable, the analyses presented herein are not well-suited for this task and, indeed, was not their purpose.

Strengths of the analyses presented in this Chapter included the use of germline genetic variants that proxy for gene expression levels, which should have reduced the influence of confounding and bias through reverse causation. Furthermore, these variants were obtained from relatively large meta-analyses allowing for increased statistical power and more precise estimates. The addition of subtype diagnoses and tissue-specific data allowed for a deeper investigation into the risk profile of glioma.

The methodology employed throughout was also a strength of these analyses. The MR analyses used in this Chapter, which were less liable to sources of confounding and bias, provided evidence which strengthens causal inferences for gene expression and glioma risk. Combining the MR results with colocalisation provided supporting evidence further strengthening causal inference by determining whether gene expression and glioma risk share a single, causal variant – a necessary condition for causality. The method of colocalisation used in this Chapter has an inherent limitation whereby it assumes there is only a single causal variant within the genomic region being tested [113]; however, this was addressed by the inclusion of a conditional analysis before the colocalisation analysis which conditioned upon the lead SNP(s) and allowed for the single causal variant assumption to hold. Follow-up analyses in different tissues also acted as a replication study providing further evidence of causality. However, despite providing evidence of causality, this study did not prove causality and further studies would be required to determine this.

These analyses were not without limitations. Despite the relatively large datasets used to extract data relating to gene expression and glioma, analyses were still likely suffer from low statistical power arising from small sizes. This was potentially evidenced in the non-GBM subtype analyses. All but one of the main results were instrumented by a single SNP which limited the ability to undertake common MR sensitivity analyses to detect, for example, horizontal pleiotropy. Colocalisation has been proffered as a sensitivity analysis that can at least eliminate spurious associations that have arisen due to horizontal pleiotropy because a shared causal variant for two traits is necessary, though not sufficient, for them to be causally related [140]. Despite this, horizontal pleiotropy remains a concern for QTL studies due to instruments generally consisting of single SNPs. Another limitation of these analyses was that MR provides effect estimates for lifetime exposure to gene expression whereas expression levels of genes changes frequently. Overall, whilst the results presented in this Chapter were consistent across the sensitivity analyses, there remains a possibility that they were biased through horizontal pleiotropy.

4.5 Summary

This Chapter has demonstrated the effectiveness of MR and colocalisation to identify putatively causal genes for glioma susceptibility. Analyses revealed causal evidence for three novel genes (*RETREG2*, *FAM178B* and *MVB12B*) whose genetically predicted expression associated with glioma risk. Follow-up analyses also showed that there was no large difference between germline genetic variants associated with gene expression and glioma subtype risk. The brain and blood tissue-specific analyses suggested that the causal estimates for glioma are different based on whether the gene is expressed in brain or blood tissue. Finally, the tissue-specific analyses highlighted five candidate tissues (cerebellum, cortex, and the putamen, caudate and nucleus accumbens basal ganglia) and four genes (*JAK1*, *STMN3*, *PICK1* and *EGFR*) which had causal evidence for affecting glioma risk.

CHAPTER 5

INTEGRATING MULTI-OMICS DATA TO IDENTIFY POTENTIAL CHEMOPREVENTIVE TARGETS FOR GLIOMA

The work presented in the previous Chapter analysed expression quantitative trait loci (eQTL) data to determine how germline genetic variation in gene expression levels affect glioma risk. The work presented in this Chapter builds on the previous analyses by combining the results from the eQTL analyses with other molecular data, namely protein QTLs (pQTL), and with evidence from drug target identification and prioritisation platforms to determine potential chemoprevention targets for glioma.

5.1 Introduction

Drug development remains an incredibly difficult and expensive undertaking despite increased investment from pharmaceutical companies. To demonstrate this point, a recent 2020 study by Wouters, *et al.* estimated the median cost to bring a modern drug to market was US $985 million [388]. Furthermore, the entire process of bringing a drug to the clinic – from identifying a target to gaining approval from a governing body – is beset by a high failure rate: as much as 90% of drugs fail during clinical development, particularly during phase II and III clinical trials to determine efficacy in human subjects [148, 185, 261, 275]. Therefore, pharmaceutical companies and academic institutions, which are increasingly involved in drug discovery and clinical trials [342], are investing in new methodologies to identify and prioritise potentially successful targets. For example, leveraging human genetics to identify genetically supported target-indication pairs, i.e. a drug which will target a given gene, has been estimated to double the success rate of drugs in clinical development [232]. Therefore, 'omics data – such as transcriptomics and proteomics – are increasingly utilised to provide evidence of a target to mitigate high failure rates.

In regards to glioma, the drug treatment regimen has remained relatively static since temozolomide (TMZ) was approved through the early-to-mid 2000s [105], though there are drugs used off-label to treat gliomas, particularly glioblastoma (GBM), such as bevacizumab and afatinib (described in Section 1.7.3 - *Chemotherapy*). Inherent challenges related to treating glioma means drug discovery and re-purposing efforts are often difficult and unsuccessful. Such challenges involve crossing or bypassing the selective membrane called the blood-brain barrier (BBB) and the blood-brain tumour barrier (BBTB), both of which mean drug delivery is frequently difficult [351], and the highly heterogeneous nature of high-grade glioma, both intra- and inter-tumourally.

Analyses presented in this Chapter leveraged human genetic data and analysed these within a combined Mendelian randomisation (MR) and colocalisation framework. Germline genetic variants associated with protein abundance levels were used to i) strengthen causal inferences for protein levels influencing glioma risk, and ii) inform potential drugs targets and identify drugs which may be re-purposed for glioma chemoprevention by combining sources of evidence with orthogonal sources of bias. These analyses could form the early part of a drug discovery pipeline with the aim of translation to glioma prevention and treatment and, eventually, improved patient outcomes.

5.2 Methods

5.2.1 Data

The analysis plan for this Chapter closely followed the analysis plan used in the previous eQTL Chapter. Summary-level statistics for pQTL data were obtained from two studies: the first was a pooled analysis of five plasma pQTL studies conducted by Zheng, *et al.* [400] and contained data for 2,113 pQTLs associated with 1,699 proteins with an effective sample size of 3,301; the second dataset was a publicly available brain-based pQTL resource as described by Robins, *et al.* [289] and contained data for 7,901 proteins available in a sample size of 144. These data were cleaned, subjected to quality control protocols and analysed separately, without meta-analysis, to retain the ability to separately examine how risk for glioma compared across tissue type (i.e. whole blood or brain). These datasets were described in detail in Section 3.4.4 - *Whole Blood pQTL Pooled Analysis* and Section 3.4.5 - *Brain pQTL Dataset*. Before use in the analysis, pQTLs were designated as either *cis*- or *trans*-acting depending on the position of the pQTL relative to the protein-coding gene region (inside and outside a 500Kbp window, respectively). Finally, the full summary-level data for glioma, consisting of 12,496 cases and 18,190 controls, were eligible for use in this analysis. This also included a breakdown of 6,191 GBM and 5,819 non-GBM cases.

5.2.2 Instrument Selection

The approximate F-statistic was calculated for each SNP, which has been found to be a good approximation for the F-statistic in single-SNP instruments [37]. This is a measure of how liable

a SNP is to induce weak-instrument bias and was described and derived by Bowden, *et al.* [37]. This was calculated using

$$F \approx \frac{\beta^2}{SE^2} \tag{5.1}$$

where β is the SNP-protein association and SE is the standard error for the SNP-protein association. Bowden, *et al.* recommended a threshold for inclusion at $F \geq 10$; however, a threshold of $F \geq 15$ was specified instead due to utilising an approximation. SNPs were also removed if in high linkage disequilibrium (LD) to ensure the selection of independent SNPs. This was achieved using the MR-Base R package function "clump_data" with a clumping $r^2 = 0.01$ [142].

pQTLs were excluded from the plasma pQTL dataset based on the categorisations as described in Zheng, *et al.* [400] – specifically, SNPs designated as "tier 3" were excluded from analyses; these were SNPs which were deemed inappropriate for use in an MR analysis due to increased liability to horizontal pleiotropy because these pQTLs were found to instrument for more than five proteins. The tiers derived by the authors of the original pooled analysis were described in detail in Section 3.4.4 - *Whole Blood pQTL Pooled Analysis*. Furthermore, data from the brain pQTL dataset were obtained from a paper by Kibinge, *et al.*, where the authors provided a pre-cleaned and ready-for-MR data download [173]. The steps they undertook to clean the data were similar to that described above, including LD clumping and selection based on the F-statistic – though these steps were independently repeated for this study as validation of their quality control and cleaning process.

All SNPs which passed the instrument selection protocols described thus far were eligible for inclusion into the MR analysis. SNPs were categorised into either *cis*– or *trans*–acting SNPs (inside and outside a 500Kbp window of the protein-coding region, respectively) to limit the influence of horizontal pleiotropy in the MR analysis. A *cis*–pQTL-only MR analysis was first conducted followed by a complementary MR analysis consisting of all SNPs. The rationale behind excluding *trans*–acting pQTLs in the first MR analysis was similar for eQTLs in the previous Chapter, whereby *trans* SNPs are more prone to be susceptible to horizontal pleiotropy due to their distal nature on the genome. However, their inclusion can be beneficial where power is low because the pQTL would have more than one associated SNP and would allow for further heterogeneity sensitivity analyses. Furthermore, by including *trans* SNPs, the variance explained in the protein levels would increase and thus too would the power of the MR analysis. Unlike in the previous Chapter, there was also less concern about the burden of correction for multiple testing when there was a much smaller number of proteins which were potentially instrumentable compared to genes in the previous datasets. However, by including *trans*–acting SNPs, this exposes the analyses to more potential for pleiotropy which, even with including *trans*–acting SNPs, was not possible to interrogate with robust MR methods, e.g. MR-Egger. Therefore, care needed to be taken in interpreting the results from these analyses. There were no *trans*–acting SNPs reported in the brain pQTL dataset, therefore including *trans*–acting SNPs only applies to the plasma pQTL dataset.

The initial MR analysis investigating the effect of independent, *cis*–only plasma and brain proteins and glioma risk consisted of 730 tests. Therefore, a Bonferroni-corrected P value threshold of 6.35×10^{-5} (0.05/730) was specified for MR analyses conducted in this Chapter. A heuristic threshold for suggestive evidence was set at $P < 1.00\times10^{-3}$. This threshold was justified for two reasons: i) to enable follow-up analyses for results obtained in an otherwise likely low powered dataset by comparing to other findings with different sources of bias, thus reducing the overall likelihood of erroneous inferences; and ii) the SNP selection process was rigorous to ensure only relatively high-powered SNPs robustly associated with protein abundance levels were eligible for inclusion as instruments.

5.2.3 Identifying the Influence of the Plasma and Brain Proteome on Subtype Risk

The MR analyses were conducted using the MR-Base R package [142] to investigate how genetically predicted plasma and brain protein levels affected risk of glioma, and GBM and non-GBM subtypes. Proteins that were instrumented by single SNPs were analysed using the Wald ratio (WR) and likewise proteins instrumented by multi-SNP pQTLs by the inverse variance weighted (IVW) method. These methodologies weres described in detail in Section 2.3.3 - *Mendelian Randomisation*.

The two pQTL datasets used different, unknown units of measurement for protein abundance levels, meaning that the SNP-protein associations and the MR causal estimates between the plasma and brain pQTL datasets could not be directly compared. Therefore, to plot and and present the results in the main body of this thesis, the pQTL results for the brain and blood datasets were scaled by one standard deviation (SD) of that dataset to allow for comparative presentations. The unscaled results are presented in Appendix D. However, given the unknown units of measurement, it was more informative to interpret the direction of effects, rather than magnitude of effects, for results in this Chapter.

Results from the MR analysis which met at least the suggestive P value threshold of $P < 1.00\times10^{-3}$ were eligible for inclusion into the colocalisation analysis. All SNPs associated with the protein of interest within a ±500Kbp window of the lead SNP used in the MR analyses were extracted for this analysis. Each region was conditioned upon the marginal associations in the region of the lead SNP using the Genome-wide Complex Trait Analysis Conditional and Joint analysis (GCTA-COJO, [391]) tool. Colocalisation analyses were conducted using the coloc R package [113]. The colocalisation methodology was explained in depth in Section 2.3.2 - *Statistical Colocalisation*. Standard parameters were used for these analyses.

The full summary statistics for the brain pQTL dataset were not available and so a formal colocalisation analysis could not be conducted for results from this dataset. Instead, an approximate colocalisation analysis was conducted which involved extracting SNPs from the region of interest from the glioma data. Each region was similarly subjected to a conditional analysis using

GCTA-COJO [391]. The LD of the pQTL and each of the conditionally independent SNPs from the outcome dataset were then systematically compared such that, if the pQTL and one of these SNPs showed evidence of high LD, then this was approximate evidence of colocalisation. This method was employed by Zheng, *et al.* in their plasma proteome analysis, where full summary statistics were not available for some datasets [400].

As a sensitivity analysis, all results from the MR analysis were subjected to Steiger filtering to test the correct orientation of the causal estimate from genetically predicted altered protein levels to glioma subtype risk. Briefly, Steiger filtering works by estimating the variance explained by the SNPs in the plasma proteins and the glioma datasets. If the variance is higher in the plasma proteins dataset than the glioma dataset, then the causal estimate is correctly orientated [141]. Steiger filtering determines whether the direction of the causal effect is correctly orientated from the exposure to the outcome dataset with a corresponding P value. If an association had a Steiger $P < 0.05$ and the direction of effect was correctly orientated then this result was categorised as "true"; if an association had a Steiger $P < 0.05$ and the direction of effect was incorrectly orientated then this result was categorised as "false"; finally, if an association had a Steiger $P \geq 0.05$ then this result was categorised as "uncertain".

5.2.4 Correlation with Transcriptomics Results

To explore whether there was correlation between causal estimates for genetically predicted protein and gene expression levels, all MR associations from the current Chapter and the previous Chapter were systematically linked based on gene and that gene's protein product. Firstly, correlations for the causal estimates for all QTLs were calculated based on molecular trait (protein levels or gene expression), tissue of derivation (blood or brain) and subtype diagnosis, resulting in four datasets per subtype to compare. Pearson correlation coefficients were calculated using the "cor" R command and displayed as a heat map. Secondly, causal estimates for QTLs with an MR $P < 0.05$ were plotted with correlations similarly calculated for each subtype using the "cor" R command.

5.2.5 Triangulation of Evidence to Build the Case of a Drug Target

Many approved targets for drugs are proteins due to the key role they play in many biological mechanisms [149]. This analysis investigated causal evidence for variance in protein abundance levels influencing glioma risk. However, these analyses neither prove causality nor prove the use of a protein as a target that could reduce glioma occurrence if correctly agonised or antagonised. Nonetheless, results from these analyses could aid in the drug target identification and prioritisation process by highlighting proteins of interest for follow-up *in vivo* or clinical studies. Therefore, to provide greater confidence in the results presented here, triangulation of evidence was integrated into the study design. This is a concept that uses multiple sources of data to

Table 5.1: Sources of evidence that the OT Platform uses to calculate an association score for a drug target and a disease. The OT website gives a detailed breakdown on how this evidence is synthesised and weighted to form the overall association score (`https://docs.targetvalidation.org/getting-started/scoring`, accessed 2020-10).

Data Type	**Data sources and factors that affect the relative strength of the evidence scores**
Genetic Associations	ClinVar; GWAS and PheWASCatalog; Genomics England PanelApp; UniProt; Gene2Phenotype; ClinGen.
Somatic Mutations	Cancer Gene Census; ClinVar somatic; IntOgen; UniProt somatic.
Known Drugs	ChEMBL.
Affected Pathways	Reactome; Sysbio; SLAPenrich; PROGENy; CRISPR evidence.
RNA Expression	Expression Atlas score.
Text Mining	Europe PMC.
Animal Models	PhenoDigm.

answer the same question [197]. These data sources had differing sources of bias and increased confidence in a result when they all supported the same inference.

Two open access, publicly available sources of information were used to enhance the results from the MR and colocalisation analyses. The first, Open Targets Platform (OT, `http://targetvalidation.org`) is a platform born of public-private collaboration that aims to synthesise evidence from myriad sources to inform drug target discovery efforts [52]. OT provides an association score for each target on their platform. This association score summarises the evidence for the use of a target on the disease of interest and ranges from 0, meaning no evidence, to 1, meaning substantial evidence. An overview of these evidence sources is given in Table 5.1.

The second data source that was used is the Drug-Genome Interaction database (DGIdb, `https://www.dgidb.org/`), which collates evidence related to the druggable genome – meaning genes or their products that are predicted to influence disease outcomes – and whether the target is clinically actionable – that is, the target is used to inform clinical actions, such as with *MGMT* promoter methylation informing TMZ treatment in glioma patients [63]. DGIdb also provides a look-up function for targets, which have either known or suspected associated drugs, including those undergoing clinical trials. Whilst DGIdb and OT do contain overlapping data (i.e., from text mining), specifically non-overlapping, complementary sources of information were used to inform on the potential use of a protein/gene as a druggable target by including only data on the druggable genome and for clinically actionable targets from DGIdb, and text mining information from the OT platform (weighted as part of the association score).

Each of the main results from both this Chapter and the previous transcriptomics Chapter were systematically linked to evidence from the OT Platform and DGIdb. For potentially viable drugs outlined by DGIdb (i.e., existing drugs that are available for the target of interest), a

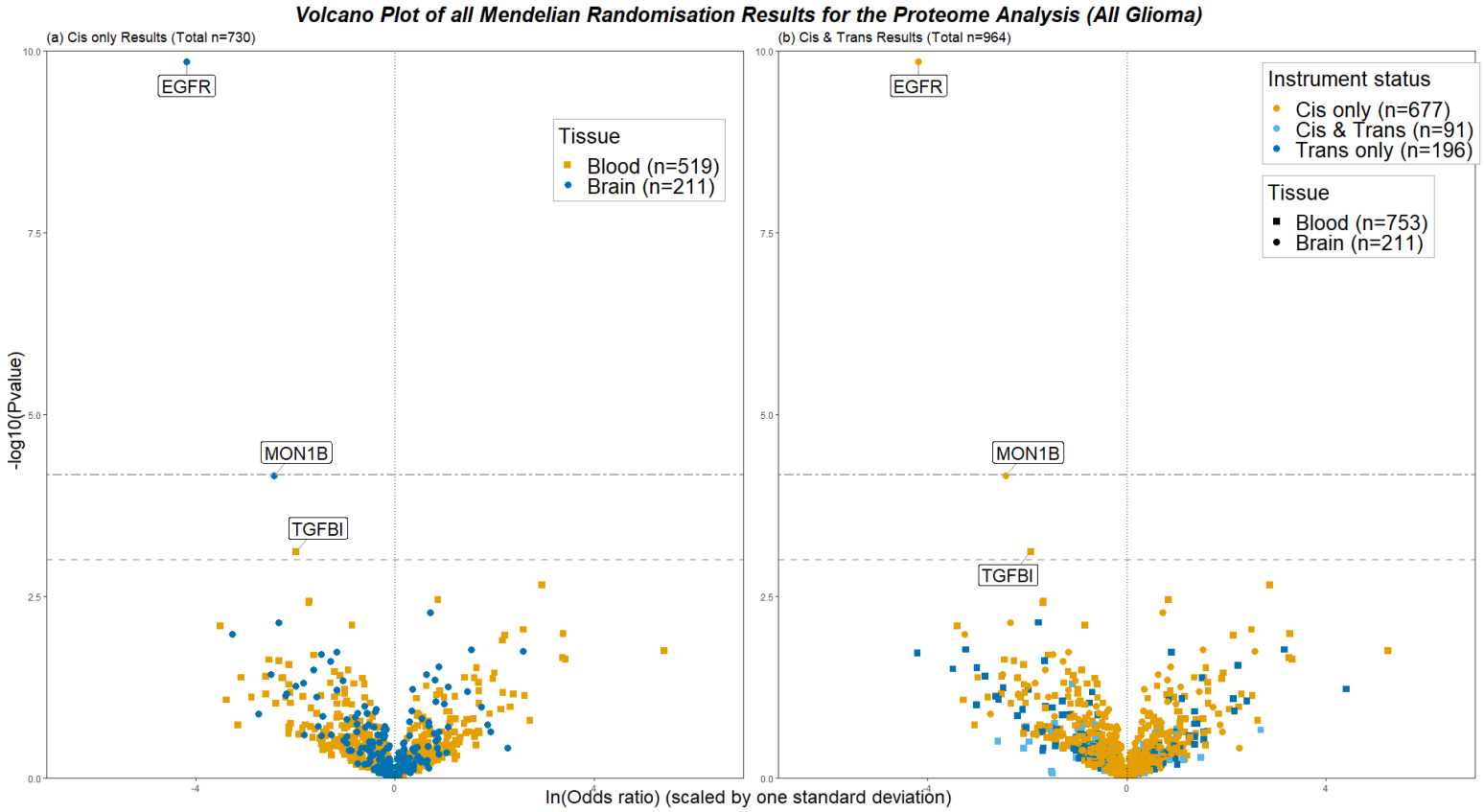
non-exhaustive literature search was performed to determine how these drugs currently relate to glioma therapies. Due to a lack of preventative therapies available for glioma, given the nature of the disease, the literature search included drugs which are used post-diagnosis for treatment. It does not necessarily follow that molecular traits associated with risk are also associated with progression or prognosis, and therefore drug targets identified by these analyses cannot be used to inform on such treatment strategies without further follow-up studies. This limitation is discussed more in Section 5.4 - *Discussion*.

5.3 Results

5.3.1 Statistical Results

MR analyses were used to estimate the association between 730 proteins, instrumented by *cis* pQTLs, and all glioma risk. Associations for two proteins instrumented in brain tissue met the Bonferroni-corrected P value threshold and one instrumented in blood tissue met at least the suggestive P value threshold (Figure 5.1 (a)). Inclusion of *trans* SNPs allowed for analysis of 196 more proteins and increased power for 91 proteins all from the plasma pQTL dataset. Despite this, no other associations met either the stringent or suggestive P value threshold. These results are shown in Figure 5.1 (b). When considering subtype risk for GBM (6,191 cases) and non-GBM (5,819 cases), and using both *cis*-only and *cis* and *trans* instruments, one association met the Bonferroni-corrected P value threshold and two associations met the suggestive P value threshold, all instrumented in brain tissue. These results are shown in Table 5.2. The unabridged and unscaled MR results are given the Online Appendix, Table S5-6 (see Appendix K).

Figure 5.1: Volcano plot of all results from the MR analysis estimating the causal effects of genetically predicted protein levels and all glioma risk using *cis* and *trans* SNPs. The horizontal dashed lines shows the *P* value thresholds. Points are shaped and coloured depending on the tissue type and if the instrument was constructed with *cis* or *trans* only SNPs, or both. MR ln(OR) were scaled by one SD of the MR results within each dataset, such that the results from the two datasets could be plotted on the same scale.



Only one result was subjected to the formal, statistical colocalisation analysis, which was for TGFBI, and revealed high evidence of colocalisation between altered protein abundance levels and all glioma risk (82%). Furthermore, the Steiger filtering sensitivity analysis identified the direction of effect was correctly orientated (plasma protein abundance affects glioma risk and not vice versa) (Steiger $P = 8.69 \times 10^{-64}$). The remaining five results from brain tissue were subjected to the approximate colocalisation analysis, with one result showing high ($r^2 > 0.80$) LD between the pQTL and the outcome dataset (MON1B and all glioma, $r^2 = 0.83$). The remaining four results from the brain pQTL dataset were in moderate LD with the subtype dataset ($r^2 = 0.40$ to 0.83). For the brain pQTL results, the Steiger filtering analysis showed an uncertain result ($P > 0.66$), hence little evidence for the direction of effect between these proteins and glioma risk, except in the case of TAGLN (Steiger $P = 0.02$), where evidence of the correct orientation was stronger.

Correlation analyses were conducted to determine how the causal estimates for QTLs compared across tissue types and glioma subtypes. Firstly, the correlation for all QTLs were calculated, without a selection threshold and found that there was a general weak-to-moderate correlation between each of the four QTL datasets ($-0.38 \leq$ Pearson correlation ≤ 0.15, Figure 5.2). Secondly, MR results for QTLs were subjected to a selection threshold of $P < 0.05$. There was a strong correlation for non-GBM results (Pearson correlation = 0.94), a moderate correlation for all glioma (Pearson correlation = 0.67) and a weak correlation for GBM (Pearson correlation = 0.35). The number of results which passed the selection threshold of $P < 0.05$ was quite low overall: five for all glioma, eight for GBM and 11 for non-GBM. These results are shown in Figure 5.3.

Table 5.2: Main results from the MR and colocalisation analysis. Odds ratios (OR) and confidence intervals (CI) are given per SD increase in the protein abundance levels. The "coloc" column shows the results from the colocalisation analyses. A statistical colocalisation analysis was conducted if the number is followed by a % sign, which references the H_4 coloc model [113]. If the number is not followed by a % sign, then the approximate colocalisation analysis technique was used instead. This number quantified the linkage disequilibrium (r^2) between the pQTL and the lead SNP in the outcome dataset. The higher the number, the more evidence of LD and therefore potential that the two datasets would colocalise in this region. The highest r^2 was shown as the likeliest candidate for colocalisation. Steiger results were categorised for ease of reading: "uncertain" means the Steiger P did not met the threshold of 0.05 and "true" means the direction of effect was correctly orientated from protein abundance to glioma risk and Steiger $P < 0.05$. An expanded version of this table is given in Appendix D.

Protein	**SNP**	**Tissue**	**Subtype**	**OR (95% CI) per SD**	**P value**	**Coloc**	**Steiger**
EGFR	rs138154852	Brain	GBM	3.02×10^{-3} (6.35×10^{-4}, 1.33×10^{-2})	3.07×10^{-13}	0.50 [a]	Uncertain
			All glioma	0.02 (4.24×10^{-3}, 0.06)	1.43×10^{-10}	0.50 [a]	Uncertain
MON1B	rs111987476	Brain	All glioma	0.09 (0.03, 0.29)	6.96×10^{-5}	0.76 [a]	Uncertain
			GBM	0.08 (0.02, 0.36)	9.21×10^{-4}	0.83 [a]	Uncertain
TAGLN	rs79627789	Brain	Non-GBM	2.19 (1.38, 3.50)	9.53×10^{-4}	0.40 [a]	True
TGFBI	rs13159365	Blood	All glioma	0.14 (0.04, 0.44)	7.70×10^{-4}	82%	True

[a] Approximate colocalisation was used to derive these results.

Figure 5.2: Heatmap showing the correlation between the MR associations for the different QTL analyses on glioma subtype risk. Heatmaps were split into subtype shown on the x-axis. The e- and p- suffix denotes expression and protein QTLs, respectively. eQTL results were taken from the work discussed in Chapter 4. The number of included QTLs is also given (n).

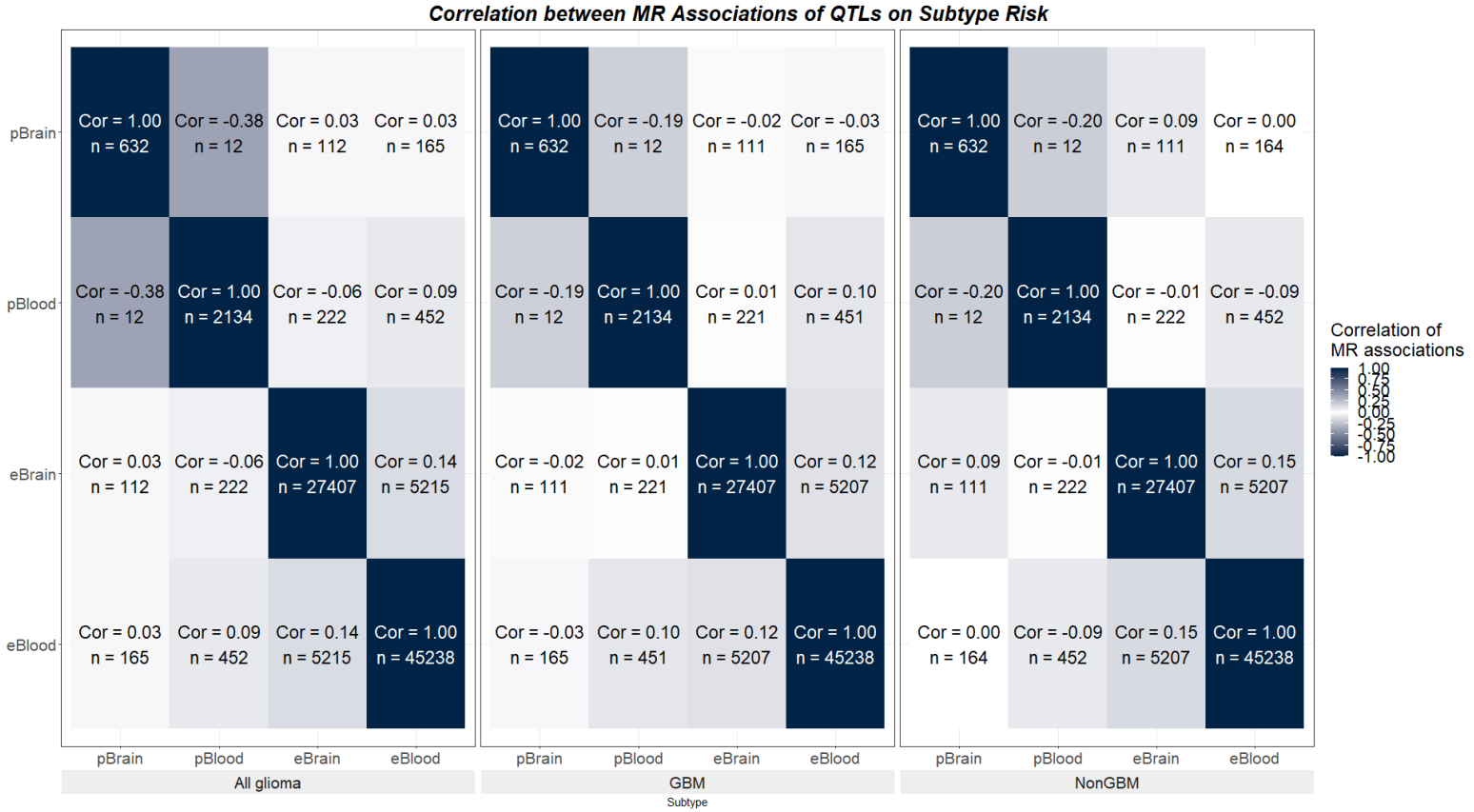
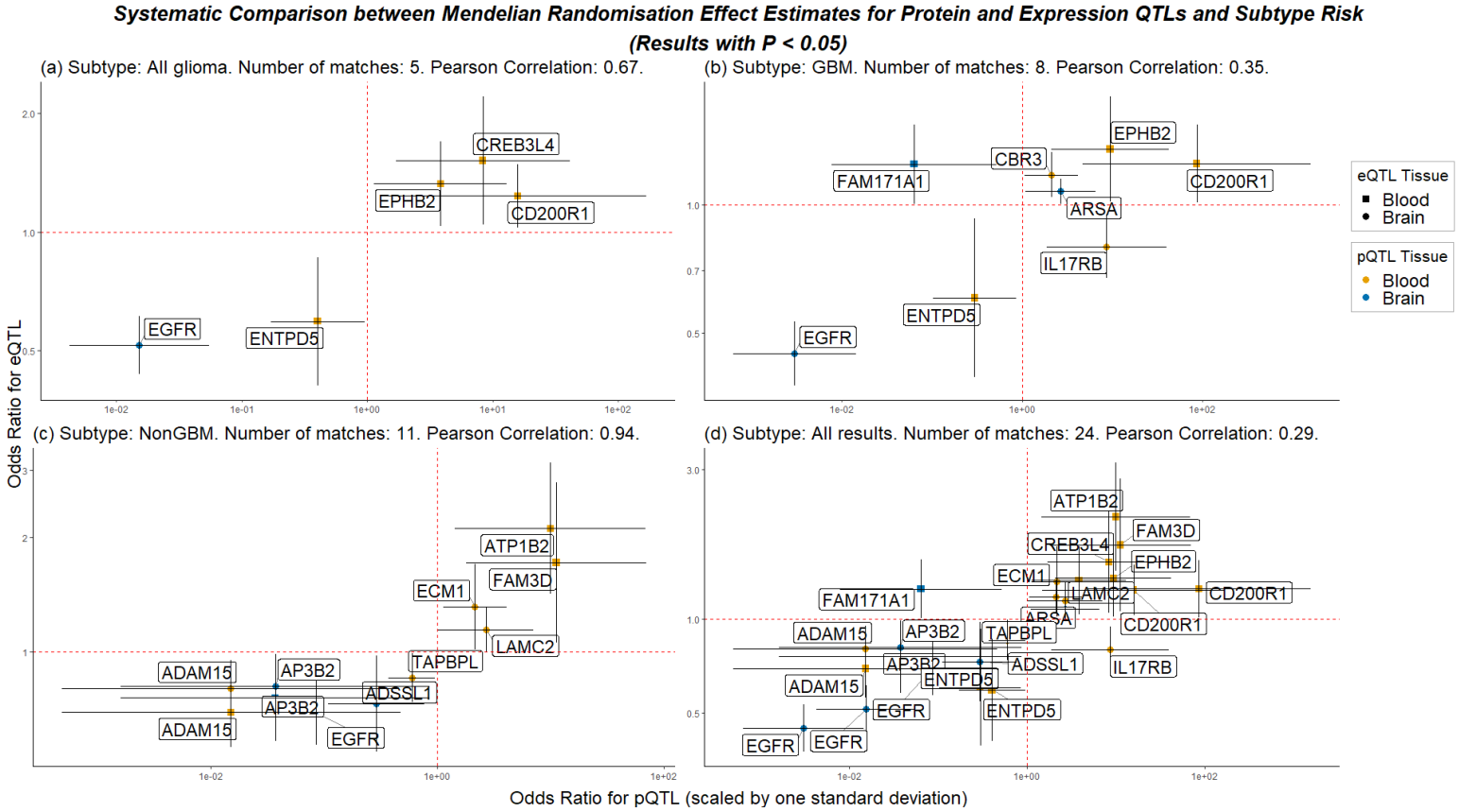


Figure 5.3: Graph showing the systematic comparison of the MR associations between pQTLs and eQTLs from both brain and blood tissues. Breakdown is given for subtypes: (a) all glioma, (b) GBM, (c) non-GBM and (d) all results agnostic of subtype. Results are shown if the MR $P < 0.05$.



5.3.2 Triangulation of Evidence

The main results from this and the previous Chapters were systematically linked to evidence from the OT Platform and DGIdb to highlight putative targets for future studies. Although the majority of the main results were mostly genes highlighted in the transcriptomics Chapter, hereon protein products are generally referred to (in a non-italicised font face, as standard) within the context of identification of drug targets, as proteins are more likely to be the actionable target for a drug.

Evidence from the MR analyses thus far were presented in Table 5.3, which showed how a number of potential targets may be potentially used as drug targets for glioma prevention. In total, three targets had an OT association score of 1.00, which were EGFR, JAK1 and TP53. Six targets had druggable genome evidence, whereby a gene or its product is known or predicted to have an interaction with at least one drug [63]. This included EGFR, JAK1 and TP53 as well as ABCB6, PRLR and TGFBI. Finally, three targets had evidence of being used to inform clinical decisions: EGFR, JAK1 and TP53. EGFR and TP53 are well-known somatic mutations and histological biomarkers in glioma biology (as described in Section 1.6.4 - *TP53 Mutation* and Section 1.6.6 - *EGFR Amplification*), though germline mutations for both have been implicated in glioma biology as well [112, 189, 220]. JAK1 has also been previously implicated in glioma biology due to its part in the janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway [41, 220].

A search on DGIdb provided evidence for existing drugs which have a documented interaction with any of the genes or proteins that form the main results, and either have been linked to glioma previously (e.g., through clinical trials) or may affect glioma through suspected pathways. The following subsections detail a break-down for each potential target. All of these drugs were implicated or studied in the context of glioma treatment, and not prevention which the MR analyses presented above have investigated. This is important to note and discussed further in Section 5.4 - *Discussion*.

5.3.2.1 EGFR-targeting Drugs

There are myriad drugs that either target or affect EGFR in some way; DGIdb lists 181 potential drug-protein interactions as of October 2020. Listed below are some of these drugs, which show promise for glioma.

5.3.2.1.1 Afatinib Afatinib is an anti-neoplastic, kinase inhibitor used primarily in treating non-small cell lung cancer by targeting mutations in *EGFR* and *HER2*. A 2019 *in vivo* study found that afatinib in combination with temozolomide (TMZ) inhibited GBM tumour growth more than TMZ monotherapy [354]. A completed phase II trial (NCT00875433) and phase I/II trial (NCT00727506) conducted by Boehringer Ingelheim reported that, whilst afatinib was tolerable in recurrent GBM patients, efficacy was limited compared to TMZ-only treatment, particularly

Table 5.3: Summary of evidence from the MR analyses linked to the OT Platform and DGIdb. All OT association scores are in relation to glioma, except those marked with *, where the highest association score was for GBM. Effect on risk is given with respect to increased gene expression or protein abundance levels. Evidence of the "druggable genome" is from Hopkins, *et al.* [149]. Evidence for "clinically actionable" targets is from Rhodes, *et al.* [285].

Target	**Gene exp. evidence**	**Protein evidence**	**Effect on risk**	**OT score**	**Druggable genome**	**Clinically actionable**
ABCB6	Brain	No	Decrease	0.07	Yes	No
BAIAP2L2	Brain	No	Decrease	0.27 *	No	No
EGFR	Brain	Brain	Decrease	1.00	Yes	Yes
FAM178B	Brain	No	Increase	0.00	No	No
JAK1	Blood & Brain	No	Increase	1.00	Yes	Yes
MVB12B	Brain	No	Increase	0.00	No	No
PANK4	Blood	No	Decrease	0.00	No	No
PICK1	Brain	No	Increase	0.04	No	No
PRLR	Brain	No	Decrease	0.03 *	Yes	No
RETREG2	Brain	No	Decrease	0.00	No	No
STMN3	Brain	No	Decrease	0.18	No	No
TGFBI	No	Blood	Decrease	0.13 *	Yes	No
TP53	Blood	No	Decrease	1.00	Yes	Yes

in tumours without *EGFRvIII* mutations [282]. Another phase I trial determined that afatinib was tolerable for patients with newly diagnosed GBM when combined with radiotherapy and TMZ (NCT00977431) [305]. Finally, there is another phase I trial underway to determine the pharmacokinetics of afatinib by analysing cerebrospinal fluid and blood samples of patients with brain cancer (NCT02423525). Overall, afatinib appears to be safely tolerated by brain tumour patients, however, efficacy is less certain, likely due to the complexity and intra-tumoural heterogeneity involving *EGFR* mutations in brain tumours.

5.3.2.1.2 Osimertinib Another kinase inhibitor primarily used to treat non-small cell lung cancer with *EGFR* T790M mutations, osimertinib is a drug that is showing recent promise in the field of glioma treatment. One case study in 2019 presented a patient with *EGFR*-mutant, *IDH* wild-type GBM who undertook osimertinib as a monotherapy after standard treatment regimens; however, despite a complete response at one tumour site, another lesion progressed significantly, highlighting that tumour heterogeneity may play an important part in EGFR-targeting treatments [214]. A recent study showed that osimertinib suppressed GBM tumour proliferation in cell lines and also in patient-derived xenografts mice [56]. Furthermore, there is both human and murine evidence, albeit at an early stage, that osimbertinib can penetrate the blood-brain barrier and effectively treat brain tumours [15, 54, 206]. There are currently three clinical trials investigating osimertinib usage and brain tumours: the first is a phase I trial

investigating adjuvant osimertinib with stereotactic radiosurgery in non-small cell lung cancer metastatic brain tumours that are *EGFR*-mutant (NCT03535363); the second is a phase II trial to determine whether osimertinib reduces glucose utilisation of *EGFR*-active recurrent GBM tumours using fludeoxyglucose F-18 positron emission tomography (PET) (NCT03732352); finally, the third is a phase I trial investigating the safety profile of implantable microdevices in primary brain tumours using drugs commonly used in the clinic for treatment, including osimertinib (NCT04135807). Osimiterinib is still undergoing early clinical trials and thus it is still too early to determine whether the drug will be both safely tolerated by patients and prove an efficacious therapeutical option.

5.3.2.1.3 Erlotinib Erlotinib is another non-small cell lung cancer, tyrosine kinase inhibitor that targets EGFR. Similar to afatinib, erlotinib has been determined to be safe for brain tumour patients but efficacy has been limited due to resistance to the drug, likely mediated by *EGFRvIII* mutations [278, 315, 397]. Overall, erlotinib has shown poor efficacy in treating glioma despite relatively promising results from pre-clinical studies as noted in a review by Li, *et al.*[201].

5.3.2.1.4 Other EGFR-targeting Drugs Many EGFR-targeting drugs trialled for glioma treatment are re-purposed drugs used as a treatment option for non-small cell lung cancer, though some are also used in the treatment of colorectal and breast cancers and, whilst evidence has pointed to EGFR as a promising target in the treatment regimen for glioma, clinical results have not reflected the same predicted efficacy in pre-clinical trials when these drugs are used to treat glioma and other brain tumour patients. These include the aforementioned three drugs and others such as gefitinib (limited activity in high-grade glioma patients) [101], dacomitinib (limited activity as monotherapy for recurrent GBM with EGFR amplification) [321], cetuximab (limited activity taken post-surgery, radiotherapy and chemotherapy in patients with recurrent, progressive high-grade glioma) [233], lapatinib (not efficacious for treatment for recurrent GBM patients with and without concomitant enzyme-inducing anti-epileptic drugs) [345] and panitumumab (given to patients with malignant gliomas and combined with irinotecan, a chemotherapeutic agent used for treatment of colorectal cancer; the trial for this drug was terminated early due to lack of efficacy) (NCT01017653).

5.3.2.2 FAM178B-targeting Drugs

The family with sequence similarity 178 member B (FAM178B) protein has a yet-unknown function. RNA expression analyses have been shown *FAM178B* to be highly expressed in skeletal, male and brain tissues in the Genotype-Tissue Expression (GTEx) and Human Protein Atlas projects [121, 346]. It has also been phenotypically associated with bipolar disorder and lithium response [161]. For that reason, lithium has been associated as a FAM178B-targeting drug. Trials have already investigated whether lithium is efficacious in glioma treatment, likely due to its

ability to bypass the blood-brain barrier rather than due to any association with FAM178B. A pre-clinical study showed that lithium can inhibit invasion of glioma cells through inhibition of glycogen synthase kinase-3 (GSK-3), highlighting that pathway and lithium as avenues of further research [241]. Another pre-clinical study found that lithium enhances the effect of TMZ in TP53 wildtype GBM cells due to inhibition of GSK-3, which then induces glioma cell death through activation of nuclear factor of activated T cells (NFAT)1/Fas ligand (FasL) signalling [127]. Despite this evidence, efficacy has yet to be tested in a clinical setting; a phase II trial that sought to investigate whether TMZ, bevacizumab and lithium, combined with radiotherapy, could be used to treat high-grade glioma was terminated early due to lack of participants; however, preliminary results suggested the treatment regimen inhibited GBM invasion [230]. Lithium remains a drug of interest for glioma treatment, with its ability to inhibit GSK-3 and induce glioma cell death as the main mechanistic pathway of interest.

5.3.2.3 JAK1-targeting Drugs

Janus kinase 1 (JAK1) is a tyrosine kinase protein involved in the JAK-signal transducers and activators of transcription (STAT) signalling pathway and is essential in cytokine signal transduction (mainly interferon (IFN)-α/β and -γ). The JAK-STAT pathway plays a key role in many downstream biological functions with studies focusing on the STAT family of molecules, which have already been implicated as tumour suppressors (STAT1) or potent oncogenes due to aberrant activation (*STAT3*, *STAT5*) [340]. However, recently there has been interest in determining the effects of the JAK proteins on gliomagenesis and whether these may be targeted for treatment.

5.3.2.3.1 Ruxolitinib Ruxolitinib is a small molecule used to treat myelofibrosis by selectively inhibiting JAK1/JAK2. The drug has also been found to block downstream activation of the suspected oncogene *STAT3* in clinical trials relating to myelofibrosis [356]. An *in vivo* study found that ruxolitinib may be therapeutically relevant for glioma patients due to decreased invasion and tumourigenesis of GBM cells, possibly due to the interplay between the JAK-STAT pathway and IFN-α and -γ [74]. There is a phase I clinical trial currently underway to determine the safety of ruxolitinib when combined with the standard treatment regimen of brain cancer in grade III glioma and GBM patients (NCT03514069).

5.3.2.3.2 Other JAK1-targeting Drugs There exist other JAK1-targeting drugs however these are relatively new, with many trials and studies having been conducted within the last five to 10 years. Currently, there are drugs undergoing various trials to determine safety and efficacy in other diseases, for example, itacitinib and non-small cell lung cancer, upadacitinib and immune-related inflammatory diseases such as Crohn's disease and atopic dermatitis and peficitinib, baricitinib and filgotinib for treatment of rheumatoid arthritis. Currently, there

appears to be no trials or pre-clinical studies that have examined the effect of these drugs in relation to glioma or other brain cancers.

5.3.2.4 PRLR-targeting Drugs

Prolactin (PRL) and its related receptor, PRLR, play a key role in the immune system, are integral to lactation and are key upstream activators of the JAK-STAT pathway through the binding of PRL to the receptor resulting in the activation of JAK2 [277]. PRLR is associated with endocrine system disorders and has been previously investigated as a potential target for the treatment of breast and prostate cancers [251]. Whilst PRLR has been detected in GBM biopsies and implicated elsewhere in glioma biology, there have been few pre-clinical and no clinical trials investigating PRLR as a potential therapeutic option for brain cancer treatment. One such agent of interest would be LFA102, a humanised monoclonal antibody, which, although did show limited anti-tumoural activity in a phase I trial for patients with PRLR-positive metastatic breast and castration-resistant prostate cancers [2], may prove to be effective for PRLR-positive gliomas.

5.3.2.5 TP53-targeting Drugs

The tumour protein 53 (*TP53*) gene encodes the important tumour suppressor protein, p53, that responds to cellular stress and may induce cell death through apoptosis, DNA repair and other important cellular housekeeping functions. P53 is commonly deregulated in glioma and cancer in general and, whilst it is used in a diagnostic context for glioma, little has been made of p53 as a therapeutic target until recently, in part due to the complex nature of *TP53* mutations, the paradoxical nature of p53 – whereby treatment may induce p53-related toxicity in both normal and cancerous cells [12] – and the difficulty in treating for loss of function. However, there have been some recent advances in this area, particularly looking at gain of function treatments. One such product is APR-246 (PRIMA-1MET) a small molecule that induces apoptosis in p53-mutant tumour cells which has passed initial safety trials [76, 199]. An *in vivo* pre-clinical study showed that GBM cells were sensitive to APR-246, though the authors of the study highlight this may be due to *MGMT* targeting, as sensitivity was independent of *TP53* status [260]. There is also interest in agents that target upstream inhibitors of p53, namely the agent SAR405838 and the gene it targets, mouse double minute 2 (*MDM2*), which has shown to have an acceptable safety profile in phase I trials [71, 72], and may form a valid therapeutical target for downstream p53-related functions.

5.4 Discussion

The large-scale systematic MR analysis presented in this Chapter investigated the effect of the human plasma and brain proteome on glioma risk. Analyses identified variants associated with alteration of protein abundance for four proteins that affected glioma risk. Linkage with

results from previous transcriptomics analyses revealed novel insights into how the causal estimates between plasma and brain gene expression and protein abundance levels are correlated. Triangulation with drug target prioritisation and identification platforms strengthened the inference of these analyses and may inform further pre-clinical studies to determine potential novel targets for drug re-purposing.

Three of the four proteins that formed the results from this Chapter were instrumented by pQTLs in brain tissue. These were: EGFR, a well-known prognostic biomarker for glioma and many cancers; vascular fusion protein MON1 homolog B (MON1B), a potential prognostic biomarker for renal and endometrial cancer [346]; and transgelin (TAGLN), a protein involved in smooth muscle differentiation, which was initially a suspected tumour suppressor [9] though further studies have provided contradictory evidence that its expression may increase tumourigenesis [88]. Increased abundance of EGFR and MON1B were associated with decreased risk of all glioma and GBM, whilst increased abundance of TAGLN was associated with increased risk of non-GBM. The final result that reached the suggestive P value threshold was transforming growth factor beta induced (TGFBI), instrumented by a pQTL in blood, and was associated with reduced risk of all glioma (OR = 0.14 per SD; 95% CI: 0.04, 0.44 per SD). The TGFBI protein is induced by transforming growth factor beta (TGF-β), is involved in cell adhesion and plays a role corneal dystrophy. Abundance of the protein has also been implicated in glioma cell proliferation and migration [122] and overexpression of the gene has been observed within mesenchymal subtypes of high-grade gliomas, where survival times were markedly shorter than in patients with normal expression of *TGFBI*, highlighting the potential of the protein as a diagnostic and prognostic biomarker [254].

As the full summary-level statistics were not available for the brain tissue pQTL dataset, statistical colocalisation analyses were not possible for these results. Therefore, approximate colocalisation analyses were conducted instead for these results to gain insight into whether the variant of interest was the shared, causal variant driving the effect for both protein abundance levels and glioma subtype risk. Only the MON1B result had strong evidence of high LD, and thus provided suggestive evidence for colocalisation with the variant in the GBM dataset ($r^2 = 0.83$). The rest of the results from the brain pQTL dataset showed moderate evidence of LD with associated variants in the region in the subtype risk dataset of interest ($r^2 = 0.40$ to 0.76). The TGFBI result was eligible for the statistical colocalisation analysis and showed strong evidence of colocalising with all glioma ($H_4 = 82\%$). Whilst conclusions were limited for the brain pQTLs given the approximate colocalisation analysis, TGFBI did show strong evidence of colocalisation indicating further evidence for this protein being causally implicated in glioma risk and may be more reliable as colocalisation can discount some potentially unreliable associations [140].

Steiger filtering showed that the direction of effect was correctly orientated for each of the results, but four of the six results failed to meet the Steiger P threshold of 0.05. The results for TAGLN and TGFBI were both correctly orientated and met the Steiger P value threshold. The

four results for which the Steiger result was uncertain came from the brain pQTL dataset, which had a relatively low sample size (n = 144) compared to the glioma outcome dataset (12,496 cases and 18,190 controls). The reason for the uncertain result may be twofold: i) the smaller sample size of the brain pQTL dataset inherently meant that the statistical power was decreased and thus the estimation of the variance explained calculated by the Steiger filtering method may be lower than in reality; and ii) the authors of the pQTL dataset study used a non-standard unit measurement for protein abundance levels which may cause an underestimation of the variance explained and thus influence the result of the analysis [141]. Altogether, it was unlikely that glioma onset affected germline variance of protein abundance levels in brain tissue; however, the dataset available was likely to be too underpowered to correctly interrogate this and these results should not be taken at face value.

The analysis which investigated the correlation between the protein abundance and gene expression results provided insight into how the causal estimates between these molecular data compared across tissue types and subtypes. The first analysis sought to determine whether MR results for QTLs were correlated when not subjected to a selection threshold based on P value. It appeared that, across all glioma subtypes, molecular data, and tissue types, results were only weakly correlated. This result implied that, given currently available data, no one dataset can sufficiently proxy for another (e.g., by using higher-powered eQTLs in brain to inform on protein abundance levels in brain). A P value threshold for selection ($P < 0.05$) was applied to determine how QTLs compared when the MR results are more reliable. This analysis showed higher correlations across each subtype diagnosis: all glioma (Pearson correlation = 0.67), GBM (Pearson correlation = 0.35) and non-GBM (Pearson correlation = 0.94), though there were only a few QTLs that matched across datasets according to tissue and molecular type (n = 5 to 11). All systematically linked QTLs from all glioma subtypes combined, showed a generally weak correlation (Pearson correlation = 0.28, n = 24) which implied that subtype diagnosis was important. For instance, GBM showed lower correlation likely due to two QTLs whose directions of effect differed across molecular type (associated with FAM171A1 and IL7RB). Overall, these results strengthened the conclusion from the previous Chapter, which was that blood QTLs cannot proxy for brain QTLs and vice versa. Furthermore, these analyses showed that eQTLs and pQTLs did not always necessarily correlate with one another in regard to glioma subtype risk, and that eQTLs and pQTLs, even in the same tissue, did not always correlate with one another.

Triangulation of evidence from drug target identification and prioritisation platforms highlighted six potentially interesting genes/proteins which had evidence of translatability: ABCB6, EGFR, JAK1, PRLR, TGFBI and TP53. These six targets formed part of the druggable genome – where either the gene or gene product form known targets for existing drugs – and EGFR, JAK1 and TP53 were considered to be clinically actionable – that is, these biomarkers may be used to determine clinical actions or outcomes. The rest of the QTL results did not have evidence of forming a drug target currently, but this does not necessarily discount their importance in glioma

biology. Likewise, there are well-known drug targets which did not appear in this analysis, not necessarily because of evidence of lack of an effect, but because they were not instrumented by the data used. For example, ipatasertib, an AKT pathway inhibitor, has shown to be effective in some cancers with PTEN loss [69, 176]. As discussed further on, results presented in this Chapter should not lead to the de-prioritisation of drugs or targets, like ipatasertib, over others. Instead, evidence presented herein highlights those targets as interesting candidates for follow up. These analyses have recapitulated important biomarkers and targets which are well-known and well-studied in relation to glioma, lending credence to the methodologies and their results herein. These biomarkers were causally implicated in glioma risk and will make for important candidates for follow up studies, either within the same framework with larger sample sizes, which will allow for more robust inferences, or for experimental, lab-based studies that could determine potential functionality or pathways of interest.

Finally, a non-exhaustive literature search was conducted to link potential drug targets to evidence for existing drugs at any stage of development. In total, five genes or proteins had such evidence. These targets were also investigated within the context of glioma treatment previously. Each of these drugs were postulated to improve prognosis of glioma and generally showed acceptable safety profiles but efficacy was almost uniformly lower than expected. Many of the drugs did not progress past phase II trials despite strong evidence from pre-clinical studies in some cases. This could be due to the heterogeneous nature of glioma, both intra- and inter-tumourally, and has been posited as the main reason for the failure for many EGFR-targeting drugs to meet expectations [31, 95, 360]. Whilst it is unlikely that data will be granular enough to investigate such questions within an MR framework, it can at least guide drug target identification and prioritisation by highlighting important molecular markers that can guide future research.

Similar to the previous Chapter, using a stringent threshold for the MR results may have missed proteins which may be otherwise important in the physiology of glioma risk. However, unlike the eQTL datasets used in the previous Chapter, the pQTL datasets were much more likely to be underpowered due to smaller sample sizes. Therefore a lenient threshold was set at $P < 1 \times 10^{-3}$. Despite this, there were still only four proteins which had MR evidence of an effect on glioma risk. Combining evidence from different sources which point to the same conclusion can reduce the risk of spurious findings when using lenient thresholds; however, even with linkage to the previous Chapter's results and to evidence from the literature, the protein results should still be interpreted with caution due to low power and the potential of pleiotropy. Furthermore, the study from which the pQTL dataset was derived used a smaller sample size and with fewer proteins assayed than genes in the eQTL datasets. Therefore, it is unlikely that a stringent threshold missed potentially interesting results because of the underpowered nature of the dataset. Assaying more proteins in larger populations will help future studies to uncover putative causal relationships between the human proteome and complex traits using an MR framework.

Interpretation of the proteomic results should be done with caution. It is difficult to know how much emphasis should be placed on both the direction and magnitude of effect of MR studies using QTLs which appear to have discordant effects in different tissues or across QTL types. Until further studies are done at a more focused level to determine how best to interpret such results, evidence from MR should be integrated into existing pipelines to provide further evidence for the prioritisation of a putative drug target. Even should a target show discordant directions of effect, this can still be informative for physiological effects on complex traits [225]. However, results from MR should not necessarily form the only source of evidence for the prioritisation of a target and may not be informative for the de-prioritisation thereof, even when producing seemingly contradictory results. Altogether, more research is required in this area to aid interpretation of these results and how they can inform on clinical or pharmaceutical applications.

Strengths of the analyses presented in this Chapter included the use of germline genetic variants that were associated with protein abundance levels. This, combined with the MR methodology, reduced the influence of confounding and bias through reverse causation on causal estimates. Drug targets with genetic evidence are twice as likely to be approved and pass phase II and III clinical trials [98, 232], highlighting how MR and colocalisation can be used to improve drug discovery pipelines. Combining results with evidence from studies and databases with orthogonal sources of bias, namely from the OT Platform and DGIdb, helped strengthen inferences. Similarly, evidence collated from *in vivo* studies and clinical trials showed some of the highlighted targets had drugs already implicated in glioma treatment and had generally acceptable safety profiles.

This research was not without limitations, however. Small sample sizes, particularly within the brain pQTL dataset, limited robust causal inference and interrogative sensitivity analyses. Many proteins were instrumented by single SNPs, further limiting analyses to examine the presence of horizontal pleiotropy, though this is inherent for QTL studies generally. Inclusion of *trans*-SNPs for the blood pQTL dataset was intended to increase statistical power and potentially allow for such analyses but many proteins were still only instrumented by a singular SNP. Including *trans*-SNPs in this way may seem to unnecessarily increase the risk of pleiotropy in a dataset that is unable to deeply analyse its presence, but in their phenome-wide MR study of the plasma proteome on complex diseases, Zheng, *et al.* concluded that inclusion of trans SNPs increased the variance explained for associated proteins and, importantly for this research, identified potentially important therapeutic targets that were not instrumented by *cis*-SNPs [400]. Overall, all of the proteins in the main results for this Chapter were instrumented by single *cis* SNPs. There were also limitations present inherent to the methodologies used; MR, in particular, provides estimates for lifetime exposure to protein abundance levels whereas, in reality, protein levels will frequently change over much smaller time courses. Finally, care should be taken if directly interpreting the results of the pQTL analyses because the unit of measurement in the brain-derived pQTL dataset was unknown.

The MR analyses in this Chapter provided causal estimates for variance in protein abundance levels on glioma risk. The evidence collated as part of Section 5.3.2 - *Triangulation of Evidence* uniformly consisted of drugs and targets investigated within the context of glioma treatment. As stated previously in this Chapter, it does not necessarily follow that factors and targets for disease risk translate to factors and targets for disease progression. For the MR analyses to investigate this, a GWAS for progression would be required but data in this area are severely lacking [259]. Altogether, the supporting evidence from the literature search provided limited support for a putative drug target's efficacy for prevention of glioma. However, the collated evidence generally showed drugs whose pharmacodynamics targetted the highlighted proteins and were generally tolerable for glioma patients. Therefore, although evidence for efficacy was limited by different investigated outcomes (i.e., risk versus progression), the evidence did support: i) the drug as a potential candidate to prevent glioma, and ii) the drug as safe and tolerable for glioma patients.

5.5 Summary

This Chapter provided evidence that supported causal inferences for the effect of germline variants associated with protein abundance levels on glioma risk. Correlation of MR causal estimates were examined between gene expression and protein abundance QTLs and glioma subtype risk, and revealed that QTLs of one type did not inform on QTLs of another across either brain or blood tissues. Finally, evidence combined from drug identification and prioritisation platforms and databases highlighted genes and proteins as existing targets for glioma chemoprevention. These analyses showed that a combined MR-colocalisation framework can identify drug targets for glioma chemoprevention; and, with integration of more descriptive datasets and evidence, this framework could also identify novel targets for glioma or other diseases. Further *in vivo* or pharmaco-epidemiological studies can then interrogate these results before undertaking clinical trials to determine safety and efficacy in patients, providing much needed evidence to increase the success rates of such trials.

CHAPTER 6

Role of Treatments for Hyperlipidaemia and Diabetes in Risk and Mortality of Primary and Secondary Brain Tumours

This project was undertaken as a part of a pre-published piece of work [291], under review at the European Journal of Epidemiology. The original idea of the project was devised by Yoav Ben-Shlomo, Richard Martin, Kathreena Kurian and Maria Theresa Redaniel. Permissions to conduct the analysis were obtained by the aforementioned co-authors. Data were extracted from CPRD by Maria Theresa Redaniel and Martha Elwenspoek. My roles in this analysis consisted of cleaning and linkage of CPRD data, conducting all analyses and interpreting results, with supervision from Yoav Ben-Shlomo, Richard Martin, and Kathreena Kurian. I wrote the initial and follow-up drafts of the paper along with supervision from Yoav Ben-Shlomo, Richard Martin and Kathreena Kurian, which included producing all tables and figures.

6.1 Introduction

Drug companies are increasingly withdrawing from the development of novel drugs due to significant costs and high failure rates, which is especially true for cancer therapeutics [157]. However, re-purposing existing approved drugs for different diseases than originally designed provides significant advantages over how drugs are traditionally developed saving substantial time and costs involved in conducting new clinical trials. Furthermore, re-purposing drugs with approval for treatment in other diseases means that the safety profile of the agent is relatively well-known and results in more successful early clinical trials during phases I and II. As discussed in Section 1.7.3 - *Chemotherapy*, the treatment regime for glioma mostly consists of temozolomide (TMZ), a chemotherapeutic agent which was granted approval from the National

Institute of Health and Care Excellence (NICE) in 2001. There is a desperate need for more drugs and treatments that can be used to treat BT patients, with two areas of particularly urgent need. The first is to slow or stop progression to improve mortality rates once the patient has a BT. The second area is regarding preventing secondary and metastatic tumours, which are commonly found from primary sites such as breast, lung and skin (though others as well). Incidence of metastatic BT are increasing, due to a variety of reasons, e.g., better survival rates amongst primary cancers which metastasise to the brain [68]. Overall, any drugs which may be re-purposed to prevent incidence or improve mortality of BT would be important for patient outcomes.

Nuclear hormone receptors (NHR) have crucial roles in cellular homeostasis and have been implicated in the development of cancer. Peroxisome proliferator-activated receptors (PPARs) are a family of ligand-activated transcription factors that regulate the expression of genes and, as a subtype of NHRs, are involved in the control of proliferation and differentiation of cells [181, 271, 295]. This has marked PPARs as potential candidates in the treatment of cancer, with some studies showing beneficial outcomes in animal models and some early-stage human trials [18, 75, 228, 331]. There are three isoforms of PPARs which are designed as -α, -β/δ and -γ. The genes regulated by these transcription factors are also involved in the transport, metabolism and storage of fatty acids, inflammation and diabetes mellitus [183, 271, 353]. PPAR-α and -γ are of considerable clinical significance due to the existence of agonistic compounds that these bind to, namely hypolipidaemic fibrates and anti-diabetic thiazolidinediones (glitazones) respectively. Importantly for brain cancer, and other brain-related diseases, both drugs have been observed to cross the blood-brain barrier [80, 97].

PPAR-α expression has been seen to be enriched within the classical glioblastoma (GBM) subtype and constituted an independent prognostic marker for improved overall survival [137]. PPAR-γ is increased in human breast cancer, and ligand activation of this receptor results in a more differentiated and thus less malignant state of the disease [227] and reduced growth of colon cancer cells [306]. Pioglitazone, a PPAR-γ agonist, showed a dose-dependent reduction of glioma tumour invasion in murine glioma models when combined with 6-OH-11-O-hydroxyfenantrene (IIF) [257].

Human evidence is limited with a meta-analysis of 17 randomised placebo-controlled trials, and a separate phase II trial, suggesting that PPAR-α agonists may not succeed as anti-cancer agents in general [32, 293]. However, the first of these studies did not look directly at central nervous system (CNS) cancer and the second of these studies did not have a large cohort of CNS cancer cases ($n = 97$). A previous case control study found that diabetic GBM patients were less likely to be treated with a PPAR-γ agonist when compared to a hip fracture control population [119], though the study was potentially biased as thiazolidinedione use may be over-represented amongst people with fractures [85, 219].

Analyses presented in this Chapter explored whether there was evidence for the re-purposing

of two types of drugs, fibrates and glitazones, in reducing risk and in improving mortality rates of primary and secondary BT when compared with other drug treatments for hyperlipidaemia and type 2 diabetes. It is important to note that analyses in this Chapter use BT as outcome, and not solely gliomas, to increase the number of cases available. This is described further in Section 6.2 - *Methods*.

6.2 Methods

Two nested case-control studies and a case-only clinical cohort study were undertaken to examine the effects of exposure to fibrates and glitazones on BT risk and mortality. Fibrates are anti-hyperlipidaemic medications and glitazones are anti-type 2 diabetic medications. Data were obtained from the UK Clinical Practice Research Datalink (CPRD). CPRD is a primary care database with clinical information on over 11 million people from more than 670 UK GP practices [145]. The CPRD is a well-described and validated database of healthcare records, and representative of the general UK population [145, 146].

The study was approved by the Independent Scientific Advisory Committee (ISAC) for the UK Medicines and Healthcare Products Regulatory Agency (ISAC protocol number: 18_149R). The data were extracted from CPRD GOLD and linked to Office of National Statistics (ONS) death registration data and census data on area deprivation.

6.2.1 Participants

In total, there were 15,538,338 participants available in the CPRD Gold, August 2018 snapshot. Participants were eligible for inclusion into this study if they were 18 years or older and registered within the CPRD between 1 January 2000 to 31 December 2016, the former being the first year that glitazones were licensed within the UK. Follow-up was stopped when one of the following occurred: death; BT diagnosis; end of registration at a CPRD GP practice; or the end-date of the study. It was specified for inclusion that BT patients must have received their diagnosis after registration at a CPRD participating GP practice due to the possibility that their prescription history may be missing or incomplete.

All participants in the nested case-control studies had to be treated with either an anti-hyperlipidaemic or anti-diabetic medication before the end date of the study to reduce the likelihood of confounding by indication (explained further in Section 6.2.4 - *Exposures*), a type of confounding which can distort the association between exposure and outcome when an indication (in this studies, this is the reason to prescribe) is also associated with the exposure and outcome. Participants treated solely with insulin therapy were dropped from the analyses as the indication of interest was oral treatments for diabetes. Participants with diabetes who started on an oral medication but at some point received insulin therapy were still eligible for inclusion. Participants on combination therapy (for example, both glitazone plus another anti-diabetic medication) were

not excluded so long as their treatment included drugs of interest. Similarly, some people were exposed to both fibrates and glitazones, in which case these were considered exposed to both drugs in both case-control analyses. Sensitivity analysis also investigated the effects of monotherapy.

In the cohort study, all BT patients in CPRD GOLD were included, regardless of fibrate or glitazone drug exposure were included so long as they had a minimum of one year of follow-up of observation prior to censoring. This was to be analogous to a trial of newly diagnosed BT patients, some of whom are treated and some are not – however, this is a non-randomsed comparison and thus may be biased and limits causal inference.

6.2.2 Cases

Cases were defined as those patients who were diagnosed with a brain tumour (primary or secondary) using Read codes based on histopathological analysis according to World Health Organisation (WHO) guidelines. Appendix E lists Read codes and CPRD descriptions used to define cases. Secondary tumours were included as fibrates and glitazones may theoretically effect risk of other tumours which spread to the brain. Only incident brain tumour patients were included, i.e. first recorded diagnosis after registration at a GP practice, otherwise there is the possibility that their prescription history may be missing or incomplete. Finally, this definition of cases was used for both the nested case-control and cohort studies.

6.2.3 Controls

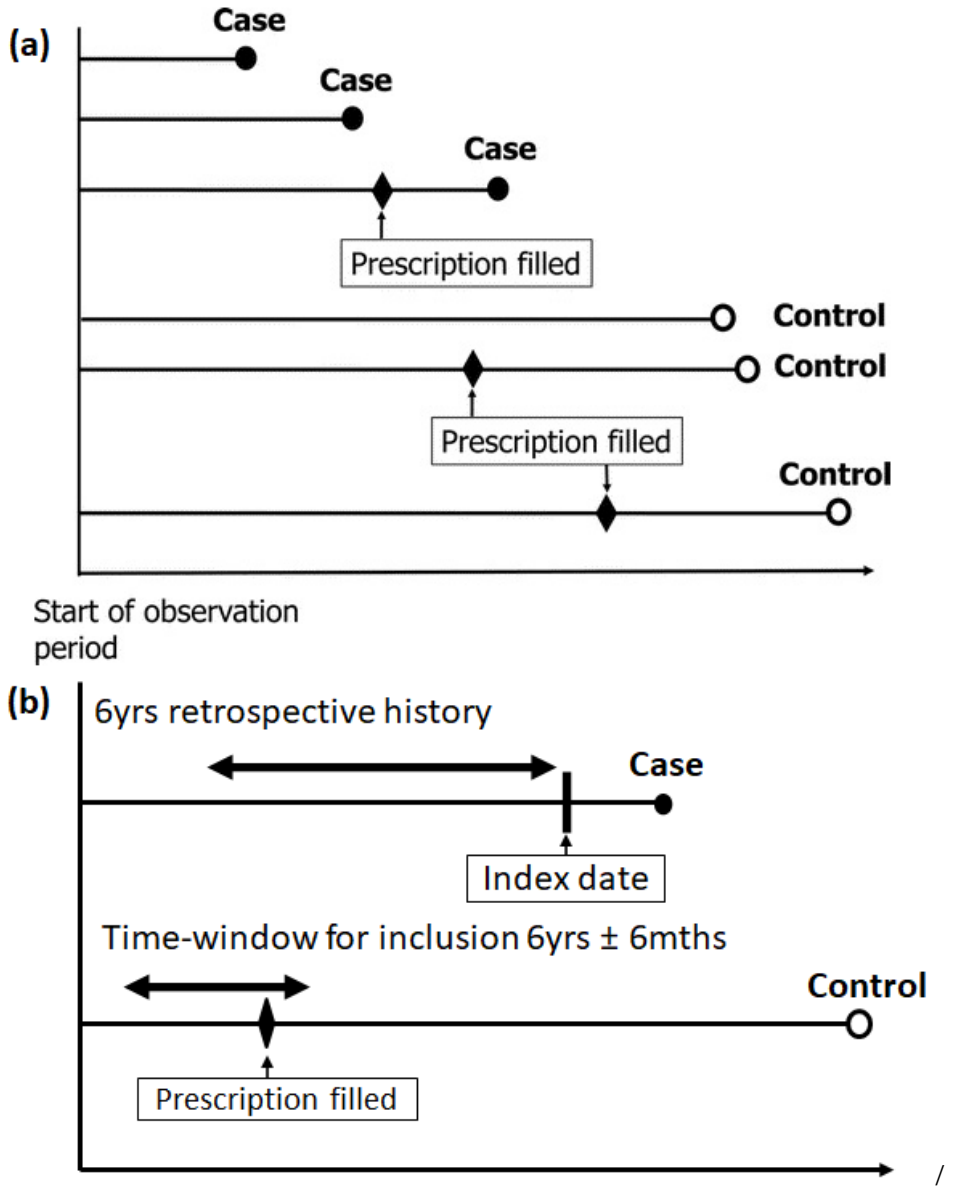
In the nested case-control studies, potential controls were identified and then randomly sampled such that up to four controls were selected per case using the following criteria. First, controls were defined as those participants without diagnosis of a brain tumour and contemporaneously registered within the CPRD. Controls were then subject to the same selection criteria as cases and were strata matched by age group (< 20, $20-29$, $30-39$, $40-49$, $50-59$, $60-69$, $70-79$ and ≥ 80) and sex.

Immortal time and time-window biases are common in pharmaco-epidemiology studies. Immortal time bias occurs when participants cannot experience the outcome during the study follow-up period and so appear immortal, and time-window bias occurs when different methods to select controls inherently means a longer potential exposure duration than cases, thereby inducing a spurious negative association between the exposure and outcome [337]. To address these biases, controls were required to have the same retrospective duration of potential exposure (any drug treatment for either hyperlipidaemia or diabetes) within the CPRD as cases, based on the case index date plus or minus six months. For example, if a case had 6.3 years of any retrospective drug history from their index date in 2013, then only controls who also had between 5.8 and 6.8 years of any drug exposure over the same secular time period were sampled. This meant that both cases and controls had the same potential for recorded exposure to any anti-hyperlipidaemic or anti-diabetic drug. This bias is illustrated in Figure 6.1. As baseline was

defined as the first date of exposure to any drug treatment for either hyperlipidaemia or diabetes, including first line treatments such as statins or metformin, immortal time bias should be avoided which would otherwise have been induced if follow-up started instead from the use of fibrates or glitazones (which are commonly second line therapies). Finally, incident users were sampled, and not prevalent users, as baseline was defined at first treatment in the nested case-control studies to mimic the design of a trial.

In the cohort study, all BT patients in CPRD GOLD were included from 2000-2016, regardless of fibrate or glitazone drug exposure, as long as they had a minimum of one year of follow-up of observation prior to censoring. The baseline for start of follow-up was set at date of diagnosis as recorded in CPRD. This analytical strategy is analogous to a trial of newly diagnosed BT patients, some of whom are treated and some who are not, though it is a non-randomised comparison and thus may be biased.

Figure 6.1: Portrayal of time-window bias [337]. The bias may be induced due to the inherently longer observation duration controls will have compared to cases. Cases will therefore have a shorter duration in which they may be exposed to the drug of interest compared to controls. This can lead to the incorrect conclusion of a protective effect when there is none. Allowing for a similar or the same time-window in which cases and controls have equal chances of being exposed can therefore account for this bias (see (b)). (a) is adapted from Suissa, *et al.* [337].

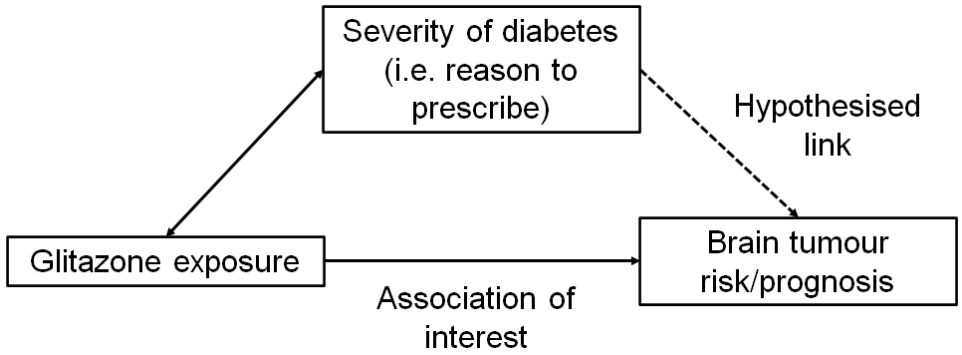


6.2.4 Exposures

Participants were classified as exposed if they received a fibrate or glitazone prescription for the treatment of hyperlipidaemia or type 2 diabetes and were compared to those who received another drug for the management of those diseases (classified as unexposed). Product codes used to determine exposure status are given in Appendix F & G.

As stated in Section 6.2.1 - *Participants*, it was specified all participants had to be prescribed *any* medication to treat hyperlipidaemia or type 2 diabetes, not just fibrates and glitazones. This was done with the purpose to reduce the risk of confounding by indication, whereby the indication of a drug (a pharmacological term to describe the reason to prescribe) can induce a distorted association between the exposure and outcome should it also be associated with both itself. Figure 6.2 shows an example of this confounding in the glitazone analyses. The way to reduce the risk of inducing confounding by indication is to include many different indications – if the association between exposure and outcome is consistent across indications then that association can be deemed unlikely to have arisen due to confounding by indication. By including different fibrates and glitazones with different indications this should go some way to account for presence of this type of confounding. For example, not just fenofibrate was included in the analyses, but also bezafibrate, ciprofibrate and gemfibrozil. However, despite steps taken to reduce the influence of this confounding on the results, it was not possible to completely eliminate the potential for confounding by indication.

Figure 6.2: Graph indicating how confounding by indication may be induced in these analyses, with an example given for the glitazone study. Severity of diabetes, i.e., the indication of glitazone prescription, is associated with exposure to glitazone and may also be associated with BT risk and prognosis [320]. Glitazone exposure is inherently be associated with severity of diabetes. As this study sought to understand the association between glitazone exposure and BT risk and prognosis, confounding by indication may be induced should the hypothesised link between severity of diabetes and BT risk/prognosis exist.



Three types of exposure variables were created, based on prescription data: i) a binary variable (yes/no) to indicate if the participant was ever exposed to a prescription for fibrates or glitazones as compared to any other drug for the management of hyperlipidaemia or type 2 diabetes. This variable was used in both the nested case-control and cohort studies; ii) the total uninterrupted

time each participant was prescribed either fibrates or glitazones before censoring was calculated (referred to as "longest exposure duration"). Each uninterrupted prescription duration (defined as no breaks of 90 days or more between prescriptions) was summed, and the longest of these durations was used as the longest exposure duration; iii) the total prescription duration was also calculated, with or without interruptions, referred to as "total exposure duration". Exposure duration for both longest and total exposure duration were categorised into unexposed and yearly categories: exposure ≤ 1 year, >1 and ≤ 2 years, >2 and ≤ 3 years, etc.. Due to small numbers of participants reaching longer exposure durations, any duration exposure periods over 6 years were considered in the longest duration category in the glitazone analysis, and over 7 years as the longest duration category in the fibrate analysis.

6.2.5 Confounders and Covariates

The following variables were considered as potential confounders that might influence both risk of developing or dying from a brain tumour, as well as potentially influencing the choice of drug agent that a doctor might prescribe: age, sex and socio-economic status (SES). An ecological proxy measure, the Index of Multiple Deprivation 2015 (IMD) [78], was used to proxy for SES. This is a commonly used measure in the UK that uses census data on a wide variety of economic and health factors to derive a postcode-based deprivation score so that that a higher score indicates less deprivation. IMD scores were grouped into five equal sized groups (quintiles) from most to least deprived areas, so that adjustment for non-linear associations was possible.

In the analyses for diabetic medications, it was important to adjust for severity or degree of diabetic control as a potential confounder as this may determine choice of anti-diabetic medications and possible cancer risk and progression. Therefore, where available, mean glycosylated haemoglobin (HbA1c) levels were calculated and linked to each participant. Units were converted and standardised to the International Federation of Clinical Chemistry (IFCC) units in mmols/mol. An ordinal variable was constructed with three levels: 1, indicating well controlled diabetes for levels ≤ 58 mmols/mol; 2, indicating sub-optimally controlled diabetes for levels > 58 mmols/mol and ≤ 75 mmols/mol; and 3, indicating poorly controlled diabetes for levels > 75 mmols/mol. If no measures of the participant's HbA1c levels were identified in the database, it was initially assumed they had mild or well controlled diabetes and these participants were assigned into the lowest ordinal level based on the assumption that the clinician did not feel it necessary to monitor the participant's glycaemic control. Sensitivity analyses were conducted to test this assumption, described in . Post-baseline HbA1c readings were included, as attempting to restrict HbA1c to only pre-baseline measurements meant there was a large amount of missing data.

Potential confounders described thus far were considered to be the same in both the nested case-control and cohort studies. For the cohort study only, it was also assumed participants with more co-morbidities would be more likely to die and this may also influence the choice of

medication by the prescribing GP. Therefore, to assess the affect of exposure and BT prognosis, a measure of each participant's co-morbidities was included as a covariate. The Charlson co-morbidity index score [126] was derived using a list of Read codes from Khan, *et al.*[172]. This resulted in an ordinal score, which was calibrated to 4 units (half the interquartile range) so that the model coefficient is for a 4-unit change in Charlson co-morbidity index score.

6.2.6 Statistical Methods

The fibrate and glitazone nested case-control study analyses were conducted using logistic regression to compute odds ratios (OR) and 95% confidence intervals (CI) for exposure status to fibrate and glitazone drugs and exposure duration to the two types of drugs and case-control status. Unadjusted and multivariable models were ran and adjusted for age, sex, IMD quintile score, retrospective prescription history duration for both the fibrate and glitazone studies. Measurements of HbA1c levels were included in models in the glitazone study. Retrospective prescription history duration, which was the variable constructed to ensure cases and controls had the same retrospective prescription history in CPRD, thus accounting for time-window bias, was adjusted for to ensure equal chance of being exposed in both cases and controls.

Pearce, in their 2016 paper, described how running conditional (matched) logistic model may induce bias into results for studies which do not necessarily require a matched analysis [262]. As controls were matched to cases upon inclusion, matching was already inherent to this analysis. Therefore, an unmatched logistic regression should suffice so long as matching variables were treated as covariates (and were included in the multivariable models).

Dose exposure was analysed both as a continuous ordinal variable and also as a "dummy" (categorical based on per year categories) variable so it was possible to check for any evidence of non-linearity in the pattern of the OR.

The cohort study analyses consisted of both unadjusted Kaplan-Meier graphs and adjusted Cox proportional hazards models with estimated hazard ratios (HR) and 95% CIs. The proportional hazards assumption was tested by examination of the scaled Schoenfeld residuals.

6.2.7 Sensitivity Analyses

Several sensitivity analyses were conducted to check the robustness of the initial findings: i) first drug exposure: exposure was re-classified as being only if the participant's first medication was a fibrate or glitazone. Hence someone whose first treatment was another drug but was later switched to a glitazone or fibrate was classified as unexposed in this analysis. This was done to examine for the potential of selection bias; ii) dual exposures: participants were dropped from the analysis if they were exposed to both a fibrate and a glitazone; iii) latency periods: analyses were repeated excluding any exposure for six, 12 and 24 months prior to the index date to test for a potential latency period between exposure and physiological effects. If exposure suppresses tumour development, it is biologically implausible to see an effect within a very short time

period (i.e. no latency). If however, exposure slows the rate of growth of an existing sub-clinical tumour, then there may be no latency period; iv) glycaemic control: the assumption that missing HbA1c levels should not have been allocated to the well-controlled group was tested by using multiple imputation to predict missing HbA1c levels based on: case-control status, age, IMD, sex, retrospective prescription history, ever exposure to a glitazone, total glitazone exposure and number of consultations, defined as each day the patient had at least one in-person consultation. 55 datasets were generated, roughly equal to the missingness of the data, and these were then combined using Rubin's rule. Analyses were also repeated without HbA1c adjustment and on the complete case subset, so that any participants with missing HbA1c levels were dropped; v) primary and secondary tumours: the case-control analyses were repeated stratified by primary and secondary BT to see if there was any evidence of heterogeneity of effect. This was not possible for the cohort study where only a few participants were exposed to the drugs overall, and hence afforded little power (76 primary and 56 secondary-diagnosed participants were exposed to a fibrate and 64 primary and 35 secondary-diagnosed participants were exposed to a glitazone).

6.3 Results

After data extraction, cleaning and linkage, the study populations consisted of 9,741 participants with 129,356 person years of follow up in the fibrate case-control study, 2,400 participants with 30,871 person years of follow up in the glitazone study and 7,496 participants with 13,805 person years of follow up in the cohort study (see flowcharts for participant selection in Figure 6.3, 6.4 & 6.5).

Figure 6.3: Participant selection for the anti-hyperlipidaemic medications case-control study.

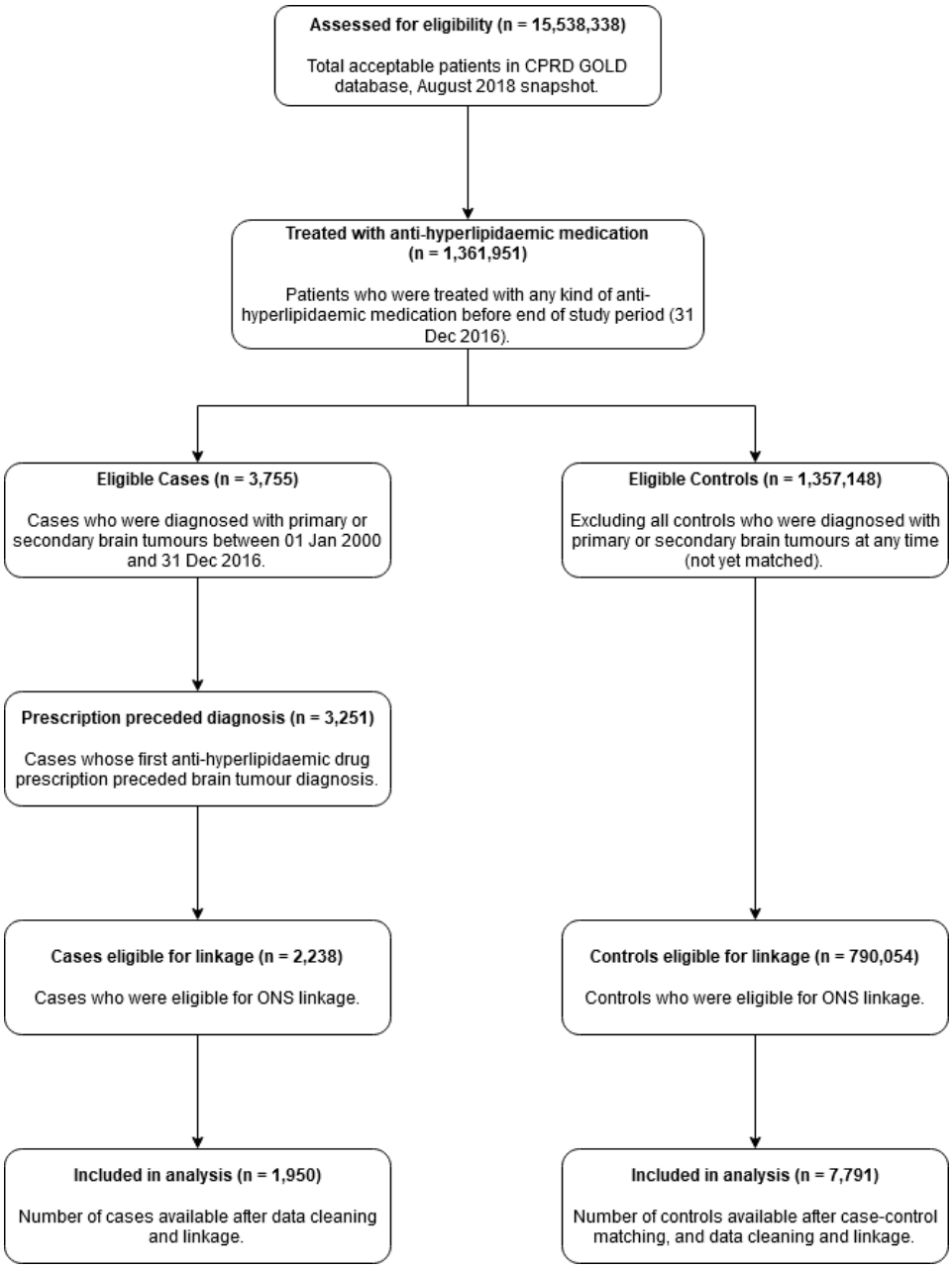


Figure 6.4: Participant selection for the anti-diabetic medications case-control study.

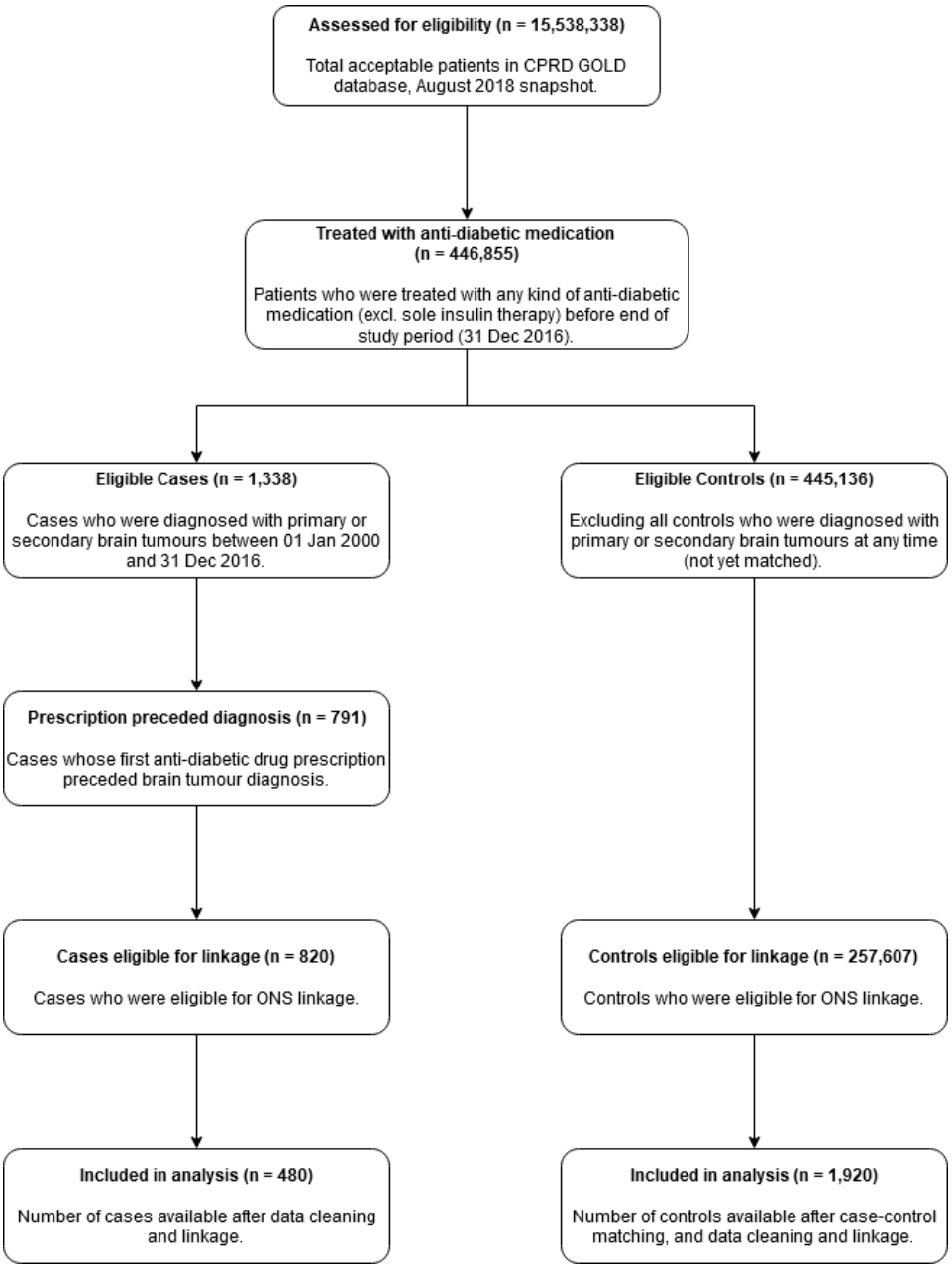
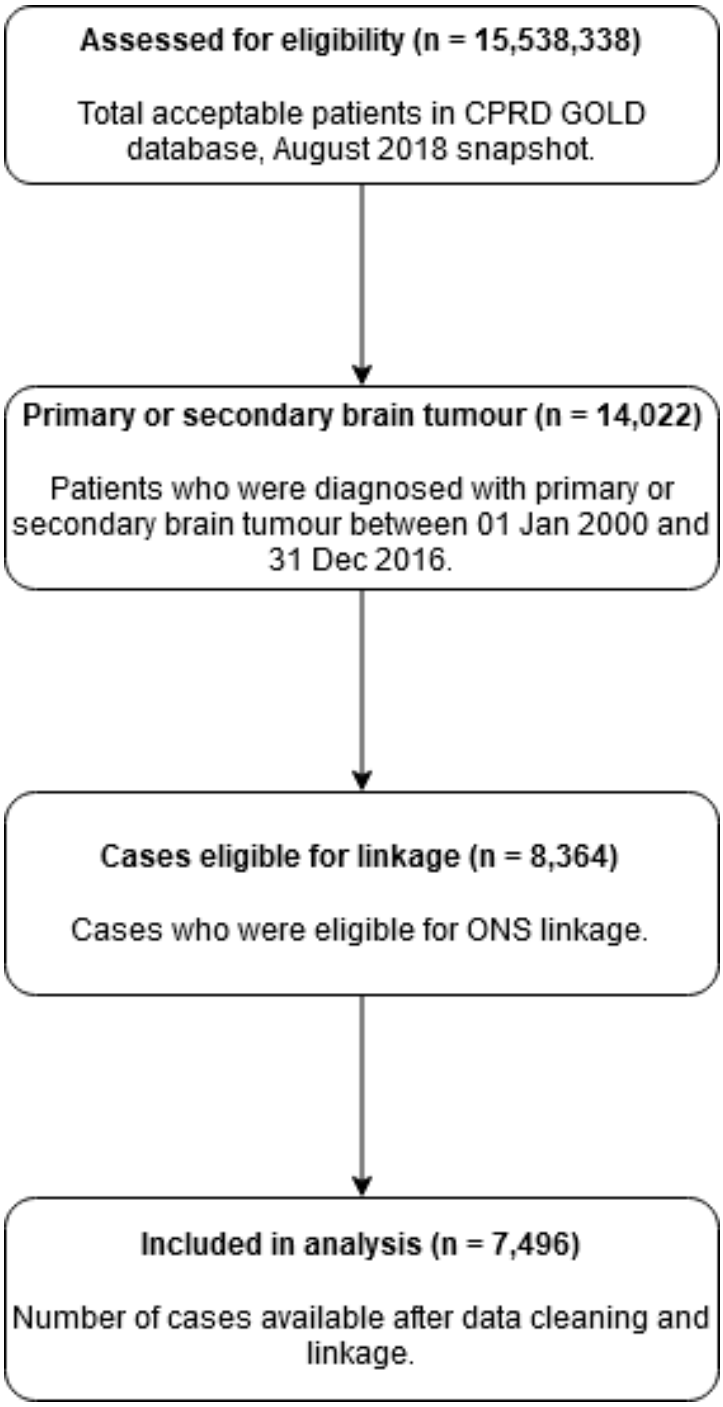


Figure 6.5: Participant selection for the cohort study.



Characteristics for the fibrate nested case-control study are presented in Table 6.1. Cases appeared to have similar probability of exposure to fibrates but less exposure to glitazones compared to controls. For fibrates, there was little evidence that ever exposed (adjusted OR (aOR) = 0.98; 95% CI: 0.78, 1.24; $P = 0.88$), per year increase in longest duration of exposure (aOR = 0.97; 95% CI: 0.91, 1.03; $P = 0.35$) or per year increase in total duration of exposure (aOR = 0.97; 95% CI: 0.92, 1.01; $P = 0.11$) was associated with risk. There was also little evidence of a dose-response relationship between fibrate exposure and BT risk either as a continuous or dummy (categorical) variable. Sensitivity analyses using different exposure variables found similar results but with less precision. Results for the fibrates nested case-control study are given in Table 6.2.

Characteristics for the glitazone nested case-control study are presented in Table 6.3. For glitazone exposure, the association with ever exposed was consistent with chance (aOR = 0.80; 95% CI: 0.63, 1.03; $P = 0.08$) but there was an inverse association with duration of longest exposure (aOR per year = 0.88; 95% CI: 0.81, 0.95; $P = 2.00\times10^{-3}$) and total exposure (aOR = 0.85; 95% CI: 0.79, 0.92; $P = 8.00\times10^{-5}$). Analysis by duration period was consistent with an effect only being observed after four years as all the ORs were less than one, however the CIs for the shorter duration periods were sufficiently wide as to not exclude any effect from earlier periods. The sensitivity analysis which only classified exposed if a glitazone was the first medication found no association but with very wide CIs due to fewer exposed cases and controls (aOR = 1.23; 95% CI: 0.40, 3.80, $P = 0.86$). These results are presented in Table 6.4.

The first step to investigate the assumption that participants with missing HbA1c levels should be categorised as well-controlled (i.e., in the lowest category for the categorical variable) was to conduct a positive control test. As glitazones are routinely used as a second line therapy in the UK, glitazone usage should therefore associate with worse glycaemic control (higher HbA1c measurements). This was observed in both the crude and adjusted models, where fibrate exposure was also tested to show the lack of an association. These results are presented in Appendix H.

After this, multiple imputation was used to how best to deal with missingness in the HbA1c variable. This sensitivity analysis showed that the inverse relation for glitazone exposure and BT risk remained when including imputing missing HbA1c values (aOR = 0.81, 95% CI: 0.66, 1.00; $P = 0.11$) and by not adjusting for HbA1c levels at all (aOR = 0.79, 95% CI: 0.62, 1.02; $P = 0.07$). However, this association was attenuated to the null in the complete case analysis (aOR = 0.91, 95% CI: 0.59, 1.40; $P = 0.67$). Similar results were also observed for the exposure duration variables. In both the imputed and complete case analyses, there was no association between HbA1c levels and case-control status suggesting that HbA1c may not be a confounder. These results are presented in Appendix H.

Analyses investigating the effects of monotherapy (i.e. participants were dropped if exposed to both a fibrate and a glitazone) and the potential of a latency period found similar results to the main case-control analyses, though generally confidence intervals were wider and results were

Table 6.1: Patient characteristics for the fibrate nested case-control study.

Variable	**Cases (%)** N = 1950	**Controls (%)** N = 7791	**Crude OR (95% CI)**	***P***
Fibrate exposure status				
Exposed	124 (6.4%)	523 (6.7%)	0.94 (0.77, 1.16)	0.58
Unexposed	1826 (93.6%)	7268 (93.3%)	Referent	-
Mean fibrate exposure duration (longest), days (SD)	743 (989)	813 (1018)	0.98 (0.93, 1.03) [a]	0.44
Mean fibrate exposure duration (total), days (SD)	925 (1202)	1141 (1319)	0.97 (0.92, 1.01) [a]	0.11
First drug exposure status				
Fibrate	61 (3.1%)	265 (3.4%)	0.92 (0.69, 1.22)	0.54
Other anti- hyperlipidaemia drug	1889 (96.9%)	7526 (96.6%)	Referent	-
Sex				
Male	1118 (57.3%)	4469 (57.4%)	Referent	-
Female	832 (42.7%)	3322 (42.6%)	1.00 (0.91, 1.11)	0.98
Age, years				
<20	0 (0.0%)	0 (0.0%)	-	-
20 - 29	0 (0.0%)	0 (0.0%)	-	-
30 - 39	2 (0.1%)	5 (0.1%)	1.58 (0.31, 8.17)	0.58
40 - 49	6 (0.3%)	29 (0.4%)	0.82 (0.34, 1.98)	0.66
50 - 59	40 (2.1%)	193 (2.5%)	0.82 (0.58, 1.16)	0.26
60 - 69	220 (11.3%)	848 (10.9%)	1.03 (0.87, 1.21)	0.75
70 - 79	561 (28.8%)	2279 (29.3%)	0.97 (0.87, 1.09)	0.65
≥ 80	1121 (57.5%)	4437 (57.0%)	Referent	-
IMD, quintile				
1, lower SES	457 (23.4%)	1768 (22.7%)	1.05 (0.91, 1.22)	0.48
2	439 (22.5%)	1790 (23.0%)	Referent	-
3	430 (22.1%)	1594 (20.5%)	1.10 (0.95, 1.28)	0.21
4	331 (17.0%)	1466 (18.8%)	0.92 (0.79, 1.08)	0.31
5, higher SES	293 (15.0%)	1173 (15.1%)	1.02 (0.86, 1.20)	0.83
Mean prescription history, years (SD)	1.98 (2.26)	1.95 (2.25)	1.00 (0.98, 1.03) [a]	0.68

[a] OR for these variables were calculated for per year increase in variable.

Table 6.2: Results from the logistic regressions for the nested case-control study to investigate the effect of fibrate exposure on BT risk. Model adjusted for sex, age, IMD and retrospective prescription history (categorised yearly).

Variable	**Cases (%)** N = 1950	**Controls (%)** N = 7791	**Adjusted OR (95% CI)**	***P***
Fibrate exposure status				
Exposed	124 (6.4%)	523 (6.7%)	0.98 (0.78, 1.24)	0.88
Unexposed	1826 (93.6%)	7268 (93.3%)	Referent	-
Mean fibrate exposure duration, days (SD)	743 (989)	813 (1018)	0.97 (0.91, 1.03) [a]	0.35
Mean fibrate exposure duration (total), days (SD)	925 (1202)	1141 (1319)	0.97 (0.92, 1.01) [a]	0.11
Per-year fibrate exposure				
Unexposed	1855 (95.1%)	7404 (95.0%)	Referent	-
≤ 1 year	51 (2.6%)	185 (2.4%)	1.10 (0.81, 1.51)	0.56
1 < years ≤ 2	17 (0.9%)	62 (0.8%)	1.10 (0.64, 1.88)	0.68
2 < years ≤ 3	5 (0.3%)	36 (0.5%)	0.55 (0.22, 1.41)	0.18
3 < years ≤ 4	4 (0.2%)	24 (0.3%)	0.67 (0.23, 1.92)	0.44
4 < years ≤ 5	5 (0.3%)	23 (0.3%)	0.87 (0.33, 2.28)	0.77
5 < years ≤ 6	3 (0.2%)	9 (0.1%)	1.33 (0.36, 4.92)	0.68
6 < years ≤ 7	4 (0.2%)	14 (0.2%)	1.16 (0.38, 3.55)	0.79
> 7 years	6 (0.3%)	34 (0.4%)	1.74 (0.45, 6.75)	0.44
First drug exposure status				
Fibrate	38 (2.0%)	158 (2.0%)	0.97 (0.67, 1.38)	0.85
Other anti- hyperlipidaemia drug	1912 (98.0%)	7633 (98.0%)	Referent	-

[a] OR for these variables were calculated for per year increase in variable.

attenuated due to smaller numbers of cases. These results are presented in Appendix I & J.

Examining how exposure to fibrates and glitazones affected risk of primary and secondary BT separately found similar results compared to the analyses using all cases combined. That is, only exposure to glitazones showed a reduction in risk of developing either primary or secondary BT. Results from these analyses were noticeably attenuated or had wider CI due to the smaller number of participants. These results are shown in Table 6.5.

Finally, in the cohort study, there was little evidence that exposure to fibrates improved survival rates when compared to the non-exposed population (unadjusted exposed median survival time = 3.31 months, 95% CI: 2.62, 4.85 months versus unadjusted unexposed median survival time = 3.77 months, 95% CI: 3.57, 4.16 months) (Figure 6.6). There was also little evidence to suggest that glitazones were associated with survival when compared to the unexposed population (unadjusted exposed median survival time = 3.87 months, 95% CI: 2.85, 6.56 months versus unadjusted unexposed median survival time = 5.11 months, 95% CI: 4.26, 5.87 months)

Table 6.3: Patient characteristics for the glitazone nested case-control study.

Variable	**Cases (%)** N = 480	**Controls (%)** N = 1920	**Crude OR (95% CI)**	***P***
Glitazone exposure status				
Exposed	97 (20.2%)	460 (24.0%)	0.80 (0.63, 1.03)	0.08
Unexposed	383 (79.8%)	1460 (76.0%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	607 (583)	872 (760)	0.88 (0.81, 0.95) [a]	0.002
Mean glitazone exposure duration (total), days (SD)	603 (720)	1002 (823)	0.85 (0.79, 0.92) [a]	0.00008
First drug exposure status				
Glitazone	4 (0.8%)	13 (0.7%)	1.23 (0.40, 3.80)	0.86
Other anti-type 2 diabetes drug	476 (99.2%)	1907 (99.3%)	Referent	-
Sex				
Male	277 (57.7%)	1108 (57.7%)	Referent	-
Female	203 (42.3%)	812 (42.3%)	1.00 (0.82, 1.22)	1.00
Age, years				
<20	1 (0.2%)	1 (0.1%)	3.93 (0.24, 63.00)	0.33
20 - 29	1 (0.2%)	7 (0.4%)	0.56 (0.07, 4.58)	0.59
30 - 39	3 (0.6%)	11 (0.6%)	1.07 (0.30, 3.87)	0.92
40 - 49	7 (1.5%)	28 (1.5%)	0.98 (0.42, 2.27)	0.97
50 - 59	15 (3.1%)	76 (4.0%)	0.78 (0.44, 1.37)	0.38
60 - 69	56 (11.7%)	233 (12.1%)	0.94 (0.68, 1.30)	0.73
70 - 79	135 (28.1%)	535 (27.9%)	0.99 (0.79, 1.25)	0.94
≥ 80	262 (54.6%)	1029 (52.6%)	Referent	-
IMD, quintile				
1, lower SES	90 (18.8%)	362 (18.9%)	1.33 (0.95, 1.86)	0.09
2	114 (23.8%)	376 (19.6%)	1.62 (1.18, 2.24)	0.003
3	109 (22.7%)	384 (20.0%)	1.52 (1.10, 2.10)	0.01
4	78 (16.3%)	418 (21.8%)	Referent	-
5, higher SES	89 (18.5%)	380 (19.8%)	1.26 (0.90, 1.75)	0.18
Mean prescription history, years (SD)	2.12 (2.41)	1.99 (2.27)	1.02 (0.98, 1.07) [a]	0.27
HbA1c				
1, well controlled	418 (87.0%)	1490 (77.6%)	Referent	-
2	42 (9.0%)	305 (15.9%)	0.49 (0.35, 0.69)	0.001
3, poorly controlled	20 (4.0%)	125 (6.5%)	0.57 (0.35, 0.93)	0.02
Missing (coded as 1)	338 (70.4%)	979 (51.0%)	-	-

[a] OR for these variables were calculated for per year increase in variable.

Table 6.4: Results from the logistic regressions for the nested case-control study to investigate the effect of glitazone exposure on BT risk. Model adjusted for sex, age, IMD, retrospective prescription history (categorised yearly) and mean HbA1c measurements.

Variable	**Cases** (%) N = 480	**Controls** (%) N = 1920	**Adjusted OR (95% CI)**	***P***
Glitazone exposure status				
Exposed	97 (20.2%)	460 (24.0%)	0.80 (0.65, 1.07)	0.14
Unexposed	383 (79.8%)	1460 (76.0%)	Referent	-
Mean glitazone exposure duration (longest exposure duration), days (SD)	607 (583)	872 (760)	0.88 (0.81, 0.95) [a]	0.002
Mean glitazone exposure duration (total exposure duration), days (SD)	603 (720)	1002 (823)	0.86 (0.79, 0.92) [a]	0.00008
Per-year glitazone exposure				
Unexposed	383 (79.8%)	1460 (76.0%)	Referent	-
≤ 1 year	41 (8.5%)	151 (7.9%)	1.11 (0.77, 1.61)	0.59
1 < years ≤ 2	23 (4.8%)	89 (4.6%)	1.06 (0.66, 1.71)	0.82
2 < years ≤ 3	29 (4.0%)	66 (3.4%)	1.13 (0.67, 1.92)	0.65
3 < years ≤ 4	8 (1.7%)	61 (3.2%)	0.52 (0.25, 1.10)	0.07
4 < years ≤ 5	2 (0.4%)	41 (2.1%)	0.18 (0.04, 0.77)	0.003
5 < years ≤ 6	2 (0.4%)	23 (1.2%)	0.30 (0.07, 1.30)	0.06
> 6 years	2 (0.4%)	29 (1.5%)	0.44 (0.10, 1.94)	0.23
First drug exposure status				
Glitazone	4 (0.8%)	13 (0.7%)	1.10 (0.35, 3.44)	0.87
Other anti-type 2 diabetes drug	476 (99.2%)	1907 (99.3%)	Referent	-

[a] OR for these variables were calculated for per year increase in variable.

(Figure 6.7). Similarly, the results from the Cox's proportional hazards models showed little evidence of an association between either fibrate or glitazone exposure and risk of death, both unadjusted and adjusted for other covariates (Table 6.6).

Table 6.5: Follow-up analysis to determine how exposure to fibrates and glitazones differentially affected primary and secondary brain tumour risk. Model adjusted for sex, age, IMD and retrospective prescription history. HbA1c adjustment were made in the glitazones analyses.

Variable	**Cases (%)**	**Controls (%)**	**Adjusted OR (95% CI)**	***P***
	Fibrates - Primary cases only			
	N = 1094	N = 4375		
Fibrate exposure status				
Exposed	68 (6.2%)	278 (6.4%)	0.99 (0.75, 1.30)	0.94
Unexposed	1026 (93.8%)	4097 (93.7%)	Referent	-
Mean fibrate exposure duration (longest), days (SD)	799 (1050)	755 (913)	1.01 (0.94, 1.09) [a]	0.76
Mean fibrate exposure duration (total), days (SD)	969 (1183)	1073 (1233)	0.99 (0.93, 1.05) [a]	0.64
	Fibrates - Secondary cases only			
	N = 856	N = 3416		
Fibrate exposure status				
Exposed	56 (6.5%)	245 (7.2%)	0.91 (0.67, 1.23)	0.55
Unexposed	800 (93.5%)	3171 (92.8%)	Referent	-
Mean fibrate exposure duration (longest), days (SD)	676 (913)	878 (1123)	0.95 (0.87, 1.03) [a]	0.16
Mean fibrate exposure duration (total), days (SD)	871 (1232)	1217 (1408)	0.95 (0.88, 1.01) [a]	0.08
	Glitazones - Primary cases only			
	N = 278	N = 1112		
Glitazone exposure status				
Exposed	61 (21.9%)	261 (23.5%)	0.96 (0.69, 1.33)	0.81
Unexposed	217 (78.1%)	851 (76.5%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	551 (599)	851 (754)	0.89 (0.80, 0.98) [a]	0.02
Mean glitazone exposure duration (total), days (SD)	684 (695)	1000 (809)	0.90 (0.82, 0.99) [a]	0.02
	Glitazones - Secondary cases only			
	N = 202	N = 808		
Glitazone exposure status				
Exposed	36 (17.8%)	199 (24.6%)	0.68 (0.46, 1.02)	0.06
Unexposed	166 (82.2%)	609 (75.4%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	701 (550)	899 (767)	0.87 (0.77, 0.99) [a]	0.02
Mean glitazone exposure duration (total), days (SD)	845 (681)	1073 (842)	0.88 (0.79, 0.98) [a]	0.01

[a] OR for these variables were calculated for per year increase in variable.

Figure 6.6: Kaplain-Meier curve for the results from the cohort analysis investigating the effects of fibrate exposure on brain tumour survival.

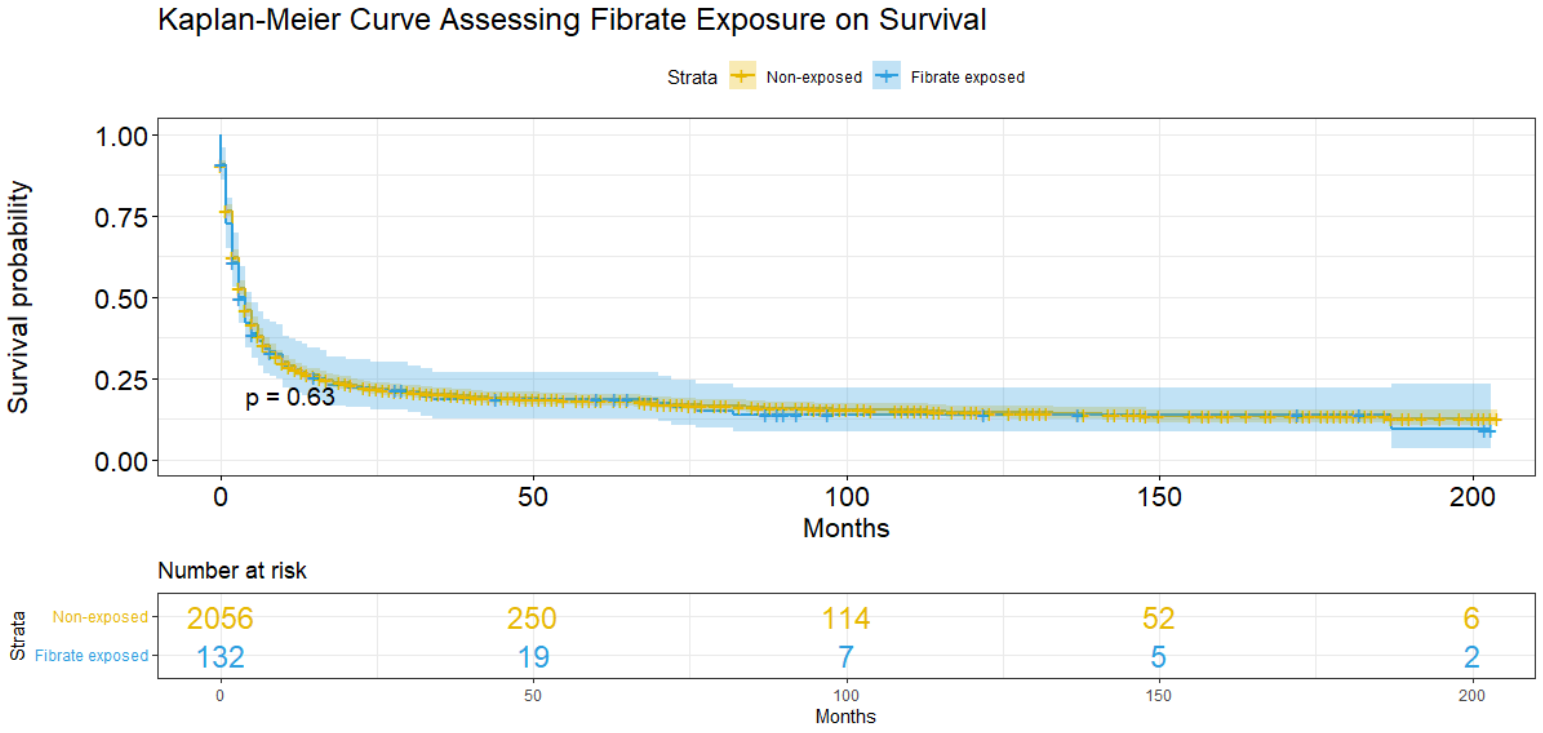
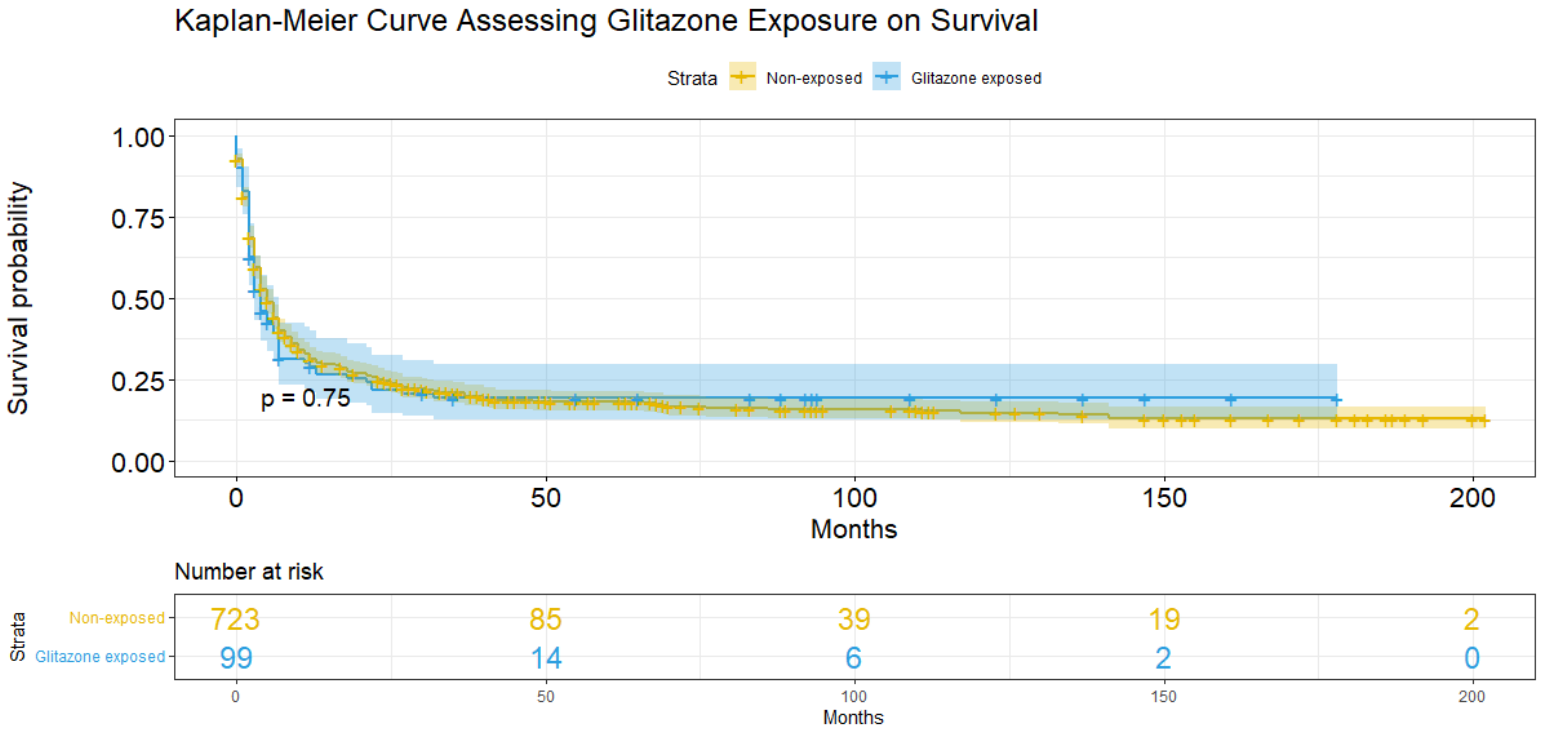


Figure 6.7: Kaplain-Meier curve for the results from the cohort analysis investigating the effects of glitazone exposure on brain tumour survival.



6.4 Discussion

This Chapter presented a pharmaco-epidemiological study which examined the effects of fibrate and glitazone exposure on risk and survival of BT compared to other treatments for hyperlipidaemia and type 2 diabetes. There have been various *in vivo* studies that have suggested fibrate exposure may be protective for BT, specifically gliomas, by modulating PPAR-α inhibition [30, 137, 382, 383]. There was similar evidence available for the consideration of glitazones as a treatment option from *in vivo* studies [119, 120, 264, 265]. However, prior to this study, there have been no other pharmaco-epidemiological studies that examined whether fibrates or glitazones affected BT risk and survival.

The case-control analyses presented herein found little evidence that exposure to fibrates was associated with BT risk, in contrast to some previous literature [255]. Sensitivity analyses, investigating a potential dose-response effect, or exposure status based on whether fibrates were the first prescribed drug for treatment, also showed little evidence of an association between fibrate exposure and BT risk. The latency period analysis had a weak suggestion of a modest protective effect, but this may have been due to chance.

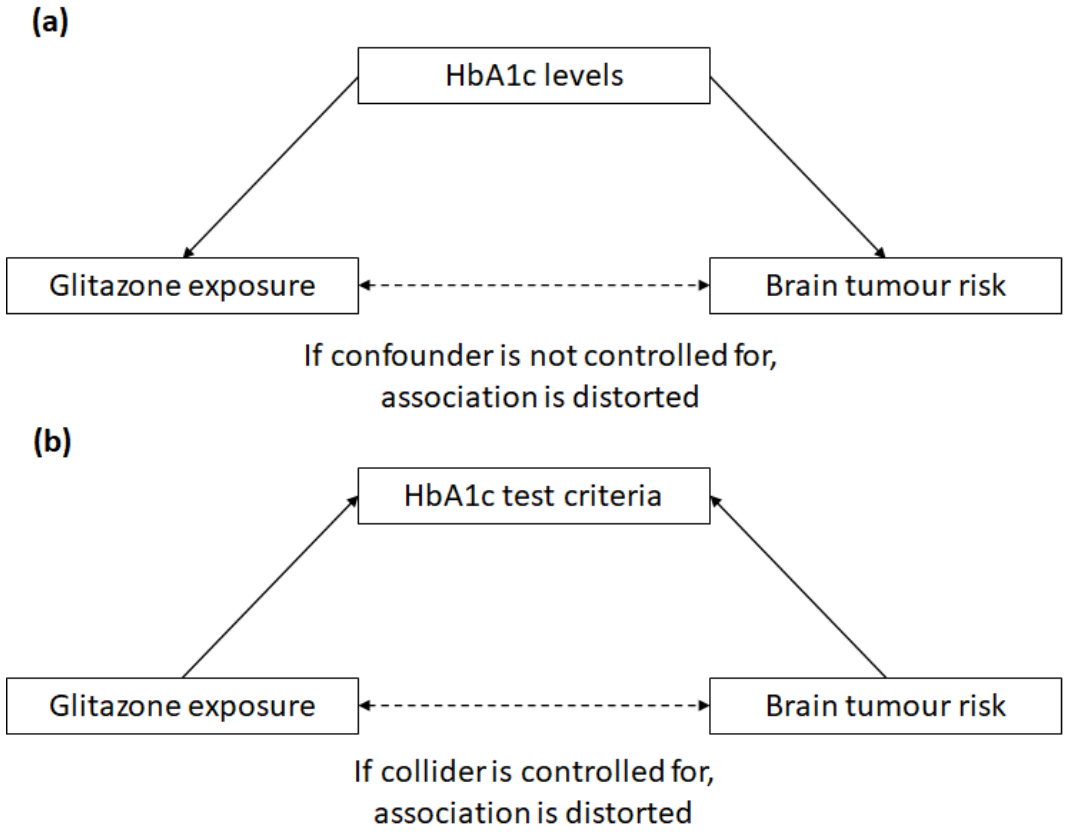
In the glitazone case-control analysis, there was an inverse effect observed for ever exposure to glitazones, which was also seen in the dose-response analysis (which seemed most marked for exposure durations of four or more years) and in both univariable and multivariable models. Repeating the analyses and accounting for various latency periods showed a slightly stronger protective effect, but given the wide CI this should be interpreted with caution. Similar inverse associations were seen for both primary and secondary BT, but limited power meant analyses could not test whether there were differences by type of cancer for those subjects with secondary BT.

The assumption that missing HbA1c levels should be assumed to reflect good glycaemic control was tested using follow-up sensitivity analyses to ensure this did not bias the potential negative association between glitazone exposure and BT risk. In the univariable analysis, it was observed that HbA1c levels representing worse glycaemic control had a negative association with BT risk. However, in both the multiple imputation and complete case analysis there was no association found. Three of the four methods of dealing with missing HbA1c levels showed no consistent inverse associations except in the complete case analysis, where the effect was attenuated. A potential explanation for this difference is collider bias [61], whereby controlling or selecting (e.g. complete case) on a common effect of both the exposure (glitazone exposure) and the outcome (BT) distorts the association between the two (see Figure 6.8). This would distort the association between treatment and BT if glitazones and BT are associated with having a HbA1c test. This is possible as, in the UK, glitazones are recommended as second line therapy for diabetes [344] and hence given to less easy-to-control diabetics who are more likely to require greater monitoring. BT risk may also be associated with testing if risk factors or confounders, such as age, are associated with the probability of testing.

Table 6.6: Results from the fibrates and glitazones Cox's proportional hazards model. Model was adjusted for sex, age, IMD, retrospective prescription history and measure of Charlson comorbidity index. HbA1c levels were included in the glitazones analysis models.

Variable	**Cases (%)**	**Adjusted HR (95% CI)**	***P***
Fibrates - All cases			
	N = 2188		
Fibrate exposure status			
Exposed	132 (6.0%)	0.96 (0.79, 1.17)	0.71
Unexposed	2056 (94.0%)	Referent	-
Fibrates - Primary cases only			
	N = 1288		
Fibrate exposure status			
Exposed	76 (5.9%)	0.96 (0.73, 1.27)	0.8
Unexposed	1212 (94.1%)	Referent	-
Fibrates - Secondary cases only			
	N = 900		
Fibrate exposure status			
Exposed	56 (6.2%)	1.19 (0.90, 1.56)	0.22
Unexposed	844 (93.8%)	Referent	-
Glitazones - All cases			
	N = 822		
Glitazone exposure status			
Exposed	99 (12.0%)	0.99 (0.77, 1.27)	0.93
Unexposed	723 (88.0%)	Referent	-
Glitazones - Primary cases only			
	N = 504		
Glitazone exposure status			
Exposed	64 (12.7%)	1.08 (0.78, 1.48)	0.65
Unexposed	440 (87.3%)	Referent	-
Glitazones - Secondary cases only			
	N = 318		
Glitazone exposure status			
Exposed	35 (11.0%)	1.08 (0.75, 1.56)	0.69
Unexposed	283 (89.0%)	Referent	-

Figure 6.8: Directed acyclic graph showing how HbA1c testing may be a collider in the glitazone case-control study. HbA1c levels are inherently associated with glitazone exposure. However, if HbA1c levels are also associated in some way with BT risk, then HbA1c levels will form a confounder that must be controlled for in the models. This is represented in (a). On the other hand, if glitazone exposure and BT risk are associated with HbA1c testing criteria, then controlling for HbA1c levels will induce collider bias and distort the association between glitazone exposure and BT risk, as represented in (b). Further studies are required to determine this potential interplay between HbA1c levels, glitazone exposure and BT risk. Adapted from Catalogue of Bias [198].



In the cohort analyses, there was little evidence for an association with fibrate or glitazone exposure and all-cause mortality. This was observed in both the unadjusted Kaplan-Meier curves and the multivariate Cox's proportional hazard models.

Strengths of this study included using CPRD, a large and well-established database which has been validated by numerous sources [145, 146, 170]. It should be free of selection bias as almost all UK residents are registered with a general practitioner and the population captured by CPRD practices is representative of the general population [145]. Exposure was well measured as it was recorded from medical systems that used the record to print out prescribed drugs for redemption at a pharmacist. The exposure data were collected prospectively, prior to diagnosis, which will avoid measurement error and recall bias. However, there were no data on drug

compliance and intake or therapeutic levels. Despite this, response to therapies were monitored by the participant's general practitioner and so dosage is likely to be adjusted in response to good or bad control of each participant's lipids and blood sugar levels. Exposure was classified from first prescription of any drug for lipids or diabetes, avoiding immortal time bias and controls were sampled to have similar total prescription duration which was also adjusted for in the multivariable models. This should have prevented time-window bias.

Risk of confounding by indication was minimised by only sampling cases and controls who could have been exposed to fibrates or glitazones because of a clinical indication. However, this reduced the number of available cases and statistical power to detect even modest effects. Furthermore, statins are first-line medications for hyperlipidaemia and have been shown to potentially improve prevention and survival of some cases [48, 108, 272], though the evidence has been mixed [64]. This may have resulted in not being able to detect any beneficial effects of fibrates on survival but the effects of statins on brain cancer specifically is still contradictory. It was not possible to investigate this further as it fell outside of the scope of the initial proposal.

A further limitation was that a considerable amount of HbA1c data were missing. Multiple imputation was used to take this into account and seemed to provide a less biased estimate than the complete case analysis. HbA1c measurements were required to be taken post-baseline due to limited data; however, taking the mean of all HbA1c measurements attempted to reduce this issue. Adjustment for SES was at an ecological rather than individual level, but if anything is likely to bias the associations to the null due to non-differential measurement error. Due to insufficient data, it was not possible to test whether specific types of primary cancers, such as lung cancer, were more, less or equally likely to show a reduced risk due to limited statistical power.

Results presented in this Chapter were not subjected to thresholds for selection, unlike for the previous Chapters. The reasons for this were twofold. Firstly, analyses within this Chapter were hypothesis driven due to pre-existing evidence for the potential of an effect between the drugs of interest and BT risk and survival. Including a threshold for multiple testing across the analyses, even for the different measurements of exposure, would likely result in the (possibly incorrect) rejection of the results from these analyses due to an overly stringent threshold. Correction for multiple testing is beneficial when different questions are being assessed; however, each of the exposure variables attempted to answer the same question in a slightly different way and, indeed, the results from these analyses were broadly consistent across each of the exposure variables. A study reported by Ridker, *et al.*, which sought to determine whether rosuvastatin could prevent vascular events in participants, is a good example where application of a stringent threshold for multiple testing would have erroneously resulted in the rejection of an otherwise consistent result [287]. Secondly, it is common for P values to remain unreported in observational studies, where consistency and size and direction of the effect are deemed more important. The American Statistical Association goes as far as to advise that P values should not be reported for

an observational study [366]. P values are reported for analyses in this Chapter for completeness sake, however, consistency of results and the size and direction of observed effects only are used to inform on conclusions.

6.5 Summary

This study failed to show strong evidence that fibrates have any effect on BT risk or prognosis. Longer exposure to glitazones was associated with a decreased risk of being diagnosed with primary or secondary BT, but not survival after diagnosis. Further research needs to replicate this finding using independent datasets preferably large in size and/or with better data on glycaemic control and confounders. If the glitazone association is biologically causal, this may lead to a better understanding of the pathophysiological mechanisms and potential therapies for the prevention of brain cancers. For example, patients with specific aggressive cancers that have a high probability of brain metastases, could be treated with glitazones as a tertiary prevention strategy to prevent secondary BT. Such a hypothesis could be tested in a future double blind clinical trial if stronger evidence emerges from other datasets, given the safety and current use of glitazones for the management of diabetes.

CHAPTER 7

DISCUSSION

7.1 Introduction

The aims of this thesis were to:

1. Leverage germline genetic variation associated with glioma to elucidate mediating molecular pathways, via gene expression, on gliomagenesis.
2. In a similar way, leverage germline genetic variation associated glioma to elucidate mediating molecular pathways, via protein abundance levels, on gliomagenesis.
3. Utilise multi-omic data from the previous two aims, i.e. for gene expression and protein abundance levels, to identify novel targets for intervention.
4. Determine whether anti-hyperlipidaemia and anti-diabetes drugs can be repurposed to reduce incidence or improve prognosis of brain tumours.

This Chapter provides a summary for each of the results Chapters presented earlier and how they bear relevance to the aims set out at the beginning of this thesis. Strengths and limitations are discussed for analyses used throughout the research presented herein, and finally avenues of future work are discussed.

7.2 General Discussion

This Section summarises and discusses the key findings from each of the results Chapters presented in this thesis, and includes clinical implications and potential translational uses.

Table 7.1: Summary of results presented in this thesis. Transcriptomics results from Chapter 4 are those which passed all sensitivity analyses in that Chapter. Proteomics results from Chapter 5 are for all of the MR results. Finally, results from the pharmaco-epidemiological study in Chapter 6 are presented for the main analyses. Outcome phenotypes are for glioma or glioma subtype in the MR analyses or brain tumour (BT) risk or mortality in the pharmaco-epidemiological study. Ratios are odds ratios (OR) except where noted with * which are given as hazard ratios (HR). OR and confidence intervals (CI) for the proteomics results are scaled per one SD to allow for comparison between the brain and whole blood derived protein quantitative trait loci (pQTL) datasets.

Exposure	**Outcome**	**Odds Ratio (95% CI)**	**P value**
Chapter 4 Transcriptomics results			
ABCB6	All glioma	0.57 (0.44, 0.74)	2.20×10^{-5}
BAIAP2L2	All glioma	0.65 (0.55, 0.78)	1.65×10^{-6}
	GBM	0.60 (0.49, 0.73)	2.85×10^{-7}
EGFR	GBM	0.45 (0.38, 0.53)	9.99×10^{-20}
FAM178B	All glioma	1.47 (1.23, 1.77)	3.59×10^{-5}
JAK1	All glioma	1.21 (1.13, 1.29)	6.95×10^{-8}
	GBM	1.27 (1.17, 1.37)	1.56×10^{-9}
MVB12B	All glioma	1.24 (1.12, 1.38)	5.27×10^{-5}
PANK4	All glioma	0.46 (0.32, 0.67)	4.30×10^{-5}
PICK1	All glioma	1.72 (1.39, 2.14)	8.82×10^{-7}
	GBM	1.96 (1.54, 2.51)	6.60×10^{-8}
PRLR	All glioma	0.66 (0.54, 0.82)	9.33×10^{-5}
RETREG2	All glioma	0.68 (0.57, 0.80)	9.54×10^{-6}
	GBM	0.67 (0.55, 0.81)	6.13×10^{-5}
STMN3	All glioma	0.36 (0.29, 0.46)	1.44×10^{-16}
	GBM	0.29 (0.22, 0.38)	4.55×10^{-19}
TP53	Non-GBM	0.17 (0.09, 0.32)	9.61×10^{-8}
Chapter 5 Proteomics results			
EGFR	GBM	3.02×10^{-3} (6.35×10^{-4}, 1.33×10^{-2})	3.07×10^{-13}
	All glioma	0.02 (4.24×10^{-3}, 0.06)	1.43×10^{-10}
MON1B	All glioma	0.09 (0.03, 0.29)	6.96×10^{-5}
	GBM	0.08 (0.02, 0.36)	9.21×10^{-4}
TAGLN	Non-GBM	2.19 (1.38, 3.50)	9.53×10^{-4}
TGFBI	All glioma	0.14 (0.04, 0.44)	7.70×10^{-4}
Chapter 6 Pharmaco-epidemiological results			

Fibrates:			
Ever exposure	Risk	0.98 (0.78, 1.24)	0.88
Longest exposure (yearly)	Risk	0.97 (0.92, 1.01)	0.11
Ever exposure	Mortality	0.60 (0.79, 1.17) *	0.71
Glitazones:			
Ever exposure	Risk	0.80 (0.65, 1.07)	0.14
Longest exposure (yearly)	Risk	0.86 (0.79, 0.92)	8.00×10^{-5}
Ever exposure	Mortality	0.99 (0.77, 1.27) *	0.93

Table 7.1 provides an overview results that were presented in previous Chapters.

7.2.1 Exploring the Effect of Gene Expression on Glioma Risk

Chapter 4 leveraged a combined Mendelian randomisation (MR) and colocalisation framework to explore the causal role of genetically predicted gene expression levels on glioma risk and how such an analysis may build upon current knowledge of the genetic susceptibility of glioma. In total, genetic variants associated with 12 genes were robustly associated with glioma risk across the combined MR-colocalisation framework and follow-up sensitivity analyses. Included in the 12 genes were important oncogenes previously implicated in glioma biology, such as *TP53*, *EGFR* and *JAK1*. Furthermore, there was evidence that expression of novel genes *RETREG2*, *FAM178B* and *MVB12B* were causally implicated in glioma risk.

Follow-up analyses were conducted to determine how the causal estimates for the highlighted 12 genes affected glioma subtype risk, split between high-grade glioblastoma (GBM) and low-grade, non-GBM gliomas. No large differences were observed in either direction or magnitude of effect for each gene's influence on risk. This implies that, although there are demonstrable differences between the expression profiles for different brain tumours (BT) – particularly between high- and low-grade gliomas – within these analyses and future analyses of similar nature, the gain in statistical power from including all cases in the same analysis outweighs any loss of granularity in determining subtype susceptibility using germline genetic variants associated with gene expression. However, investigators should be cautious as this conclusion may not be applicable to other BT types or diseases.

Finally, expression quantitative trait loci (eQTLs) from the Genotype-Tissue Expression (GTEx) v8 project [3] were used to analysis how expression of each of the 12 genes in 13 different brain tissues differentially influenced glioma risk. Five tissues were responsible for 56% of the MR associations, including two tissues where gliomas are generally seen (cortex and cerebellum) and three tissues from the deep brain in the basal ganglia where gliomas are rarely seen, namely the putamen, caudate and nucleus accumbens. Although gliomas are rarer in these tissues,

variance in these gene's expression specific to these tissues of the brain may drive gliomagenesis, though further in vivo studies are required to determine this.

The results from this Chapter provided insight that is complimentary to results from genome-wide association studies (GWAS). By highlighting novel oncogenes and providing causal evidence for variance in gene expression influencing glioma risk, these analyses have built upon knowledge of the genetic susceptibility of glioma.

7.2.2 Exploring the Effect of Protein Abundance Levels on Glioma Risk

In Chapter 5, analyses utilised the same analytical pipeline of combined MR and colocalisation to explore how genetically predicted protein abundance levels affected glioma risk. Results from the initial analysis highlighted three proteins whose altered levels associated with reduced risk of glioma, GBM or non-GBM (EGFR, MON1B and TGFBI) and one protein whose altered levels associated with increased risk (TAGLN). Three of these proteins were instrumented by pQTLs derived from brain tissue (EGFR, MON1B and TAGLN) whilst TGFBI was instrumented by a pQTL from blood. Brain-specific proteins and brain-derived pQTLs are more likely to be relevant for glioma drug target identification due to presence of the blood-brain barrier, a highly selective membrane, that limits protein transfer between brain tissue and blood. Furthermore, brain tissue would be the targeted tissue for any potential therapeutic agent, meaning associations from data derived from brain tissue would be more relevant for drug target identification, which is especially so for estimating efficacy. The use of brain tissue-derived pQTLs in this analysis was, therefore, more informative for highlighting proteins of interest that may form targets of interest for glioma prevention.

Formal statistical colocalisation analyses were limited in this Chapter because the brain-derived pQTL dataset was not made wholly publicly available and statistical colocalisation cannot be conducted without the full summary statistics. Instead, approximate colocalisation analyses were conducted that examined the linkage disequilibrium (LD) structure between the pQTL and glioma datasets. Approximate colocalisation in this way allowed for the same inference of underlying causality in the association between protein abundance levels and glioma risk [400].

Comparison with the eQTL results from the previous Chapter revealed no strong correlation between the causal estimates for e- and pQTLs systematically linked according to their associated gene and protein. This has important implications for drug discovery pipelines which use eQTLs as evidence to support a drug target because these results show that, for glioma specifically, the direction and magnitude of an eQTL's causal estimate may not agree with that of the pQTL at the some locus. This was evidenced in Figure 5.3 in Section 5.3.1 - *Statistical Results*, where the direction of the causal estimates for QTLs were in opposite directions, e.g., for FAM171A1 and IL17RB. Studies have shown drug targets with genetic evidence are two times as likely to receive approval from a governing body than those without [232], where genetic evidence has been so far defined as causal genes linked to coding variants [178, 232]. Whilst these studies

provide justification for the use of genetic evidence to support a drug target, results from the MR analyses presented in this thesis showed that genetic evidence consisting of gene expression data may not translate to targettable proteins for glioma prevention; whether this is the case for other diseases is unclear and should be considered in such studies.

7.2.3 Identifing Drugs for Potential Drug Targets of Interest

Also within Chapter 5, results from the eQTL and pQTL analyses were linked to supporting evidence from two publicly available drug target identification and prioritisation platforms – the Open Targets (OT) platform [52] and the Drug Genome Interaction database (DGIdb) [63]. In total, six genes or proteins showed evidence of forming a drug target for glioma, including important biomarkers and oncogenes EGFR, JAK1 and TP53.

Further to this, a non-exhaustive literature search was conducted for existing drugs that target any of the proteins from the MR analyses as mechanisms for effect. MR results were included which had robust causal evidence from the combined MR and colocalisation analyses and were robust to sensitivity analyses. This search included existing approved agents or agents undergoing clinical trials. Drugs were found which target five proteins, including the protein for novel gene *FAM178B*. Many of the drugs have already been examined within *in vivo* or other pre-clinical settings in the context of glioma. Furthermore, many drugs had undergone or were undergoing early-stage (phase I or II) clinical trials to assess whether the drug could be re-purposed for glioma therapy.

It is important to note that existing drugs were uniformly studied within the context of glioma treatment, i.e., to improve patient outcomes post diagnosis. The MR analyses conducted in this thesis highlighted putative drug targets which could be informative for chemopreventive therapies, i.e., to reduce risk of getting glioma. Therefore, evidence gathered from the literature search should not be viewed as providing support for a putative target's efficacy. Instead, this evidence highlighted potential drugs which: i) could be re-purposed to reduce risk in clinically high-risk populations, and ii) are safe and tolerable for glioma patients to take. Altogether, evidence collected in this way can lend further support to the identification and prioritisation of drug targets as highlighted by MR analyses, so long as such evidence is carefully and correctly interpreted.

7.2.4 Determining whether Anti-Hyperlipidaemic and -Diabetic Medications may be Re-purposed for Brain Tumour Treatment

Chapter 6 presented a pharmaco-epidemiological study which sought to determine whether anti-hyperlipidaemic and anti-diabetic medications can be re-purposed to reduce incidence and improve mortality for BT. The medications of interest were fibrates and glitazones, respectively used to treat hyperlipidaemia and type 2 diabetes. There exists pre-clinical evidence that supports the use of these medications for BT treatment, particularly through their pharmacodynamics of

targeting PPAR-α and -γ, respectively. It is important to note that this project was not limited to gliomas specifically and included all BT cases available in the Clinical Practice Research Datalink (CPRD) which allowed for increased sample sizes and hence statistical power.

There was little evidence that exposure to fibrates was associated with BT risk or mortality. Conversely, increase in per-year exposure to glitazones was associated with a decreased risk of BT, but there was little evidence that exposure to glitazones was associated with mortality. The evidence provided by this study supports further research into whether glitazones will form an effective treatment option for BT.

Particularly interesting was that increased exposure to glitazones reduced risk of both primary and secondary BT compared to other treatments for type 2 diabetes, indicating the drug could be prescribed to patients whose tumours commonly metastasise to the brain to reduce risk of metastatic lesions; likewise, glitazones may be used to improve outcomes for patients with recurrent GBM. Beyond the clinical implications of these results, pharmaco-epidemiological studies should be considered important evidence generators which can increase support for targets indicated by genetic studies like those conducted in Chapters 4 and 5. In particular, pharmaco-epidemiological studies can be a powerful tool that increases support for commonly prescribed agents where large datasets exist and can be leveraged to provide support for drug re-purposing.

7.2.5 Overall Summary

Analyses which used a combined MR-colocalisation framework provided causal evidence for germline genetic variants associated with putative oncogenes and oncoproteins related to glioma risk, including expression of novel genes *RETREG2*, *FAM178B* and *MVB12B*. This analytical pipeline also provided support for proteins as potential drug targets for glioma prevention, namely ABCB6, EGFR, JAK1, PRLR, TGFBI and TP53. Evidence from drug target prioritisation platforms further supports EGFR, FAM178B, JAK1, PRLR and TP53-targeting drugs as candidates for re-purposing to help prevent glioma, though more evidence and research is required to determine potential efficacy before undergoing clinical trials. One such study design which can aid with drug re-purposing efforts is a pharmaco-epidemiological study, which was applied to the case of two existing families of drugs which have *in vivo* evidence of potential efficacy for treating and preventing BT. That study found evidence that of the two families of medications, only glitazones may have a protective effect against BT risk – both for primary and secondary tumours – that highlights this family of drugs as a candidate for further study.

7.3 Strengths and Limitations

Each Chapter has provided a discussion for strengths and limitations of the methodologies and datasets relevant to analyses in that Chapter; however, a number of these are also relevant

across this entire thesis. These strengths and limitations are discussed below, including how limitations were overcame.

7.3.1 Study Designs

The MR analyses conducted throughout this thesis have a reduced risk of bias due to reverse causation as temporality between exposures, proxied by genetic variants, and outcomes is established. That is, glioma diagnosis cannot precede germline mutation inheritance, but reverse causation can still occur if, for example, SNPs are unknowingly related more to glioma risk than the exposure of interest. Furthermore, the combination of MR and colocalisation together in the same framework increased the robustness of causal inferences due to the ability of colocalisation to eliminate some spurious associations which may arise due to the underlying LD structure of the examined locus.

In Chapter 6, case-control and cohort study designs were employed in a pharmacoepidemiological context. The cohort study design and Cox's proportional hazards models allowed for assessment of causality which provided evidence that exposure to fibrate and glitazone medications did not improve BT mortality rates. However, in the case-control studies which leveraged logistic regressions, evidence was found that exposure to glitazones, especially for per-year increase in exposure, reduced risk of both primary and secondary BT – though, inherent to this study design, causal inferences could not be made.

For analyses in that Chapter, efforts were made at the study design stage to account for common sources of bias and confounding in pharmaco-epidemiological studies. For example, using CPRD data meant reduced risk of, e.g., recall bias affecting the results. Furthermore, careful identification and selection of controls based on similar characteristics to the cases minimised selection bias and time-related biases, such as immortal time and time-window bias [337]. Confounding by indication has been shown to be a concern in many pharmaco-epidemiological studies [188, 284, 314]. To address this, participants were eligible for inclusion into the study if they were treated with any anti-hyperlipidaemia or anti-diabetic medication. Increasing the number of indications present in the study in this way reduced the influence of confounding by indication but inherently meant smaller sample sizes, leading to more accurate but less precise results.

7.3.2 Subtype Diagnosis

As gliomas, and BT in general, are heterogeneous diseases, the ability to differentiate between glioma subtypes in Chapters 4 and 5 allow for nuanced conclusions which are more relevant for clinical applications. Despite this, results in the transcriptomics Chapter (Chapter 4) generally agreed in direction of effect on risk for all glioma, GBM and non-GBM tumours though were somewhat attenuated in the subtype analyses, possibly due to smaller sample sizes to detect an effect. A similar pattern was observed in the proteomics analysis in Chapter 5. Therefore, this led

to the conclusion in both these Chapters that, due to available studies and sample sizes, it was more beneficial in an MR context to combine glioma GWAS data due to the increase in sample size with no loss of granularity in inference for the specific subtypes.

Furthermore, observational and *in vivo* studies have associated germline variants with particular somatically defined tumour subtypes [90, 112, 116, 189]. The results presented in Chapters 4 and 5 generally provided little evidence to support this conclusion in the context of germline variants associated with gene expression and protein abundance levels, and the broad categorisations of GBM and non-GBM gliomas. This was especially true for results which were robust to sensitivity analyses. However, as stated, non-GBM results tended to show slightly attenuated results due to containing more heterogeneous tumour types compared to the GBM subtype, for example. Further studies are required to investigate whether germline variants associate with both gene expression and protein abundance levels and histological subtypes – in particular, a better and more granularly defined "non-GBM" glioma category would benefit such an analysis, along with greater sample sizes.

In Chapter 6, analyses specifically examined how exposure to fibrates and glitazones differentially affected risk and mortality for both primary and secondary BT. A drug that would reduce risk for secondary BT is particularly clinically attractive due to increasing survival rates for primary tumours that commonly metastasise to the brain (discussed in Section 1.5 - *Risk Factors*). Therefore, the ability to examine how exposure affected secondary tumour diagnosis is more translatable and impactful on patient outcomes where many malignant gliomas like GBM are also likely to re-occur. Examining primary and secondary BT separately like this reduced the case numbers available for the analysis but allowed for clinically relevant conclusions that can inform further studies.

7.3.3 Horizontal Pleiotropy

The MR analyses that were conducted were not immune to horizontal pleiotropy. A description and visual representation of horizontal pleiotropy was given in Chapter 2, Figure 2.7. Many methods supplemental to MR have been developed to identify and adjust for the presence of horizontal pleiotropy, such as MR-Egger [36] and the MR pleiotropy residual sum and outlier (MR-PRESSO) test [355] though these methods inherently require multi-SNP instruments to detect potential outliers whose effects are driven by horizontal pleiotropy. However, as is common for 'omics analyses, many of the instruments constructed for the analyses in this thesis consisted of single SNPs, meaning the aforementioned sensitivity analyses could not be conducted. Although colocalisation can potentially discount some associations which have arisen due to bias or confounding [140], an alternative method that could have used to explore the presence of bias due to horizontal pleiotropy would have been to perform a systematically look-up of SNPs in the GWAS Catalogue to find if they were associated with other pathways or mechanisms which link to glioma indirectly. However, there are only two risk factors associated with glioma

(previous exposure to ionising radiation and certain genetic disorders, as detailed in Chapter 1) so it is not immediately clear whether doing such a search would have aided in accounting for the presence of horizontal pleiotropy. Even considering putative risk factors discussed briefly in Section 1.8.5 - *Strengthening Causal Inference for Putative Glioma Risk Factors using Mendelian Randomisation*, there was no overlap between variants associated with those risk factors and gene expression and protein abundance levels. A phenome-wide search, which would expand such a search to include more traits, could potentially help in elucidating relevant pathways and mechanisms through which the SNP of interest influences glioma risk and would make for an interesting project; this is discussed further Section 7.4 - *Future Work* and how such a strategy could also be beneficial when applied to a drug discovery pipeline.

7.3.4 Sample Size

An issue inherent to working with a rare disease like glioma is limited sample size. Although analyses in Chapter 5 had access to the largest meta-analysis of glioma GWAS, the sample size is still fairly low at 12,496 cases and 18,190 controls. Furthermore, only a subset of this dataset were used for analyses in Chapter 4 due to data sharing agreements which limited an already small sample size further to 7,400 cases and 8,257 controls. The exposure datasets were also of small sample sizes as GWAS of molecular data tend to be, though attempts were made to overcome this by, where possible, using meta-analysis of studies. As sample sizes increase, either due to larger studies or more studies available for meta-analysis, this will lead to an increase in statistical power for MR studies.

In Chapter 6, analyses used case-control and cohort study designs that were well suited for the study of rare diseases; however, steps were taken at the study design stage to minimise the influence of bias and confounding inherently meant reduced sample sizes, e.g., by ensuring that all participants must have been prescribed any medication used to treat hyperlipidaemia or type 2 diabetes so that the risk of confounding by indication was reduced.

Altogether, despite using some of the largest and most comprehensive datasets currently available, inferences were still limited due to small sample sizes which may not have allowed for powerful enough analysis to detect effects; however, as sample sizes increase due to more data collection then low power to detect effects will become less of a limitation on analyses.

7.4 Future Work

7.4.1 Extensions to my Thesis Work

Within this thesis, Chapters 4 and 5 are hypothesis generating and so lend themselves well to informing on further analyses. To start with, there are new methodologies that are being derived that can aid analytical drug discovery pipelines. For example, a recently published paper detailed one such method called MR joint tissue imputation (MR-JTI) which integrated multi-tissue

Table 7.2: List of eQTL studies that could be potentially included in a new brain eQTL meta-analysis. Estimated effective n = 3,217. Studies are non-exhaustive; identification of study inclusion is on-going.

Study	**Data Available**	**Sample Size**
GTEx v8 [3]	13 regions; described in Table 3.3.	114 to 209, depending on tissue
Braineac (UKBEC) [280]	10 regions: cerebellar cortex, frontal cortex, hippocampus, inferior olivary nucleus, occipital cortex, putamen, substantia nigra, temporal cortex, thalamus and intralobular white matter	134
BrainSeq phase 2 [62]	Hippocampus and dorsolateral prefrontal cortex	551 (286 with schizophrenia)
PsychENCODE [365]	Prefrontal and temporal cortices and cerebellum	1,362 (including samples with schizophrenia, bipolar disorder and autism spectrum disorder)
Religious order study and memory and ageing project (ROSMAP) [234]	Cortex	467 (some samples with dementia)
CommonMind Consortium (CMC) [106]	Dorsolateral prefrontal cortex	494 (258 with schizophrenia)

eQTLs to increase power and strengthen causal inferences in transcriptome MR studies [402]. There are also increasingly numerous studies that have derived eQTL data in many different tissues types which could also be analysed using this new MR-JTI methodology. The brain-derived eQTL meta-analysis conducted by Qi, *et al.* in 2018 that was used in Chapter 4 has a markedly larger sample size than any single eQTL study (estimated effective n = 1,194) [276]. One ongoing project that I am currently leading will seek to meta-analysis these new eQTL studies with the aim to create a new brain eQTL meta-analysis that almost triples the estimated effective sample size ($n = 3,217$) compared to the Qi, *et al.* resource. As of January 2021, this project is currently at the data collation stage where Table 7.2 shows a list of studies for which data request applications have been accepted or are undergoing review.

Whilst the work in Chapters 4 and 5 focused on drug target discovery and identification, the analyses in these Chapters could be expanded to also inform on drug safety profiles through what is called a phenome-wide association study (PheWAS). PheWAS seek to determine if an exposure is associated with many different outcomes or vice versa. For example, if a SNP associated with protein abundance levels in the proteomics analysis was associated with glioma risk, analyses

Table 7.3: Results from the exploratory MR analysis investigating the effects of SNPs in PPAR-α and -γ associated with lipid levels affected glioma risk (12,496 cases and 18,190 controls). Results are shown for lipid measurements where these proteins could be instrumented: low-density lipoprotein cholesterol (LDLc), triglycerides and total cholesterol.

Protein	**Associated lipid**	**Subtype**	**SNPs**	**OR (95% CI)**	***P***
PPAR-α	LDLc	All glioma	1	0.80 (0.13, 4.81)	0.81
		GBM	1	0.58 (0.06, 5.28)	0.63
		Non-GBM	1	1.24 (0.12, 12.97)	0.86
PPAR-γ	LDLc	All glioma	2	2.45 (0.89, 6.76)	0.08
		GBM	2	1.74 (0.50, 6.04)	0.38
		Non-GBM	2	3.52 (0.93, 13.39)	0.06
	Triglycerides	All glioma	2	1.41 (0.23, 8.67)	0.71
		GBM	2	1.04 (0.18, 6.04)	0.96
		Non-GBM	2	1.84 (0.43, 7.95)	0.41
	Total cholesterol	All glioma	5	1.66 (0.84, 3.28)	0.14
		GBM	5	1.53 (0.64, 3.65)	0.34
		Non-GBM	5	1.78 (0.78, 4.02)	0.17

could leverage a PheWAS to identify other potential pathways, traits or diseases which could be impacted (if combined with MR in an MR-PheWAS framework) were this target to be drugged. These associations may be taken into consideration before clinical trials and inform on whether a drug would have a tolerable safety profile.

The MR analysis pipeline used in Chapters 4 and 5 was also directly applicable to the pharmaco-epidemiological study in Chapter 6. GWAS summary-level data were obtained from the Global Lipids Genetics Consortium [385] for this exploratory analysis, the rationale being to determine whether the mechanism of reduced risk from exposure to fibrates and glitazones occurd through lipid-related pathways. Therefore, instruments were constructed for both PPAR-α and -γ genes, whose protein products are the targets for fibrates and glitazones respectively. The exploratory MR analysis appeared to be underpowered (evidenced by the wide CI) but indicated SNPs in PPAR-γ associated with increased low-density lipoprotein cholesterol (LDLc) levels may increase risk of glioma (Table 7.3); however, given the overall wide confidence intervals and large P values, it appears that the MR analysis did not have sufficient power (e.g., 12% empirical power for PPAR-γ and LDLc, calculated using Burgess' method [47]). This makes for an interesting avenue of follow-up that could provide evidence for genetic variants associated with lipid-related pathways being implicated in glioma risk.

7.4.2 Future Directions

Beyond extensions directly applicable to the work presented in this thesis, there are also ways in which to expand the scope of the work to better improve the drug discovery pipeline by integration

of either new methods or more molecular data.

Such methods include two-step or multivariable MR, originally proposed as a strategy to better understand and explore the causal role of the epigenome in mediating exposure-outcome associations and to highlight causal pathways in disease risk [283]. Leveraging two-step MR, for example, could elucidate on how specific pathways mediated through a drug target of interest affects the disease of interest. Likewise, the colocalisation analyses could be expanded to integrate multiple sources of data and determine if these all colocalise within the region of interest, in what is called multi-trait colocalisation (moloc) [114]. Originally applied to a multi-omics analysis of expression and methylation QTLs (mQTLs), this could also be expanded to other molecular data, including pQTLs.

Finally, integrating more sources of molecular data could help determine potentially causal variants for glioma risk and therefore elucidate on putative drug targets. This could mean inclusion of the aforementioned mQTLs, which would elucidate on potential epigenetic pathways of interest. There also exist splicing QTLs (sQTLs), which are genetic variants associated with alternative splicing, a key function that increases transcript variation and proteome diversity and is frequently observed to be aberrant in cancer. To show how fast the field is moving, there are also newly derived cell-type interaction expression and splicing QTLs (ie/isQTLs), which are eQTLs and sQTLs that map to specific cell types [177]. These are only some of the different types of molecular data that could be integrated into a drug discovery pipeline; there are ever-expanding amounts of QTL datasets being derived that may be relevant to drug discovery efforts [401].

7.5 Summary

The results presented in this thesis indicated putatively causal variants associated with altered expression and abundance levels of oncogenes and oncoproteins for glioma risk. Analyses provided evidence for three novel genes *RETREG2*, *FAM178B* and *MVB12B*, and validated known genes so far associated with glioma through GWAS and other observational studies. These results also informed potential druggable targets for glioma prevention as highlighted by evidence from a literature search. There was also evidence that agents for these targets were generally tolerable for glioma patients and may inform future studies to determine whether these agents could be re-purposed for glioma chemopreventive therapies. Results from the pharmaco-epidemiological studies provided evidence that glitazones, an anti-type 2 diabetic medication, may also be re-purposed to reduce risk of primary and secondary BT, should these results be validated in further studies. Altogether, analyses in this thesis were combined to form a pipeline for the identification, prioritisation and re-purposing of drug targets with the aim of improving patients outcomes in the clinic. Future work conducting analyses in larger sample sizes, and with integration of more datasets, will lead to improved power that will enable deeper exploration into how molecular phenotypes affect glioma risk and increased ability to highlight potentially translatable drug

targets to improve glioma patient outcomes.

APPENDIX A

APPENDIX A

All Mendelian randomisation results for all three glioma subtypes for the risk factors analysis presented in Chapter 1, Table 1.3. Odds ratios (OR) and 95% confidence intervals (CI) are presented for the MR analysis results.

Risk Factor	**Subtype**	**OR (95% CI)**	***P***
Alcohol consumption	All	4.42 (1.07, 18.30)	0.04
	GBM	8.37 (1.69, 41.54)	0.009
	Non-GBM	1.39 (0.18, 10.71)	0.75
Allergic disease	All	1.23 (0.98, 1.54)	0.079
	GBM	1.29 (1.00, 1.67)	0.048
	Non-GBM	1.13 (0.88, 1.47)	0.34
LDLc	All	0.98 (0.82, 1.17)	0.84
	GBM	1.08 (0.88, 1.33)	0.47
	Non-GBM	0.79 (0.63, 0.99)	0.04
Obesity (childhood extreme)	All	1.11 (1.02, 1.21)	0.016
	GBM	1.12 (1.02, 1.22)	0.021
	Non-GBM	1.09 (0.97, 1.22)	0.13
Telomere length	All	4.09 (1.13, 14.86)	0.032
	GBM	4.39 (0.87, 22.22)	0.074
	Non-GBM	4.05 (1.72, 9.56)	0.001
Triglycerides	All	1.00 (0.83, 1.21)	0.99
	GBM	1.13 (0.91, 1.39)	0.27
	Non-GBM	0.77 (0.59, 1.00)	0.049

APPENDIX B

APPENDIX B

Results for all genes which passed at least the suggestive P value threshold in the Mendelian randomisation (MR) analyses presented in Chapter 4, including in subtypes and tissues which did not meet that threshold. Odds ratios (OR) and corresponding 95% confidence intervals (CI) are presented for the MR analysis results. For the colocalisation analyses, the number of SNPs included in the analysis and the corresponding H4 % are given. For the Steiger filtering analysis, the direction of effect is given (a check mark signifies the correct direction, from exposure to outcome), corresponding P value and a flag signifying whether that analysis supported the MR analysis results: i) a check mark shows the correct direction of effect with $P < 0.05$; ii) a cross mark shows the wrong direction of effect with $P < 0.05$; iii) a question mark if $P \geq 0.05$. Finally, a check or cross mark shows whether that results passed all analyses or failed at any stage, respectively.

			Mendelian randomisation			Coloc		Steiger filtering			
Gene	**Subtype**	**Tissue**	**SNPs**	***P***	**OR (95% CI)**	**SNPs**	***H*$_4$ (%)**	**Dir**	***P***	**Flag**	**Pass All?**
	All	Brain	1	2.20×10^{-5}	0.57 (0.44, 0.74)	1521	97	✓	7.41×10^{-6}	✓	✓
ABCB6	GBM	Brain	1	1.53×10^{-4}	0.57 (0.43, 0.76)	1518	91	✓	7.95×10^{-6}	✓	✗
	Non-GBM	Brain	1	1.31×10^{-3}	0.57 (0.41, 0.80)	1519	69	✓	7.58×10^{-6}	✓	✗
	All	Brain	1	1.62×10^{-6}	0.65 (0.55, 0.78)	1809	96	✓	2.36×10^{-10}	✓	✓
BAIAP2L2	GBM	Brain	1	2.85×10^{-7}	0.60 (0.49, 0.73)	1808	81	✓	1.25×10^{-9}	✓	✓
	Non-GBM	Brain	1	1.82×10^{-2}	0.76 (0.60, 0.95)	1808	18	✓	1.18×10^{-11}	✓	✗
	All	Brain	1	3.31×10^{-11}	0.38 (0.29, 0.51)	2225	40	✓	2.30×10^{-3}	✓	✗
CDKN2B	GBM	Brain	1	1.92×10^{-12}	0.32 (0.23, 0.44)	2222	39	✓	7.16×10^{-3}	✓	✗
	Non-GBM	Brain	1	6.81×10^{-5}	0.47 (0.32, 0.68)	2222	49	✓	6.19×10^{-4}	✓	✗
	All	Blood	1	6.96×10^{-5}	0.23 (0.11, 0.48)	3878	5	✓	1.54×10^{-1}	?	✗
DLGAP5	GBM	Blood	1	1.69×10^{-2}	0.30 (0.11, 0.80)	3855	10	✓	5.75×10^{-2}	?	✗
	Non-GBM	Blood	1	1.27×10^{-1}	0.40 (0.13, 1.29)	3852	4	✓	1.92×10^{-2}	✓	✗
	All	Brain	2	2.58×10^{-14}	0.52 (0.43, 0.61)	3136	80	✓	6.46×10^{-7}	✓	✗
EGFR	GBM	Brain	2	9.99×10^{-20}	0.45 (0.38, 0.53)	2798	81	✓	3.53×10^{-6}	✓	✓
	Non-GBM	Brain	2	2.75×10^{-4}	0.69 (0.57, 0.84)	2798	20	✓	1.18×10^{-8}	✓	✗
	All	Brain	1	3.59×10^{-5}	1.47 (1.23, 1.77)	904	94	✓	1.97×10^{-16}	✓	✓
FAM178B	GBM	Brain	1	1.83×10^{-4}	1.48 (1.21, 1.82)	904	85	✓	2.47×10^{-16}	✓	✗
	Non-GBM	Brain	1	5.69×10^{-3}	1.40 (1.10, 1.78)	904	46	✓	6.16×10^{-17}	✓	✗
	All	Blood	1	2.42×10^{-11}	0.37 (0.27, 0.49)	4667	0	✓	1.91×10^{-12}	✓	✗
GMEB2	GBM	Blood	1	1.68×10^{-11}	0.32 (0.23, 0.45)	4643	0	✓	2.59×10^{-15}	✓	✗
	Non-GBM	Blood	1	4.77×10^{-5}	0.45 (0.31, 0.66)	4647	62	✓	1.27×10^{-10}	✓	✗

HELZ2	All	Blood	1	5.31×10^{-5}	0.34 (0.20, 0.57)	4628	0	✓	6.75×10^{-5}	✓	✗
	GBM	Blood	1	3.94×10^{-6}	0.25 (0.14, 0.45)	4604	0	✓	9.29×10^{-4}	✓	✗
	Non-GBM	Blood	1	1.53×10^{-2}	0.41 (0.20, 0.84)	4609	1	✓	8.42×10^{-6}	✓	✗
JAK1	All	Blood	1	1.71×10^{-5}	5.27 (2.47, 11.23)	3168	38	✓	8.28×10^{-2}	?	✗
	GBM	Blood	1	3.54×10^{-7}	9.25 (3.93, 21.77)	3145	75	✓	4.41×10^{-1}	?	✗
	Non-GBM	Blood	1	2.58×10^{-1}	1.77 (0.66, 4.75)	3118	4	✓	3.59×10^{-4}	✓	✗
	All	Brain	1	6.95×10^{-8}	1.21 (1.13, 1.29)	2064	81	✓	6.89×10^{-139}	✓	✓
	GBM	Brain	1	1.56×10^{-9}	1.27 (1.17, 1.37)	2063	95	✓	1.84×10^{-134}	✓	✓
	Non-GBM	Brain	1	4.57×10^{-2}	1.10 (1.00, 1.20)	2063	11	✓	2.57×10^{-147}	✓	✗
MVB12B	All	Brain	1	5.27×10^{-5}	1.24 (1.12, 1.38)	2400	97	✓	2.53×10^{-23}	✓	✓
	GBM	Brain	1	1.28×10^{-3}	1.27 (1.10, 1.47)	2400	69	✓	6.05×10^{-23}	✓	✗
	Non-GBM	Brain	1	3.30×10^{-3}	1.29 (1.09, 1.52)	2400	43	✓	1.04×10^{-22}	✓	✗
PANK4	All	Blood	1	4.30×10^{-5}	0.46 (0.32, 0.67)	4175	97	✓	4.62×10^{-10}	✓	✓
	GBM	Blood	1	6.17×10^{-4}	0.48 (0.32, 0.73)	4134	80	✓	1.13×10^{-9}	✓	✗
	Non-GBM	Blood	1	2.21×10^{-3}	0.47 (0.29, 0.76)	4155	59	✓	9.44×10^{-10}	✓	✗
PICK1	All	Brain	1	8.82×10^{-7}	1.72 (1.39, 2.14)	1865	97	✓	4.34×10^{-7}	✓	✓
	GBM	Brain	1	6.60×10^{-8}	1.96 (1.54, 2.51)	1864	92	✓	2.13×10^{-6}	✓	✓
	Non-GBM	Brain	1	1.97×10^{-2}	1.41 (1.06, 1.88)	1864	18	✓	2.94×10^{-8}	✓	✗
PRLR	All	Brain	1	9.33×10^{-5}	0.66 (0.54, 0.82)	2090	91	✓	1.13×10^{-7}	✓	✓
	GBM	Brain	1	2.72×10^{-4}	0.65 (0.51, 0.82)	2088	89	✓	1.55×10^{-7}	✓	✗
	Non-GBM	Brain	1	8.32×10^{-4}	0.62 (0.47, 0.82)	2088	74	✓	2.41×10^{-7}	✓	✗
RAVER2	All	Brain	1	7.78×10^{-6}	1.16 (1.09, 1.24)	2058	4	✓	5.17×10^{-129}	✓	✗
	GBM	Brain	1	1.80×10^{-7}	1.22 (1.13, 1.31)	2057	29	✓	5.58×10^{-125}	✓	✗
	Non-GBM	Brain	1	8.53×10^{-2}	1.08 (0.99, 1.18)	2057	6	✓	2.53×10^{-135}	✓	✗
RETREG2	All	Brain	1	9.54×10^{-6}	0.68 (0.57, 0.80)	1511	98	✓	7.90×10^{-11}	✓	✓

	GBM	Brain	1	6.13×10^{-5}	0.67 (0.55, 0.81)	1508	95	✓	9.91×10^{-11}	✓	✓
	Non-GBM	Brain	1	1.83×10^{-3}	0.69 (0.55, 0.87)	1509	62	✓	4.46×10^{-11}	✓	✗
	All	Brain	1	1.44×10^{-16}	0.36 (0.29, 0.46)	3119	96	✓	1.50×10^{-3}	✓	✓
STMN3	GBM	Brain	1	4.55×10^{-19}	0.29 (0.22, 0.38)	3116	97	✓	7.88×10^{-3}	✓	✓
	Non-GBM	Brain	1	1.61×10^{-5}	0.51 (0.37, 0.69)	3116	77	✓	9.41×10^{-5}	✓	✗
	All	Blood	1	1.12×10^{-13}	0.16 (0.10, 0.26)	3733	98	✓	6.36×10^{-2}	?	✗
TP53	GBM	Blood	1	9.87×10^{-13}	0.13 (0.07, 0.22)	3712	99	✓	1.19×10^{-1}	?	✗
	Non-GBM	Blood	1	9.61×10^{-8}	0.17 (0.09, 0.32)	3713	98	✓	3.35×10^{-2}	✓	✓

APPENDIX C

APPENDIX C

Full results from the colocalisation analyses presented in Appendix B and Chapter 4. LocusZoom regional plots of these regions are provided in the Online Appendix.

Gene	Subtype	Tissue	SNPs	H_0	H_1	H_2	H_3	H_4
	All Glioma	Brain	1521	0.04%	0.93%	0.09%	2.12%	96.82%
ABCB6	GBM	Brain	1518	0.17%	4.36%	0.18%	4.42%	90.86%
	Non-GBM	Brain	1519	1.00%	25.09%	0.18%	4.51%	69.22%
	All Glioma	Brain	1809	0.00%	0.11%	0.00%	3.89%	96.01%
BAIAP2L2	GBM	Brain	1808	0.00%	0.01%	0.00%	18.68%	81.31%
	Non-GBM	Brain	1808	0.00%	69.24%	0.00%	13.07%	17.69%
	All Glioma	Brain	2225	0.00%	0.00%	0.20%	60.03%	39.77%
CDKN2B	GBM	Brain	2222	0.00%	0.00%	0.21%	61.17%	38.62%
	Non-GBM	Brain	2222	0.00%	0.02%	0.18%	51.40%	48.40%
	All Glioma	Brain	3878	0.00%	1.23%	0.00%	93.91%	4.86%
DLGAP5	GBM	Brain	3855	0.00%	34.40%	0.00%	56.05%	9.54%
	Non-GBM	Brain	3852	0.00%	62.17%	0.00%	34.00%	3.83%
	All Glioma	Brain	3136	0.00%	0.00%	0.00%	19.87%	80.13%
EGFR	GBM	Brain	2798	0.00%	0.00%	0.00%	19.10%	80.90%
	Non-GBM	Brain	2798	0.00%	64.75%	0.00%	14.90%	20.35%

	All Glioma	Brain	904	0.00%	1.29%	0.00%	4.59%	94.12%
FAM178B	GBM	Brain	904	0.00%	4.42%	0.00%	10.16%	85.42%
	Non-GBM	Brain	904	0.00%	48.50%	0.00%	5.34%	46.16%
	All Glioma	Blood	4667	0.00%	0.00%	0.00%	99.98%	0.02%
GMEB2	GBM	Blood	4643	0.00%	0.00%	0.00%	100.00%	0.00%
	Non-GBM	Blood	4647	0.00%	0.99%	0.00%	37.50%	61.51%
	All Glioma	Blood	4628	0.00%	0.00%	0.00%	100.00%	0.00%
HELZ2	GBM	Blood	4604	0.00%	0.00%	0.00%	100.00%	0.00%
	Non-GBM	Blood	4609	0.00%	2.54%	0.00%	96.35%	1.11%
	All Glioma	Blood	3168	0.00%	0.02%	0.00%	62.01%	37.97%
	GBM	Blood	3145	0.00%	0.00%	0.00%	24.97%	75.03%
	Non-GBM	Blood	3118	0.00%	71.26%	0.00%	24.60%	4.14%
JAK1	All Glioma	Brain	2064	0.00%	0.01%	0.00%	19.05%	80.95%
	GBM	Brain	2063	0.00%	0.00%	0.00%	4.89%	95.11%
	Non-GBM	Brain	2063	0.00%	71.98%	0.00%	16.70%	11.32%
	All Glioma	Brain	2400	0.00%	2.47%	0.00%	0.74%	96.79%
MVB12B	GBM	Brain	2400	0.00%	26.29%	0.00%	4.40%	69.31%
	Non-GBM	Brain	2400	0.00%	33.96%	0.00%	23.29%	42.74%
	All Glioma	Blood	4175	0.00%	1.67%	0.00%	1.39%	96.93%
PANK4	GBM	Blood	4134	0.00%	14.53%	0.00%	5.99%	79.48%
	Non-GBM	Blood	4155	0.00%	30.46%	0.00%	11.07%	58.47%
	All Glioma	Brain	1865	0.00%	0.07%	0.00%	2.59%	97.34%
PICK1	GBM	Brain	1864	0.00%	0.01%	0.00%	8.52%	91.47%
	Non-GBM	Brain	1864	0.00%	69.13%	0.00%	13.32%	17.56%
	All Glioma	Brain	2090	0.00%	2.72%	0.00%	6.79%	90.49%
PRLR	GBM	Brain	2088	0.00%	5.22%	0.00%	6.20%	88.58%
	Non-GBM	Brain	2088	0.00%	18.62%	0.00%	7.29%	74.09%
	All Glioma	Brain	2058	0.00%	0.03%	0.00%	95.69%	4.28%
RAVER2	GBM	Brain	2057	0.00%	0.00%	0.00%	71.74%	28.26%
	Non-GBM	Brain	2057	0.00%	76.49%	0.00%	17.73%	5.79%
	All Glioma	Brain	1511	0.00%	0.53%	0.00%	1.17%	98.30%
RETREG2	GBM	Brain	1508	0.00%	2.60%	0.00%	2.59%	94.81%
	Non-GBM	Brain	1509	0.00%	32.60%	0.00%	5.87%	61.53%
	All Glioma	Brain	3119	0.00%	0.00%	0.00%	3.62%	96.38%
STMN3	GBM	Brain	3116	0.00%	0.00%	0.00%	3.57%	96.43%

	Non-GBM	Brain	3116	0.00%	0.58%	0.00%	21.98%	77.44%
	All Glioma	Blood	3733	0.00%	0.00%	0.00%	2.50%	97.50%
TP53	GBM	Blood	3712	0.00%	0.00%	0.00%	0.41%	99.59%
	Non-GBM	Blood	3713	0.00%	0.04%	0.00%	2.15%	97.81%

APPENDIX D

APPENDIX D

Results for the proteomic Mendelian randomisation analyses presented in Chapter 5. Odds ratios (OR) presented herein are usncaled, whereas the OR presented in the main body of this thesis are scaled by one SD of that dataset.

			Mendelian randomisation			Coloc		Steiger filtering			
Protein	**Subtype**	**Tissue**	**SNPs**	***P***	**OR (95% CI)**	**SNPs**	**H4**	**Dir**	***P***	**Flag**	**Pass All?**
EGFR	All glioma	Brain	1	1.43×10^{-10}	6.29×10^{-9} (1.96×10^{-11}, 2.02×10^{-6})	-	0.50 [a]	✓	0.80	?	✗
	GBM	Brain	1	3.07×10^{-13}	4.24×10^{-12} (3.72×10^{-15}, 4.84×10^{-9})	-	0.50 [a]	✓	0.91	?	✗
	Non-GBM	Brain	1	0.005	1.50×10^{-5} (6.50×10^{-9}, 0.03)	-	0.50 [a]	✓	0.68	?	✗
MON1B	All glioma	Brain	1	6.96×10^{-5}	1.70×10^{-5} (7.39×10^{-8}, 0.004)	-	0.76 [a]	✓	0.66	?	✗
	GBM	Brain	1	9.21×10^{-4}	1.26×10^{-5} (1.58×10^{-8}, 0.01)	-	0.83 [a]	✓	0.67	?	✗
	Non-GBM	Brain	1	0.008	6.14×10^{-5} (4.62×10^{-8}, 0.08)	-	0.76 [a]	✓	0.74	?	✗
TAGLN	All glioma	Brain	1	0.06	4.80 (0.93, 24.64)	-	0.002 [a]	✓	0.01	✓	✗
	GBM	Brain	1	0.80	1.29 (0.18, 9.31)	-	0.01 [a]	✓	0.01	✓	✗
	Non-GBM	Brain	1	0.001	34.84 (4.24, 286.39)	-	0.40 [a]	✓	0.02	✓	✓
TGFBI	All glioma	Blood	1	0.0008	0.87 (0.80, 0.94)	2687	82%	✓	8.69×10^{-64}		✓
	GBM	Blood	1	0.004	0.86 (0.78, 0.95)	2585	37%	✓	1.76×10^{-63}	✓	✗
	Non-GBM	Blood	1	0.014	0.87 (0.78, 0.97)	2519	18%	✓	4.74×10^{-64}	✓	✗

[a] Approximate colocalisation method was used to generate these results.

APPENDIX E

APPENDIX E

List of Read codes and Clinical Practice Research Datalink (CPRD) descriptions used to define cases for analyses in Chapter 6.

Read Code	Description	CPRD Medical Code	CPRD Description
B506.00	Malignant neoplasm of choroid	15991	Malignant neoplasm of choroid
B51..00	Malignant neoplasm of brain	18617	Malignant neoplasm of brain
B51..11	Cerebral tumour - malignant	10851	Cerebral tumour - malignant
B510.00	Malignant neoplasm cerebrum (excluding lobes and ventricles)	15711	Malignant neoplasm cerebrum (excluding lobes and ventricles)
B510000	Malignant neoplasm of basal ganglia	48073	Malignant neoplasm of basal ganglia
B510100	Malignant neoplasm of cerebral cortex	61399	Malignant neoplasm of cerebral cortex
B510300	[Not provided]	99913	Malignant neoplasm of globus pallidus
B510400	Malignant neoplasm of hypothalamus	70942	Malignant neoplasm of hypothalamus

B510500	Malignant neoplasm of thalamus	62126	Malignant neoplasm of thalamus
B510z00	Malignant neoplasm of cerebrum NOS	54133	Malignant neoplasm of cerebrum NOS
B511.00	Malignant neoplasm of frontal lobe	42426	Malignant neoplasm of frontal lobe
B512.00	Malignant neoplasm of temporal lobe	46792	Malignant neoplasm of temporal lobe
B512000	Malignant neoplasm of hippocampus	67236	Malignant neoplasm of hippocampus
B512z00	Malignant neoplasm of temporal lobe NOS	47556	Malignant neoplasm of temporal lobe NOS
B513.00	Malignant neoplasm of parietal lobe	19226	Malignant neoplasm of parietal lobe
B514.00	Malignant neoplasm of occipital lobe	39088	Malignant neoplasm of occipital lobe
B515.00	Malignant neoplasm of cerebral ventricles	52511	Malignant neoplasm of cerebral ventricles
B515000	Malignant neoplasm of choroid plexus	46789	Malignant neoplasm of choroid plexus
B516.00	Malignant neoplasm of cerebellum	45154	Malignant neoplasm of cerebellum
B517.00	Malignant neoplasm of brain stem	44089	Malignant neoplasm of brain stem
B517000	Malignant neoplasm of cerebral peduncle	64557	Malignant neoplasm of cerebral peduncle
B517100	Malignant neoplasm of medulla oblongata	49132	Malignant neoplasm of medulla oblongata
B517200	Malignant neoplasm of midbrain	93537	Malignant neoplasm of midbrain
B517300	Malignant neoplasm of pons	91240	Malignant neoplasm of pons
B517z00	Malignant neoplasm of brain stem NOS	68641	Malignant neoplasm of brain stem NOS
B51y.00	Malignant neoplasm of other parts of brain	71139	Malignant neoplasm of other parts of brain
B51y000	Malignant neoplasm of corpus callosum	59170	Malignant neoplasm of corpus callosum

B51y200	Malignant neoplasm, overlapping lesion of brain	65241	Malignant neoplasm, overlapping lesion of brain
B51yz00	[Not provided]	100733	Malignant neoplasm of other part of brain NOS
B51z.00	Malignant neoplasm of brain NOS	41520	Malignant neoplasm of brain NOS
B521.00	Malignant neoplasm of cerebral meninges	28919	Malignant neoplasm of cerebral meninges
B521200	[Not provided]	109473	Malignant neoplasm of cerebral pia mater
B521z00	Malignant neoplasm of cerebral meninges NOS	70104	Malignant neoplasm of cerebral meninges NOS
B52X.00	Malignant neoplasm of meninges, unspecified	49875	Malignant neoplasm of meninges, unspecified
B542.00	Malignant neoplasm pituitary gland and craniopharyngeal duct	59823	Malignant neoplasm pituitary gland and craniopharyngeal duct
B542000	Malignant neoplasm of pituitary gland	8550	Malignant neoplasm of pituitary gland
B542z00	Malig neop pituitary gland or craniopharyngeal duct NOS	59718	Malig neop pituitary gland or craniopharyngeal duct NOS
B543.00	Malignant neoplasm of pineal gland	42460	Malignant neoplasm of pineal gland
B583.00	Secondary malignant neoplasm of brain and spinal cord	33843	Secondary malignant neoplasm of brain and spinal cord
B583000	Secondary malignant neoplasm of brain	5198	Secondary malignant neoplasm of brain
B583200	Cerebral metastasis	5199	Cerebral metastasis
B583z00	Secondary malignant neoplasm of brain or spinal cord NOS	59375	Secondary malignant neoplasm of brain or spinal cord NOS
B8yy300	Carcinoma in situ of pituitary gland	45909	Carcinoma in situ of pituitary gland
BA06.00	Neoplasm of unspecified nature of brain	1044	Neoplasm of unspecified nature of brain

BB9G.00	[M]Infiltrating ductular carcinoma	7319	[M]Infiltrating ductular carcinoma
BBa3.00	[M]Pineoblastoma	50151	[M]Pineoblastoma
BBb..00	[M]Gliomas	12309	[M]Gliomas
BBb0.00	[M]Glioma, malignant	31574	[M]Glioma, malignant
BBb0.11	[M]Glioma NOS	8523	[M]Glioma NOS
BBb0.12	[M]Gliosarcoma	34252	[M]Gliosarcoma
BBb1.00	[M]Gliomatosis cerebri	38551	[M]Gliomatosis cerebri
BBb2.00	[M]Mixed glioma	68808	[M]Mixed glioma
BBb2.11	[M]Mixed glioma	39386	[M]Mixed glioma
BBb3.00	[M]Subependymal glioma	94267	[M]Subependymal glioma
BBb3.11	[M]Subependymal astrocytoma NOS	90487	[M]Subependymal astrocytoma NOS
BBb3.12	[M]Subependymal astrocytoma NOS	28344	[M]Subependymal astrocytoma NOS
BBb4.00	[M]Subependymal giant cell astrocytoma	49168	[M]Subependymal giant cell astrocytoma
BBb8.00	[M]Ependymoma, anaplastic type	52751	[M]Ependymoma, anaplastic type
BBb8.11	[M]Ependymoblastoma	46769	[M]Ependymoblastoma
BBbA.00	[M]Myxopapillary ependymoma	43114	[M]Myxopapillary ependymoma
BBbB.00	[M]Astrocytoma NOS	8547	[M]Astrocytoma NOS
BBbB.11	[M]Astrocytic glioma	27748	[M]Astrocytic glioma
BBbC.00	[M]Astrocytoma, anaplastic type	8328	[M]Astrocytoma, anaplastic type
BBbE.00	[M]Gemistocytic astrocytoma	45531	[M]Gemistocytic astrocytoma
BBbF.00	[M]Fibrillary astrocytoma	27846	[M]Fibrillary astrocytoma
BBbG.00	[M]Pilocytic astrocytoma	30273	[M]Pilocytic astrocytoma
BBbG.11	[M]Juvenile astrocytoma	61783	[M]Juvenile astrocytoma
BBbG.12	[M]Piloid astrocytoma	98800	[M]Piloid astrocytoma
BBbH.00	[Not provided]	103047	[M]Spongioblastoma NOS
BBbK.00	[M]Astroblastoma	50235	[M]Astroblastoma

BBbL.00	[M]Glioblastoma NOS	23083	[M]Glioblastoma NOS
BBbL.11	[M]Glioblastoma multiforme	9575	[M]Glioblastoma multiforme
BBbM.00	[M]Giant cell glioblastoma	66064	[M]Giant cell glioblastoma
BBbQ.00	[M]Oligodendroglioma NOS	27744	[M]Oligodendroglioma NOS
BBbR.00	[M]Oligodendroglioma, anaplastic type	49186	[M]Oligodendroglioma, anaplastic type
BBbS.00	[M]Oligodendroblastoma	46404	[M]Oligodendroblastoma
BBbT.00	[M]Medulloblastoma NOS	34763	[M]Medulloblastoma NOS
BBbU.00	[M]Desmoplastic medulloblastoma	65952	[M]Desmoplastic medulloblastoma
BBbV.00	[M]Medullomyoblastoma	31767	[M]Medullomyoblastoma
BBbW.00	[M]Cerebellar sarcoma NOS	37473	[M]Cerebellar sarcoma NOS
BBbZ.00	[M]Pleomorphic xanthoastrocytoma	67587	[M]Pleomorphic xanthoastrocytoma
BBba.00	[M]Primitive neuroectodermal tumour	41695	[M]Primitive neuroectodermal tumour
BBbz.00	[M]Glioma NOS	27653	[M]Glioma NOS
BBc6.00	[M]Ganglioglioma	31629	[M]Ganglioglioma
BBc7.11	[M]Neuroastrocytoma	68479	[M]Neuroastrocytoma
BBd1.00	[M]Meningiomatosis NOS	98677	[M]Meningiomatosis NOS
BBd1.11	[M]Diffuse meningiomatosis	95108	[M]Diffuse meningiomatosis
BBd2.00	[M]Meningioma, malignant	27363	[M]Meningioma, malignant
BBd2.11	[M]Leptomeningeal sarcoma	60347	[M]Leptomeningeal sarcoma
BBd2.12	[M]Meningothelial sarcoma	96798	[M]Meningothelial sarcoma
BBd5.00	[M]Psammomatous meningioma	38870	[M]Psammomatous meningioma
BBd7.11	[M]Angioblastic meningioma	46490	[M]Angioblastic meningioma

BBdB.00	[Not provided]	106134	[M]Meningeal sarcomatosis
BBdz.00	[M]Meningioma NOS	47848	[M]Meningioma NOS
BBm0.00	[M]Microglioma	63973	[M]Microglioma
ByuA200	[X]Malignant neoplasm of meninges, unspecified	63925	[X]Malignant neoplasm of meninges, unspecified
ByuA300	[X]Malig neopl, overlap lesion brain & other part of CNS	47633	[X]Malig neopl, overlap lesion brain & other part of CNS

APPENDIX F

APPENDIX F

List of product codes and names used to treat hyperlipidaemia. Used in the fibrate analyses in Chapter 6.

Product Code	**Product Name**	**PPAR-α Targeting?**
25	Simvastatin 20mg tablets	No
28	Atorvastatin 10mg tablets	No
42	Simvastatin 10mg tablets	No
51	Simvastatin 40mg tablets	No
75	Atorvastatin 20mg tablets	No
184	Bezafibrate 200mg tablets	Yes
379	Fluvastatin 20mg capsules	No
490	Pravastatin 10mg tablets	No
602	Bezafibrate 400mg modified-release tablets	Yes
644	Colestyramine 4g oral powder sachets	No
653	Ezetimibe 10mg tablets	No
713	Rosuvastatin 10mg tablets	No
730	Pravastatin 20mg tablets	No
745	Atorvastatin 40mg tablets	No
802	Simvador 40mg tablets (Discovery Pharmaceuticals)	No
818	Simvastatin 20mg/5ml oral solution sugar free	No
1212	Colestipol 5g granules sachets sugar free	No
1214	Bezalip 400mg Tablet (Roche Products Ltd)	Yes

1215	Fenofibrate 100mg Capsule	Yes
1217	Lipantil micro 200 200mg Capsule (Fournier Pharmaceuticals Ltd)	Yes
1219	Pravastatin 40mg tablets	No
1221	Lipostat 10mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)	No
1223	Lipostat 40mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)	No
1322	Clofibrate 500mg capsules	Yes
1324	Bezalip 200mg Tablet (Roche Products Ltd)	Yes
1477	Atromid -s 500mg Capsule (AstraZeneca UK Ltd)	Yes
1716	Questran 4g oral powder sachets (Bristol-Myers Squibb Pharmaceuticals Ltd)	No
1764	Questran Light 4g oral powder sachets (Bristol-Myers Squibb Pharmaceuticals Ltd)	No
2137	Fluvastatin 40mg capsules	No
2215	Lopid 300mg capsules (Pfizer Ltd)	Yes
2435	Lipantil 100mg Capsule (Fournier Pharmaceuticals Ltd)	Yes
2718	Zocor 10mg tablets (Merck Sharp & Dohme Ltd)	No
2955	Lipitor 40mg tablets (Pfizer Ltd)	No
3089	Ciprofibrate 100mg tablets	Yes
3159	Fenofibrate 200mg capsules	Yes
3318	Gemfibrozil 300mg capsules	Yes
3411	Lipitor 10mg tablets (Pfizer Ltd)	No
3690	Lipostat 20mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)	No
4062	Lopid 600mg tablets (Pfizer Ltd)	Yes
4067	Olbetam 250mg capsules (Pfizer Ltd)	No
4920	Fenofibrate micronised 200mg capsules	Yes
4928	Lipantil Micro 200 capsules (BGP Products Ltd)	Yes
5148	Simvastatin 80mg tablets	No
5216	Bezalip mono 400mg Modified-release tablet (Roche Products Ltd)	Yes
5390	Fenofibrate micronised 267mg capsules	Yes
5564	Colestid Orange sachets (Pharmacia Ltd)	No
5775	Atorvastatin 80mg tablets	No

5985	Lescol XL 80mg tablets (Novartis Pharmaceuticals UK Ltd)	No
6120	Ezetrol 10mg tablets (Merck Sharp & Dohme Ltd)	No
6155	Colestyramine with aspartame 4g sugar free powder	No
6168	Zocor 40mg tablets (Merck Sharp & Dohme Ltd)	No
6213	Rosuvastatin 20mg tablets	No
6365	Colestid 5g granules sachets plain (Pfizer Ltd)	No
7196	Zocor 20mg tablets (Merck Sharp & Dohme Ltd)	No
7347	Crestor 10mg tablets (AstraZeneca UK Ltd)	No
7374	Lipitor 20mg tablets (Pfizer Ltd)	No
7540	Lipantil Micro 267 capsules (BGP Products Ltd)	Yes
7544	Niaspan 750mg modified-release tablets (Abbott Laboratories Ltd)	No
7551	Niaspan 1g modified-release tablets (Abbott Laboratories Ltd)	No
7552	Simvastatin 20mg / Ezetimibe 10mg tablets	No
7554	Rosuvastatin 5mg tablets	No
8082	Gemfibrozil 600mg tablets	Yes
8104	Acipimox 250mg capsules	No
8380	Lescol 20mg capsules (Novartis Pharmaceuticals UK Ltd)	No
8706	Modalim 100mg tablets (Sanofi)	Yes
9153	Lescol 40mg capsules (Novartis Pharmaceuticals UK Ltd)	No
9491	Fenofibrate micronised 67mg capsules	Yes
9639	Fenofibrate micronised 160mg tablets	Yes
9716	Supralip 160mg tablets (BGP Products Ltd)	Yes
9897	Rosuvastatin 40mg tablets	No
9920	Simvador 20mg tablets (Discovery Pharmaceuticals)	No
9930	Crestor 40mg tablets (AstraZeneca UK Ltd)	No
10172	Simvastatin 40mg / Ezetimibe 10mg tablets	No
10183	Simvastatin 40mg with ezetimibe 10mg tablet	No
10206	Simvastatin 80mg with ezetimibe 10mg tablet	No
11627	Fluvastatin 80mg modified-release tablets	No
11785	Colestyramine 4g oral powder sachets sugar free	No
11815	Simvastatin 20mg with ezetimibe 10mg tablet	No

11976	Niaspan 500mg modified-release tablets (Abbott Laboratories Ltd)	No
12211	Nicotinic acid 50mg tablets	No
13041	Simvador 10mg tablets (Discovery Pharmaceuticals)	No
14219	Simvastatin 80mg / Ezetimibe 10mg tablets	No
14379	Lipantil Micro 67 capsules (BGP Products Ltd)	Yes
14963	Nicotinic acid 500mg modified-release tablets	No
15252	Crestor 20mg tablets (AstraZeneca UK Ltd)	No
16186	Inegy 10mg/80mg tablets (Merck Sharp & Dohme Ltd)	No
17059	Inegy 10mg/40mg tablets (Merck Sharp & Dohme Ltd)	No
17614	Zimbacol XL 400mg tablets (Archimedes Pharma UK Ltd)	Yes
17683	Lipitor 80mg tablets (Pfizer Ltd)	No
17688	Crestor 5mg tablets (AstraZeneca UK Ltd)	No
17813	Nicotinic acid 100mg Tablet	No
17824	Nicotinic acid 25mg Tablet	No
18081	Colestid Orange 5g granules sachets (Pfizer Ltd)	No
18098	Nicotinic acid 375mg + 500mg + 750mg Modified-release tablet	No
18126	Nicotinic acid 1g modified-release tablets	No
19938	Colestipol with aspartame granules	No
21020	Inegy 10mg/20mg tablets (Merck Sharp & Dohme Ltd)	No
22579	Zocor 80mg tablets (Merck Sharp & Dohme Ltd)	No
23153	Liparol 400 XL tablets (Ashbourne Pharmaceuticals Ltd)	Yes
23153	Liparol 400 XL tablets (Ashbourne Pharmaceuticals Ltd)	No
23634	GEMFIBROZIL	Yes
24009	NICOTINIC ACID 500 MG TAB	No
24084	Colestyramine 4g oral powder sachets sugar free (PLIVA Pharma Ltd)	No
24509	SIMVASTATIN	No
24583	Nicotinic acid 750mg modified-release tablets	No
25018	BEZAFIBRATE	Yes
29213	Bezagen XL 400mg tablets (Mylan Ltd)	Yes
29328	Bezafibrate 200mg tablets (A A H Pharmaceuticals Ltd)	Yes
29438	SIMVASTATIN	No
31221	Bezafibrate 200mg tablets (Mylan Ltd)	Yes

31783	Fenogal 200mg capsules (Genus Pharmaceuticals Ltd)	Yes
31930	Zocor heart-pro 10mg Tablet (McNeil Products Ltd)	No
32110	Colestyramine 4g Sachets (Dominion Pharma)	No
32909	Simvastatin 80mg tablets (A A H Pharmaceuticals Ltd)	No
32921	Pravastatin 10mg Tablet (Dr Reddy's Laboratories (UK) Ltd)	No
33082	Simvastatin 20mg tablets (A A H Pharmaceuticals Ltd)	No
33603	Fibrazate XL 400mg tablets (Sandoz Ltd)	Yes
33944	Bezafibrate 200mg tablets (Teva UK Ltd)	Yes
34181	Bezafibrate 400mg Modified-release tablet (Hillcross Pharmaceuticals Ltd)	Yes
34201	Colestyramine 4g oral powder sachets sugar free (Actavis UK Ltd)	No
34277	Gemfibrozil 600mg tablets (Teva UK Ltd)	Yes
34312	Simvastatin 20mg tablets (Mylan Ltd)	No
34316	Simvastatin 20mg tablets (Teva UK Ltd)	No
34353	Simvastatin 40mg tablets (Mylan Ltd)	No
34366	Simvastatin 20mg tablets (IVAX Pharmaceuticals UK Ltd)	No
34376	Simvastatin 40mg tablets (Teva UK Ltd)	No
34381	Simvastatin 40mg tablets (IVAX Pharmaceuticals UK Ltd)	No
34476	Simvastatin 20mg Tablet (Ratiopharm UK Ltd)	No
34481	Simvastatin 10mg tablets (IVAX Pharmaceuticals UK Ltd)	No
34502	Simvastatin 40mg tablets (A A H Pharmaceuticals Ltd)	No
34535	Simvastatin 10mg tablets (Mylan Ltd)	No
34545	Simvastatin 40mg Tablet (Ratiopharm UK Ltd)	No
34560	Simvastatin 10mg Tablet (Ratiopharm UK Ltd)	No
34746	Simvastatin 20mg Tablet (Niche Generics Ltd)	No
34814	Simvastatin 20mg tablets (Wockhardt UK Ltd)	No
34820	Pravastatin 40mg tablets (A A H Pharmaceuticals Ltd)	No
34879	Simvastatin 40mg Tablet (Niche Generics Ltd)	No
34891	Simvastatin 20mg tablets (Kent Pharmaceuticals Ltd)	No
34907	Simvastatin 40mg tablets (Wockhardt UK Ltd)	No
34955	Simvastatin 10mg tablets (A A H Pharmaceuticals Ltd)	No
34969	Simvastatin 40mg tablets (Actavis UK Ltd)	No
36377	Pravastatin 20mg tablets (Teva UK Ltd)	No

37266	Colesevelam 625mg tablets	No
37434	Simvastatin 40mg tablets (Sandoz Ltd)	No
37953	Cholestagel 625mg tablets (Sanofi)	No
39060	Simvastatin 20mg tablets (Dexcel-Pharma Ltd)	No
39420	Bezalip Mono 400mg modified-release tablets (Teva UK Ltd)	Yes
39576	Bezalip 200mg tablets (Teva UK Ltd)	Yes
39652	Simvastatin 40mg/5ml oral solution sugar free	No
39675	Simvastatin 20mg/5ml Oral suspension (Martindale Pharmaceuticals Ltd)	No
39870	Simvador 80mg tablets (Discovery Pharmaceuticals)	No
40340	Simvastatin 10mg tablets (Teva UK Ltd)	No
40382	Pravastatin 20mg tablets (A A H Pharmaceuticals Ltd)	No
40601	Simvastatin 20mg tablets (Ranbaxy (UK) Ltd)	No
40729	Tredaptive 1000mg/20mg modified-release tablets (Merck Sharp & Dohme Ltd)	No
40885	Nicotinic acid 1g / Laropiprant 20mg modified-release tablets	No
41396	Fenofibrate micronised 200mg capsules (A A H Pharmaceuticals Ltd)	Yes
41657	Simvastatin 80mg tablets (Teva UK Ltd)	No
42801	Bezafibrate xl 400mg Modified-release tablet (Generics (UK) Ltd)	Yes
43218	Pravastatin 10mg tablets (Teva UK Ltd)	No
44528	Simvastatin 20mg/5ml oral suspension sugar free (Rosemont Pharmaceuticals Ltd)	No
44650	Simvastatin 40mg tablets (Dexcel-Pharma Ltd)	No
44878	Ranzolont 10mg tablets (Ranbaxy (UK) Ltd)	No
45219	Simvastatin 40mg tablets (Kent Pharmaceuticals Ltd)	No
45235	Simvastatin 20mg tablets (Sandoz Ltd)	No
45245	Simvastatin 20mg tablets (Actavis UK Ltd)	No
45346	Simvastatin 40mg tablets (Arrow Generics Ltd)	No
46878	Simvastatin 40mg tablets (Almus Pharmaceuticals Ltd)	No
46956	Simvastatin 80mg tablets (Arrow Generics Ltd)	No
47023	Omega-3 fish oil with glycerol Emulsion for infusion	No
47065	Atorvastatin 20mg chewable tablets sugar free	No
47090	Atorvastatin 10mg chewable tablets sugar free	No
47630	Lipitor 20mg chewable tablets (Pfizer Ltd)	No

47721	Lipitor 10mg chewable tablets (Pfizer Ltd)	No
47774	Simvastatin 10mg tablets (Arrow Generics Ltd)	No
47935	Fenofibrate 200mg Capsule (Teva UK Ltd)	Yes
47948	Simvastatin 10mg tablets (Tillomed Laboratories Ltd)	No
47988	Pravastatin 40mg tablets (Mylan Ltd)	No
48018	Simvastatin 20mg tablets (Arrow Generics Ltd)	No
48051	Simvastatin 10mg tablets (Kent Pharmaceuticals Ltd)	No
48058	Simvastatin 10mg tablets (Ranbaxy (UK) Ltd)	No
48078	Simvastatin 10mg tablets (Actavis UK Ltd)	No
48097	Pravastatin 40mg tablets (Teva UK Ltd)	No
48221	Simvastatin 20mg/5ml oral suspension sugar free	No
48346	Atorvastatin 60mg tablets	No
48431	Simvastatin 40mg/5ml oral suspension sugar free	No
48518	Atorvastatin 10mg/5ml oral solution	No
48585	Nicotinic acid 500mg capsules	No
48867	Simvastatin 40mg tablets (Alliance Healthcare (Distribution) Ltd)	No
48973	Atorvastatin 30mg tablets	No
49061	Simvastatin 40mg tablets (Bristol Laboratories Ltd)	No
49062	Simvastatin 20mg tablets (Alliance Healthcare (Distribution) Ltd)	No
49558	Atorvastatin 20mg tablets (A A H Pharmaceuticals Ltd)	No
49587	Simvastatin 80mg tablets (Almus Pharmaceuticals Ltd)	No
49609	Bezafibrate 400mg Modified-release tablet (Sandoz Ltd)	Yes
49751	Atorvastatin 40mg tablets (Alliance Healthcare (Distribution) Ltd)	No
50071	Fenofibrate 160mg Tablet (Teva UK Ltd)	Yes
50236	Atorvastatin 10mg tablets (Zentiva)	No
50272	Atorvastatin 40mg tablets (Pfizer Ltd)	No
50483	Simvastatin 40mg tablets (Relonchem Ltd)	No
50564	Simvastatin 20mg tablets (Relonchem Ltd)	No
50670	Simvastatin 40mg tablets (Aurobindo Pharma Ltd)	No
50703	Simvastatin 40mg tablets (Accord Healthcare Ltd)	No
50754	Simvastatin 20mg tablets (Medreich Plc)	No
50788	Atorvastatin 20mg tablets (Pfizer Ltd)	No
50790	Atorvastatin 20mg tablets (Dexcel-Pharma Ltd)	No
50882	Simvastatin 40mg tablets (Somex Pharma)	No
50925	Pravastatin 10mg tablets (Sigma Pharmaceuticals Plc)	No

50963	Atorvastatin 40mg tablets (Teva UK Ltd)	No
51085	Simvastatin 10mg tablets (Medreich Plc)	No
51134	Atorvastatin 10mg tablets (A A H Pharmaceuticals Ltd)	No
51155	Natures Aid Omega-3 Fish Oil 1000mg softgels capsules (Natures Aid Ltd)	No
51166	Simvastatin 40mg tablets (Medreich Plc)	No
51200	Atorvastatin 40mg tablets (Arrow Generics Ltd)	No
51233	Simvastatin 10mg tablets (Alliance Healthcare (Distribution) Ltd)	No
51359	Atorvastatin 20mg tablets (Arrow Generics Ltd)	No
51483	Simvastatin 20mg tablets (Aurobindo Pharma Ltd)	No
51622	Atorvastatin 20mg tablets (Consilient Health Ltd)	No
51676	Pravastatin 40mg tablets (Medreich Plc)	No
51715	Simvastatin 10mg tablets (Sigma Pharmaceuticals Plc)	No
51822	Natures Aid Omega-3 Fish Oil 500mg capsules (Natures Aid Ltd)	No
51876	Atorvastatin 40mg tablets (Consilient Health Ltd)	No
51890	Pravastatin 20mg tablets (Medreich Plc)	No
52097	Atorvastatin 40mg tablets (Wockhardt UK Ltd)	No
52098	Simvastatin 40mg tablets (Ranbaxy (UK) Ltd)	No
52168	Atorvastatin 20mg tablets (Aspire Pharma Ltd)	No
52211	Atorvastatin 20mg tablets (Actavis UK Ltd)	No
52257	Simvastatin 20mg tablets (Accord Healthcare Ltd)	No
52397	Atorvastatin 40mg tablets (Dr Reddy's Laboratories (UK) Ltd)	No
52398	Atorvastatin 40mg tablets (A A H Pharmaceuticals Ltd)	No
52459	Atorvastatin 80mg tablets (Actavis UK Ltd)	No
52460	Atorvastatin 40mg tablets (Aspire Pharma Ltd)	No
52625	Simvastatin 10mg tablets (Wockhardt UK Ltd)	No
52676	Simvastatin 10mg/5ml oral suspension	No
52755	Pravastatin 20mg tablets (Alliance Healthcare (Distribution) Ltd)	No
52812	Simvastatin 20mg tablets (Sigma Pharmaceuticals Plc)	No
52814	Bezafibrate 400mg modified-release tablets (Alliance Healthcare (Distribution) Ltd)	Yes
52821	Atorvastatin 80mg tablets (Dr Reddy's Laboratories (UK) Ltd)	No

52953	Simvastatin 20mg tablets (Bristol Laboratories Ltd)	No
52962	Simvastatin 80mg tablets (Medreich Plc)	No
53087	Simvastatin 20mg tablets (Somex Pharma)	No
53250	Modalim 100mg tablets (Lexon (UK) Ltd)	Yes
53340	Zocor 40mg tablets (Lexon (UK) Ltd)	No
53415	Simvastatin 10mg tablets (Aurobindo Pharma Ltd)	No
53460	Crestor 10mg tablets (DE Pharmaceuticals)	No
53594	Lipitor 80mg tablets (Mawdsley-Brooks & Company Ltd)	No
53676	Simvastatin 20mg tablets (Tillomed Laboratories Ltd)	No
53770	Fluvastatin 40mg capsules (A A H Pharmaceuticals Ltd)	No
53772	Atorvastatin 80mg tablets (Alliance Healthcare (Distribution) Ltd)	No
53822	Simvastatin 10mg tablets (Bristol Laboratories Ltd)	No
53887	Atorvastatin 40mg tablets (Actavis UK Ltd)	No
53890	Atorvastatin 80mg tablets (Pfizer Ltd)	No
53908	Simvastatin 10mg tablets (Dexcel-Pharma Ltd)	No
53966	Simvastatin 40mg tablets (Phoenix Healthcare Distribution Ltd)	No
54240	Simvastatin 40mg tablets (Sigma Pharmaceuticals Plc)	No
54266	Simvastatin 20mg/5ml oral suspension	No
54435	Pravastatin 40mg tablets (Almus Pharmaceuticals Ltd)	No
54493	Simvastatin 10mg tablets (Relonchem Ltd)	No
54535	Atorvastatin 10mg tablets (Pfizer Ltd)	No
54606	Simvastatin 20mg/5ml oral suspension sugar free (A A H Pharmaceuticals Ltd)	No
54607	Pravastatin 20mg tablets (Almus Pharmaceuticals Ltd)	No
54655	Simvastatin 10mg tablets (Accord Healthcare Ltd)	No
54819	Simvastatin 40mg/5ml oral suspension sugar free (Rosemont Pharmaceuticals Ltd)	No
54947	Simvastatin 20mg tablets (Almus Pharmaceuticals Ltd)	No
54976	Simvastatin 10mg tablets (Somex Pharma)	No
54985	Simvastatin 40mg/5ml oral suspension	No
54992	Atorvastatin 10mg/5ml oral suspension	No
55032	Atorvastatin 10mg tablets (Dexcel-Pharma Ltd)	No
55034	Atorvastatin 40mg/5ml oral suspension	No
55444	Atorvastatin 40mg tablets (Zentiva)	No

55452	Simvastatin 20mg tablets (Phoenix Healthcare Distribution Ltd)	No
55727	Atorvastatin 10mg tablets (Actavis UK Ltd)	No
55912	Pravastatin 40mg tablets (Alliance Healthcare (Distribution) Ltd)	No
56065	Simvastatin 20mg/5ml oral suspension sugar free (Waymade Healthcare Plc)	No
56097	Atorvastatin 10mg chewable tablets sugar free	No
56146	Pravastatin 10mg tablets (Waymade Healthcare Plc)	No
56165	Atorvastatin 20mg chewable tablets sugar free	No
56182	Atorvastatin 80mg tablets (Zentiva)	No
56248	Atorvastatin 20mg tablets (Sigma Pharmaceuticals Plc)	No
56481	Zocor 10mg tablets (Sigma Pharmaceuticals Plc)	No
56494	Zocor 20mg tablets (Sigma Pharmaceuticals Plc)	No
56564	Atorvastatin 20mg tablets (Almus Pharmaceuticals Ltd)	No
56607	Pravastatin 20mg tablets (Waymade Healthcare Plc)	No
56735	Pravastatin 20mg tablets (Mylan Ltd)	No
56841	Atorvastatin 40mg tablets (Dexcel-Pharma Ltd)	No
56893	Pravastatin 40mg tablets (Accord Healthcare Ltd)	No
56916	Pravastatin 40mg tablets (PLIVA Pharma Ltd)	No
57108	Pravastatin 40mg tablets (Waymade Healthcare Plc)	No
57117	Atorvastatin 80mg tablets (Waymade Healthcare Plc)	No
57137	Pravastatin 10mg tablets (Almus Pharmaceuticals Ltd)	No
57219	Fenofibrate micronised 200mg capsules (Sandoz Ltd)	Yes
57296	Pravastatin 20mg tablets (Phoenix Healthcare Distribution Ltd)	No
57329	Simvastatin 25mg/5ml oral suspension	No
57348	Atorvastatin 10mg tablets (Consilient Health Ltd)	No
57397	Pravastatin 10mg tablets (Accord Healthcare Ltd)	No
57489	Ciprofibrate 100mg tablets (Zentiva)	Yes
57568	Zocor 10mg tablets (Lexon (UK) Ltd)	No
57763	Rosuvastatin 10mg tablets (Waymade Healthcare Plc)	No
57834	Atorvastatin 40mg tablets (DE Pharmaceuticals)	No
57836	Atorvastatin 80mg tablets (Teva UK Ltd)	No
57999	Crestor 40mg tablets (Lexon (UK) Ltd)	No
58041	Atorvastatin 20mg tablets (Teva UK Ltd)	No
58110	Atorvastatin 20mg tablets (Zentiva)	No

58315	Simvastatin 20mg tablets (Waymade Healthcare Plc)	No
58394	Atorvastatin 20mg tablets (Alliance Healthcare (Distribution) Ltd)	No
58418	Atorvastatin 80mg tablets (A A H Pharmaceuticals Ltd)	No
58617	Rosuvastatin 20mg/5ml oral suspension	No
58635	Bezalip Mono 400mg modified-release tablets (DE Pharmaceuticals)	Yes
58742	Atorvastatin 80mg tablets (Arrow Generics Ltd)	No
58755	Simvastatin 10mg tablets (Phoenix Healthcare Distribution Ltd)	No
58834	Atorvastatin 10mg tablets (DE Pharmaceuticals)	No
58868	Atorvastatin 10mg tablets (Sigma Pharmaceuticals Plc)	No
59002	Bezafibrate 400mg modified-release tablets (DE Pharmaceuticals)	Yes
59272	Atorvastatin 20mg tablets (Waymade Healthcare Plc)	No
59278	Fluvastatin 20mg capsules (Zentiva)	No
59331	Lipitor 10mg tablets (DE Pharmaceuticals)	No
59357	Atorvastatin 10mg tablets (Ranbaxy (UK) Ltd)	No
59446	Atorvastatin 40mg tablets (Almus Pharmaceuticals Ltd)	No
59447	Crestor 20mg tablets (Waymade Healthcare Plc)	No
59452	Rosuvastatin 5mg tablets (Waymade Healthcare Plc)	No
59508	Pravastatin 20mg tablets (Accord Healthcare Ltd)	No
59776	Atorvastatin 80mg tablets (Aspire Pharma Ltd)	No
59859	Atorvastatin 10mg tablets (Teva UK Ltd)	No
60101	Colestyramine 4g oral powder sachets sugar free (Teva UK Ltd)	No
60160	Rosuvastatin 5mg tablets (Mawdsley-Brooks & Company Ltd)	No
60251	Pravastatin 10mg tablets (Sandoz Ltd)	No
60342	Berocca effervescent tablets tropical (Bayer Plc)	No
60385	Bezalip Mono 400mg modified-release tablets (Lexon (UK) Ltd)	Yes
60464	Atorvastatin 20mg/5ml oral suspension	No
60511	Atorvastatin 40mg tablets (Ranbaxy (UK) Ltd)	No
60607	Atorvastatin 80mg tablets (DE Pharmaceuticals)	No
60788	Fenofibrate micronised 267mg capsules (Zentiva)	Yes

60989	Atorvastatin 80mg tablets (Phoenix Healthcare Distribution Ltd)	No
61087	Questran Light 4g oral powder sachets (Mawdsley-Brooks & Company Ltd)	No
61134	Pravastatin 20mg tablets (Sigma Pharmaceuticals Plc)	No
61149	Atorvastatin 10mg tablets (Waymade Healthcare Plc)	No
61155	Simvastatin 40mg/5ml oral suspension sugar free (A A H Pharmaceuticals Ltd)	No
61321	Simvastatin 10mg tablets (Sandoz Ltd)	No
61360	Simvastatin 10mg tablets (Almus Pharmaceuticals Ltd)	No
61665	Simvastatin 10mg tablets (Waymade Healthcare Plc)	No
62137	Simvastatin 40mg tablets (Waymade Healthcare Plc)	No
62148	Fluvastatin 20mg capsules (Actavis UK Ltd)	No
62219	Atorvastatin 20mg tablets (DE Pharmaceuticals)	No
62429	Atorvastatin 20mg tablets (DE Pharmaceuticals)	No
62476	Atorvastatin 80mg tablets (Almus Pharmaceuticals Ltd)	No
62979	Pravastatin 40mg tablets (Kent Pharmaceuticals Ltd)	No
63074	Pravastatin 20mg tablets (PLIVA Pharma Ltd)	No
63140	Atorvastatin 10mg tablets (Alliance Healthcare (Distribution) Ltd)	No
63249	Atorvastatin 80mg tablets (Consilient Health Ltd)	No
63343	Generic Crampex tablets	No
63469	Atorvastatin 30mg tablets (Consilient Health Ltd)	No
63737	Fenofibrate micronised 267mg capsules (Sigma Pharmaceuticals Plc)	Yes
63787	Pravastatin 10mg tablets (Tillomed Laboratories Ltd)	No
64067	Atorvastatin 20mg/5ml oral solution	No
64104	Simvastatin 20mg tablets (Crescent Pharma Ltd)	No
64180	Simvastatin 10mg tablets (Crescent Pharma Ltd)	No
64307	Simvastatin 40mg tablets (Crescent Pharma Ltd)	No
64503	Bezalip Mono 400mg modified-release tablets (Waymade Healthcare Plc)	Yes
64702	Atorvastatin 30mg tablets (A A H Pharmaceuticals Ltd)	No
64810	Atorvastatin 40mg tablets (Phoenix Healthcare Distribution Ltd)	No
64825	Atorvastatin 10mg tablets (Phoenix Healthcare Distribution Ltd)	No

64868	Atorvastatin 40mg tablets (Sigma Pharmaceuticals Plc)	No
64933	Fenofibrate micronised 267mg capsules (Ranbaxy (UK) Ltd)	Yes
64968	Simvastatin 10mg tablets (DE Pharmaceuticals)	No
64984	Fenofibrate micronised 160mg tablets (Phoenix Healthcare Distribution Ltd)	Yes
65181	Simvastatin 40mg tablets (DE Pharmaceuticals)	No
65193	Atorvastatin 20mg tablets (Ranbaxy (UK) Ltd)	No
65572	Fenofibrate micronised 160mg tablets (Genus Pharmaceuticals Ltd)	Yes
65679	Simvastatin 20mg tablets (DE Pharmaceuticals)	No
65901	Simvastatin 40mg tablets (Zentiva)	No
65925	Simvastatin 20mg/5ml oral suspension sugar free (Alliance Healthcare (Distribution) Ltd)	No
66087	Modalim 100mg tablets (Mawdsley-Brooks & Company Ltd)	Yes
66425	Bezafibrate 400mg modified-release tablets (A A H Pharmaceuticals Ltd)	Yes
66505	Fenofibrate 145mg / Simvastatin 40mg tablets	Yes
66564	Bezafibrate 400mg modified-release tablets (Phoenix Healthcare Distribution Ltd)	Yes
66780	Fenofibrate 145mg / Simvastatin 20mg tablets	Yes
66963	Atorvastatin 80mg tablets (Sigma Pharmaceuticals Plc)	No
67098	Simvastatin 10mg tablets (Brown & Burk UK Ltd)	No
67157	Fenofibrate micronised 200mg capsules (Phoenix Healthcare Distribution Ltd)	Yes
67328	Lescol XL 80mg tablets (Mawdsley-Brooks & Company Ltd)	No
67329	Lipantil Micro 267 capsules (DE Pharmaceuticals)	Yes
67402	Atorvastatin 40mg tablets (Kent Pharmaceuticals Ltd)	No
67573	Atorvastatin 10mg tablets (DE Pharmaceuticals)	No
67660	Atorvastatin 80mg tablets (Ranbaxy (UK) Ltd)	No
67745	Simvastatin 10mg tablets (Zentiva)	No
67773	Simvastatin 20mg tablets (Zentiva)	No
67829	Pravastatin 20mg tablets (Sandoz Ltd)	No
67846	Atorvastatin 10mg tablets (Almus Pharmaceuticals Ltd)	No
67883	Lomitapide 5mg capsules	No

68023	Atorvastatin 10mg tablets (Aspire Pharma Ltd)	No
68048	Atorvastatin 20mg tablets (Phoenix Healthcare Distribution Ltd)	No
68156	Pravastatin 10mg tablets (A A H Pharmaceuticals Ltd)	No
68386	Colestyramine 4g oral powder sachets (J M McGill Ltd)	No
68467	Atorvastatin 20mg tablets (Kent Pharmaceuticals Ltd)	No
68563	Simvastatin 40mg tablets (Brown & Burk UK Ltd)	No
68686	Simvastatin 20mg tablets (Genesis Pharmaceuticals Ltd)	No
68785	Atorvastatin 10mg tablets (Mylan Ltd)	No
68827	Atorvastatin 20mg tablets (Mylan Ltd)	No
69093	Atorvastatin 80mg tablets (Wockhardt UK Ltd)	No
69413	Simvastatin 20mg tablets (Brown & Burk UK Ltd)	No
69427	Atorvastatin 40mg tablets (Mylan Ltd)	No

APPENDIX G

APPENDIX G

List of product codes and names used to treat type II diabetes (oral medications only). Used in the glitazone analyses in Chapter 6.

Product Code	**Product Name**	**PPAR-γ Targeting?**
23	Metformin 500mg tablets	No
32	Gliclazide 80mg tablets	No
93	Metformin 850mg tablets	No
240	Rifampicin 150mg capsules	No
469	Rosiglitazone 4mg tablets	Yes
479	Acarbose 50mg tablets	No
547	Glipizide 2.5mg tablets	No
548	Pioglitazone 15mg tablets	Yes
735	Metformin 100mg/ml Oral solution	No
1254	Glibenclamide 5mg tablets	No
1964	Diamicron 80mg tablets (Servier Laboratories Ltd)	No
1965	Tolbutamide 500mg tablets	No
2219	Glibenclamide 2.5mg tablets	No
4862	Diabetamide 2.5mg tablets (Ashbourne Pharmaceuticals Ltd)	No
5174	Acarbose 100mg tablets	No
5227	Rosiglitazone 8mg tablets	Yes
5276	Glimepiride 1mg tablets	No

5316	Glimepiride 4mg tablets	No
5353	Glimepiride 2mg tablets	No
5621	Glucobay 50mg tablets (Bayer Plc)	No
5627	Gliclazide 30mg modified-release tablets	No
5636	Glipizide 5mg tablets	No
5678	Nateglinide 120mg tablets	No
5989	Nateglinide 180mg tablets	No
6337	Glimepiride 3mg tablets	No
6855	Avandamet 2mg/500mg tablets (GlaxoSmithKline UK Ltd)	Yes
7048	Metformin 500mg modified-release tablets	No
7166	Glucophage 500mg tablets (Merck Serono Ltd)	No
7284	Amaryl 2mg tablets (Zentiva)	No
7325	Avandamet 4mg/1000mg tablets (GlaxoSmithKline UK Ltd)	Yes
7332	Amaryl 1mg tablets (Zentiva)	No
7375	Rosiglitazone 4mg / Metformin 1g tablets	Yes
7409	Amaryl 3mg tablets (Zentiva)	No
7610	Glucophage 850mg tablets (Merck Serono Ltd)	No
7744	Daonil 5mg tablets (Sanofi)	No
7818	Rifater tablets (Sanofi)	No
7912	Semi-Daonil 2.5mg tablets (Sanofi)	No
8976	Euglucon 2.5mg tablets (Aventis Pharma)	No
9105	Glucobay 100mg tablets (Bayer Plc)	No
9662	Avandia 4mg tablets (GlaxoSmithKline UK Ltd)	Yes
9691	Rifampicin with isoniazid & pyrazinamide tablet	No
9699	Pioglitazone 30mg tablets	Yes
9707	Repaglinide 1mg tablets	No
9748	Repaglinide 2mg tablets	No
9865	Repaglinide 500microgram tablets	No
10051	Pioglitazone 45mg tablets	Yes
11284	Amaryl 4mg tablets (Zentiva)	No
11316	NovoNorm 500microgram tablets (Novo Nordisk Ltd)	No
11321	NovoNorm 1mg tablets (Novo Nordisk Ltd)	No
11366	NovoNorm 2mg tablets (Novo Nordisk Ltd)	No
11483	Nateglinide 60mg tablets	No
11601	Rosiglitazone 2mg / Metformin 500mg tablets	Yes
11604	Rosiglitazone 1mg / Metformin 500mg tablets	Yes

11609	Metformin with rosiglitazone 500mg + 1mg Tablet	Yes
11610	Metformin with rosiglitazone 500mg + 2mg Tablet	Yes
11695	Diamicron 30mg MR tablets (Servier Laboratories Ltd)	No
11717	Rosiglitazone 2mg / Metformin 1g tablets	Yes
11737	Metformin with rosiglitazone 1000mg + 4mg Tablet	Yes
11760	Metformin with rosiglitazone 1000mg + 2mg Tablet	Yes
11946	Tolbutamide 50mg/ml Injection	No
11990	Metformin 500mg/5ml oral solution sugar free	No
12455	Rastinon 500mg Tablet (Hoechst Marion Roussel)	No
12513	Glibenese 5mg tablets (Pfizer Ltd)	No
13331	Euglucon 5mg tablets (Sanofi)	No
13628	Romozin 400mg Tablet (Glaxo Wellcome UK Ltd)	Yes
14164	Avandamet 2mg/1000mg tablets (GlaxoSmithKline UK Ltd)	Yes
15232	Avandia 8mg tablets (GlaxoSmithKline UK Ltd)	Yes
15374	Gliclazide 40mg/5ml oral suspension	No
15955	Starlix 120mg tablets (Novartis Pharmaceuticals UK Ltd)	No
16044	Glucophage SR 500mg tablets (Merck Serono Ltd)	No
16602	Calabren 2.5mg Tablet (Berk Pharmaceuticals Ltd)	No
17343	Gliclazide 80mg tablets (A A H Pharmaceuticals Ltd)	No
17580	Avandamet 1mg/500mg tablets (GlaxoSmithKline UK Ltd)	Yes
17698	Minodiab 5mg tablets (Pfizer Ltd)	No
17706	Minodiab 2.5mg tablets (Pfizer Ltd)	No
17770	Glucagon novo 10mg Injection (Novo Nordisk Ltd)	No
18220	Pioglitazone 15mg / Metformin 850mg tablets	Yes
19472	Actos 45mg tablets (Takeda UK Ltd)	Yes
20287	Actos 15mg tablets (Takeda UK Ltd)	Yes
20889	Actos 30mg tablets (Takeda UK Ltd)	Yes
21424	Glibenclamide 5mg/5ml oral suspension	No
21564	Gliclazide 80mg tablets (Wockhardt UK Ltd)	No
21832	Diabetamide 5mg tablets (Ashbourne Pharmaceuticals Ltd)	No
21892	Diaglyk 80mg tablets (Ashbourne Pharmaceuticals Ltd)	No
22239	Glucagon lilly 1mg Injection (Eli Lilly and Company Ltd)	No

23945	Starlix 60mg tablets (Novartis Pharmaceuticals UK Ltd)	No
25636	Libanil 2.5mg Tablet (Approved Prescription Services Ltd)	No
25678	Glucamet 500mg Tablet (Opus Pharmaceuticals Ltd)	No
26218	Calabren 5mg Tablet (Berk Pharmaceuticals Ltd)	No
26258	Glucamet 850mg Tablet (Opus Pharmaceuticals Ltd)	No
27125	Starlix 180mg tablets (Novartis Pharmaceuticals UK Ltd)	No
27501	Orabet 500mg Tablet (Lagap)	No
28708	Malix 2.5mg Tablet (Lagap)	No
29326	Glipizide 5mg tablets (Mylan Ltd)	No
29939	Gliclazide 80mg tablets (Mylan Ltd)	No
30316	Metformin with pioglitazone 850mg + 15mg Tablet	Yes
30460	Malix 5mg Tablet (Lagap)	No
31077	Competact 15mg/850mg tablets (Takeda UK Ltd)	Yes
31146	Metsol 500mg/5ml oral solution (Kappin Ltd)	No
31212	Gliclazide 80mg tablets (Actavis UK Ltd)	No
31474	Libanil 5mg Tablet (Approved Prescription Services Ltd)	No
33087	Metformin 500mg tablets (Actavis UK Ltd)	No
33562	Duclazide 80mg Tablet (Dumex Ltd)	No
33673	Tolbutamide 500mg tablets (Actavis UK Ltd)	No
33674	Metformin 850mg tablets (A A H Pharmaceuticals Ltd)	No
34004	Metformin 500mg tablets (IVAX Pharmaceuticals UK Ltd)	No
34020	Metformin 850mg tablets (IVAX Pharmaceuticals UK Ltd)	No
34135	Metformin 500mg Tablet (M & A Pharmachem Ltd)	No
34323	Metformin 500mg tablets (A A H Pharmaceuticals Ltd)	No
34399	Gliclazide 80mg tablets (IVAX Pharmaceuticals UK Ltd)	No
34504	Metformin 500mg tablets (Wockhardt UK Ltd)	No
34507	Glibenclamide 2.5mg tablets (Wockhardt UK Ltd)	No
34563	Glibenclamide 5mg tablets (Wockhardt UK Ltd)	No
34598	Metformin 500mg tablets (Mylan Ltd)	No
34676	Glibenclamide 2.5mg tablets (A A H Pharmaceuticals Ltd)	No

34697	Metformin 850mg tablets (Wockhardt UK Ltd)	No
34706	Glibenclamide 2.5mg tablets (IVAX Pharmaceuticals UK Ltd)	No
34742	Metformin 850mg tablets (Teva UK Ltd)	No
34802	Glipizide 5mg tablets (IVAX Pharmaceuticals UK Ltd)	No
34836	Metformin 850mg tablets (Actavis UK Ltd)	No
34917	Metformin 500mg tablets (Teva UK Ltd)	No
34932	Gliclazide 80mg tablets (Genus Pharmaceuticals Ltd)	No
34957	Tolbutamide 500mg tablets (A A H Pharmaceuticals Ltd)	No
35022	Sitagliptin 100mg tablets	No
35144	Byetta 5micrograms/0.02ml solution for injection 1.2ml pre-filled disposable devices (AstraZeneca UK Ltd)	No
35149	Exenatide 10micrograms/0.04ml solution for injection 2.4ml pre-filled disposable devices	No
35150	Byetta 10micrograms/0.04ml solution for injection 2.4ml pre-filled disposable devices (AstraZeneca UK Ltd)	No
35251	Exenatide 5micrograms/0.02ml solution for injection 1.2ml pre-filled disposable devices	No
35462	Januvia 100mg tablets (Merck Sharp & Dohme Ltd)	No
35561	Prandin 2mg tablets (Novo Nordisk Ltd)	No
36774	Prandin 1mg tablets (Novo Nordisk Ltd)	No
36856	Gliclazide 80mg tablets (Sandoz Ltd)	No
36948	Prandin 0.5mg tablets (Novo Nordisk Ltd)	No
37617	Rosiglitazone 2mg tablet	Yes
37874	Vildagliptin 50mg / Metformin 850mg tablets	No
37875	Vildagliptin 50mg tablets	No
37902	Vildagliptin 50mg / Metformin 1g tablets	No
38355	Metformin 750mg modified-release tablets	No
38400	Glucophage SR 750mg tablets (Merck Serono Ltd)	No
38551	Eucreas 50mg/1000mg tablets (Novartis Pharmaceuticals UK Ltd)	No
39149	Galvus 50mg tablets (Novartis Pharmaceuticals UK Ltd)	No
39203	Eucreas 50mg/850mg tablets (Novartis Pharmaceuticals UK Ltd)	No
39560	Bolamyn SR 500mg tablets (Teva UK Ltd)	No

39598	Metformin 1g modified-release tablets	No
39729	Glucophage SR 1000mg tablets (Merck Serono Ltd)	No
39988	Metformin 500mg oral powder sachets sugar free	No
40007	Glucophage 1000mg oral powder sachets (Merck Serono Ltd)	No
40110	Glucophage 500mg oral powder sachets (Merck Serono Ltd)	No
40233	Metformin 1g oral powder sachets sugar free	No
40365	Glimepiride 1mg tablets (Actavis UK Ltd)	No
40425	Nazdol MR 30mg tablets (Teva UK Ltd)	No
40642	Victoza 6mg/ml solution for injection 3ml pre-filled pen (Novo Nordisk Ltd)	No
40693	Liraglutide 6mg/ml solution for injection 3ml pre-filled disposable devices	No
41204	Saxagliptin 5mg tablets	No
41431	Onglyza 5mg tablets (AstraZeneca UK Ltd)	No
41558	Glibenclamide 5mg tablets (Teva UK Ltd)	No
41559	Glibenclamide 5mg tablets (A A H Pharmaceuticals Ltd)	No
41593	Glibenclamide 2.5mg tablets (Teva UK Ltd)	No
42161	Orabet 500mg Tablet (Sandoz Ltd)	No
42790	Gliclazide 80mg Tablet (Merck Generics (UK) Ltd)	No
43065	Gliclazide 40mg tablets	No
43270	Metformin 500mg/5ml oral solution sugar free (Rosemont Pharmaceuticals Ltd)	No
43465	Zicron 40mg tablets (Bristol Laboratories Ltd)	No
43619	Metformin 1g / Sitagliptin 50mg tablets	No
43684	Janumet 50mg/1000mg tablets (Merck Sharp & Dohme Ltd)	No
44250	Metformin 500mg/5ml Oral solution (Hillcross Pharmaceuticals Ltd)	No
44304	Glyconon 500mg Tablet (DDSA Pharmaceuticals Ltd)	No
44473	Edicil MR 30mg tablets (Teva UK Ltd)	No
44738	Niddaryl 1mg tablets (Dee Pharmaceuticals Ltd)	No
45215	Gliclazide 80mg Tablet (Neo Laboratories Ltd)	No
45581	Metabet SR 500mg tablets (Morningside Healthcare Ltd)	No
45775	Saxagliptin 2.5mg tablets	No

45821	Onglyza 2.5mg tablets (AstraZeneca UK Ltd)	No
45831	Dacadis MR 30mg tablets (Mylan Ltd)	No
46458	Exenatide 2mg powder and solvent for prolonged-release suspension for injection vials	No
46469	Bydureon 2mg powder and solvent for prolonged-release suspension for injection vials (AstraZeneca UK Ltd)	No
46665	Linagliptin 5mg tablets	No
46716	Trajenta 5mg tablets (Boehringer Ingelheim Ltd)	No
46927	Tolbutamide 500mg tablets (Teva UK Ltd)	No
46989	Metabet SR 1000mg tablets (Morningside Healthcare Ltd)	No
47074	Gliclazide 80mg/5ml oral suspension	No
47894	Nazdol MR 30mg tablets (Consilient Health Ltd)	No
47939	Glucient SR 500mg tablets (Consilient Health Ltd)	No
48056	Gliclazide 80mg tablets (Sovereign Medical Ltd)	No
48139	Pioglitazone 30mg tablets (A A H Pharmaceuticals Ltd)	Yes
48149	Metformin 500mg tablets (Almus Pharmaceuticals Ltd)	No
48401	Sitagliptin 50mg tablets	No
48533	Sitagliptin 25mg tablets	No
49502	Glucophage SR 500mg tablets (Mawdsley-Brooks & Company Ltd)	No
49738	Metformin 1g modified-release tablets (A A H Pharmaceuticals Ltd)	No
50087	Januvia 50mg tablets (Merck Sharp & Dohme Ltd)	No
50124	Januvia 25mg tablets (Merck Sharp & Dohme Ltd)	No
50570	Glucophage SR 500mg tablets (Lexon (UK) Ltd)	No
50821	Metformin 850mg tablets (Pfizer Ltd)	No
50970	Metformin 500mg tablets (Bristol Laboratories Ltd)	No
51080	Metabet SR 1000mg tablets (Actavis UK Ltd)	No
51135	Metformin 500mg modified-release tablets (A A H Pharmaceuticals Ltd)	No
51527	Metformin 500mg tablets (Boston Healthcare Ltd)	No
51955	Gliclazide 80mg tablets (Accord Healthcare Ltd)	No
52203	Enyglid 0.5mg tablets (Consilient Health Ltd)	No
52221	Diagemet XL 500mg tablets (Genus Pharmaceuticals Ltd)	No
52442	Metformin 500mg tablets (Pfizer Ltd)	No
52634	Glucophage SR 500mg tablets (DE Pharmaceuticals)	No

53288	Gliclazide 30mg modified-release tablets (A A H Pharmaceuticals Ltd)	No
53478	Metformin 500mg modified-release tablets (Kent Pharmaceuticals Ltd)	No
53774	Metabet SR 500mg tablets (Actavis UK Ltd)	No
53867	Metformin 500mg tablets (Zentiva)	No
54150	Jentadueto 2.5mg/850mg tablets (Boehringer Ingelheim Ltd)	No
54182	Dapagliflozin 10mg tablets	No
54203	Forxiga 10mg tablets (AstraZeneca UK Ltd)	No
54265	Dapagliflozin 5mg tablets	No
54442	Metformin (roi) 1000mg Tablet	No
54480	Forxiga 5mg tablets (AstraZeneca UK Ltd)	No
54764	Gliclazide 80mg tablets (Arrow Generics Ltd)	No
54891	Saxagliptin 2.5mg / Metformin 1g tablets	No
54898	Metformin 850mg tablets (Almus Pharmaceuticals Ltd)	No
55270	Duformin 500mg Tablet (Dumex Ltd)	No
55711	Metformin 500mg tablets (Alliance Healthcare (Distribution) Ltd)	No
55739	Metformin 500mg tablets (Tillomed Laboratories Ltd)	No
55862	Gliclazide Oral solution	No
56008	Gliclazide 80mg tablets (Almus Pharmaceuticals Ltd)	No
56208	Pioglitazone 15mg tablets (A A H Pharmaceuticals Ltd)	Yes
56376	Rosiglitazone 4mg with glimepiride 4mg tablet	No
56437	Gliclazide 60mg modified-release tablets	No
56965	Komboglyze 2.5mg/1000mg tablets (AstraZeneca UK Ltd)	No
57147	Bolamyn SR 1000mg tablets (Teva UK Ltd)	No
57457	Metformin 500mg tablets (Aurobindo Pharma Ltd)	No
57601	Daonil 5mg tablets (Dowelhurst Ltd)	No
57659	Pioglitazone 30mg tablets (Actavis UK Ltd)	Yes
57830	Gliclazide 30mg modified-release tablets (Alliance Healthcare (Distribution) Ltd)	No
58051	Metformin 500mg/5ml oral solution	No
58607	Metformin 500mg/5ml oral solution sugar free (Zentiva)	No
59620	Glucophage SR 500mg tablets (Waymade Healthcare Plc)	No
59809	Alogliptin 6.25mg tablets	No

60012	Dapagliflozin 5mg / Metformin 1g tablets	No
60074	Metformin 1g modified-release tablets (Waymade Healthcare Plc)	No
60211	Canagliflozin 100mg tablets	No
60286	Metformin 500mg/5ml oral suspension	No
60430	Invokana 100mg tablets (Janssen-Cilag Ltd)	No
60495	Gliclazide 80mg tablets (Teva UK Ltd)	No
60643	Xigduo 5mg/1000mg tablets (AstraZeneca UK Ltd)	No
60968	Metformin 500mg modified-release tablets (Actavis UK Ltd)	No
61043	Sukkarto SR 1000mg tablets (Morningside Healthcare Ltd)	No
61311	Glimepiride 4mg tablets (Sigma Pharmaceuticals Plc)	No
61559	Sukkarto SR 500mg tablets (Morningside Healthcare Ltd)	No
61925	NovoNorm 500microgram tablets (Waymade Healthcare Plc)	No
61957	Gliclazide 40mg tablets (A A H Pharmaceuticals Ltd)	No
62014	Glimepiride 2mg tablets (Accord Healthcare Ltd)	No
62034	Laaglyda MR 60mg tablets (Consilient Health Ltd)	No
62144	Metformin 500mg modified-release tablets (DE Pharmaceuticals)	No
62265	Metformin 500mg modified-release tablets (Mawdsley-Brooks & Company Ltd)	No
62326	Vipidia 6.25mg tablets (Takeda UK Ltd)	No
62426	Pioglitazone 30mg tablets (Accord Healthcare Ltd)	Yes
62426	Pioglitazone 30mg tablets (Accord Healthcare Ltd)	No
62605	Metformin 850mg tablets (Kent Pharmaceuticals Ltd)	No
62661	Bydureon 2mg powder and solvent for prolonged-release suspension for injection pre-filled pen (AstraZeneca UK Ltd)	No
62824	Metformin 1g modified-release tablets (Actavis UK Ltd)	No
63045	Metformin 850mg tablets (Relonchem Ltd)	No
63046	Pioglitazone 45mg tablets (A A H Pharmaceuticals Ltd)	Yes
63048	Gliclazide 80mg tablets (Alliance Healthcare (Distribution) Ltd)	No
63107	Pioglitazone 45mg tablets (Waymade Healthcare Plc)	Yes

63131	Ziclaseg 30mg modified-release tablets (Lupin (Europe) Ltd)	No
63336	Trulicity 1.5mg/0.5ml solution for injection pre-filled pen (Eli Lilly and Company Ltd)	No
63401	Trulicity 0.75mg/0.5ml solution for injection pre-filled pen (Eli Lilly and Company Ltd)	No
63421	Pioglitazone 30mg tablets (Teva UK Ltd)	Yes
63516	Forxiga 10mg tablets (Waymade Healthcare Plc)	No
63785	Dulaglutide 0.75mg/0.5ml solution for injection pre-filled disposable devices	No
63823	Dulaglutide 1.5mg/0.5ml solution for injection pre-filled disposable devices	No
64217	Jardiance 25mg tablets (Boehringer Ingelheim Ltd)	No
64900	Glidipion 30mg tablets (Actavis UK Ltd)	Yes
64939	Glucient SR 1000mg tablets (Consilient Health Ltd)	No
65083	Synjardy 5mg/1000mg tablets (Boehringer Ingelheim Ltd)	No
65344	Empagliflozin 5mg / Metformin 850mg tablets	No
65562	Pioglitazone 30mg tablets (Alliance Healthcare (Distribution) Ltd)	Yes
65563	Pioglitazone 15mg tablets (Alliance Healthcare (Distribution) Ltd)	Yes
65694	Metformin 500mg modified-release tablets (Waymade Healthcare Plc)	No
65923	Metformin 1g modified-release tablets (Mawdsley-Brooks & Company Ltd)	No
66008	Synjardy 12.5mg/1000mg tablets (Boehringer Ingelheim Ltd)	No
66136	Glucophage SR 1000mg tablets (Waymade Healthcare Plc)	No
66399	Glimepiride 2mg tablets (A A H Pharmaceuticals Ltd)	No
67056	Amaryl 1mg tablets (Lexon (UK) Ltd)	No
67781	Gliclazide 80mg tablets (Milpharm Ltd)	No
68203	Metformin 500mg modified-release tablets (Almus Pharmaceuticals Ltd)	No
68214	Metformin 500mg/5ml oral solution sugar free (A A H Pharmaceuticals Ltd)	No
68289	Glimepiride 4mg tablets (Waymade Healthcare Plc)	No

68389	Metformin 500mg/5ml oral solution sugar free (Pinewood Healthcare)	No
68415	Gliclazide 30mg modified-release tablets (Phoenix Healthcare Distribution Ltd)	No
68636	Metformin 850mg/5ml oral solution sugar free	No
68675	Glimepiride 4mg tablets (Somex Pharma)	No
68819	Gliclazide 80mg tablets (Bristol Laboratories Ltd)	No

APPENDIX H

APPENDIX H

Results from the sensitivity analyses investigating the assumption that missing HbA1c levels should be coded in the well-controlled group. The first table shows how glitazone and fibrate exposure associates with HbA1c levels. This analysis acted as a positive control showing a positive association between worse-controlled diabetes, as proxied by HbA1c level measurements, and glitazone exposure. For this analysis, all of the cases and controls in the glitazone nested case-control study were used. Participants with missing HbA1c measurements were dropped from this analysis. This analysis may have been biased because the data were extracted to test the association between glitazone use and brain tumour risk and prognosis. Therefore, extraction may have resulted in an artefactual relationship between these variables which may have biased the analysis. Furthermore, more variables would need to be extracted and included in the model for a more accurate result (e.g., measure of BMI or history of heart conditions); however, these could not be extracted as these were not relevant to the initial proposal accepted by CPRD. Despite this, the direction of effect showed that higher HbA1c levels were positively associated with glitazone exposure in both the crude and adjusted models, though the magnitude of the effect should be treated with caution.

Variable	Exposed (%)	Unexposed (%)	Crude OR (95% CI)	*P*	Adjusted [a] OR (95% CI)	*P*
			Glitazone exposed vs unexposed			
	N = 265	N = 818				
HbA1c [b]						
1	110 (41.5%)	481 (58.8%)	Referent	-	Referent	-
2	102 (38.5%)	245 (30.0%)	1.82 (1.33, 2.48)	0.0002	1.70 (1.23, 2.35)	0.001
3	53 (20.0%)	92 (11.2%)	2.52 (1.69, 3.74)	0.00005	2.38 (1.56, 3.62)	0.00005
			Fibrate exposed vs unexposed			
	N = 20	N = 1063				
HbA1c [b]						
1	9 (45.0%)	582 (54.8%)	Referent	-		
2	10 (50.0%)	337 (31.7%)	1.92 (0.77, 4.77)	0.16	2.02 (0.78, 5.20)	0.15
3	1 (5.0%)	144 (13.5%)	0.45 (0.06, 3.57)	0.45	0.45 (0.06, 3.75)	0.47

[a] Adjusted for sex, age, IMD, retrospective prescription history and brain tumour case status.
[b] 1 represents well-controlled diabetes, 2 moderately controlled and 3 poorly controlled.

Variable	Cases (%)	Controls (%)	Crude OR (95% CI)	*P*
		Imputed HbA1c levels		
	N = 480	N = 1920		
HbA1c [a]				
1	290 (60.4%)	1075 (56.0%)	Referent	-
2	127 (28.6%)	600 (31.3%)	0.85 (0.68, 1.06)	0.22
3	53 (11.0%)	245 (12.8%)	0.95 (0.69, 1.32)	0.8
		Complete case (missing HbA1c levels dropped)		
	N = 142	N = 941		
HbA1c [a]				
1	80 (56.3%)	511 (54.3%)	Referent	-
2	42 (29.6%)	305 (32.4%)	0.88 (0.59, 1.31)	0.52
3	20 (14.1%)	125 (13.3%)	1.02 (0.60, 1.73)	0.8

[a] 1 represents well-controlled diabetes, 2 moderately controlled and 3 poorly controlled.

The following tables show characteristics, crude and adjusted results from the follow-up analyses which imputed missing HbA1c levels and used only complete cases (i.e., participants with missing HbA1c levels were dropped from the analysis). Case-control numbers in the imputed sensitivity analysis were given for one of the 55 generated datasets; however, OR were derived from analyses which used all datasets and pooled the results. Variables used to generate the imputed datasets were: case-control status, sex, age, IMD, retrospective prescription history, glitazone exposure status and length of exposure, and number of consultations (defined as days with at least one in-person consultation).

Variable	Cases (%)	Controls (%)	Adjusted OR (95% CI) [a]	*P*
Imputed HbA1c levels				
	N = 480	N = 1920		
Glitazone exposure status				
Exposed	97 (20.2%)	460 (24.0%)	0.81 (0.66, 1.00)	0.11
Unexposed	383 (79.8%)	1460 (76.0%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	607 (583)	872 (760)	0.87 (0.82, 0.93) [b]	0.001
Mean glitazone exposure duration (total), days (SD)	604 (720)	1002 (823)	0.85 (0.79, 0.91) [b]	0.00005
Complete case (Missing HbA1c levels dropped)				
	N = 142	N = 941		
Glitazone exposure status				
Exposed	32 (22.5%)	708 (24.8%)	0.91 (0.59, 1.40)	0.67
Unexposed	110 (77.5%)	223 (75.2%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	797 (635)	874 (752)	0.96 (0.85, 1.08) [b]	0.5
Mean glitazone exposure duration (total), days (SD)	833 (838)	1026 (822)	0.93 (0.83, 1.05) [b]	0.23
No HbA1c adjustment				
	N = 480	N = 1920		
Glitazone exposure status				
Exposed	97 (20.2%)	460 (24.0%)	0.79 (0.62, 1.02)	0.07
Unexposed	383 (79.8%)	1460 (76.0%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	607 (583)	872 (760)	0.87 (0.80, 0.94) [b]	0.0008
Mean glitazone exposure duration (total), days (SD)	604 (720)	1002 (823)	0.85 (0.78, 0.92) [b]	0.00004

[a] Adjusted for sex, age, IMD and retrospective prescription history.
[b] OR calculated for per year increase in variable.

APPENDIX I

APPENDIX I

Results from the monotherapy analyses, whereby participants in the nested case-control studies were dropped if they were also exposed to the other agent of interest.

Variable	Cases (%)	Controls (%)	Adjusted OR (95% CI) [a]	*P*
Fibrates - Monotherapy (No glitazone exposure)				
	N = 1936	N = 7742		
Fibrate exposure status				
Exposed	110 (5.7%)	474 (6.1%)	0.92 (0.74, 1.14)	0.46
Unexposed	1826 (94.3%)	7268 (93.9%)	Referent	-
Mean fibrate exposure duration (longest), days (SD)	698 (945)	788 (1013)	0.97 (0.90, 1.03) [b]	0.32
Mean fibrate exposure duration (total), days (SD)	878 (1185)	1106 (1315)	0.96 (0.92, 1.01) [b]	0.08
Glitazones - Monotherapy (No fibrate exposure)				
	N = 467	N = 1887		
Glitazone exposure status				
Exposed	84 (18.0%)	427 (22.6%)	0.75 (0.60, 1.01)	0.06
Unexposed	383 (82.0%)	1460 (77.4%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	541 (452)	876 (767)	0.85 (0.78, 0.93) [b]	0.0004
Mean glitazone exposure duration (total), days (SD)	556 (634)	1006 (832)	0.83 (0.76, 0.90) [b]	0.00002

[a] Adjusted for sex, age, IMD and retrospective prescription history. HbA1c adjustment made in the glitazone analysis.

[b] OR calculated for per year increase in variable.

APPENDIX J

APPENDIX J

To test for a latency period between exposure and physiological effects, exposures were excluded if they fell within six, 12 and 24 months prior to the case index date. Results here are presented for those analyses, separately by exposure of interest in the nested case-control studies.

Variable	Cases (%) N = 1950	Controls (%) N = 7791	Adjusted OR (95% CI) [a]	*P*
Fibrates - 6 months latency period				
Fibrate exposure status				
Exposed	123 (6.3%)	518 (6.7%)	0.95 (0.77, 1.16)	0.59
Unexposed	1827 (93.7%)	7273 (93.4%)	Referent	-
Mean fibrate exposure duration (longest), days (SD)	726 (978)	727 (849)	0.99 (0.94, 1.05) [b]	0.8
Mean total fibrate exposure duration (total), days (SD)	904 (1180)	1115 (1281)	0.97 (0.92, 1.01) [b]	0.12
Fibrates - 12 months latency period				
Fibrate exposure status				
Exposed	121 (6.2%)	511 (6.6%)	0.94 (0.77, 1.16)	0.57
Unexposed	1829 (93.8%)	7280 (93.4%)	Referent	-
Mean fibrate exposure duration (longest), days (SD)	707 (961)	722 (839)	0.99 (0.93, 1.05) [b]	0.71
Mean total fibrate exposure duration (total), days (SD)	883 (1155)	1096 (1247)	0.96 (0.92, 1.01) [b]	0.1
Fibrates - 24 months latency period				
Fibrate exposure status				
Exposed	114 (5.9%)	499 (6.4%)	0.91 (0.74, 1.12)	0.37
Unexposed	1836 (94.2%)	7292 (93.6%)	Referent	-
Mean fibrate exposure duration (longest), days (SD)	685 (934)	707 (819)	0.98 (0.92, 1.04) [b]	0.44
Mean total fibrate exposure duration (total), days (SD)	851 (1104)	1057 (1185)	0.95 (0.91, 1.01) [b]	0.06

[a] Adjusted for sex, age, IMD and retrospective prescription history.
[b] OR calculated for per year increase in variable.

Variable	Cases (%) N = 480	Controls (%) N = 1920	Adjusted OR (95% CI) [a]	*P*
Glitazones - 6 months latency period				
Glitazone exposure status				
Exposed	95 (19.8%)	460 (24.0%)	0.82 (0.64, 1.05)	0.12
Unexposed	385 (80.2%)	1460 (76.0%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	574 (581)	770 (651)	0.88 (0.81, 0.96) [b]	0.002
Mean glitazone exposure duration (total), days (SD)	771 (693)	988 (796)	0.89 (0.83, 0.96) [b]	0.0015
Glitazones - 12 months latency period				
Glitazone exposure status				
Exposed	92 (19.2%)	457 (23.8%)	0.79 (0.61, 1.02)	0.07
Unexposed	388 (80.8%)	1463 (76.2%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	540 (572)	753 (642)	0.87 (0.79, 0.95) [b]	0.002
Mean glitazone exposure duration (total), days (SD)	670 (690)	958 (774)	0.87 (0.80, 0.94) [b]	0.0003
Glitazones - 24 months latency period				
Glitazone exposure status				
Exposed	86 (17.9%)	449 (23.4%)	0.74 (0.57, 0.96)	0.02
Unexposed	394 (82.1%)	1471 (76.6%)	Referent	-
Mean glitazone exposure duration (longest), days (SD)	478 (557)	720 (625)	0.83 (0.75, 0.92) [b]	0.0004
Mean glitazone exposure duration (total), days (SD)	593 (672)	901 (730)	0.84 (0.76, 0.91) [b]	7.00×10^{-6}

[a] Adjusted for sex, age, HbA1c levels, IMD and retrospective prescription history.
[b] OR calculated for per year increase in variable.

APPENDIX K

APPENDIX K

This thesis also contains an online supplementary found at the following digital object identifier (DOI) URL `https://doi.org/10.5281/zenodo.5084895`. This includes six supplementary tables and LocusZoom regional plots for each of the colocalisation analyses conducted. The tables are as follows:

- Table S1: Full Mendelian randomisation (MR) results from analyses presented in Chapter 4. All MR results for both brain and blood expression quantitative trait loci (eQTLs) are presented.
- Table S2: Full colocalisation results for genes which passed the previous MR analysis in Chapter 4.
- Table S3: Full Steiger filtering results for genes which passed the previous MR analysis in Chapter 4.
- Table S4: Expanded results for the tissue-specific analyses presented in Chapter 4.
- Table S5: Full MR results from analyses presented in Chapter 5. All MR results for both brain and blood protein QTLs (pQTLs) are presented. Odds ratios (OR) presented in this table are unscaled, whereas are presented scaled within the main body of this thesis.
- Table S6: Full Steiger filtering results for proteins which passed the previous MR analysis in Chapter 5.

Finally, LocusZoom plots are arranged according to dataset where the region was extracted, which was either exposure, which could be the eQTL or pQTL datasets, or glioma subtype (all glioma, glioblastoma (GBM) or non-GBM).

BIBLIOGRAPHY

[1] M. G. Abdelwahab, K. E. Fenton, M. C. Preul, J. M. Rho, A. Lynch, P. Stafford, and A. C. Scheck.
The ketogenic diet is an effective adjuvant to radiation therapy for the treatment of malignant glioma.
PLoS One, 7(5):e36197, 2012.

[2] N. Agarwal, J. P. Machiels, C. Suarez, N. Lewis, M. Higgins, K. Wisinski, A. Awada, M. Maur, M. Stein, A. Hwang, R. Mosher, E. Wasserman, G. Wu, H. Zhang, R. Zieba, and M. Elmeliegy.
Phase I study of the prolactin receptor antagonist LFA102 in metastatic breast and castration-resistant prostate cancer.
Oncologist, 21(5):535–6, 2016.

[3] F. Aguet, A. N. Barbeira, R. Bonazzola, A. Brown, S. E. Castel, B. Jo, S. Kasela, S. Kim-Hellmuth, Y. Liang, M. Oliva, P. E. Parsana, E. Flynn, L. Fresard, E. R. Gaamzon, A. R. Hamel, Y. He, F. Hormozdiari, P. Mohammadi, M. Munoz-Aguirre, Y. Park, A. Saha, A. V. Segrc, B. J. Strober, X. Wen, V. Wucher, S. Das, D. Garrido-Martin, N. R. Gay, R. E. Handsaker, P. J. Hoffman, S. Kashin, A. Kwong, X. Li, D. MacArthur, J. M. Rouhana, M. Stephens, E. Todres, A. Vinuela, G. Wang, Y. Zou, C. D. Brown, N. Cox, E. Dermitzakis, B. E. Engelhardt, G. Getz, R. Guigo, S. B. Montgomery, B. E. Stranger, H. K. Im, A. Battle, K. G. Ardlie, and T. Lappalainen.
The GTEx consortium atlas of genetic regulatory effects across human tissues.
bioRxiv, 2019.

[4] M. Ameratunga, N. Pavlakis, H. Wheeler, R. Grant, J. Simes, and M. Khasraw.
Anti-angiogenic therapy for high-grade glioma.
Cochrane Database Syst Rev, 11:CD008218, 2018.

[5] E. S. Amirian, R. Zhou, M. R. Wrensch, S. H. Olson, M. E. Scheurer, D. Il'yasova, D. Lachance, G. N. Armstrong, L. S. McCoy, C. C. Lau, E. B. Claus, J. S. Barnholtz-Sloan, J. Schildkraut, F. Ali-Osman, S. Sadetzki, C. Johansen, R. S. Houlston, R. B. Jenkins, J. L. Bernstein, R. T. Merrell, F. G. Davis, R. Lai, S. Shete, C. I. Amos, B. S. Melin, and M. L. Bondy.

Approaching a scientific consensus on the association between allergies and glioma risk: A report from the glioma international case-control study.
Cancer Epidemiol Biomarkers Prev, 25(2):282–90, 2016.

[6] C. I. Amos, J. Dennis, Z. Wang, J. Byun, F. R. Schumacher, S. A. Gayther, G. Casey, D. J. Hunter, T. A. Sellers, S. B. Gruber, A. M. Dunning, K. Michailidou, L. Fachal, K. Doheny, A. B. Spurdle, Y. Li, X. Xiao, J. Romm, E. Pugh, G. A. Coetzee, D. J. Hazelett, S. E. Bojesen, C. Caga-Anan, C. A. Haiman, A. Kamal, C. Luccarini, D. Tessier, D. Vincent, F. Bacot, D. J. Van Den Berg, S. Nelson, S. Demetriades, D. E. Goldgar, F. J. Couch, J. L. Forman, G. G. Giles, D. V. Conti, H. Bickeboller, A. Risch, M. Waldenberger, I. Bruske-Hohlfeld, B. D. Hicks, H. Ling, L. McGuffog, A. Lee, K. Kuchenbaecker, P. Soucy, J. Manz, J. M. Cunningham, K. Butterbach, Z. Kote-Jarai, P. Kraft, L. FitzGerald, S. Lindstrom, M. Adams, J. D. McKay, C. M. Phelan, S. Benlloch, L. E. Kelemen, P. Brennan, M. Riggan, T. A. O'Mara, H. Shen, Y. Shi, D. J. Thompson, M. T. Goodman, S. F. Nielsen, A. Berchuck, S. Laboissiere, S. L. Schmit, T. Shelford, C. K. Edlund, J. A. Taylor, J. K. Field, S. K. Park, K. Offit, M. Thomassen, R. Schmutzler, L. Ottini, R. J. Hung, J. Marchini, A. Amin Al Olama, U. Peters, R. A. Eeles, M. F. Seldin, E. Gillanders, D. Seminara, A. C. Antoniou, P. D. Pharoah, G. Chenevix-Trench, S. J. Chanock, J. Simard, and D. F. Easton.
The OncoArray Consortium: A network for understanding the genetic architecture of common cancers.
Cancer Epidemiol Biomarkers Prev, 26(1):126–135, 2017.

[7] H. Arita, K. Yamasaki, Y. Matsushita, T. Nakamura, A. Shimokawa, H. Takami, S. Tanaka, A. Mukasa, M. Shirahata, S. Shimizu, K. Suzuki, K. Saito, K. Kobayashi, F. Higuchi, T. Uzuka, R. Otani, K. Tamura, K. Sumita, M. Ohno, Y. Miyakita, N. Kagawa, N. Hashimoto, R. Hatae, K. Yoshimoto, N. Shinojima, H. Nakamura, Y. Kanemura, Y. Okita, M. Kinoshita, K. Ishibashi, T. Shofuda, Y. Kodama, K. Mori, Y. Tomogane, J. Fukai, K. Fujita, Y. Terakawa, N. Tsuyuguchi, S. Moriuchi, M. Nonaka, H. Suzuki, M. Shibuya, T. Maehara, N. Saito, M. Nagane, N. Kawahara, K. Ueki, T. Yoshimine, E. Miyaoka, R. Nishikawa, T. Komori, Y. Narita, and K. Ichimura.
A combination of TERT promoter mutation and MGMT methylation status predicts clinically relevant subgroups of newly diagnosed glioblastomas.
Acta Neuropathologica Communications, 4(1):79, 2016.

[8] J. K. Aronson, C. Bankhead, K. R. Mahtani, and D. Nunan.
Catalogue of bias collaboration: Confounding by indication.
https://catalogofbias.org/biases/confounding-by-indication.
Accessed: 2020-08.

[9] S. J. Assinder, J. A. Stanton, and P. D. Prasad.

Transgelin: an actin-binding protein and tumour suppressor.
Int J Biochem Cell Biol, 41(3):482–6, 2009.

[10] I. Atkins, B. Kinnersley, Q. T. Ostrom, K. Labreche, D. Il'yasova, G. N. Armstrong, J. E. Eckel-Passow, M. J. Schoemaker, M. M. Nöthen, J. S. Barnholtz-Sloan, A. J. Swerdlow, M. Simon, P. Rajaraman, S. J. Chanock, J. Shildkraut, J. L. Bernstein, P. Hoffmann, K.-H. Jöckel, R. K. Lai, E. B. Claus, S. H. Olson, C. Johansen, M. R. Wrensch, B. Melin, R. B. Jenkins, M. Sanson, M. L. Bondy, and R. S. Houlston.
Transcriptome-wide association study identifies new candidate susceptibility genes for glioma.
Cancer Res, 79(8):2065–2071, 2019.

[11] A. Auvinen, M. Hietanen, R. Luukkonen, and R. S. Koskela.
Brain tumors and salivary gland cancers among cellular telephone users.
Epidemiology, 13(3):356–9, 2002.

[12] Y. Aylon and M. Oren.
The paradox of p53: What, how, and why?
Cold Spring Harb Perspect Med, 6(10), 2016.

[13] E. E. Bailey, H. H. Pfeifer, and E. A. Thiele.
The use of diet in the treatment of epilepsy.
Epilepsy Behav, 6(1):4–8, 2005.

[14] M. N. Bainbridge, G. N. Armstrong, M. M. Gramatges, A. A. Bertuch, S. N. Jhangiani, H. Doddapaneni, L. Lewis, J. Tombrello, S. Tsavachidis, Y. Liu, A. Jalali, S. E. Plon, C. C. Lau, D. W. Parsons, E. B. Claus, J. Barnholtz-Sloan, D. Il'yasova, J. Schildkraut, F. Ali-Osman, S. Sadetzki, C. Johansen, R. S. Houlston, R. B. Jenkins, D. Lachance, S. H. Olson, J. L. Bernstein, R. T. Merrell, M. R. Wrensch, K. M. Walsh, F. G. Davis, R. Lai, S. Shete, K. Aldape, C. I. Amos, P. A. Thompson, D. M. Muzny, R. A. Gibbs, B. S. Melin, M. L. Bondy, and Gliogene Consortium.
Germline mutations in shelterin complex genes are associated with familial glioma.
J Natl Cancer Inst, 107(1):384, 2015.

[15] P. Ballard, J. W. Yates, Z. Yang, D. W. Kim, J. C. Yang, M. Cantarini, K. Pickup, A. Jordan, M. Hickey, M. Grist, M. Box, P. Johnstrom, K. Varnas, J. Malmquist, K. S. Thress, P. A. Janne, and D. Cross.
Preclinical comparison of osimertinib with other EGFR-TKIs in EGFR-mutant NSCLC brain metastases models, and early evidence of clinical brain metastases activity.
Clin Cancer Res, 22(20):5130–5140, 2016.

[16] K. V. Ballman.

Biomarker: Predictive or prognostic?
J Clin Oncol, 33(33):3968–3971, 2015.

[17] Z. S. Bao, M. Y. Li, J. Y. Wang, C. B. Zhang, H. J. Wang, W. Yan, Y. W. Liu, W. Zhang, L. Chen, and T. Jiang.
Prognostic value of a nine-gene signature in glioma patients based on mRNA expression profiling.
CNS Neurosci Ther, 20(2):112–8, 2014.

[18] Y. Barak, D. Liao, W. He, E. S. Ong, M. C. Nelson, J. M. Olefsky, R. Boland, and R. M. Evans.
Effects of peroxisome proliferator-activated receptor delta on placentation, adiposity, and colorectal cancer.
Proc Natl Acad Sci USA, 99(1):303–8, 2002.

[19] K. W. Baranano and A. L. Hartman.
The ketogenic diet: Uses in epilepsy and other neurologic illnesses.
Curr Treat Options in Neurol, 10(6):410–419, 2008.

[20] C. A. Barker, A. J. Bishop, M. Chang, K. Beal, and T. A. Chan.
Valproic acid use during radiation therapy for glioblastoma associated with improved survival.
Int J Radiat Oncol Biol Phys, 86(3):504–9, 2013.

[21] I. Barker, Fred G., M. L. Simmons, S. M. Chang, M. D. Prados, D. A. Larson, P. K. Sneed, W. M. Wara, M. S. Berger, P. Chen, M. A. Israel, and K. D. Aldape.
EGFR overexpression and radiation response in glioblastoma multiforme.
International Journal of Radiation Oncology, Biology, Physics, 51(2):410–418, 2001.

[22] B. G. Baumert, M. E. Hegi, M. J. van den Bent, A. von Deimling, T. Gorlia, K. Hoang-Xuan, A. A. Brandes, G. Kantor, M. J. B. Taphoorn, M. B. Hassel, C. Hartmann, G. Ryan, D. Capper, J. M. Kros, S. Kurscheid, W. Wick, R. Enting, M. Reni, B. Thiessen, F. Dhermain, J. E. Bromberg, L. Feuvret, J. C. Reijneveld, O. Chinot, J. M. M. Gijtenbeek, J. P. Rossiter, N. Dif, C. Balana, J. Bravo-Marques, P. M. Clement, C. Marosi, T. Tzuk-Shina, R. A. Nordal, J. Rees, D. Lacombe, W. P. Mason, and R. Stupp.
Temozolomide chemotherapy versus radiotherapy in high-risk low-grade glioma (EORTC 22033-26033): a randomised, open-label, phase 3 intergroup study.
Lancet Oncol, 17(11):1521–1532, 2016.

[23] R. Beasley, T. Clayton, J. Crane, E. von Mutius, C. K. W. Lai, S. Montefort, and A. Stewart.

Association between paracetamol use in infancy and childhood, and risk of asthma, rhinoconjunctivitis, and eczema in children aged 6-7 years: analysis from phase three of the ISAAC programme.
Lancet, 372(9643):1039–1048, 2008.

[24] E. H. Bell, P. Zhang, E. G. Shaw, J. C. Buckner, G. R. Barger, D. E. Bullard, M. P. Mehta, M. R. Gilbert, P. D. Brown, K. J. Stelzer, J. P. McElroy, J. l. Fleming, C. D. Timmers, A. P. Becker, A. L. Salavaggione, Z. Liu, K. Aldape, D. G. Brachman, S. Z. Gertler, A. D. Murtha, C. J. Schultz, D. Johnson, N. N. Laack, G. K. Hunter, I. R. Crocker, M. Won, and A. Chakravarti.
Comprehensive genomic analysis in NRG Oncology/RTOG 9802: A phase III trial of radiation versus radiation plus procarbazine, lomustine (CCNU), and vincristine in high-risk low-grade glioma.
J Clin Oncol, 38(29):3407–3417, 2020.
PMID: 32706640.

[25] C. Benner, C. C. Spencer, A. S. Havulinna, V. Salomaa, S. Ripatti, and M. Pirinen.
FINEMAP: efficient variable selection using summary data from genome-wide association studies.
Bioinformatics, 32(10):1493–501, 2016.

[26] V. S. Benson, K. Pirie, J. Green, D. Casabonne, V. Beral, and Million Women Study Collaborators.
Lifestyle factors and primary glioma and meningioma tumours in the million women study cohort.
Br J Cancer, 99(1):185–90, 2008.

[27] V. S. Benson, K. Pirie, J. Schuz, G. K. Reeves, V. Beral, J. Green, and Million Women Study Collaborators.
Mobile phone use and risk of brain neoplasms and other cancers: prospective study.
Int J Epidemiol, 42(3):792–802, 2013.

[28] G. Berg-Beckhoff, J. Schuz, M. Blettner, E. Munster, K. Schlaefer, J. Wahrendorf, and B. Schlehofer.
History of allergic disease and epilepsy and risk of glioma and meningioma (INTERPHONE study group, germany).
Eur J Epidemiol, 24(8):433–40, 2009.

[29] C. Bettegowda, N. Agrawal, Y. Jiao, M. Sausen, L. D. Wood, R. H. Hruban, F. J. Rodriguez, D. P. Cahill, R. McLendon, G. Riggins, V. E. Velculescu, S. M. Oba-Shinjo, S. K. Marie, B. Vogelstein, D. Bigner, H. Yan, N. Papadopoulos, and K. W. Kinzler.

Mutations in *CIC* and *FUBP1* contribute to human oligodendroglioma.
Science, 333(6048):1453–5, 2011.

[30] E. Binello, E. Mormone, L. Emdad, H. Kothari, and I. M. Germano.
Characterization of fenofibrate-mediated anti-proliferative pro-apoptotic effects on high-grade gliomas and anti-invasive effects on glioma stem cells.
J Neurooncol, 117(2):225–34, 2014.

[31] R. Bonavia, M. M. Inda, W. K. Cavenee, and F. B. Furnari.
Heterogeneity maintenance in glioblastoma: a social network.
Cancer Res, 71(12):4055–60, 2011.

[32] S. Bonovas, G. K. Nikolopoulos, and P. G. Bagos.
Use of fibrates and cancer risk: a systematic review and meta-analysis of 17 long-term randomized placebo-controlled trials.
PLoS One, 7(9):e45259, 2012.

[33] J. A. Bovi.
Prevention of brain metastases.
Front Neurol, 9:758, 2018.

[34] J. Bowden, G. Davey Smith, and S. Burgess.
Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression.
Int J Epidemiol, 44(2):512–25, 2015.

[35] J. Bowden, G. Davey Smith, P. C. Haycock, and S. Burgess.
Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator.
Genet Epidemiol, 40(4):304–14, 2016.

[36] J. Bowden, M. F. Del Greco, C. Minelli, G. Davey Smith, N. A. Sheehan, and J. R. Thompson.
Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic.
Int J Epidemiol, 45(6):1961–1974, 2016.

[37] J. Bowden, M. F. Del Greco, C. Minelli, Q. Zhao, D. A. Lawlor, N. A. Sheehan, J. Thompson, and G. Davey Smith.
Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption.
Int J Epidemiol, 48(3):728–742, 2019.

[38] M. Bower, E. S. Newlands, N. M. Bleehen, M. Brada, R. J. Begent, H. Calvert, I. Colquhoun, P. Lewis, and M. H. Brampton.
Multicentre CRC phase II trial of temozolomide in recurrent or progressive high-grade glioma.
Cancer Chemother Pharmacol, 40(6):484–8, 1997.

[39] M. Z. Braganza, C. M. Kitahara, A. Berrington de Gonzalez, P. D. Inskip, K. J. Johnson, and P. Rajaraman.
Ionizing radiation and the risk of brain and central nervous system tumors: a systematic review.
Neuro Oncol, 14(11):1316–24, 2012.

[40] M. Z. Braganza, P. Rajaraman, Y. Park, P. D. Inskip, N. D. Freedman, A. R. Hollenbeck, A. B. de Gonzalez, and C. M. Kitahara.
Cigarette smoking, alcohol intake, and risk of glioma in the NIH-AARP diet and health study.
Br J Cancer, 110(1):242–8, 2014.

[41] E. C. Brantley and E. N. Benveniste.
Signal transducer and activator of transcription-3: a molecular hub for signaling pathways in gliomas.
Mol Cancer Res, 6(5):675–84, 2008.

[42] A. V. Brenner, M. S. Linet, H. A. Fine, W. R. Shapiro, R. G. Selker, P. M. Black, and P. D. Inskip.
History of allergies and autoimmune diseases and risk of brain tumors in adults.
Int J Cancer, 99(2):252–9, 2002.

[43] P. D. Brown, J. C. Buckner, J. R. O'Fallon, N. L. Iturria, C. A. Brown, B. P. O'Neill, B. W. Scheithauer, R. P. Dinapoli, R. M. Arusell, W. J. Curran, R. Abrams, and E. G. Shaw.
Effects of radiotherapy on cognitive function in patients with low-grade glioma measured by the folstein mini-mental state examination.
J Clin Oncol, 21(13):2519–24, 2003.

[44] T. J. Brown, D. A. Bota, M. J. van Den Bent, P. D. Brown, E. Maher, D. Aregawi, L. M. Liau, J. C. Buckner, M. Weller, M. S. Berger, and M. Glantz.
Management of low-grade glioma: a systematic review and meta-analysis.
Neurooncol Pract, 6(4):249–258, 2019.

[45] J. C. Buckner, E. G. Shaw, S. L. Pugh, A. Chakravarti, M. R. Gilbert, G. R. Barger, S. Coons, P. Ricci, D. Bullard, P. D. Brown, K. Stelzer, D. Brachman, J. H. Suh, C. J. Schultz, J.-P.

Bahary, B. J. Fisher, H. Kim, A. D. Murtha, E. H. Bell, M. Won, M. P. Mehta, and W. J. Curran.
Radiation plus procarbazine, CCNU, and vincristine in low-grade glioma.
N Engl J Med, 374(14):1344–1355, 2016.
PMID: 27050206.

[46] A. Buniello, J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, A. McMahon, J. Morales, E. Mountjoy, E. Sollis, D. Suveges, O. Vrousgou, P. L. Whetzel, R. Amode, J. A. Guillen, H. S. Riat, S. J. Trevanion, P. Hall, H. Junkins, P. Flicek, T. Burdett, L. A. Hindorff, F. Cunningham, and H. Parkinson.
The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.
Nucleic Acids Res, 47(D1):D1005–D1012, 2019.

[47] S. Burgess.
Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome.
Int J Epidemiol, 43(3):922–929, 03 2014.

[48] M. J. Campbell, L. J. Esserman, Y. Zhou, M. Shoemaker, M. Lobo, E. Borman, F. Baehner, A. S. Kumar, K. Adduci, C. Marx, E. F. Petricoin, L. A. Liotta, M. Winters, S. Benz, and C. C. Benz.
Breast cancer growth prevention by statins.
Cancer Res, 66(17):8707–14, 2006.

[49] Cancer Genome Atlas Research Network, D. J. Brat, R. G. Verhaak, K. D. Aldape, W. K. Yung, S. R. Salama, L. A. Cooper, E. Rheinbay, C. R. Miller, M. Vitucci, O. Morozova, A. G. Robertson, H. Noushmehr, P. W. Laird, A. D. Cherniack, R. Akbani, J. T. Huse, G. Ciriello, L. M. Poisson, J. S. Barnholtz-Sloan, M. S. Berger, C. Brennan, R. R. Colen, H. Colman, A. E. Flanders, C. Giannini, M. Grifford, A. Iavarone, R. Jain, I. Joseph, J. Kim, K. Kasaian, T. Mikkelsen, B. A. Murray, B. P. O'Neill, L. Pachter, D. W. Parsons, C. Sougnez, E. P. Sulman, S. R. Vandenberg, E. G. Van Meir, A. von Deimling, H. Zhang, D. Crain, K. Lau, D. Mallery, S. Morris, J. Paulauskis, R. Penny, T. Shelton, M. Sherman, P. Yena, A. Black, J. Bowen, K. Dicostanzo, J. Gastier-Foster, K. M. Leraas, T. M. Lichtenberg, C. R. Pierson, N. C. Ramirez, C. Taylor, S. Weaver, L. Wise, E. Zmuda, T. Davidsen, J. A. Demchok, G. Eley, M. L. Ferguson, C. M. Hutter, K. R. Mills Shaw, B. A. Ozenberger, M. Sheth, H. J. Sofia, R. Tarnuzzer, Z. Wang, L. Yang, J. C. Zenklusen, B. Ayala, J. Baboud, S. Chudamani, M. A. Jensen, J. Liu, T. Pihl, R. Raman, Y. Wan, Y. Wu, A. Ally, J. T. Auman, M. Balasundaram, S. Balu, S. B. Baylin, R. Beroukhim, M. S. Bootwalla, R. Bowlby, C. A. Bristow, D. Brooks, Y. Butterfield, R. Carlsen, S. Carter, L. Chin, et al.

Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas.
N Engl J Med, 372(26):2481–98, 2015.

[50] Cancer Research UK.
Brain, other CNS, and intracranial tumours.
https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/brain-other-cns-and-intracranial-tumours/incidence.
Accessed: 2020-09.

[51] D. Capper, D. T. W. Jones, M. Sill, V. Hovestadt, D. Schrimpf, D. Sturm, C. Koelsche, F. Sahm, L. Chavez, D. E. Reuss, A. Kratz, A. K. Wefers, K. Huang, K. W. Pajtler, L. Schweizer, D. Stichel, A. Olar, N. W. Engel, K. Lindenberg, P. N. Harter, A. K. Braczynski, K. H. Plate, H. Dohmen, B. K. Garvalov, R. Coras, A. Hoelsken, E. Hewer, M. Bewerunge-Hudler, M. Schick, R. Fischer, R. Beschorner, J. Schittenhelm, O. Staszewski, K. Wani, P. Varlet, M. Pages, P. Temming, D. Lohmann, F. Selt, H. Witt, T. Milde, O. Witt, E. Aronica, F. Giangaspero, E. Rushing, W. Scheurlen, C. Geisenberger, F. J. Rodriguez, A. Becker, M. Preusser, C. Haberler, R. Bjerkvig, J. Cryan, M. Farrell, M. Deckert, J. Hench, S. Frank, J. Serrano, K. Kannan, A. Tsirigos, W. Brueck, S. Hofer, S. Brehmer, M. Seiz-Rosenhagen, D. Hanggi, V. Hans, S. Rozsnoki, J. R. Hansford, P. Kohlhof, B. W. Kristensen, M. Lechner, B. Lopes, C. Mawrin, R. Ketter, A. Kulozik, Z. Khatib, F. Heppner, A. Koch, A. Jouvet, C. Keohane, H. Muehleisen, W. Mueller, U. Pohl, M. Prinz, A. Benner, M. Zapatka, N. G. Gottardo, P. H. Driever, C. M. Kramm, H. L. Mueller, S. Rutkowski, K. von Hoff, M. C. Fruehwald, A. Gnekow, G. Fleischhack, S. Tippelt, G. Calaminus, C.-M. Monoranu, A. Perry, C. Jones, et al.
DNA methylation-based classification of central nervous system tumours.
Nature, 555(7697):469–474, 2018.

[52] D. Carvalho-Silva, A. Pierleoni, M. Pignatelli, C. Ong, L. Fumis, N. Karamanis, M. Carmona, A. Faulconbridge, A. Hercules, E. McAuley, A. Miranda, G. Peat, M. Spitzer, J. Barrett, D. G. Hulcoop, E. Papa, G. Koscielny, and I. Dunham.
Open targets platform: new developments and updates two years on.
Nucleic Acids Res, 47(D1):D1056–D1065, 2019.

[53] A. M. Chacko, C. Li, D. A. Pryma, S. Brem, G. Coukos, and V. Muzykantov.
Targeted delivery of antibody-based therapeutic and imaging agents to CNS tumors: crossing the blood-brain barrier divide.
Expert Opin Drug Deliv, 10(7):907–26, 2013.

[54] G. Chagoya, S. G. Kwatra, C. W. Nanni, C. M. Roberts, S. M. Phillips, S. Nullmeyergh, S. P. Gilmore, I. Spasojevic, D. L. Corcoran, C. C. Young, K. V. Ballman, R. Ramakrishna, D. A. Cross, J. M. Markert, M. Lim, M. R. Gilbert, G. J. Lesser, and M. M. Kwatra.

Efficacy of osimertinib against EGFRvIII+ glioblastoma.
Oncotarget, 11(22):2074–2082, 2020.

[55] B. Chen, T. Liang, P. Yang, H. Wang, Y. Liu, F. Yang, and G. You.
Classifying lower grade glioma cases according to whole genome gene expression.
Oncotarget, 7(45):74031–74042, 2016.

[56] C. Chen, C. D. Cheng, H. Wu, Z. W. Wang, L. Wang, Z. R. Jiang, A. L. Wang, C. Hu, Y. F. Dong, W. X. Niu, S. Qi, Z. P. Qi, J. Liu, W. C. Wang, C. S. Niu, and Q. S. Liu.
Osimertinib successfully combats EGFR-negative glioblastoma cells by inhibiting the MAPK pathway.
Acta Pharmacol Sin, 2020.

[57] C. Chen, T. Xu, J. Chen, J. Zhou, Y. Yan, Y. Lu, and S. Wu.
Allergy and risk of glioma: a meta-analysis.
Eur J Neurol, 18(3):387–95, 2011.

[58] H. Chen, M. H. Ward, K. L. Tucker, B. I. Graubard, R. D. McComb, N. A. Potischman, D. D. Weisenburger, and E. F. Heineman.
Diet and risk of adult glioma in eastern Nebraska, United States.
Cancer Causes Control, 13(7):647–655, 2002.

[59] O. L. Chinot, W. Wick, W. Mason, R. Henriksson, F. Saran, R. Nishikawa, A. F. Carpentier, K. Hoang-Xuan, P. Kavan, D. Cernea, A. A. Brandes, M. Hilton, L. Abrey, and T. Cloughesy.
Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma.
N Engl J Med, 370(8):709–22, 2014.

[60] S. Chun, A. Casparino, N. A. Patsopoulos, D. C. Croteau-Chonka, B. A. Raby, P. L. De Jager, S. R. Sunyaev, and C. Cotsapas.
Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types.
Nat Genet, 49(4):600–605, 2017.

[61] S. R. Cole, R. W. Platt, E. F. Schisterman, H. Chu, D. Westreich, D. Richardson, and C. Poole.
Illustrating bias due to conditioning on a collider.
Int J Epidemiol, 39(2):417–420, 2010.

[62] L. Collado-Torres, E. E. Burke, A. Peterson, J. Shin, R. E. Straub, A. Rajpurohit, S. A. Semick, W. S. Ulrich, C. BrainSeq, A. J. Price, C. Valencia, R. Tao, A. Deep-Soboslay, T. M. Hyde, J. E. Kleinman, D. R. Weinberger, and A. E. Jaffe.
Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia.

Neuron, 103(2):203–216 e8, 2019.

[63] K. C. Cotto, A. H. Wagner, Y. Y. Feng, S. Kiwala, A. C. Coffman, G. Spies, A. Wollam, N. C. Spies, O. L. Griffith, and M. Griffith.
DGIdb 3.0: a redesign and expansion of the drug-gene interaction database.
Nucleic Acids Res, 46(D1):D1068–D1073, 2018.

[64] K. M. Dale, C. I. Coleman, N. N. Henyan, J. Kluger, and C. M. White.
Statins and cancer risk: a meta-analysis.
JAMA, 295(1):74–80, 2006.

[65] S. Das, L. Forer, S. Schonherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P. R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, and C. Fuchsberger.
Next-generation genotype imputation service and methods.
Nat Genet, 48(10):1284–1287, 2016.

[66] G. Davey Smith.
Data dredging, bias, or confounding.
BMJ, 325(7378):1437–1438, 2002.

[67] G. Davey Smith and S. Ebrahim.
'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?
Int J Epidemiol, 32(1):1–22, 2003.

[68] F. G. Davis, T. A. Dolecek, B. J. McCarthy, and J. L. Villano.
Toward determining the lifetime occurrence of metastatic brain tumors estimated from 2007 United States cancer incidence data.
Neuro Oncol, 14(9):1171–1177, 09 2012.

[69] J. S. de Bono, U. De Giorgi, D. N. Rodrigues, C. Massard, S. Bracarda, A. Font, J. A. Arranz Arija, K. C. Shih, G. D. Radavoi, N. Xu, W. Y. Chan, H. Ma, S. Gendreau, R. Riisnaes, P. H. Patel, D. J. Maslyar, and V. Jinga.
Randomized phase II study evaluating akt blockade with ipatasertib, in combination with abiraterone, in patients with metastatic prostate cancer with and without PTEN loss.
Clin Cancer Res, 25(3):928–936, 2019.

[70] P. L. De Jager, Y. Ma, C. McCabe, J. Xu, B. N. Vardarajan, D. Felsky, H.-U. Klein, C. C. White, M. A. Peters, B. Lodgson, et al.
A multi-omic atlas of the human frontal cortex for aging and alzheimer,Äôs disease research.
Sci Data, 5(1):1–13, 2018.

[71] M. de Jonge, V. A. de Weger, M. A. Dickson, M. Langenberg, A. Le Cesne, A. J. Wagner, K. Hsu, W. Zheng, S. Mace, G. Tuffal, K. Thomas, and J. H. Schellens.
A phase I study of SAR405838, a novel human double minute 2 (HDM2) antagonist, in patients with solid tumours.
Eur J Cancer, 76:144–151, 2017.

[72] V. A. de Weger, M. de Jonge, M. H. G. Langenberg, J. H. M. Schellens, M. Lolkema, A. Varga, B. Demers, K. Thomas, K. Hsu, G. Tuffal, S. Goodstal, S. Mace, and E. Deutsch.
A phase I study of the HDM2 antagonist SAR405838 combined with the MEK inhibitor pimasertib in patients with advanced solid tumours.
Br J Cancer, 120(3):286–293, 2019.

[73] P. C. De Witt Hamer, S. G. Robles, A. H. Zwinderman, H. Duffau, and M. S. Berger.
Impact of intraoperative stimulation brain mapping on glioma surgery outcome: a meta-analysis.
J Clin Oncol, 30(20):2559–65, 2012.

[74] E. Delen and O. Doganlar.
The dose dependent effects of ruxolitinib on the invasion and tumorigenesis in gliomas cells via inhibition of interferon gamma-depended JAK/STAT signaling pathway.
J Korean Neurosurg Soc, 63(4):444–454, 2020.

[75] G. D. Demetri, C. D. Fletcher, E. Mueller, P. Sarraf, R. Naujoks, N. Campbell, B. M. Spiegelman, and S. Singer.
Induction of solid tumor differentiation by the peroxisome proliferator-activated receptor-gamma ligand troglitazone in patients with liposarcoma.
Proc Natl Acad Sci U S A, 96(7):3951–6, 1999.

[76] S. Deneberg, H. Cherif, V. Lazarevic, P. O. Andersson, M. von Euler, G. Juliusson, and S. Lehmann.
An open-label phase I dose-finding study of APR-246 in hematological malignancies.
Blood Cancer J, 6(7):e447, 2016.

[77] P. Dent, D. B. Reardon, J. S. Park, G. Bowers, C. Logsdon, K. Valerie, and R. Schmidt-Ullrich.
Radiation-induced release of transforming growth factor a activates the epidermal growth factor receptor and mitogen-activated protein kinase pathway in carcinoma cells, leading to increased proliferation and protection from radiation-induced cell death.
Mol Biol Cell, 10(8):2493–2506, 1999.

[78] Department of Communities and Local Government.
The English indices of deprivation 2015, 2015.

Accessed: 2020-08.

[79] Department of Health and Social Care.
Report of the task and finish working group on brain tumour research, 2018.
Accessed: 2021-01.

[80] D. Deplanque, P. Gelé, O. Pétrault, I. Six, C. Furman, M. Bouly, S. Nion, B. Dupuis, D. Leys, J.-C. Fruchart, R. Cecchelli, B. Staels, P. Duriez, and R. Bordet.
Peroxisome proliferator-activated receptor-alpha activation as a mechanism of preventive neuroprotection induced by chronic fenofibrate treatment.
J Neurosci, 23(15):6264–6271, 2003.

[81] J. A. DiMasi, L. Feldman, A. Seckler, and A. Wilson.
Trends in risks associated with new drug development: success rates for investigational drugs.
Clin Pharmacol Ther, 87(3):272–7, 2010.

[82] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen.
Innovation in the pharmaceutical industry: New estimates of R&D costs.
J Health Econ, 47:20–33, 2016.

[83] L. Disney-Hogg, A. J. Cornish, A. Sud, P. J. Law, B. Kinnersley, D. I. Jacobs, Q. T. Ostrom, K. Labreche, J. E. Eckel-Passow, G. N. Armstrong, E. B. Claus, D. Il'yasova, J. Schildkraut, J. S. Barnholtz-Sloan, S. H. Olson, J. L. Bernstein, R. K. Lai, M. J. Schoemaker, M. Simon, P. Hoffmann, M. M. Nothen, K. H. Jockel, S. Chanock, P. Rajaraman, C. Johansen, R. B. Jenkins, B. S. Melin, M. R. Wrensch, M. Sanson, M. L. Bondy, and R. S. Houlston.
Impact of atopy on risk of glioma: a mendelian randomisation study.
BMC Med, 16(1):42, 2018.

[84] L. Disney-Hogg, A. Sud, P. J. Law, A. J. Cornish, B. Kinnersley, Q. T. Ostrom, K. Labreche, J. E. Eckel-Passow, G. N. Armstrong, E. B. Claus, D. Il'yasova, J. Schildkraut, J. S. Barnholtz-Sloan, S. H. Olson, J. L. Bernstein, R. K. Lai, A. J. Swerdlow, M. Simon, P. Hoffmann, M. M. Nothen, K. H. Jockel, S. Chanock, P. Rajaraman, C. Johansen, R. B. Jenkins, B. S. Melin, M. R. Wrensch, M. Sanson, M. L. Bondy, and R. S. Houlston.
Influence of obesity-related risk factors in the aetiology of glioma.
Br J Cancer, 118(7):1020–1027, 2018.

[85] C. R. Dormuth, G. Carney, B. Carleton, K. Bassett, and J. M. Wright.
Thiazolidinediones and fractures in men and women.
Arch Intern Med, 169(15):1395–402, 2009.

[86] L. Douw, M. Klein, S. S. A. A. Fagel, J. van den Heuvel, M. J. B. Taphoorn, N. K. Aaronson, T. J. Postma, W. P. Vandertop, J. J. Mooij, R. H. Boerman, G. N. Beute, J. D. Sluimer, B. J. Slotman, J. C. Reijneveld, and J. J. Heimans.
Cognitive and radiological effects of radiotherapy in patients with low-grade glioma: long-term follow-up.
Lancet Neurol, 8(9):810–818, 2009.

[87] R. Dubrow, A. S. Darefsky, Y. Park, S. T. Mayne, S. C. Moore, B. Kilfoy, A. J. Cross, R. Sinha, A. R. Hollenbeck, A. Schatzkin, and M. H. Ward.
Dietary components related to n-nitroso compound formation: a prospective study of adult glioma.
Cancer Epidemiol Biomark Prev, 19(7):1709–22, 2010.

[88] M. Dvorakova, R. Nenutil, and P. Bouchal.
Transgelins, cytoskeletal proteins implicated in different aspects of cancer development.
Expert Rev Proteomics, 11(2):149–65, 2014.

[89] J. M. L. Ebos, C. R. Lee, and R. S. Kerbel.
Tumor and host-mediated pathways of resistance and disease progression in response to antiangiogenic therapy.
Clin Cancer Res, 15(16):5020–5025, 2009.

[90] J. E. Eckel-Passow, P. A. Decker, M. L. Kosel, T. M. Kollmeyer, A. M. Molinaro, T. Rice, A. A. Caron, K. L. Drucker, C. E. Praska, M. Pekmezci, H. M. Hansen, L. S. McCoy, P. M. Bracci, B. J. Erickson, C. F. Lucchinetti, J. L. Wiemels, J. K. Wiencke, M. L. Bondy, B. Melin, T. C. Burns, C. Giannini, D. H. Lachance, M. R. Wrensch, and R. B. Jenkins.
Using germline variants to estimate glioma and subtype risks.
J Neuro-Oncol, 21(4):451–461, 01 2019.

[91] S. L. Edwards, J. Beesley, J. D. French, and A. M. Dunning.
Beyond GWASs: illuminating the dark road from association to function.
Am J Hum Genet, 93(5):779–97, 2013.

[92] V. Emilsson, M. Ilkov, J. R. Lamb, N. Finkel, E. F. Gudmundsson, R. Pitts, H. Hoover, V. Gudmundsdottir, S. R. Horman, T. Aspelund, L. Shu, V. Trifonov, S. Sigurdsson, A. Manolescu, J. Zhu, O. Olafsson, J. Jakobsdottir, S. A. Lesley, J. To, J. Zhang, T. B. Harris, L. J. Launer, B. Zhang, G. Eiriksdottir, X. Yang, A. P. Orth, L. L. Jennings, and V. Gudnason.
Co-regulatory networks of human serum proteins link genetics to disease.
Science, 361(6404):769–773, 2018.

[93] M. Esteller, J. Garcia-Foncillas, E. Andion, S. N. Goodman, O. F. Hidalgo, V. Vanaclocha, S. B. Baylin, and J. G. Herman.
Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents.
N Engl J Med, 343(19):1350–4, 2000.

[94] S. Everhard, G. Kaloshi, E. Criniere, A. Benouaich-Amiel, J. Lejeune, Y. Marie, M. Sanson, M. Kujas, K. Mokhtari, K. Hoang-Xuan, J. Y. Delattre, and J. Thillet.
MGMT methylation: a marker of response to temozolomide in low-grade gliomas.
Ann Neurol, 60(6):740–3, 2006.

[95] C. Faulkner, A. Palmer, H. Williams, C. Wragg, H. R. Haynes, P. White, R. M. DeSouza, M. Williams, K. Hopkins, and K. M. Kurian.
EGFR and EGFRvIII analysis in glioblastoma as therapeutic biomarkers.
Br J Neurosurg, 29(1):23–29, 2015.

[96] C. Fernandes, A. Costa, L. Osorio, R. C. Lago, P. Linhares, B. Carvalho, and C. Caeiro.
Current standards of care in glioblastoma therapy.
Codon Publications, 2017.

[97] W. T. Festuccia, S. Oztezcan, M. Laplante, M. Berthiaume, C. Michel, S. Dohgu, R. G. Denis, M. N. Brito, N. A. Brito, D. S. Miller, W. A. Banks, T. J. Bartness, D. Richard, and Y. Deshaies.
Peroxisome proliferator-activated receptor-gamma-mediated positive energy balance in the rat is associated with reduced sympathetic drive to adipose tissues and thyroid status.
Endocrinology, 149(5):2121–2130, 05 2008.

[98] C. Finan, A. Gaulton, F. A. Kruger, R. T. Lumbers, T. Shah, J. Engmann, L. Galver, R. Kelley, A. Karlsson, R. Santos, J. P. Overington, A. D. Hingorani, and J. P. Casas.
The druggable genome and support for target identification and validation in drug development.
Sci Transl Med, 9(383), 2017.

[99] L. Folkersen, E. Fauman, M. Sabater-Lleal, R. J. Strawbridge, M. Franberg, B. Sennblad, D. Baldassarre, F. Veglia, S. E. Humphries, R. Rauramaa, U. de Faire, A. J. Smit, P. Giral, S. Kurl, E. Mannarino, S. Enroth, A. Johansson, S. B. Enroth, S. Gustafsson, L. Lind, C. Lindgren, A. P. Morris, V. Giedraitis, A. Silveira, A. Franco-Cereceda, E. Tremoli, I. s. group, U. Gyllensten, E. Ingelsson, S. Brunak, P. Eriksson, D. Ziemek, A. Hamsten, and A. Malarstig.
Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease.
PLoS Genet, 13(4):e1006706, 2017.

[100] B. D. Fox, V. J. Cheung, A. J. Patel, D. Suki, and G. Rao.
Epidemiology of metastatic brain tumors.
Neurosurg Clin N Am, 22(1):1–6, v, 2011.

[101] E. Franceschi, G. Cavallo, S. Lonardi, E. Magrini, A. Tosoni, D. Grosso, L. Scopece, V. Blatt, B. Urbini, A. Pession, G. Tallini, L. Crino, and A. A. Brandes.
Gefitinib in patients with progressive high-grade gliomas: a multicentre phase II study by gruppo italiano cooperativo di neuro-oncologia (GICNO).
Br J Cancer, 96(7):1047–51, 2007.

[102] J. M. Freeman, E. H. Kossoff, and A. L. Hartman.
The ketogenic diet: one decade later.
Pediatrics, 119(3):535–43, 2007.

[103] P. Frei, A. H. Poulsen, C. Johansen, J. H. Olsen, M. Steding-Jessen, and J. Schuz.
Use of mobile phones and risk of brain tumours: update of danish cohort study.
BMJ, 343:d6387, 2011.

[104] W. A. Freije, F. E. Castro-Vargas, Z. Fang, S. Horvath, T. Cloughesy, L. M. Liau, P. S. Mischel, and S. F. Nelson.
Gene expression profiling of gliomas strongly predicts survival.
Cancer Res, 64(18):6503–10, 2004.

[105] H. S. Friedman, T. Kerby, and H. Calvert.
Temozolomide and treatment of malignant glioma.
Clin Cancer Res, 6(7):2585–2597, 2000.

[106] M. Fromer, P. Roussos, S. K. Sieberts, J. S. Johnson, D. H. Kavanagh, T. M. Perumal, D. M. Ruderfer, E. C. Oh, A. Topol, H. R. Shah, L. L. Klei, R. Kramer, D. Pinto, Z. H. Gumus, A. E. Cicek, K. K. Dang, A. Browne, C. Lu, L. Xie, B. Readhead, E. A. Stahl, J. Xiao, M. Parvizi, T. Hamamsy, J. F. Fullard, Y. C. Wang, M. C. Mahajan, J. M. Derry, J. T. Dudley, S. E. Hemby, B. A. Logsdon, K. Talbot, T. Raj, D. A. Bennett, P. L. De Jager, J. Zhu, B. Zhang, P. F. Sullivan, A. Chess, S. M. Purcell, L. A. Shinobu, L. M. Mangravite, H. Toyoshiba, R. E. Gur, C. G. Hahn, D. A. Lewis, V. Haroutunian, M. A. Peters, B. K. Lipska, J. D. Buxbaum, E. E. Schadt, K. Hirai, K. Roeder, K. J. Brennand, N. Katsanis, E. Domenici, B. Devlin, and P. Sklar.
Gene expression elucidates functional impact of polygenic risk for schizophrenia.
Nat Neurosci, 19(11):1442–1453, 2016.

[107] S. B. Gabriel, S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly, and D. Altshuler.

The structure of haplotype blocks in the human genome.
Science, 296(5576):2225–9, 2002.

[108] D. Gaist, L. Andersen, J. Hallas, H. T. Sorensen, H. D. Schroder, and S. Friis.
Use of statins and risk of glioma: a nationwide case-control study in Denmark.
Br J Cancer, 108(3):715–20, 2013.

[109] M. D. Gallagher and A. S. Chen-Plotkin.
The post-GWAS era: From association to function.
Am J Hum Genet, 102(5):717–730, 2018.

[110] Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, and G. A. McVean.
A map of human genome variation from population-scale sequencing.
Nature, 467(7319):1061–73, 2010.

[111] M.-M. Georgescu and A. Olar.
Genetic and histologic spatiotemporal evolution of recurrent, multifocal, multicentric and metastatic glioblastoma.
Acta Neuropathol Commun, 8(1):10, 2020.

[112] S. Ghasimi, C. Wibom, A. M. Dahlin, T. Brännström, I. Golovleva, U. Andersson, and B. Melin.
Genetic risk variants in the cdkn2a/b, rtel1 and egfr genes are associated with somatic biomarkers in glioma.
J Neuro-Oncol, 127(3):483–492, 2016.

[113] C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, and V. Plagnol.
Bayesian test for colocalisation between pairs of genetic association studies using summary statistics.
PLoS Genet, 10(5):e1004383, 2014.

[114] C. Giambartolomei, J. Zhenli Liu, W. Zhang, M. Hauberg, H. Shi, J. Boocock, J. Pickrell, A. E. Jaffe, CommonMind Consortium, B. Pasaniuc, and P. Roussos.
A bayesian framework for multiple trait colocalization from summary association statistics.
Bioinformatics, 34(15):2538–2545, 2018.

[115] E. L. Goode, A. M. Dunning, B. Kuschel, C. S. Healey, N. E. Day, B. A. J. Ponder, D. F. Easton, and P. P. D. Pharoah.
Effect of germ-line genetic variation on breast cancer survival in a population-based study.
Cancer Res, 62(11):3052–3057, 2002.

[116] M. L. Goodenberger and R. B. Jenkins.
Genetics of adult glioma.
Cancer Genet, 205(12):613 – 621, 2012.

[117] K. Gousias, M. Markou, V. Arzoglou, S. Voulgaris, G. Vartholomatos, A. Kostoula, P. Voulgari, K. Polyzoidis, and A. P. Kyritsis.
Frequent abnormalities of the immune system in gliomas and correlation with the WHO grading system of malignancy.
J Neuroimmunol, 226(1-2):136–42, 2010.

[118] L. A. Gravendeel, M. C. Kouwenhoven, O. Gevaert, J. J. de Rooi, A. P. Stubbs, J. E. Duijm, A. Daemen, F. E. Bleeker, L. B. Bralten, N. K. Kloosterhof, B. De Moor, P. H. Eilers, P. J. van der Spek, J. M. Kros, P. A. Sillevis Smitt, M. J. van den Bent, and P. J. French.
Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology.
Cancer Res, 69(23):9065–72, 2009.

[119] C. Grommes, D. S. Conway, A. Alshekhlee, and J. S. Barnholtz-Sloan.
Inverse association of PPARgamma agonists use and high grade glioma development.
J Neurooncol, 100(2):233–9, 2010.

[120] C. Grommes, G. E. Landreth, M. Sastre, M. Beck, D. L. Feinstein, A. H. Jacobs, U. Schlegel, and M. T. Heneka.
Inhibition of in vivo glioma growth and invasion by peroxisome proliferator-activated receptor γ agonist treatment.
Mol Pharmacol, 70(5):1524–1533, 2006.

[121] G.TEx Consortium.
The genotype-tissue expression (GTEx) project.
Nat Genet, 45(6):580–5, 2013.

[122] S. K. Guo, M. F. Shen, H. W. Yao, and Y. S. Liu.
Enhanced expression of TGFBI promotes the proliferation and migration of glioma cells.
Cell Physiol Biochem, 49(3):1097–1109, 2018.

[123] K. Gupta and P. Salunke.
Molecular markers of glioma: an update on recent progress and perspectives.
J Cancer Res Clin Oncol, 138(12):1971–81, 2012.

[124] P. H. Gutin, F. M. Iwamoto, K. Beal, N. A. Mohile, S. Karimi, B. L. Hou, S. Lymberis, Y. Yamada, J. Chang, and L. E. Abrey.
Safety and efficacy of bevacizumab with hypofractionated stereotactic irradiation for recurrent malignant gliomas.
Int J Radiat Oncol Biol Phys, 75(1):156–63, 2009.

[125] D. A. Haas-Kogan, M. D. Prados, T. Tihan, D. A. Eberhard, N. Jelluma, N. D. Arvold, R. Baumber, K. R. Lamborn, A. Kapadia, M. Malec, M. S. Berger, and D. Stokoe.
Epidermal growth factor receptor, protein kinase B/Akt, and glioma response to erlotinib.
J Natl Cancer Inst, 97(12):880–7, 2005.

[126] W. H. Hall, R. Ramachandran, S. Narayan, A. B. Jani, and S. Vijayakumar.
An electronic application for rapidly calculating Charlson comorbidity score.
BMC Cancer, 4:94, 2004.

[127] S. Han, L. Meng, Y. Jiang, W. Cheng, X. Tie, J. Xia, and A. Wu.
Lithium enhances the antitumour effect of temozolomide against TP53 wild-type glioblastoma cells via NFAT1/FasL signalling.
Br J Cancer, 116(10):1302–1311, 2017.

[128] L. Hardell and M. Carlberg.
Mobile phones, cordless phones and the risk for brain tumours.
Int J Oncol, 35(1):5–17, 2009.

[129] L. Hardell and M. Carlberg.
Mobile phone and cordless phone use and the risk for glioma - analysis of pooled case-control studies in Sweden, 1997-2003 and 2007-2009.
Pathophysiology, 22(1):1–13, 2015.

[130] L. Hardell, M. Carlberg, and K. Hansson Mild.
Use of cellular telephones and brain tumour risk in urban and rural areas.
Occup Environ Med, 62(6):390–4, 2005.

[131] L. Hardell, A. Hallquist, K. H. Mild, M. Carlberg, A. Pahlson, and A. Lilja.
Cellular and cordless telephones and the risk for brain tumours.
Eur J Cancer Prev, 11(4):377–386, 2002.

[132] L. Hardell, K. H. Mild, and M. Carlberg.
Further aspects on cellular and cordless telephones and brain tumours.
Int J Oncol, 22(2):399–407, 2003.

[133] L. Hardell, A. Nasman, A. Pahlson, A. Hallquist, and K. Hansson Mild.
Use of cellular telephones and the risk for brain tumours: A case-control study.
International Journal of Oncology, 1999.

[134] J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, et al.
GENCODE: the reference human genome annotation for The ENCODE Project.
Genome Res, 22(9):1760–1774, 2012.

[135] M. G. Hart, R. Garside, G. Rogers, K. Stein, and R. Grant.
Temozolomide for high grade glioma.
Cochrane Database Syst Rev, (4):CD007415, 2013.

[136] C. Hartmann, J. Meyer, J. Balss, D. Capper, W. Mueller, A. Christians, J. Felsberg, M. Wolter, C. Mawrin, W. Wick, M. Weller, C. Herold-Mende, A. Unterberg, J. W. Jeuken, P. Wesseling, G. Reifenberger, and A. von Deimling.
Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas.
Acta Neuropathol, 118(4):469–74, 2009.

[137] H. R. Haynes, P. White, K. M. Hares, J. Redondo, K. C. Kemp, W. G. B. Singleton, C. L. Killick-Cole, J. R. Stevens, K. Garadi, S. Guglani, A. Wilkins, and K. M. Kurian.
The transcription factor PPARalpha is overexpressed and is associated with a favourable prognosis in IDH-wildtype primary glioblastoma.
Histopathology, 70(7):1030–1043, 2017.

[138] M. E. Hegi, A. C. Diserens, T. Gorlia, M. F. Hamou, N. de Tribolet, M. Weller, J. M. Kros, J. A. Hainfellner, W. Mason, L. Mariani, J. E. Bromberg, P. Hau, R. O. Mirimanoff, J. G. Cairncross, R. C. Janzer, and R. Stupp.
MGMT gene silencing and benefit from temozolomide in glioblastoma.
N Engl J Med, 352(10):997–1003, 2005.

[139] A. B. Heimberger, R. Hlatky, D. Suki, D. Yang, J. Weinberg, M. Gilbert, R. Sawaya, and K. Aldape.
Prognostic effect of epidermal growth factor receptor and EGFRvIII in glioblastoma multiforme patients.
Clin Cancer Res, 11(4):1462–1466, 2005.

[140] G. Hemani, J. Bowden, and G. Davey Smith.
Evaluating the potential role of pleiotropy in mendelian randomization studies.
Hum Mol Genet, 27(R2):R195–R208, 2018.

[141] G. Hemani, K. Tilling, and G. Davey Smith.
Orienting the causal relationship between imprecisely measured traits using GWAS summary data.
PLoS Genet, 13(11):e1007081, 2017.

[142] G. Hemani, J. Zheng, B. Elsworth, K. H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J. Bowden, R. Langdon, V. Y. Tan, J. Yarmolinsky, H. A. Shihab, N. J. Timpson, D. M. Evans, C. Relton, R. M. Martin, G. Davey Smith, T. R. Gaunt, and P. C. Haycock.
The MR-Base platform supports systematic causal inference across the human phenome.

Elife, 7, 2018.

[143] C. B. Henderson, F. M. Filloux, S. C. Alder, J. L. Lyon, and D. A. Caplin.
Efficacy of the ketogenic diet as a treatment option for epilepsy: meta-analysis.
J Child Neurol, 21(3):193–8, 2006.

[144] D. G. Hernandez, M. A. Nalls, M. Moore, S. Chong, A. Dillman, D. Trabzuni, J. R. Gibbs, M. Ryten, S. Arepalli, M. E. Weale, A. B. Zonderman, J. Troncoso, R. O'Brien, R. Walker, C. Smith, S. Bandinelli, B. J. Traynor, J. Hardy, A. B. Singleton, and M. R. Cookson.
Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain.
Neurobiol Dis, 47(1):20–8, 2012.

[145] E. Herrett, A. M. Gallagher, K. Bhaskaran, H. Forbes, R. Mathur, T. van Staa, and L. Smeeth.
Data resource profile: Clinical practice research datalink (CPRD).
Int J Epidemiol, 44(3):827–36, 2015.

[146] E. Herrett, A. D. Shah, R. Boggon, S. Denaxas, L. Smeeth, T. van Staa, A. Timmis, and H. Hemingway.
Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study.
BMJ, 346:f2350, 2013.

[147] L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio.
Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.
Proc Natl Acad Sci U S A, 106(23):9362–7, 2009.

[148] A. D. Hingorani, V. Kuan, C. Finan, F. A. Kruger, A. Gaulton, S. Chopade, R. Sofat, R. J. MacAllister, J. P. Overington, H. Hemingway, S. Denaxas, D. Prieto, and J. P. Casas.
Improving the odds of drug development success through human genomics: modelling study.
Sci Rep, 9(1):18911, 2019.

[149] A. L. Hopkins and C. R. Groom.
The druggable genome.
Nat Rev Drug Discov, 1(9):727–30, 2002.

[150] F. Hormozdiari, M. van de Bunt, A. V. Segre, X. Li, J. W. J. Joo, M. Bilow, J. H. Sul, S. Sankararaman, B. Pasaniuc, and E. Eskin.

Colocalization of GWAS and eQTL signals detects target genes.
Am J Hum Genet, 99(6):1245–1260, 2016.

[151] S. Horn, A. Figl, P. S. Rachakonda, C. Fischer, A. Sucker, A. Gast, S. Kadel, I. Moll, E. Nagore, K. Hemminki, D. Schadendorf, and R. Kumar.
TERT promoter mutations in familial and sporadic melanoma.
Science, 339(6122):959–961, 2013.

[152] A. E. Howell, J. W. Robinson, R. E. Wootton, A. McAleenan, S. Tsavachidis, Q. T. Ostrom, M. Bondy, G. Armstrong, C. Relton, P. Haycock, R. M. Martin, J. Zheng, and K. M. Kurian.
Testing for causality between systematically identified risk factors and glioma: a Mendelian randomization study.
BMC Cancer, 20(1):508, 2020.

[153] A. E. Howell, J. Zheng, P. C. Haycock, A. McAleenan, C. Relton, R. M. Martin, and K. M. Kurian.
Use of Mendelian randomization for identifying risk factors for brain tumors.
Front Genet, 9:525, 2018.

[154] B. N. Howie, P. Donnelly, and J. Marchini.
A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.
PLoS Genet, 5(6):e1000529, 2009.

[155] J. B. Hsu, T. H. Chang, G. A. Lee, T. Y. Lee, and C. Y. Chen.
Identification of potential biomarkers related to glioma survival by gene expression profile analysis.
BMC Med Genomics, 11(Suppl 7):34, 2019.

[156] F. W. Huang, E. Hodis, M. J. Xu, G. V. Kryukov, L. Chin, and L. A. Garraway.
Highly recurrent TERT promoter mutations in human melanoma.
Science, 339(6122):957–959, 2013.

[157] L. Hutchinson and R. Kirk.
High drug attrition rates–where are we going wrong?
Nat Rev Clin Oncol, 8(4):189–90, 2011.

[158] P. D. Inskip, A. J. Sigurdson, L. Veiga, P. Bhatti, C. Ronckers, P. Rajaraman, H. Boukheris, M. Stovall, S. Smith, S. Hammond, T. O. Henderson, T. C. Watt, A. C. Mertens, W. Leisenring, K. Stratton, J. Whitton, S. S. Donaldson, G. T. Armstrong, L. L. Robison, and J. P. Neglia.

Radiation-related new primary solid cancers in the childhood cancer survivor study: Comparative radiation dose response and modification of treatment effects.
Int J Radiat Oncol Biol Phys, 94(4):800–7, 2016.

[159] Intergroup Radiation Therapy Oncology Group Trial, G. Cairncross, B. Berkey, E. Shaw, R. Jenkins, B. Scheithauer, D. Brachman, J. Buckner, K. Fink, L. Souhami, N. Laperierre, M. Mehta, and W. Curran.
Phase III trial of chemotherapy plus radiotherapy compared with radiotherapy alone for pure and mixed anaplastic oligodendroglioma: Intergroup radiation therapy oncology group trial 9402.
J Clin Oncol, 24(18):2707–14, 2006.

[160] International Agency for Research on Cancer.
IARC monographs on the identificaiton of carcinogenic hazards to humans.
https://monographs.iarc.fr/list-of-classifications.
Accessed: 2020-08.

[161] International Consortium on Lithium Genetics, A. T. Amare, K. O. Schubert, L. Hou, S. R. Clark, S. Papiol, U. Heilbronner, F. Degenhardt, F. Tekola-Ayele, Y. H. Hsu, T. Shekhtman, M. Adli, N. Akula, K. Akiyama, R. Ardau, B. Arias, J. M. Aubry, L. Backlund, A. K. Bhattacharjee, F. Bellivier, A. Benabarre, S. Bengesser, J. M. Biernacka, A. Birner, C. Brichant-Petitjean, P. Cervantes, H. C. Chen, C. Chillotti, S. Cichon, C. Cruceanu, P. M. Czerski, N. Dalkner, A. Dayer, M. Del Zompo, J. R. DePaulo, B. Etain, P. Falkai, A. J. Forstner, L. Frisen, M. A. Frye, J. M. Fullerton, S. Gard, J. S. Garnham, F. S. Goes, M. Grigoroiu-Serbanescu, P. Grof, R. Hashimoto, J. Hauser, S. Herms, P. Hoffmann, A. Hofmann, S. Jamain, E. Jimenez, J. P. Kahn, L. Kassem, P. H. Kuo, T. Kato, J. Kelsoe, S. Kittel-Schneider, S. Kliwicki, B. Konig, I. Kusumi, G. Laje, M. Landen, C. Lavebratt, M. Leboyer, S. G. Leckband, A. Tortorella, M. Manchia, L. Martinsson, M. J. McCarthy, S. McElroy, F. Colom, M. Mitjans, F. M. Mondimore, P. Monteleone, C. M. Nievergelt, M. M. Nothen, T. Novak, C. O'Donovan, N. Ozaki, U. Osby, A. Pfennig, J. B. Potash, A. Reif, E. Reininghaus, G. A. Rouleau, J. K. Rybakowski, M. Schalling, P. R. Schofield, B. W. Schweizer, G. Severino, P. D. Shilling, K. Shimoda, C. Simhandl, C. M. Slaney, A. Squassina, T. Stamm, P. Stopkova, M. Maj, et al.
Association of polygenic score for schizophrenia and HLA antigen and inflammation genes with response to lithium in bipolar affective disorder: A genome-wide association study.
JAMA Psychiatry, 75(1):65–74, 2018.

[162] International HapMap Consortium, K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal, S. Pasternak, D. A. Wheeler, T. D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao,

H. Zhao, J. Zhou, S. B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R. C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, Y. Shen, W. Sun, H. Wang, Y. Wang, Y. Wang, X. Xiong, L. Xu, M. M. Waye, S. K. Tsui, H. Xue, J. T. Wong, L. M. Galver, J. B. Fan, K. Gunderson, S. S. Murray, A. R. Oliphant, M. S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J. F. Olivier, M. S. Phillips, S. Roumy, C. Sallee, A. Verner, T. J. Hudson, P. Y. Kwok, D. Cai, D. C. Koboldt, R. D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L. C. Tsui, W. Mak, Y. Q. Song, P. K. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, et al.

A second generation human haplotype map of over 3.1 million SNPs.

Nature, 449(7164):851–61, 2007.

[163] Interphone Study Group.

Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study.

Int J Epidemiol, 39(3):675–94, 2010.

[164] C. Irwin, M. Hunn, G. Purdie, and D. Hamilton.

Delay in radiotherapy shortens survival in patients with high grade glioma.

J Neurooncol, 85(3):339–43, 2007.

[165] A. S. Jakola, K. S. Myrmel, R. Kloster, S. H. Torp, S. Lindal, G. Unsgard, and O. Solheim.

Comparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas.

JAMA, 308(18):1881–8, 2012.

[166] A. Jalali, E. S. Amirian, M. N. Bainbridge, G. N. Armstrong, Y. Liu, S. Tsavachidis, S. N. Jhangiani, S. E. Plon, C. C. Lau, E. B. Claus, J. S. Barnholtz-Sloan, D. Il'yasova, J. Schildkraut, F. Ali-Osman, S. Sadetzki, C. Johansen, R. S. Houlston, R. B. Jenkins, D. Lachance, S. H. Olson, J. L. Bernstein, R. T. Merrell, M. R. Wrensch, F. G. Davis, R. Lai, S. Shete, K. Aldape, C. I. Amos, D. M. Muzny, R. A. Gibbs, B. S. Melin, and M. L. Bondy.

Targeted sequencing in chromosome 17q linkage region identifies familial glioma candidates in the Gliogene Consortium.

Sci Rep, 5:8278, 2015.

[167] A. Jelenkovic, R. Sund, Y.-M. Hur, Y. Yokoyama, J. v. B. Hjelmborg, S. Moeller, C. Honda, P. K. E. Magnusson, N. L. Pedersen, S. Ooki, S. Aaltonen, M. A. Stazi, C. Fagnani, C. D,ÄôIppolito, D. L. Freitas, J. A. Maia, F. Ji, F. Ning, Z. Pang, E. Rebato, A. Busjahn, C. Kandler, K. J. Saudino, K. L. Jang, W. Cozen, A. E. Hwang, T. M. Mack, W. Gao,

C. Yu, L. Li, R. P. Corley, B. M. Huibregtse, C. A. Derom, R. F. Vlietinck, R. J. F. Loos, K. Heikkila, J. Wardle, C. H. Llewellyn, A. Fisher, T. A. McAdams, T. C. Eley, A. M. Gregory, M. He, X. Ding, M. Bjerregaard-Andersen, H. Beck-Nielsen, M. Sodemann, A. D. Tarnoki, D. L. Tarnoki, A. Knafo-Noam, D. Mankuta, L. Abramson, S. A. Burt, K. L. Klump, J. L. Silberg, L. J. Eaves, H. H. Maes, R. F. Krueger, M. McGue, S. Pahlen, M. Gatz, D. A. Butler, M. Bartels, T. C. E. M. van Beijsterveldt, J. M. Craig, R. Saffery, L. Dubois, M. Boivin, M. Brendgen, G. Dionne, F. Vitaro, N. G. Martin, S. E. Medland, G. W. Montgomery, G. E. Swan, R. Krasnow, P. Tynelius, P. Lichtenstein, C. M. A. Haworth, R. Plomin, G. Bayasgalan, D. Narandalai, K. P. Harden, E. M. Tucker-Drob, T. Spector, M. Mangino, G. Lachance, L. A. Baker, C. Tuvblad, G. E. Duncan, D. Buchwald, G. Willemsen, A. Skytthe, K. O. Kyvik, K. Christensen, S. Y. OEncel, F. Aliev, F. Rasmussen, J. H. Goldberg, T. I. A. Sorensen, et al.

Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts.

Sci Rep, 6(1):28496, 2016.

[168] R. B. Jenkins, Y. Xiao, H. Sicotte, P. A. Decker, T. M. Kollmeyer, H. M. Hansen, M. L. Kosel, S. Zheng, K. M. Walsh, T. Rice, P. Bracci, L. S. McCoy, I. Smirnov, J. S. Patoka, G. Hsuang, J. L. Wiemels, T. Tihan, A. R. Pico, M. D. Prados, S. M. Chang, M. S. Berger, A. A. Caron, S. R. Fink, C. Halder, A. L. Rynearson, B. L. Fridley, J. C. Buckner, B. P. O'Neill, C. Giannini, D. H. Lachance, J. K. Wiencke, J. E. Eckel-Passow, and M. R. Wrensch.

A low-frequency variant at 8q24.21 is strongly associated with risk of oligodendroglial tumors and astrocytomas with IDH1 or IDH2 mutation.

Nat Genet, 44(10):1122–5, 2012.

[169] Y. Jiao, P. J. Killela, Z. J. Reitman, A. B. Rasheed, C. M. Heaphy, R. F. de Wilde, F. J. Rodriguez, S. Rosemberg, S. M. Oba-Shinjo, S. K. Nagahashi Marie, C. Bettegowda, N. Agrawal, E. Lipp, C. Pirozzi, G. Lopez, Y. He, H. Friedman, A. H. Friedman, G. J. Riggins, M. Holdhoff, P. Burger, R. McLendon, D. D. Bigner, B. Vogelstein, A. K. Meeker, K. W. Kinzler, N. Papadopoulos, L. A. Diaz, and H. Yan.

Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas.

Oncotarget, 3(7):709–22, 2012.

[170] S. S. Jick, J. A. Kaye, C. Vasilakis-Scaramozza, L. A. Garcia Rodriguez, A. Ruigomez, C. R. Meier, R. G. Schlienger, C. Black, and H. Jick.

Validity of the general practice research database.

Pharmacotherapy, 23(5):686–9, 2003.

[171] T. M. Johanns, C. J. Ferguson, P. M. Grierson, S. Dahiya, and G. Ansstas.

Rapid clinical and radiographic response with combined dabrafenib and trametinib in adults with BRAF-mutated high-grade glioma.
J Natl Compr Canc Netw, 16(1):4 – 10, 2018.

[172] N. F. Khan, R. Perera, S. Harper, and P. W. Rose.
Adaptation and validation of the charlson index for read/oxmis coded databases.
BMC Fam Pract, 11:1, 2010.

[173] N. K. Kibinge, C. L. Relton, T. R. Gaunt, and T. G. Richardson.
Characterizing the causal pathway for genetic variants associated with neurological phenotypes using human brain-derived proteome data.
Am J Hum Genet, 106(6):885–892, 2020.

[174] P. J. Killela, C. J. Pirozzi, P. Healy, Z. J. Reitman, E. Lipp, B. A. Rasheed, R. Yang, B. H. Diplas, Z. Wang, P. K. Greer, H. Zhu, C. Y. Wang, A. B. Carpenter, H. Friedman, A. H. Friedman, S. T. Keir, J. He, Y. He, R. E. McLendon, n. Herndon, J. E., H. Yan, and D. D. Bigner.
Mutations in IDH1, IDH2, and in the TERT promoter define clinically distinct subgroups of adult malignant gliomas.
Oncotarget, 5(6):1515–25, 2014.

[175] J. Kim, I.-H. Lee, H. J. Cho, C.-K. Park, Y.-S. Jung, Y. Kim, S. H. Nam, B. S. Kim, M. D. Johnson, D.-S. Kong, H. J. Seol, J.-I. Lee, K. M. Joo, Y. Yoon, W.-Y. Park, J. Lee, P. J. Park, and D.-H. Nam.
Spatiotemporal evolution of the primary glioblastoma genome.
Cancer Cell, 28(3):318–328, 2015.

[176] S.-B. Kim, R. Dent, S.-A. Im, M. Espie, S. Blau, A. R. Tan, S. J. Isakoff, M. Oliveira, C. Saura, M. J. Wongchenko, A. V. Kapp, W. Y. Chan, S. M. Singel, D. J. Maslyar, J. Baselga, S.-B. Kim, K. S. Lee, S.-A. Im, M. Espie, H.-C. Wang, S. Blau, R. Dent, A. Tan, J. H. Sohn, M. De Laurentiis, L. G. Estevez, C.-S. Huang, G. Romieu, M. Velez, R. Villanueva, P. F. Conte, S. Dakhil, M. Debled, A. G. Martin, S. Hurvitz, J. H. Kim, C. Levy, M. Oliveira, P. S. Rovira, J. H. Seo, V. Valero, G. Vidal, A. Wong, M. A. K. Allison, R. Figlin, D. Chan, S.-C. Chen, Y.-H. Chen, M. Cobleigh, F. De Braud, L. Dirix, V. Hansen, A. H. Bessard, N. Iannotti, S. Isakoff, W. Lawler, A. Montano, M. Salkini, and L. Seigel.
Ipatasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (LOTUS): a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial.
Lancet Oncol, 18(10):1360–1372, 2017.

[177] S. Kim-Hellmuth, F. Aguet, M. Oliva, M. Munoz-Aguirre, S. Kasela, V. Wucher, S. E. Castel, A. R. Hamel, A. Vinuela, A. L. Roberts, S. Mangul, X. Wen, G. Wang, A. N.

Barbeira, D. Garrido-Martin, B. B. Nadel, Y. Zou, R. Bonazzola, J. Quan, A. Brown, A. Martinez-Perez, J. M. Soria, G. T. Consortium, G. Getz, E. T. Dermitzakis, K. S. Small, M. Stephens, H. S. Xi, H. K. Im, R. Guigo, A. V. Segre, B. E. Stranger, K. G. Ardlie, and T. Lappalainen.
Cell type-specific genetic regulation of gene expression across human tissues.
Science, 369(6509), 2020.

[178] E. A. King, J. W. Davis, and J. F. Degner.
Are drug targets with genetic support twice as likely to be approved? revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval.
PLOS Genetics, 15(12):1–20, 12 2019.

[179] B. Kinnersley, R. S. Houlston, and M. L. Bondy.
Genome-wide association studies in glioma.
Cancer Epidemiol Biomark Prev, 27(4):418–428, 2018.

[180] B. Kinnersley, M. Labussiere, A. Holroyd, A. L. Di Stefano, P. Broderick, J. Vijayakrishnan, K. Mokhtari, J. Y. Delattre, K. Gousias, J. Schramm, M. J. Schoemaker, S. J. Fleming, S. Herms, S. Heilmann, S. Schreiber, H. E. Wichmann, M. M. Nothen, A. Swerdlow, M. Lathrop, M. Simon, M. Bondy, M. Sanson, and R. S. Houlston.
Genome-wide association study identifies multiple susceptibility loci for glioma.
Nat Commun, 6:8559, 2015.

[181] S. Kitamura, Y. Miyazaki, Y. Shinomura, S. Kondo, S. Kanayama, and Y. Matsuzawa.
Peroxisome proliferator-activated receptor gamma induces growth arrest and differentiation markers of human colon cancer cells.
Jpn J Cancer Res, 90(1):75–80, 1999.

[182] M. Klein, J. J. Heimans, N. K. Aaronson, H. M. van der Ploeg, J. Grit, M. Muller, T. J. Postma, J. J. Mooij, R. H. Boerman, G. N. Beute, G. J. Ossenkoppele, G. W. van Imhoff, A. W. Dekker, J. Jolles, B. J. Slotman, H. Struikmans, and M. J. B. Taphoorn.
Effect of radiotherapy and other treatment-related factors on mid-term to long-term cognitive sequelae in low-grade gliomas: a comparative study.
Lancet, 360(9343):1361–1368, 2002.

[183] S. A. Kliewer, S. S. Sundseth, S. A. Jones, P. J. Brown, G. B. Wisely, C. S. Koble, P. Devchand, W. Wahli, T. M. Willson, J. M. Lenhard, and J. M. Lehmann.
Fatty acids and eicosanoids regulate gene expression through direct interactions with peroxisome proliferator-activated receptors alpha and gamma.
Proc Natl Acad Sci U S A, 94(9):4318–23, 1997.

[184] P. Koivunen, S. Lee, C. G. Duncan, G. Lopez, G. Lu, S. Ramkissoon, J. A. Losman, P. Joensuu, U. Bergmann, S. Gross, J. Travins, S. Weiss, R. Looper, K. L. Ligon, R. G. Verhaak, H. Yan, and J. Kaelin, W. G.
Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation.
Nature, 483(7390):484–8, 2012.

[185] I. Kola and J. Landis.
Can the pharmaceutical industry reduce attrition rates?
Nat Rev Drug Discov, 3(8):711–5, 2004.

[186] T. N. Kreisl, L. Kim, K. Moore, P. Duic, C. Royce, I. Stroud, N. Garren, M. Mackey, J. A. Butman, K. Camphausen, J. Park, P. S. Albert, and H. A. Fine.
Phase II trial of single-agent bevacizumab followed by bevacizumab plus irinotecan at tumor progression in recurrent glioblastoma.
J Clin Oncol, 27(5):740–5, 2009.

[187] N. Kryuchkova-Mostacci and M. Robinson-Rechavi.
A benchmark of gene expression tissue-specificity metrics.
Brief Bioinform, 18(2):205–214, 2017.

[188] D. N. Kyriacou and R. J. Lewis.
Confounding by indication in clinical research.
JAMA, 316(17):1818–1819, 2016.

[189] A. P. Kyritsis, M. L. Bondy, M. Xiao, E. L. Berman, J. E. Cunningham, P. S. Lee, V. A. Levin, and H. Saya.
Germline p53 Gene Mutations in Subsets of Glioma Patients.
J Natl Cancer Inst, 86(5):344–349, 03 1994.

[190] M. Labussière, B. Boisselier, K. Mokhtari, A.-L. Di Stefano, A. Rahimian, M. Rossetto, P. Ciccarino, O. Saulnier, R. Paterra, Y. Marie, G. Finocchiaro, and M. Sanson.
Combined analysis of TERT, EGFR, and IDH status defines distinct prognostic glioblastoma classes.
Neurology, 83(13):1200–1206, 2014.

[191] D. H. Lachance, P. Yang, D. R. Johnson, P. A. Decker, T. M. Kollmeyer, L. S. McCoy, T. Rice, Y. Xiao, F. Ali-Osman, F. Wang, S. M. Stoddard, D. J. Sprau, M. L. Kosel, J. K. Wiencke, J. L. Wiemels, J. S. Patoka, F. Davis, B. McCarthy, A. L. Rynearson, J. B. Worra, B. L. Fridley, B. P. O'Neill, J. C. Buckner, D. Il'yasova, R. B. Jenkins, and M. R. Wrensch.
Associations of high-grade glioma with glioma risk alleles and histories of allergy and smoking.
Am J Epidemiol, 174(5):574–81, 2011.

[192] A. Lacourt, E. Cardis, J. Pintos, L. Richardson, L. Kincl, G. Benke, S. Fleming, M. Hours, D. Krewski, D. McLean, M. E. Parent, S. Sadetzki, K. Schlaefer, B. Schlehofer, J. Lavoue, M. van Tongeren, and J. Siemiatycki.
INTEROCC case-control study: lack of association between glioma tumors and occupational exposure to selected combustion products, dusts and other chemical agents.
BMC Public Health, 13:340, 2013.

[193] Y. Landry and J.-P. Gies.
Drugs and their molecular targets: an updated overview.
Fund Clinc Pharmacol, 22(1):1–18, 2008.

[194] N. Laperriere, L. Zuraw, and G. Cairncross.
Radiotherapy for newly diagnosed malignant glioma in adults: a systematic review.
Radiother and Oncol, 64(3):259–273, 2002.

[195] S. Larjavaara, R. Mantyla, T. Salminen, H. Haapasalo, J. Raitanen, J. Jaaskelainen, and A. Auvinen.
Incidence of gliomas by anatomic location.
Neuro Oncol, 9(3):319–25, 2007.

[196] D. A. Lawlor, R. M. Harbord, J. A. Sterne, N. Timpson, and G. Davey Smith.
Mendelian randomization: using genes as instruments for making causal inferences in epidemiology.
Stat Med, 27(8):1133–63, 2008.

[197] D. A. Lawlor, K. Tilling, and G. Davey Smith.
Triangulation in aetiological epidemiology.
Int J Epidemiol, 45(6):1866–1886, 2016.

[198] H. Lee, J. K. Aronson, and D. Nunan.
Catalogue of bias collaboration: Collider bias.
https://catalogofbias.org/biases/collider-bias.
Accessed: 2020-07.

[199] S. Lehmann, V. J. Bykov, D. Ali, O. Andren, H. Cherif, U. Tidefelt, B. Uggla, J. Yachnin, G. Juliusson, A. Moshfegh, C. Paul, K. G. Wiman, and P. O. Andersson.
Targeting p53 in vivo: a first-in-human study with p53-targeting compound APR-246 in refractory hematologic malignancies and prostate cancer.
J Clin Oncol, 30(29):3633–9, 2012.

[200] V. A. Levin, P. Silver, J. Hannigan, W. M. Wara, P. H. Gutin, R. L. Davis, and C. B. Wilson.
Superiority of post-radiotherapy adjuvant chemotherapy with CCNU, procarrazine, and vincristine (PCV) over BCNU for anaplastic gliomas: NCOG 6G61 final report.

Int J Radiat Oncol Biol Phys, 18(2):321–324, 1990.

[201] X. Li, C. Wu, N. Chen, H. Gu, A. Yen, L. Cao, E. Wang, and L. Wang.
PI3K/Akt/mTOR signaling pathway and targeted therapy for glioblastoma.
Oncotarget, 7(22):33440–50, 2016.

[202] Y. Li, J. Byun, G. Cai, X. Xiao, Y. Han, O. Cornelis, J. E. Dinulos, J. Dennis, D. Easton, I. Gorlov, M. F. Seldin, and C. I. Amos.
FastPop: a rapid principal component derived method to infer intercontinental ancestry using genetic data.
BMC Bioinformatics, 17:122, 2016.

[203] E. Linos, T. Raine, A. Alonso, and D. Michaud.
Atopy and risk of brain tumors: a meta-analysis.
J Natl Cancer Inst, 99(20):1544–50, 2007.

[204] M. S. Litwin and H.-J. Tan.
The diagnosis and treatment of prostate cancer: A review.
JAMA, 317(24):2532–2542, 2017.

[205] J. Z. Liu, F. Tozzi, D. M. Waterworth, S. G. Pillai, P. Muglia, L. Middleton, W. Berrettini, C. W. Knouff, X. Yuan, G. Waeber, P. Vollenweider, M. Preisig, N. J. Wareham, J. H. Zhao, R. J. Loos, I. Barroso, K. T. Khaw, S. Grundy, P. Barter, R. Mahley, A. Kesaniemi, R. McPherson, J. B. Vincent, J. Strauss, J. L. Kennedy, A. Farmer, P. McGuffin, R. Day, K. Matthews, P. Bakke, A. Gulsvik, S. Lucae, M. Ising, T. Brueckl, S. Horstmann, H. E. Wichmann, R. Rawal, N. Dahmen, C. Lamina, O. Polasek, L. Zgaga, J. Huffman, S. Campbell, J. Kooner, J. C. Chambers, M. S. Burnett, J. M. Devaney, A. D. Pichard, K. M. Kent, L. Satler, J. M. Lindsay, R. Waksman, S. Epstein, J. F. Wilson, S. H. Wild, H. Campbell, V. Vitart, M. P. Reilly, M. Li, L. Qu, R. Wilensky, W. Matthai, H. H. Hakonarson, D. J. Rader, A. Franke, M. Wittig, A. Schafer, M. Uda, A. Terracciano, X. Xiao, F. Busonero, P. Scheet, D. Schlessinger, D. St Clair, D. Rujescu, G. R. Abecasis, H. J. Grabe, A. Teumer, H. Volzke, A. Petersmann, U. John, I. Rudan, C. Hayward, A. F. Wright, I. Kolcic, B. J. Wright, J. R. Thompson, A. J. Balmforth, A. S. Hall, N. J. Samani, C. A. Anderson, T. Ahmad, C. G. Mathew, M. Parkes, J. Satsangi, M. Caulfield, P. B. Munroe, M. Farrall, A. Dominiczak, J. Worthington, et al.
Meta-analysis and imputation refines the association of 15q25 with smoking quantity.
Nat Genet, 42(5):436–40, 2010.

[206] X. Liu, X. Chen, L. Shi, Q. Shan, Q. Cao, C. Yue, H. Li, S. Li, J. Wang, S. Gao, M. Niu, and R. Yu.
The third-generation EGFR inhibitor AZD9291 overcomes primary resistance by continuously blocking ERK signaling in glioblastoma.

J Exp Clin Cancer Res, 38(1):219, 2019.

[207] Y. Liu, B. Elsworth, P. Erola, V. Haberland, G. Hemani, M. Lyon, J. Zheng, O. Lloyd, M. Vabistsevits, and T. R. Gaunt.
EpiGraphDB: a database and data mining platform for health data science.
Bioinformatics, 11 2020.
btaa961.

[208] K. Lote, T. Egeland, B. Hager, B. Stenwig, K. Skullerud, J. Berg-Johnsen, I. Storm-Mathisen, and H. Hirschberg.
Survival, prognostic factors, and therapeutic efficacy in low-grade glioma: a retrospective study in 379 patients.
J Clin Oncol, 15(9):3129–40, 1997.

[209] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison.
The 2016 World Health Organization classification of tumors of the central nervous system: a summary.
Acta Neuropathol, 131(6):803–20, 2016.

[210] D. N. Louis, P. Wesseling, K. Aldape, D. J. Brat, D. Capper, I. A. Cree, C. Eberhart, D. Figarella-Branger, M. Fouladi, G. N. Fuller, C. Giannini, C. Haberler, C. Hawkins, T. Komori, J. M. Kros, H. K. Ng, B. A. Orr, S. H. Park, W. Paulus, A. Perry, T. Pietsch, G. Reifenberger, M. Rosenblum, B. Rous, F. Sahm, C. Sarkar, D. A. Solomon, U. Tabori, M. J. van den Bent, A. von Deimling, M. Weller, V. A. White, and D. W. Ellison.
cIMPACT-NOW update 6: new entity and diagnostic principle recommendations of the cIMPACT-Utrecht meeting on future CNS tumor classification and grading.
Brain Pathol, 30(4):844–856, 2020.

[211] C. Lu, P. S. Ward, G. S. Kapoor, D. Rohle, S. Turcan, O. Abdel-Wahab, C. R. Edwards, R. Khanin, M. E. Figueroa, A. Melnick, K. E. Wellen, D. M. O'Rourke, S. L. Berger, T. A. Chan, R. L. Levine, I. K. Mellinghoff, and C. B. Thompson.
IDH mutation impairs histone demethylation and results in a block to cell differentiation.
Nature, 483(7390):474–8, 2012.

[212] R. M. Lucas and R. M. Rodney Harris.
On the nature of evidence and 'proving' causality: Smoking and lung cancer vs. sun exposure, vitamin D and multiple sclerosis.
Int J Environ Res Public Health, 15(8), 2018.

[213] D. Ly, H. P. Bagshaw, C. J. Anker, J. D. Tward, K. F. Grossmann, R. L. Jensen, and D. C. Shrieve.

Local control after stereotactic radiosurgery for brain metastases in patients with melanoma with and without BRAF mutation and treatment.
J Neurosurg, 123(2):395–401, 2015.

[214] I. Makhlin, R. D. Salinas, D. Zhang, F. Jacob, G. L. Ming, H. Song, D. Saxena, J. F. Dorsey, M. P. Nasrallah, J. J. Morrissette, Z. A. Binder, D. M. O'Rourke, A. S. Desai, S. Brem, and S. J. Bagley.
Clinical activity of the EGFR tyrosine kinase inhibitor osimertinib in EGFR-mutant glioblastoma.
CNS Oncol, 8(3):CNS43, 2019.

[215] T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll, and P. M. Visscher.
Finding the missing heritability of complex diseases.
Nature, 461(7265):747–53, 2009.

[216] U. M. Marigorta and A. Navarro.
High trans-ethnic replicability of gwas results implies common causal variants.
PLoS Genet, 9(6):1–13, 06 2013.

[217] M. J. McGirt, K. D. Than, J. D. Weingart, K. L. Chaichana, F. J. Attenello, A. Olivi, J. Laterra, L. R. Kleinberg, S. A. Grossman, H. Brem, and A. Quinones-Hinojosa.
Gliadel (BCNU) wafer plus concomitant temozolomide therapy after primary resection of glioblastoma multiforme.
J Neurosurg, 110(3):583–8, 2009.

[218] M. McKenzie, A. K. Henders, A. Caracella, N. R. Wray, and J. E. Powell.
Overlap of expression quantitative trait loci (eQTL) in human brain and blood.
BMC Med Genomics, 7(1):31, 2014.

[219] C. Meier, M. E. Kraenzlin, M. Bodmer, S. S. Jick, H. Jick, and C. R. Meier.
Use of thiazolidinediones and fracture risk.
Arch Intern Med, 168(8):820–5, 2008.

[220] B. S. Melin, J. S. Barnholtz-Sloan, M. R. Wrensch, C. Johansen, D. Il'yasova, B. Kinnersley, Q. T. Ostrom, K. Labreche, Y. Chen, G. Armstrong, Y. Liu, J. E. Eckel-Passow, P. A. Decker, M. Labussiere, A. Idbaih, K. Hoang-Xuan, A. L. Di Stefano, K. Mokhtari, J. Y. Delattre, P. Broderick, P. Galan, K. Gousias, J. Schramm, M. J. Schoemaker, S. J. Fleming, S. Herms, S. Heilmann, M. M. Nothen, H. E. Wichmann, S. Schreiber, A. Swerdlow,

M. Lathrop, M. Simon, M. Sanson, U. Andersson, P. Rajaraman, S. Chanock, M. Linet, Z. Wang, M. Yeager, GliomaScan Consortium, J. K. Wiencke, H. Hansen, L. McCoy, T. Rice, M. L. Kosel, H. Sicotte, C. I. Amos, J. L. Bernstein, F. Davis, D. Lachance, C. Lau, R. T. Merrell, J. Shildkraut, F. Ali-Osman, S. Sadetzki, M. Scheurer, S. Shete, R. K. Lai, E. B. Claus, S. H. Olson, R. B. Jenkins, R. S. Houlston, and M. L. Bondy.
Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors.
Nat Genet, 49(5):789–794, 2017.

[221] T. E. Merchant, L. E. Kun, S. Wu, X. Xiong, R. A. Sanford, and F. A. Boop.
Phase II trial of conformal radiation therapy for pediatric low-grade glioma.
J Clin Oncol, 27(22):3598–604, 2009.

[222] D. S. Michaud, C. N. Holick, T. T. Batchelor, E. Giovannucci, and D. J. Hunter.
Prospective study of meat intake and dietary nitrates, nitrites, and nitrosamines and risk of adult glioma.
Am J Clin Nutr, 90(3):570–7, 2009.

[223] D. Migliorini, D. Aguiar, M. I. Vargas, A. Lobrinus, and P. Y. Dietrich.
BRAF/MEK double blockade in refractory anaplastic pleomorphic xanthoastrocytoma.
Neurology, 88(13):1291–1293, 2017.

[224] J. L. Min, G. Hemani, E. Hannon, K. F. Dekkers, J. Castillo-Fernandez, R. Luijk, E. Carnero-Montoro, D. J. Lawson, K. Burrows, M. Suderman, A. D. Bretherick, T. G. Richardson, J. Klughammer, V. Iotchkova, G. Sharp, A. A. Khleifat, A. Shatunov, A. Iacoangeli, W. L. McArdle, K. M. Ho, A. Kumar, C. Söderhäll, C. Soriano-Tárraga, E. Giralt-Steinhauer, N. Kazmi, D. Mason, A. F. McRae, D. L. Corcoran, K. Sugden, S. Kasela, A. Cardona, F. R. Day, G. Cugliari, C. Viberti, S. Guarrera, M. Lerro, R. Gupta, S. Bollepalli, P. Mandaviya, Y. Zeng, T.-K. Clarke, R. M. Walker, V. Schmoll, D. Czamara, C. Ruiz-Arenas, F. I. Rezwan, R. E. Marioni, T. Lin, Y. Awaloff, M. Germain, D. Aïssi, R. Zwamborn, K. v. Eijk, A. Dekker, J. v. Dongen, J.-J. Hottenga, G. Willemsen, C.-J. Xu, G. Barturen, F. Català-Moll, M. Kerick, C. Wang, P. Melton, H. R. Elliott, J. Shin, M. Bernard, I. Yet, M. Smart, T. Gorrie-Stone, C. Shaw, A. A. Chalabi, S. M. Ring, G. Pershagen, E. Melén, J. Jiménez-Conde, J. Roquer, D. A. Lawlor, J. Wright, N. G. Martin, G. W. Montgomery, T. E. Moffitt, R. Poulton, T. Esko, L. Milani, A. Metspalu, J. R. B. Perry, K. K. Ong, N. J. Wareham, G. Matullo, C. Sacerdote, A. Caspi, L. Arseneault, F. Gagnon, M. Ollikainen, J. Kaprio, J. F. Felix, F. Rivadeneira, H. Tiemeier, M. H. v. IJzendoorn, A. G. Uitterlinden, V. W. Jaddoe, C. Haley, A. M. McIntosh, K. L. Evans, A. Murray, K. Räikkönen, J. Lahti, E. A. Nohr, T. I. Sørensen, T. Hansen, C. S. Morgen, E. B. Binder, S. Lucae, J. R. Gonzalez, M. Bustamante, J. Sunyer, J. W. Holloway, W. Karmaus, H. Zhang, I. J. Deary, N. R. Wray, J. M. Starr, M. Beekman, D. v. Heemst, P. E. Slagboom, P.-E. Morange,

D.-A. Trégouët, J. H. Veldink, G. E. Davies, E. J. d. Geus, D. I. Boomsma, J. M. Vonk, B. Brunekreef, G. H. Koppelman, M. E. Alarcón-Riquelme, R.-C. Huang, C. Pennell, J. v. Meurs, M. A. Ikram, A. D. Hughes, T. Tillin, N. Chaturvedi, Z. Pausova, T. Paus, T. D. Spector, M. Kumari, L. C. Schalkwyk, P. M. Visscher, G. D. Smith, C. Bock, T. R. Gaunt, J. T. Bell, B. T. Heijmans, J. Mill, and C. L. Relton.
Genomic and phenomic insights from an atlas of genetic effects on DNA methylation.
medRxiv, 2020.

[225] A. Mizuno and Y. Okada.
Biological characterization of expression quantitative trait loci (eQTLs) showing tissue-specific opposite directional effects.
European Journal of Human Genetics, 27(11):1745–1756, 2019.

[226] E. Morita, V. Sandrin, S. L. Alam, D. M. Eckert, S. P. Gygi, and W. I. Sundquist.
Identification of human MVB12 proteins as ESCRT-I subunits that function in HIV budding.
Cell Host Microbe, 2(1):41–53, 2007.

[227] E. Mueller, P. Sarraf, P. Tontonoz, R. M. Evans, K. J. Martin, M. Zhang, C. Fletcher, S. Singer, and B. M. Spiegelman.
Terminal differentiation of human breast cancer through PPARγ.
Molecular Cell, 1(3):465–470, 1998.

[228] E. Mueller, M. Smith, P. Sarraf, T. Kroll, A. Aiyer, D. S. Kaufman, W. Oh, G. Demetri, W. D. Figg, X. P. Zhou, C. Eng, B. M. Spiegelman, and P. W. Kantoff.
Effects of ligand activation of peroxisome proliferator-activated receptor gamma in human prostate cancer.
Proc Natl Acad Sci U S A, 97(20):10990–5, 2000.

[229] D.-H. Nam, K. Park, Y. L. Suh, and J.-H. Kim.
Expression of VEGF and brain specific angiogenesis inhibitor-1 in glioblastoma: prognostic significance.
Oncol Rep, 11(4):863–869, 2004.

[230] A. Narayana, M. M. Tam, D. B. Gruber, J. Golfinos, E. Parker, D. Zagzag, and M. L. Gruber.
A phase II trial of lithium, bevacizumab, temozolomide, and radiation for newly diagnosed glioblastomas (GBM).
J Clin Oncol, 31(15-suppl):e13033–e13033, 2013.

[231] J. P. Neglia, L. L. Robison, M. Stovall, Y. Liu, R. J. Packer, S. Hammond, Y. Yasui, C. E. Kasper, A. C. Mertens, S. S. Donaldson, A. T. Meadows, and P. D. Inskip.

New primary neoplasms of the central nervous system in survivors of childhood cancer: a report from the childhood cancer survivor study.
J Natl Cancer Inst, 98(21):1528–37, 2006.

[232] M. R. Nelson, H. Tipney, J. L. Painter, J. Shen, P. Nicoletti, Y. Shen, A. Floratos, P. C. Sham, M. J. Li, J. Wang, L. R. Cardon, J. C. Whittaker, and P. Sanseau.
The support of human genetic evidence for approved drug indications.
Nat Genet, 47(8):856–60, 2015.

[233] B. Neyns, J. Sadones, E. Joosens, F. Bouttens, L. Verbeke, J. F. Baurain, L. D'Hondt, T. Strauven, C. Chaskis, P. In't Veld, A. Michotte, and J. De Greve.
Stratified phase II trial of cetuximab in patients with recurrent high-grade glioma.
Ann Oncol, 20(9):1596–1603, 2009.

[234] B. Ng, C. C. White, H. U. Klein, S. K. Sieberts, C. McCabe, E. Patrick, J. Xu, L. Yu, C. Gaiteri, D. A. Bennett, S. Mostafavi, and P. L. De Jager.
An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome.
Nat Neurosci, 20(10):1418–1426, 2017.

[235] D. L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M. E. Dolan, and N. J. Cox.
Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS.
PLoS Genet, 6(4):e1000888, 2010.

[236] A. Nicolato, M. A. Gerosa, P. Fina, P. Iuzzolino, F. Giorgiutti, and A. Bricolo.
Prognostic factors in low-grade supratentorial astrocytomas: A uni-multivariate statistical analysis in 76 surgically treated adult patients.
Surg Neurol, 44(3):208–223, 1995.

[237] C. Nieder, O. Spanne, M. P. Mehta, A. L. Grosu, and H. Geinitz.
Presentation, patterns of care, and survival in patients with brain metastases: what has changed in the last 20 years?
Cancer, 117(11):2505–12, 2011.

[238] H. Nikbakht, E. Panditharatna, L. G. Mikael, R. Li, T. Gayden, M. Osmond, C.-Y. Ho, M. Kambhampati, E. I. Hwang, D. Faury, A. Siu, S. Papillon-Cavanagh, D. Bechet, K. L. Ligon, B. Ellezam, W. J. Ingram, C. Stinson, A. S. Moore, K. E. Warren, J. Karamchandani, R. J. Packer, N. Jabado, J. Majewski, and J. Nazarian.
Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma.
Nat Commun, 7(1):11185, 2016.

[239] T. Nitta and K. Sato.
Prognostic implications of the extent of surgical resection in patients with intracranial malignant gliomas.
Cancer, 75(11):2727–2731, 1995.

[240] A. D. Norden, G. S. Young, K. Setayesh, A. Muzikansky, R. Klufas, G. L. Ross, A. S. Ciampa, L. G. Ebbeling, B. Levy, J. Drappatz, S. Kesari, and P. Y. Wen.
Bevacizumab for recurrent malignant gliomas: efficacy, toxicity, and patterns of recurrence.
Neurology, 70(10):779–87, 2008.

[241] M. O. Nowicki, N. Dmitrieva, A. M. Stein, J. L. Cutter, J. Godlewski, Y. Saeki, M. Nita, M. E. Berens, L. M. Sander, H. B. Newton, E. A. Chiocca, and S. Lawler.
Lithium inhibits invasion of glioma cells; possible involvement of glycogen synthase kinase-3.
Neuro Oncol, 10(5):690–9, 2008.

[242] C. L. Nutt, D. Mani, R. A. Betensky, P. Tamayo, J. G. Cairncross, C. Ladd, U. Pohl, C. Hartmann, M. E. McLaughlin, and T. T. Batchelor.
Gene expression-based classification of malignant gliomas correlates better with survival than histological classification.
Cancer Res, 63(7):1602–1607, 2003.

[243] H. Ohgaki.
Epidemiology of brain tumors.
Methods Mol Biol, 472:323–42, 2009.

[244] H. Ohgaki and P. Kleihues.
Epidemiology and etiology of gliomas.
Acta Neuropathol, 109(1):93–108, 2005.

[245] H. Ohgaki and P. Kleihues.
Genetic pathways to primary and secondary glioblastoma.
Am J Pathol, 170(5):1445–53, 2007.

[246] O. I. Olopade, T. A. Grushko, R. Nanda, and D. Huo.
Advances in breast cancer: Pathways to personalized medicine.
Clin Cancer Res, 14(24):7988–7999, 2008.

[247] Q. T. Ostrom, L. Bauchet, F. G. Davis, I. Deltour, J. L. Fisher, C. E. Langer, M. Pekmezci, J. A. Schwartzbaum, M. C. Turner, K. M. Walsh, M. R. Wrensch, and J. S. Barnholtz-Sloan.
The epidemiology of glioma in adults: a "state of the science" review.
Neuro Oncol, 16(7):896–913, 2014.

[248] Q. T. Ostrom, D. J. Cote, M. Ascha, C. Kruchko, and J. S. Barnholtz-Sloan.
Adult glioma incidence and survival by race or ethnicity in the United States from 2000 to 2014.
JAMA Oncol, 4(9):1254–1262, 2018.

[249] Q. T. Ostrom, K. M. Egan, L. B. Nabors, T. Gerke, R. C. Thompson, J. J. Olson, R. LaRocca, S. Chowdhary, J. E. Eckel-Passow, G. Armstrong, J. K. Wiencke, J. L. Bernstein, E. B. Claus, D. Il'yasova, C. Johansen, D. H. Lachance, R. K. Lai, R. T. Merrell, S. H. Olson, S. Sadetzki, J. M. Schildkraut, S. Shete, R. S. Houlston, R. B. Jenkins, M. R. Wrensch, B. Melin, C. I. Amos, J. T. Huse, J. S. Barnholtz-Sloan, and M. L. Bondy.
Glioma risk associated with extent of estimated European genetic ancestry in African Americans and Hispanics.
Int J Cancer, 146(3):739–748, 2020.

[250] Q. T. Ostrom, H. Gittleman, P. Liao, T. Vecchione-Koval, Y. Wolinsky, C. Kruchko, and J. S. Barnholtz-Sloan.
CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014.
Neuro Oncol, 19(suppl-5):v1–v88, 2017.

[251] C. C. O'Sullivan and S. E. Bates.
Targeting prolactin receptor (PRLR) signaling in PRLR-positive breast and prostate cancer.
Oncologist, 21(5):523–6, 2016.

[252] S. S. Ousman and P. Kubes.
Immune surveillance in the central nervous system.
Nat Neurosci, 15(8):1096–101, 2012.

[253] E. Pan, S. B. Mitchell, and J. S. Tsai.
A retrospective study of the safety of BCNU wafers with concurrent temozolomide and radiotherapy and adjuvant temozolomide for newly diagnosed glioblastoma patients.
J Neurooncol, 88(3):353–7, 2008.

[254] Y. B. Pan, C. H. Zhang, S. Q. Wang, P. H. Ai, K. Chen, L. Zhu, Z. L. Sun, and D. F. Feng.
Transforming growth factor beta induced (TGFBI) is a potential signature gene for mesenchymal subtype high-grade glioma.
J Neurooncol, 137(2):395–407, 2018.

[255] D. Panigrahy, A. Kaipainen, S. Huang, C. E. Butterfield, C. M. Barnes, M. Fannon, A. M. Laforme, D. M. Chaponis, J. Folkman, and M. W. Kieran.
PPARalpha agonist fenofibrate suppresses tumor growth through direct and indirect angiogenesis inhibition.

Proc Natl Acad Sci U S A, 105(3):985–90, 2008.

[256] T. C. Pansuriya, R. van Eijk, P. d'Adamo, M. A. van Ruler, M. L. Kuijjer, J. Oosting, A. M. Cleton-Jansen, J. G. van Oosterwijk, S. L. Verbeke, D. Meijer, T. van Wezel, K. H. Nord, L. Sangiorgi, B. Toker, B. Liegl-Atzwanger, M. San-Julian, R. Sciot, N. Limaye, L. G. Kindblom, S. Daugaard, C. Godfraind, L. M. Boon, M. Vikkula, K. C. Kurek, K. Szuhai, P. J. French, and J. V. Bovee.
Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome.
Nat Genet, 43(12):1256–61, 2011.

[257] A. Papi, L. Tatenhorst, D. Terwel, M. Hermes, M. P. Kummer, M. Orlandi, and M. T. Heneka.
PPARgamma and RXRgamma ligands act synergistically as potent antineoplastic agents in vitro and in vivo glioma models.
J Neurochem, 109(6):1779–90, 2009.

[258] M. E. Parent, M. C. Turner, J. Lavoue, H. Richard, J. Figuerola, L. Kincl, L. Richardson, G. Benke, M. Blettner, S. Fleming, M. Hours, D. Krewski, D. McLean, S. Sadetzki, K. Schlaefer, B. Schlehofer, J. Schuz, J. Siemiatycki, M. van Tongeren, and E. Cardis.
Lifetime occupational exposure to metals and welding fumes, and risk of glioma: a 7-country population-based case-control study.
Environ Health, 16(1):90, 2017.

[259] L. Paternoster, K. Tilling, and G. Davey Smith.
Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges.
PLoS Genet, 13(10):e1006944, 2017.

[260] M. Patyka, Z. Sharifi, K. Petrecca, J. Mansure, B. Jean-Claude, and S. Sabri.
Sensitivity to PRIMA-1MET is associated with decreased MGMT in human glioblastoma cells and glioblastoma stem cells irrespective of p53 status.
Oncotarget, 7(37):60245–60269, 2016.

[261] S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, and A. L. Schacht.
How to improve R&D productivity: the pharmaceutical industry's grand challenge.
Nat Rev Drug Discov, 9(3):203–14, 2010.

[262] N. Pearce.
Analysis of matched case-control studies.
BMJ, 352:i969, 2016.

[263] A. E. Pegg, M. E. Dolan, and R. C. Moschel.
Structure, Function, and Inhibition of O6-Alkylguanine-DNA Alkyltransferase, volume 51 of *Progress in Nucleic Acid Research and Molecular Biology*.
Academic Press, 1995.

[264] J. M. Perez-Ortiz, P. Tranque, M. Burgos, C. F. Vaquero, and J. Llopis.
Glitazones induce astroglioma cell death by releasing reactive oxygen species from mitochondria: modulation of cytotoxicity by nitric oxide.
Mol Pharmacol, 72(2):407–17, 2007.

[265] J. M. Perez-Ortiz, P. Tranque, C. F. Vaquero, B. Domingo, F. Molina, S. Calvo, J. Jordan, V. Cena, and J. Llopis.
Glitazones differentially regulate primary astrocyte and glioma cell survival. Involvement of reactive oxygen species and peroxisome proliferator-activated receptor-gamma.
J Biol Chem, 279(10):8976–85, 2004.

[266] J.-B. Pingault, C. A. M. Cecil, J. Murray, M. R. Munafo, and E. Viding.
Causal inference in psychopathology: A systematic review of Mendelian randomisation studies aiming to identify environmental risk factors for psychopathology.
Psychopathology Review, a4(1):4–25, 2017.

[267] J. J. Plascak and J. L. Fisher.
Area-based socioeconomic position and adult glioma: a hierarchical analysis of surveillance epidemiology and end results data.
PLoS One, 8(4):e60910, 2013.

[268] M. Platten, L. Bunse, W. Wick, and T. Bunse.
Concepts in glioma immunotherapy.
Cancer Immunol Immunother, 65(10):1269–75, 2016.

[269] W. B. Pope, A. Lai, P. Nghiemphu, P. Mischel, and T. F. Cloughesy.
MRI in patients with high-grade gliomas treated with bevacizumab and chemotherapy.
Neurology, 66(8):1258–60, 2006.

[270] R. Porta, J. Sánchez-Torres, L. Paz-Ares, B. Massutí, N. Reguart, C. Mayo, P. Lianes, C. Queralt, V. Guillem, P. Salinas, S. Catot, D. Isla, A. Pradas, A. Gúrpide, J. de Castro, E. Polo, T. Puig, M. Tarón, R. Colomer, and R. Rosell.
Brain metastases from lung cancer responding to erlotinib: the importance of EGFR mutation.
Eur Respir J, 37(3):624–631, 2011.

[271] L. Poulsen, M. Siersbaek, and S. Mandrup.

PPARs: fatty acid sensors controlling metabolism.
Semin Cell Dev Biol, 23(6):631–9, 2012.

[272] J. N. Poynter, S. B. Gruber, P. D. Higgins, R. Almog, J. D. Bonner, H. S. Rennert, M. Low, J. K. Greenson, and G. Rennert.
Statins and the risk of colorectal cancer.
N Engl J Med, 352(21):2184–92, 2005.

[273] J. K. Pritchard, M. Stephens, and P. Donnelly.
Inference of population structure using multilocus genotype data.
155(2):945–959, 2000.

[274] R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, and C. J. Willer.
LocusZoom: regional visualization of genome-wide association scan results.
Bioinformatics, 26(18):2336–7, 2010.

[275] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham.
PLINK: a tool set for whole-genome association and population-based linkage analyses.
Am J Hum Genet, 81(3):559–75, 2007.

[276] T. Qi, Y. Wu, J. Zeng, F. Zhang, A. Xue, L. Jiang, Z. Zhu, K. Kemper, L. Yengo, Z. Zheng, eQTLGen Consortium, R. E. Marioni, G. W. Montgomery, I. J. Deary, N. R. Wray, P. M. Visscher, A. F. McRae, and J. Yang.
Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood.
Nat Commun, 9(1):2282, 2018.

[277] A. Radhakrishnan, R. Raju, N. Tuladhar, T. Subbannayya, J. K. Thomas, R. Goel, D. Telikicherla, S. M. Palapetta, B. A. Rahiman, D. D. Venkatesh, K. K. Urmila, H. C. Harsha, P. P. Mathur, T. S. Prasad, A. Pandey, C. Shemanko, and A. Chatterjee.
A pathway map of prolactin signaling.
J Cell Commun Signal, 6(3):169–73, 2012.

[278] J. J. Raizer, L. E. Abrey, A. B. Lassman, S. M. Chang, K. R. Lamborn, J. G. Kuhn, W. K. A. Yung, M. R. Gilbert, K. A. Aldape, P. Y. Wen, H. A. Fine, M. Mehta, L. M. DeAngelis, F. Lieberman, T. F. Cloughesy, H. I. Robins, J. Dancey, and M. D. Prados.
A phase II trial of erlotinib in patients with recurrent malignant gliomas and nonprogressive glioblastoma multiforme postradiation therapy.
Neuro Oncol, 12(1):95–103, 2009.

[279] P. Rajaraman, B. S. Melin, Z. Wang, R. McKean-Cowdin, D. S. Michaud, S. S. Wang, M. Bondy, R. Houlston, R. B. Jenkins, M. Wrensch, M. Yeager, A. Ahlbom, D. Albanes, U. Andersson, L. E. Freeman, J. E. Buring, M. A. Butler, M. Braganza, T. Carreon, M. Feychting, S. J. Fleming, S. M. Gapstur, J. M. Gaziano, G. G. Giles, G. Hallmans, R. Henriksson, J. Hoffman-Bolton, P. D. Inskip, C. Johansen, C. M. Kitahara, M. Lathrop, C. Liu, L. Le Marchand, M. S. Linet, S. Lonn, U. Peters, M. P. Purdue, N. Rothman, A. M. Ruder, M. Sanson, H. D. Sesso, G. Severi, X. O. Shu, M. Simon, M. Stampfer, V. L. Stevens, K. Visvanathan, E. White, A. Wolk, A. Zeleniuch-Jacquotte, W. Zheng, P. Decker, V. Enciso-Mora, B. Fridley, Y. T. Gao, M. Kosel, D. H. Lachance, C. Lau, T. Rice, A. Swerdlow, J. L. Wiemels, J. K. Wiencke, S. Shete, Y. B. Xiang, Y. Xiao, R. N. Hoover, J. Fraumeni, J. F., N. Chatterjee, P. Hartge, and S. J. Chanock.
Genome-wide association study of glioma and meta-analysis.
Hum Genet, 131(12):1877–88, 2012.

[280] A. Ramasamy, D. Trabzuni, S. Guelfi, V. Varghese, C. Smith, R. Walker, T. De, UK Brain Expression Consortium, North American Brain Expression Consortium, L. Coin, R. de Silva, M. R. Cookson, A. B. Singleton, J. Hardy, M. Ryten, and M. E. Weale.
Genetic variability in the regulation of gene expression in ten regions of the human brain.
Nat Neurosci, 17(10):1418–1428, 2014.

[281] A. Ranger and A. Szymczak.
Do intracranial neoplasms differ in Ollier disease and maffucci syndrome? An in-depth analysis of the literature.
Neurosurgery, 65(6):1106–13; discussion 1113–5, 2009.

[282] D. A. Reardon, L. B. Nabors, W. P. Mason, J. R. Perry, W. Shapiro, P. Kavan, D. Mathieu, S. Phuphanich, A. Cseh, Y. Fu, J. Cong, S. Wind, D. D. Eisenstat, and BI 1200 36 Trial Group and the Canadian Brain Tumour Consortium.
Phase I/randomized phase II study of afatinib, an irreversible ErbB family blocker, with or without protracted temozolomide in adults with recurrent glioblastoma.
Neuro Oncol, 17(3):430–9, 2015.

[283] C. L. Relton and G. Davey Smith.
Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease.
Int J Epidemiol, 41(1):161–76, 2012.

[284] C. Remschmidt, O. Wichmann, and T. Harder.
Frequency and impact of confounding by indication and healthy vaccinee bias in observational studies assessing influenza vaccine effectiveness: a systematic review.
BMC Infect Dis, 15:429, 2015.

[285] D. R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pander, and A. M. Chinnaiyan.
ONCOMINE: A cancer microarray database and integrated data-mining platform.
Neoplasia, 6(1):1–6, 2004.

[286] T. Rice, D. H. Lachance, A. M. Molinaro, J. E. Eckel-Passow, K. M. Walsh, J. Barnholtz-Sloan, Q. T. Ostrom, S. S. Francis, J. Wiemels, R. B. Jenkins, J. K. Wiencke, and M. R. Wrensch.
Understanding inherited genetic risk of adult glioma - a review.
Neurooncol Pract, 3(1):10–16, 2016.

[287] P. M. Ridker, E. Danielson, F. A. Fonseca, J. Genest, A. M. Gotto, J. J. Kastelein, W. Koenig, P. Libby, A. J. Lorenzatti, J. G. MacFadyen, B. G. Nordestgaard, J. Shepherd, J. T. Willerson, and R. J. Glynn.
Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein.
N Engl J Med, 359(21):2195–2207, 2008.
PMID: 18997196.

[288] M. J. Riemenschneider, J. W. Jeuken, P. Wesseling, and G. Reifenberger.
Molecular diagnostics of gliomas: state of the art.
Acta Neuropathol, 120(5):567–84, 2010.

[289] C. Robins, A. P. Wingo, W. Fan, D. M. Duong, J. Meigs, E. S. Gerasimov, E. B. Dammer, D. J. Cutler, P. L. D. Jager, D. A. Bennett, J. J. Lah, A. I. Levey, N. T. Seyfried, and T. S. Wingo.
Genetic control of the human brain proteome.
bioRxiv, 2019.

[290] G. W. Robinson, B. A. Orr, and A. Gajjar.
Complete clinical regression of a BRAF V600E-mutant pediatric glioblastoma multiforme after BRAF inhibitor therapy.
BMC Cancer, 14(1):258, 2014.

[291] J. W. Robinson, R. Martin, M. Ozawa, M. M. C. Elwenspoek, M. T. Redaniel, K. Kurian, and Y. Ben-Shlomo.
Role of treatments for diabetes and hyperlipidaemia in risk and mortality of primary and secondary brain tumours.
medRxiv, 2020.

[292] J. W. Robinson, R. M. Martin, S. Tsavachidis, A. E. Howell, C. L. Relton, G. N. Armstrong, M. Bondy, J. Zheng, and K. M. Kurian.

Transcriptome-wide Mendelian randomization study prioritising novel tissue-dependent genes for glioma susceptibility.
Sci Rep, 11(1):2329, 2021.

[293] N. J. Robison, F. Campigotto, S. N. Chi, P. E. Manley, C. D. Turner, M. A. Zimmerman, C. A. Chordas, A. M. Werger, J. C. Allen, S. Goldman, J. B. Rubin, M. S. Isakoff, W. J. Pan, Z. A. Khatib, M. A. Comito, A. E. Bendel, J. B. Pietrantonio, L. Kondrat, S. M. Hubbs, D. S. Neuberg, and M. W. Kieran.
A phase II trial of a multi-agent oral antiangiogenic (metronomic) regimen in children with recurrent or progressive cancer.
Pediatr Blood Cancer, 61(4):636–42, 2014.

[294] K. Rock, O. McArdle, P. Forde, M. Dunne, D. Fitzpatrick, B. O'Neill, and C. Faul.
A clinical review of treatment outcomes in glioblastoma multiforme–the validation in a non-trial population of the results of a randomised phase III clinical trial: has a more radical approach improved survival?
Br J Radiol, 85(1017):e729–33, 2012.

[295] E. D. Rosen and B. M. Spiegelman.
PPARgamma : a nuclear regulator of metabolism, differentiation, and cell growth.
J Biol Chem, 276(41):37731–4, 2001.

[296] Royal Marsden.
Low-grade glioma.
https://www.royalmarsden.nhs.uk/your-care/cancer-types/paediatric-cancers/low-grade-glioma.
Accessed: 2019-12.

[297] A. M. Ruder, M. A. Waters, T. Carreon, M. A. Butler, G. M. Calvert, K. E. Davis-King, K. M. Waters, P. A. Schulte, J. S. Mandel, R. F. Morton, D. J. Reding, K. D. Rosenman, and Brain Cancer Collaborative Study Group.
The Upper Midwest Health Study: industry and occupation of glioma cases and controls.
Am J Ind Med, 55(9):747–55, 2012.

[298] A. M. Ruder, J. H. Yiin, M. A. Waters, T. Carreon, M. J. Hein, M. A. Butler, G. M. Calvert, K. E. Davis-King, P. A. Schulte, J. S. Mandel, R. F. Morton, D. J. Reding, K. D. Rosenman, P. A. Stewart, and Brain Cancer Collaborative Study Group.
The Upper Midwest Health Study: gliomas and occupational exposure to chlorinated solvents.
Occup Environ Med, 70(2):73–80, 2013.

[299] S. Sadetzki, A. Chetrit, L. Freedman, M. Stovall, B. Modan, and I. Novikov.

Long-term follow-up for brain tumor development after childhood exposure to ionizing radiation for tinea capitis.
Radiat Res, 163(4):424–32, 2005.

[300] A. Sahgal, H. Soliman, and D. A. Larson.
Whole-brain radiation therapy of brain metastasis.
Prog Neurol Surg, 25:82–95, 2012.

[301] F. Sahm, C. Koelsche, J. Meyer, S. Pusch, K. Lindenberg, W. Mueller, C. Herold-Mende, A. von Deimling, and C. Hartmann.
CIC and FUBP1 mutations in oligodendrogliomas, oligoastrocytomas and astrocytomas.
Acta Neuropathol, 123(6):853–60, 2012.

[302] S. Sallinen, P. K. Sallinen, H. K. Haapasalo, H. J. Helin, P. T. Helen, P. Schraml, O. Kallioniemi, and J. Kononen.
Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques.
Cancer Res, 60(23):6617–6622, 2000.

[303] N. Sanai and M. S. Berger.
Glioma extent of resection and its impact on patient outcome.
Neurosurgery, 62(4):753–64; discussion 264–6, 2008.

[304] M. Sanson, F. J. Hosking, S. Shete, D. Zelenika, S. E. Dobbins, Y. Ma, V. Enciso-Mora, A. Idbaih, J. Y. Delattre, K. Hoang-Xuan, Y. Marie, B. Boisselier, C. Carpentier, X. W. Wang, A. L. Di Stefano, M. Labussiere, K. Gousias, J. Schramm, A. Boland, D. Lechner, I. Gut, G. Armstrong, Y. Liu, R. Yu, C. Lau, M. C. Di Bernardo, L. B. Robertson, K. Muir, S. Hepworth, A. Swerdlow, M. J. Schoemaker, H. E. Wichmann, M. Muller, S. Schreiber, A. Franke, S. Moebus, L. Eisele, A. Forsti, K. Hemminki, M. Lathrop, M. Bondy, R. S. Houlston, and M. Simon.
Chromosome 7p11.2 (EGFR) variation influences glioma risk.
Hum Mol Genet, 20(14):2897–904, 2011.

[305] F. Saran, A. James, C. McBain, S. Jefferies, F. Harris, A. Cseh, K. Pemberton, J. Schaible, S. Bender, and M. Brada.
Actr-38. a phase I trial of afatinib and radiotherapy (rt) with or without temozolomide (tmz) in patients with newly diagnosed glioblastoma (gbm).
Neuro Oncol, 20(suppl-6):vi20–vi20, 2018.

[306] P. Sarraf, E. Mueller, D. Jones, F. J. King, D. J. DeAngelo, J. B. Partridge, S. A. Holden, L. B. Chen, S. Singer, C. Fletcher, and B. M. Spiegelman.
Differentiation and reversal of malignant changes in colon cancer through PPARgamma.

Nat Med, 4(9):1046–52, 1998.

[307] S. Sathornsumetee, A. Desjardins, J. J. Vredenburgh, R. E. McLendon, J. Marcello, J. E. Herndon, A. Mathe, M. Hamilton, J. N. Rich, J. A. Norfleet, S. Gururangan, H. S. Friedman, and D. A. Reardon.
Phase II trial of bevacizumab and erlotinib in patients with recurrent malignant glioma.
Neuro Oncol, 12(12):1300–10, 2010.

[308] Y. Sato, S. Akiba, O. Kubo, and N. Yamaguchi.
A case-case study of mobile phone use and acoustic neuroma risk in Japan.
Bioelectromagnetics, 32(2):85–93, 2011.

[309] C. N. Saunders, A. J. Cornish, B. Kinnersley, P. J. Law, E. B. Claus, D. Il'yasova, J. Schildkraut, J. S. Barnholtz-Sloan, S. H. Olson, J. L. Bernstein, R. K. Lai, S. Chanock, P. Rajaraman, C. Johansen, R. B. Jenkins, B. S. Melin, M. R. Wrensch, M. Sanson, M. L. Bondy, and R. S. Houlston.
Lack of association between modifiable exposures and glioma risk: a Mendelian randomization analysis.
Neuro Oncol, 22(2):207–215, 2020.

[310] M. Scerrati, R. Roselli, M. Iacoangeli, A. Pompucci, and G. F. Rossi.
Prognostic factors in low grade (WHO grade II) gliomas of the cerebral hemispheres: the role of surgery.
J Neurol Neurosurg Psychiatry, 61(3):291–6, 1996.

[311] B. Schlehofer, I. Hettinger, P. Ryan, M. Blettner, S. Preston-Martin, J. Little, A. Arslan, A. Ahlbom, G. G. Giles, G. R. Howe, F. Menegoz, Y. Rodvall, W. N. Choi, and J. Wahrendorf.
Occupational risk factors for low grade and high grade glioma: results from an international case control study of adult brain tumours.
Int J Cancer, 113(1):116–25, 2005.

[312] A. F. Schmidt, C. Finan, M. Gordillo-Maranon, F. W. Asselbergs, D. F. Freitag, R. S. Patel, B. Tyl, S. Chopade, R. Faraway, M. Zwierzyna, and A. D. Hingorani.
Genetic drug target validation using Mendelian randomisation.
Nature Communications, 11(1):3255, 2020.

[313] R. K. Schmidt-Ullrich, R. B. Mikkelsen, P. Dent, D. G. Todd, K. Valerie, B. D. Kavanagh, J. N. Contessa, W. K. Rorrer, and P. B. Chen.
Radiation-induced proliferation of the human A431 squamous carcinoma cells is dependent on EGFR tyrosine phosphorylation.
Oncogene, 15(10):1191–1197, 1997.

[314] S. Schneeweiss, A. R. Patrick, T. Sturmer, M. A. Brookhart, J. Avorn, M. Maclure, K. J. Rothman, and R. J. Glynn.
Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results.
Med Care, 45(10 Supl 2):S131–42, 2007.

[315] A. Schulte, K. Liffers, A. Kathagen, S. Riethdorf, S. Zapf, A. Merlo, K. Kolbe, M. Westphal, and K. Lamszus.
Erlotinib resistance in EGFR-amplified glioblastoma cells is associated with upregulation of EGFRvIII and PI3Kp110delta.
Neuro Oncol, 15(10):1289–301, 2013.

[316] K. Schwartz, H. T. Chang, M. Nikolai, J. Pernicone, S. Rhee, K. Olson, P. C. Kurniali, N. G. Hord, and M. Noel.
Treatment of glioma patients with ketogenic diets: report of two cases treated with an IRB-approved energy-restricted ketogenic diet protocol and review of the literature.
Cancer Metab, 3:3, 2015.

[317] J. Schwartzbaum, F. Jonsson, A. Ahlbom, S. Preston-Martin, S. Lonn, K. C. Soderberg, and M. Feychting.
Cohort studies of association between self-reported allergic conditions, immune-related diagnoses and glioma and meningioma risk.
Int J Cancer, 106(3):423–8, 2003.

[318] J. A. Schwartzbaum, J. L. Fisher, K. D. Aldape, and M. Wrensch.
Epidemiology and molecular pathology of glioma.
Nat Clin Pract Neurol, 2(9):494–503; quiz 1 p following 516, 2006.

[319] C. A. Scrideli, J. Carlotti, C. G., O. K. Okamoto, V. S. Andrade, M. A. Cortez, F. J. Motta, A. K. Lucio-Eterovic, L. Neder, S. Rosemberg, S. M. Oba-Shinjo, S. K. Marie, and L. G. Tone.
Gene expression profile analysis of primary glioblastomas and non-neoplastic brain tissue: identification of potential target genes by oligonucleotide microarray and real-time quantitative pcr.
J Neurooncol, 88(3):281–91, 2008.

[320] C. Seliger, C. Ricci, C. R. Meier, M. Bodmer, S. S. Jick, U. Bogdahn, P. Hau, and M. F. Leitzmann.
Diabetes, use of antidiabetic drugs, and the risk of glioma.
Neuro Oncol, 18(3):340–9, 2016.

[321] J. M. Sepulveda-Sanchez, M. A. Vaz, C. Balana, M. Gil-Gil, G. Reynes, O. Gallego, M. Martinez-Garcia, E. Vicente, M. Quindos, R. Luque, A. Ramos, Y. Ruano, P. Perez-Segura, M. Benavides, P. Sanchez-Gomez, and A. Hernandez-Lain.
Phase II trial of dacomitinib, a pan-human EGFR tyrosine kinase inhibitor, in recurrent glioblastoma patients with EGFR amplification.
Neuro Oncol, 19(11):1522–1531, 2017.

[322] T. N. Sergentanis, G. Tsivgoulis, C. Perlepe, I. Ntanasis-Stathopoulos, I. G. Tzanninis, I. N. Sergentanis, and T. Psaltopoulou.
Obesity and risk for brain/CNS tumors, gliomas and meningiomas: A meta-analysis.
PLoS One, 10(9):e0136974, 2015.

[323] T. N. Seyfried and P. Mukherjee.
Targeting energy metabolism in brain cancer: review and hypothesis.
Nutr Metab (Lond), 2:30, 2005.

[324] S. Sharma, T. K. Kelly, and P. A. Jones.
Epigenetics in cancer.
Carcinogenesis, 31(1):27–36, 09 2009.

[325] S. Shete, C. C. Lau, R. S. Houlston, E. B. Claus, J. Barnholtz-Sloan, R. Lai, D. Il'yasova, J. Schildkraut, S. Sadetzki, C. Johansen, J. L. Bernstein, S. H. Olson, R. B. Jenkins, P. Yang, N. A. Vick, M. Wrensch, F. G. Davis, B. J. McCarthy, E. H. Leung, C. Davis, R. Cheng, F. J. Hosking, G. N. Armstrong, Y. Liu, R. K. Yu, R. Henriksson, Gliogene Consortium, B. S. Melin, and M. L. Bondy.
Genome-wide high-density SNP linkage search for glioma susceptibility loci: results from the Gliogene Consortium.
Cancer Res, 71(24):7568–75, 2011.

[326] S. Shiraishi, K. Tada, H. Nakamura, K. Makino, M. Kochi, H. Saya, J. Kuratsu, and Y. Ushio.
Influence of p53 mutations on prognosis of patients with glioblastoma.
Cancer, 95(2):249–57, 2002.

[327] R. L. Siegel, K. D. Miller, and A. Jemal.
Cancer statistics, 2018.
CA Cancer J Clin, 68(1):7–30, 2018.

[328] A. J. Sievert and M. J. Fisher.
Pediatric low-grade gliomas.
J Child Neurol, 24(11):1397–408, 2009.

[329] J. R. Silber, M. S. Bobola, S. Ghatan, A. Blank, D. D. Kolstoe, and M. S. Berger.
O6-methylguanine-DNA methyltransferase activity in adult gliomas: relation to patient and tumor characteristics.
Cancer Res, 58(5):1068–1073, 1998.

[330] J. R. Silber, B. A. Mueller, T. G. Ewers, and M. S. Berger.
Comparison of O6-methylguanine-DNA methyltransferase activity in brain tumors and adjacent normal brain.
Cancer Res, 53(14):3416–3420, 1993.

[331] R. C. Smallridge, J. A. Copland, M. S. Brose, J. T. Wadsworth, Y. Houvras, M. E. Menefee, K. C. Bible, M. H. Shah, A. W. Gramza, J. P. Klopper, L. A. Marlow, M. G. Heckman, and R. Von Roemeling.
Efatutazone, an oral PPAR-gamma agonist, in combination with paclitaxel in anaplastic thyroid cancer: results of a multicenter phase 1 trial.
J Clin Endocrinol Metab, 98(6):2392–400, 2013.

[332] J. W. Song and K. C. Chung.
Observational studies: Cohort and case-control studies.
Plast Reconstr Surg, 126(6), 2010.

[333] S. N. Stacey, P. Sulem, A. Jonasdottir, G. Masson, J. Gudmundsson, D. F. Gudbjartsson, O. T. Magnusson, S. A. Gudjonsson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K. R. Benediktsdottir, B. A. Nexo, A. Tjonneland, K. Overvad, P. Rudnai, E. Gurzau, K. Koppova, K. Hemminki, C. Corredera, V. Fuentelsaz, P. Grasa, S. Navarrete, F. Fuertes, M. D. Garcia-Prats, E. Sanambrosio, A. Panadero, A. De Juan, A. Garcia, F. Rivera, D. Planelles, V. Soriano, C. Requena, K. K. Aben, M. M. van Rossum, R. G. Cremers, I. M. van Oort, D. J. van Spronsen, J. A. Schalken, W. H. Peters, B. T. Helfand, J. L. Donovan, F. C. Hamdy, D. Badescu, O. Codreanu, M. Jinga, I. E. Csiki, V. Constantinescu, P. Badea, I. N. Mates, D. E. Dinu, A. Constantin, D. Mates, S. Kristjansdottir, B. A. Agnarsson, E. Jonsson, R. B. Barkardottir, G. V. Einarsson, F. Sigurdsson, P. H. Moller, T. Stefansson, T. Valdimarsson, O. T. Johannsson, H. Sigurdsson, T. Jonsson, J. G. Jonasson, L. Tryggvadottir, T. Rice, H. M. Hansen, Y. Xiao, D. H. Lachance, O. N. BP, M. L. Kosel, P. A. Decker, G. Thorleifsson, H. Johannsdottir, H. T. Helgadottir, A. Sigurdsson, V. Steinthorsdottir, A. Lindblom, G. Swedish Low-risk Colorectal Cancer Study, R. S. Sandler, T. O. Keku, K. Banasik, T. Jorgensen, D. R. Witte, T. Hansen, O. Pedersen, V. Jinga, D. E. Neal, W. J. Catalona, M. Wrensch, J. Wiencke, R. B. Jenkins, E. Nagore, U. Vogel, L. A. Kiemeney, R. Kumar, J. I. Mayordomo, J. H. Olafsson, et al.
A germline variant in the TP53 polyadenylation signal confers cancer susceptibility.
Nat Genet, 43(11):1098–103, 2011.

[334] J. A. Sterne and G. Davey Smith.
Sifting the evidence-what's wrong with significance tests?
BMJ, 322(7280):226–31, 2001.

[335] R. Stupp, W. P. Mason, M. J. van den Bent, M. Weller, B. Fisher, M. J. Taphoorn, K. Belanger, A. A. Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R. C. Janzer, S. K. Ludwin, T. Gorlia, A. Allgeier, D. Lacombe, J. G. Cairncross, E. Eisenhauer, R. O. Mirimanoff, European Organisation for Research, Treatment of Cancer, Brain Tumor and Radiotherapy Groups, and National Cancer Institute of Canada Clinical Trials Group.
Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.
N Engl J Med, 352(10):987–96, 2005.

[336] K. Suhre, M. Arnold, A. M. Bhagwat, R. J. Cotton, R. Engelke, J. Raffler, H. Sarwath, G. Thareja, A. Wahl, R. K. DeLisle, L. Gold, M. Pezer, G. Lauc, M. A. El-Din Selim, D. O. Mook-Kanamori, E. K. Al-Dous, Y. A. Mohamoud, J. Malek, K. Strauch, H. Grallert, A. Peters, G. Kastenmuller, C. Gieger, and J. Graumann.
Connecting genetic risk to disease end points through the human blood plasma proteome.
Nat Commun, 8:14357, 2017.

[337] S. Suissa, S. Dell'aniello, S. Vahey, and C. Renoux.
Time-window bias in case-control studies: statins and lung cancer.
Epidemiology, 22(2):228–31, 2011.

[338] B. B. Sun, J. C. Maranville, J. E. Peters, D. Stacey, J. R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M. A. Kamat, B. P. Prins, S. K. Wilcox, E. S. Zimmerman, A. Chi, N. Bansal, S. L. Spain, A. M. Wood, N. W. Morrell, J. R. Bradley, N. Janjic, D. J. Roberts, W. H. Ouwehand, J. A. Todd, N. Soranzo, K. Suhre, D. S. Paul, C. S. Fox, R. M. Plenge, J. Danesh, H. Runz, and A. S. Butterworth.
Genomic atlas of the human plasma proteome.
Nature, 558(7708):73–79, 2018.

[339] O. Surma-aho, M. Niemela, J. Vilkki, M. Kouri, A. Brander, O. Salonen, A. Paetau, M. Kallio, J. Pyykkonen, and J. Jaaskelainen.
Adverse long-term effects of brain radiotherapy in adult low-grade glioma patients.
Neurology, 56(10):1285–90, 2001.

[340] K. Swiatek-Machado and B. Kaminska.
STAT signaling in glioma cells.
Adv Exp Med Biol, 986:189–208, 2013.

[341] H. Takahashi, A. J. Cornish, A. Sud, P. J. Law, B. Kinnersley, Q. T. Ostrom, K. Labreche, J. E. Eckel-Passow, G. N. Armstrong, E. B. Claus, D. Il'yasova, J. Schildkraut, J. S.

Barnholtz-Sloan, S. H. Olson, J. L. Bernstein, R. K. Lai, M. J. Schoemaker, M. Simon, P. Hoffmann, M. M. Nothen, K. H. Jockel, S. Chanock, P. Rajaraman, C. Johansen, R. B. Jenkins, B. S. Melin, M. R. Wrensch, M. Sanson, M. L. Bondy, C. Turnbull, and R. S. Houlston.

Mendelian randomisation study of the relationship between vitamin D and risk of glioma.
Sci Rep, 8(1):2339, 2018.

[342] T. Takebe, R. Imai, and S. Ono.

The current status of drug discovery and development as originated in United States academia: The influence of industrial and academic collaboration on drug discovery and development.
Clin Transl Sci, 11(6):597–606, 2018.

[343] J. P. Thakkar, T. A. Dolecek, C. Horbinski, Q. T. Ostrom, D. D. Lightner, J. S. Barnholtz-Sloan, and J. L. Villano.

Epidemiologic and molecular prognostic review of glioblastoma.
Cancer Epidemiol Biomark Prev, 23(10):1985–96, 2014.

[344] The National Institute for Health and Care Excellence.

Type 2 diabetes in adults: management.
https://www.nice.org.uk/guidance/ng28/chapter/Key-priorities-for-implementation.
Accessed: 2020-08.

[345] B. Thiessen, C. Stewart, M. Tsao, S. Kamel-Reid, P. Schaiquevich, W. Mason, J. Easaw, K. Belanger, P. Forsyth, L. McIntosh, and E. Eisenhauer.

A phase I/II trial of GW572016 (lapatinib) in recurrent glioblastoma multiforme: clinical outcomes, pharmacokinetics and molecular correlation.
Cancer Chemother Pharmacol, 65(2):353–61, 2010.

[346] M. Uhlen, L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten.

Proteomics. tissue-based map of the human proteome.
Science, 347(6220):1260419, 2015.

[347] UK10K Consortium, K. Walter, J. L. Min, J. Huang, L. Crooks, Y. Memari, S. McCarthy, J. R. Perry, C. Xu, M. Futema, D. Lawson, V. Iotchkova, S. Schiffels, A. E. Hendricks, P. Danecek, R. Li, J. Floyd, L. V. Wain, I. Barroso, S. E. Humphries, M. E. Hurles,

E. Zeggini, J. C. Barrett, V. Plagnol, J. B. Richards, C. M. Greenwood, N. J. Timpson, R. Durbin, and N. Soranzo.
The UK10K project identifies rare variants in health and disease.
Nature, 526(7571):82–90, 2015.

[348] M. J. van den Bent, D. Afra, O. de Witte, M. B. Hassel, S. Schraub, K. Hoang-Xuan, P. O. Malmstrom, L. Collette, M. Pierart, R. Mirimanoff, and A. Karim.
Long-term efficacy of early versus delayed radiotherapy for low-grade astrocytoma and oligodendroglioma in adults: the EORTC 22845 randomised trial.
Lancet, 366(9490):985–990, 2005.

[349] M. J. van den Bent, A. A. Brandes, R. Rampling, M. C. Kouwenhoven, J. M. Kros, A. F. Carpentier, P. M. Clement, M. Frenay, M. Campone, J. F. Baurain, J. P. Armand, M. J. Taphoorn, A. Tosoni, H. Kletzl, B. Klughammer, D. Lacombe, and T. Gorlia.
Randomized phase II trial of erlotinib versus temozolomide or carmustine in recurrent glioblastoma: EORTC brain tumor group study 26034.
J Clin Oncol, 27(8):1268–74, 2009.

[350] M. J. van den Bent, A. F. Carpentier, A. A. Brandes, M. Sanson, M. J. Taphoorn, H. J. Bernsen, M. Frenay, C. C. Tijssen, W. Grisold, L. Sipos, H. Haaxma-Reiche, J. M. Kros, M. C. van Kouwenhoven, C. J. Vecht, A. Allgeier, D. Lacombe, and T. Gorlia.
Adjuvant procarbazine, lomustine, and vincristine improves progression-free survival but not overall survival in newly diagnosed anaplastic oligodendrogliomas and oligoastrocytomas: a randomized European Organisation for Research and Treatment of Cancer phase III trial.
J Clin Oncol, 24(18):2715–22, 2006.

[351] O. van Tellingen, B. Yetkin-Arik, M. C. de Gooijer, P. Wesseling, T. Wurdinger, and H. E. de Vries.
Overcoming the blood-brain tumor barrier for effective glioblastoma treatment.
Drug Resist Updat, 19:1–12, 2015.

[352] T. J. VanderWeele, E. J. Tchetgen Tchetgen, M. Cornelis, and P. Kraft.
Methodological challenges in Mendelian randomization.
Epidemiology, 25(3):427–35, 2014.

[353] T. Varga, Z. Czimmerer, and L. Nagy.
PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation.
Biochim Biophys Acta, 1812(8):1007–22, 2011.

[354] R. Vengoji, M. A. Macha, R. K. Nimmakayala, S. Rachagani, J. A. Siddiqui, K. Mallya, S. Gorantla, M. Jain, M. P. Ponnusamy, S. K. Batra, and N. Shonka.
Afatinib and temozolomide combination inhibits tumorigenesis by targeting EGFRvIII-cMet signaling in glioblastoma cells.
J Exp Clin Cancer Res, 38(1):266, 2019.

[355] M. Verbanck, C. Y. Chen, B. Neale, and R. Do.
Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases.
Nat Genet, 50(5):693–698, 2018.

[356] S. Verstovsek, H. Kantarjian, R. A. Mesa, A. D. Pardanani, J. Cortes-Franco, D. A. Thomas, Z. Estrov, J. S. Fridman, E. C. Bradley, S. Erickson-Viitanen, K. Vaddi, R. Levy, and A. Tefferi.
Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis.
N Engl J Med, 363(12):1117–27, 2010.

[357] J. Vinagre, A. Almeida, H. Populo, R. Batista, J. Lyra, V. Pinto, R. Coelho, R. Celestino, H. Prazeres, L. Lima, M. Melo, A. G. d. Rocha, A. Preto, P. Castro, L. Castro, F. Pardal, J. M. Lopes, L. L. Santos, R. M. Reis, J. Cameselle-Teijeiro, M. Sobrinho-Simoes, J. Lima, V. Maximo, and P. Soares.
Frequency of TERT promoter mutations in human cancers.
Nature Commun, 4(1):2185, 2013.

[358] P. M. Visscher.
Sizing up human height variation.
Nat Genet, 40(5):489–90, 2008.

[359] P. M. Visscher, M. A. Brown, M. I. McCarthy, and J. Yang.
Five years of GWAS discovery.
Am J Hum Genet, 90(1):7–24, 2012.

[360] I. Vivanco, H. I. Robins, D. Rohle, C. Campos, C. Grommes, P. L. Nghiemphu, S. Kubek, B. Oldrini, M. G. Chheda, N. Yannuzzi, H. Tao, S. Zhu, A. Iwanami, D. Kuga, J. Dang, A. Pedraza, C. W. Brennan, A. Heguy, L. M. Liau, F. Lieberman, W. K. Yung, M. R. Gilbert, D. A. Reardon, J. Drappatz, P. Y. Wen, K. R. Lamborn, S. M. Chang, M. D. Prados, H. A. Fine, S. Horvath, N. Wu, A. B. Lassman, L. M. DeAngelis, W. H. Yong, J. G. Kuhn, P. S. Mischel, M. P. Mehta, T. F. Cloughesy, and I. K. Mellinghoff.
Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors.
Cancer Discov, 2(5):458–71, 2012.

[361] U. Vosa, A. Claringbould, H.-J. Westra, M. J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Kasela, N. Pervjakova, I. Alvaes, M.-J. Fave, M. Agbessi, M. Christiansen, R. Jansen, I. Seppala, L. Tong, A. Teumer, K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. Sonmez, A. A. Andrew, V. Kukushkina, A. Kalnapenkis, S. Rueger, E. Porcu, J. Kronberg-Guzman, J. Kettunen, J. Powell, B. Lee, F. Zhang, W. Arindrarto, F. Beutner, H. Brugge, J. Dmitrieva, M. Elansary, B. P. Fairfax, M. Georges, B. T. Heijmans, M. Kahonen, Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M. Marigorta, H. Mei, Y. Momozawa, M. Muller-Nurasyid, M. Nauck, M. Nivard, B. Penninx, J. Pritchard, O. Raitakari, O. Rotzschke, E. P. Slagboom, C. D. A. Stehouwer, M. Stumvoll, P. Sullivan, P. A. C. t Hoen, J. Thiery, A. Tonjes, J. van Dongen, M. van Iterson, J. Veldink, U. Volker, C. Wijmenga, M. Swertz, A. Andiappan, G. W. Montgomery, S. Ripatti, M. Perola, Z. Kutalik, E. Dermitzakis, S. Bergmann, T. Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B. Pierce, T. Lehtimaki, D. Boomsma, B. M. Psaty, S. A. Gharib, P. Awadalla, L. Milani, W. H. Ouwehand, K. Downes, O. Stegle, A. Battle, J. Yang, P. M. Visscher, M. Scholz, G. Gibson, T. Esko, and L. Franke.
Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis.
bioRxiv, 2018.

[362] J. J. Vredenburgh, A. Desjardins, n. Herndon, J. E., J. M. Dowell, D. A. Reardon, J. A. Quinn, J. N. Rich, S. Sathornsumetee, S. Gururangan, M. Wagner, D. D. Bigner, A. H. Friedman, and H. S. Friedman.
Phase II trial of bevacizumab and irinotecan in recurrent malignant glioma.
Clin Cancer Res, 13(4):1253–9, 2007.

[363] J. J. Vredenburgh, A. Desjardins, n. Herndon, J. E., J. Marcello, D. A. Reardon, J. A. Quinn, J. N. Rich, S. Sathornsumetee, S. Gururangan, J. Sampson, M. Wagner, L. Bailey, D. D. Bigner, A. H. Friedman, and H. S. Friedman.
Bevacizumab plus irinotecan in recurrent glioblastoma multiforme.
J Clin Oncol, 25(30):4722–9, 2007.

[364] K. M. Walsh, V. Codd, I. V. Smirnov, T. Rice, P. A. Decker, H. M. Hansen, T. Kollmeyer, M. L. Kosel, A. M. Molinaro, L. S. McCoy, P. M. Bracci, B. S. Cabriga, M. Pekmezci, S. Zheng, J. L. Wiemels, A. R. Pico, T. Tihan, M. S. Berger, S. M. Chang, M. D. Prados, D. H. Lachance, B. P. O'Neill, H. Sicotte, J. E. Eckel-Passow, E. C. T. Group, P. van der Harst, J. K. Wiencke, N. J. Samani, R. B. Jenkins, and M. R. Wrensch.
Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk.
Nat Genet, 46(7):731–5, 2014.

[365] D. Wang, S. Liu, J. Warrell, H. Won, X. Shi, F. C. P. Navarro, D. Clarke, M. Gu, P. Emani, Y. T. Yang, M. Xu, M. J. Gandal, S. Lou, J. Zhang, J. J. Park, C. Yan, S. K. Rhie, K. Man-

akongtreecheep, H. Zhou, A. Nathan, M. Peters, E. Mattei, D. Fitzgerald, T. Brunetti, J. Moore, Y. Jiang, K. Girdhar, G. E. Hoffman, S. Kalayci, Z. H. Gumus, G. E. Crawford, E. C. Psych, P. Roussos, S. Akbarian, A. E. Jaffe, K. P. White, Z. Weng, N. Sestan, D. H. Geschwind, J. A. Knowles, and M. B. Gerstein.
Comprehensive functional genomic resource and integrative model for the human brain.
Science, 362(6420), 2018.

[366] R. L. Wasserstein and N. A. Lazar.
The ASA statement on p-values: Context, process, and purpose.
Am Stat, 70(2):129–133, 2016.

[367] K. Watanabe, O. Tachibana, K. Sato, Y. Yonekawa, P. Kleihues, and H. Ohgaki.
Overexpression of the EGF receptor and p53 mutations are mutually exclusive in the evolution of primary and secondary glioblastomas.
Brain Pathol, 6(3):217–223, 1996.

[368] T. Watanabe, S. Nobusawa, P. Kleihues, and H. Ohgaki.
IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas.
Am J Pathol, 174(4):1149–53, 2009.

[369] C. Watts, K. Ashkan, M. Jenkinson, K. Kurian, W. Qian, S. Price, T. Matys, G. Doughton, A. Machin, J. Jung, I. Jalloh, C. Harman, K. Gatley, G. Young, R. Hardy, and A. Dayimu.
Actr-59. improving the intra-operative diagnosis of high-grade glioma using a fluorescence biomarker – result of the Gala-Bidd study.
Neuro Oncol, 20(suppl-6):vi24–vi25, 2018.

[370] U. H. WEIDLE, J. NIEWÖHNER, and G. TIEFENTHALER.
The blood–brain barrier challenge for the treatment of brain cancer, secondary brain metastases, and neurological diseases.
Cancer Genom Proteom, 12(4):167–177, 2015.

[371] M. Weller, S. M. Pfister, W. Wick, M. E. Hegi, G. Reifenberger, and R. Stupp.
Molecular neuro-oncology in clinical practice: a new horizon.
Lancet Oncol, 14(9):e370–e379, 2013.

[372] M. Weller, R. Stupp, G. Reifenberger, A. A. Brandes, M. J. van den Bent, W. Wick, and M. E. Hegi.
MGMT promoter methylation in malignant gliomas: ready for personalized medicine?
Nat Rev Neurol, 6(1):39–51, 2010.

[373] M. Westphal, D. C. Hilt, E. Bortey, P. Delavault, R. Olivares, P. C. Warnke, I. R. Whittle, J. Jaaskelainen, and Z. Ram.

A phase 3 trial of local chemotherapy with biodegradable carmustine (BCNU) wafers (Gliadel wafers) in patients with primary malignant glioma.
Neuro Oncol, 5(2):79–88, 2003.

[374] M. Westphal, Z. Ram, V. Riddle, D. Hilt, E. Bortey, and Executive Committee of the Gliadel Study Group.
Gliadel wafer in initial surgery for malignant glioma: long-term follow-up of a multicenter controlled trial.
Acta Neurochir (Wien), 148(3):269–75; discussion 275, 2006.

[375] W. Wick, C. Hartmann, C. Engel, M. Stoffels, J. Felsberg, F. Stockhammer, M. C. Sabel, S. Koeppen, R. Ketter, R. Meyermann, M. Rapp, C. Meisner, R. D. Kortmann, T. Pietsch, O. D. Wiestler, U. Ernemann, M. Bamberg, G. Reifenberger, A. von Deimling, and M. Weller.
NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with procarbazine, lomustine, and vincristine or temozolomide.
J Clin Oncol, 27(35):5874–80, 2009.

[376] W. Wick, M. Weller, M. van den Bent, M. Sanson, M. Weiler, A. von Deimling, C. Plass, M. Hegi, M. Platten, and G. Reifenberger.
MGMT testing–the challenges for biomarker-based glioma treatment.
Nat Rev Neurol, 10(7):372–85, 2014.

[377] W. Wick, M. Weller, M. van den Bent, and R. Stupp.
Bevacizumab and recurrent malignant gliomas: a European perspective.
J Clin Oncol, 28(12):e188–9; author reply e190–2, 2010.

[378] H. Wickham.
ggplot2: Elegant Graphics for Data Analysis.
Springer-Verlag New York, 2016.

[379] M. K. H. Wiedmann, C. Brunborg, A. Di Ieva, K. Lindemann, T. B. Johannesen, L. Vatten, E. Helseth, and J. A. Zwart.
Overweight, obesity and height as risk factors for meningioma, glioma, pituitary adenoma and nerve sheath tumor: a large population-based prospective cohort study.
Acta Oncol, 56(10):1302–1309, 2017.

[380] J. L. Wiemels, J. K. Wiencke, J. D. Sison, R. Miike, A. McMillan, and M. Wrensch.
History of allergies among adults with glioma and controls.
Int J Cancer, 98(4):609–15, 2002.

[381] A. Wigertz, S. Lonn, P. Hall, and M. Feychting.

Non-participant characteristics and the association between socioeconomic factors and brain tumour risk.
J Epidemiol Community Health, 64(8):736–43, 2010.

[382] A. Wilk, K. Urbanska, M. Grabacka, J. Mullinax, C. Marcinkiewicz, D. Impastato, J. J. Estrada, and K. Reiss.
Fenofibrate-induced nuclear translocation of FoxO3A triggers Bim-mediated apoptosis in glioblastoma cells in vitro.
Cell Cycle, 11(14):2660–71, 2012.

[383] A. Wilk, D. Wyczechowska, A. Zapata, M. Dean, J. Mullinax, L. Marrero, C. Parsons, F. Peruzzi, F. Culicchia, A. Ochoa, M. Grabacka, and K. Reiss.
Molecular mechanisms of fenofibrate-induced metabolic catastrophe and glioblastoma cell death.
Mol Cell Biol, 35(1):182–98, 2015.

[384] C. J. Willer, Y. Li, and G. R. Abecasis.
METAL: fast and efficient meta-analysis of genomewide association scans.
Bioinformatics, 26(17):2190–1, 2010.

[385] C. J. Willer, E. M. Schmidt, S. Sengupta, G. M. Peloso, S. Gustafsson, S. Kanoni, A. Ganna, J. Chen, M. L. Buchkovich, S. Mora, J. S. Beckmann, J. L. Bragg-Gresham, H. Y. Chang, A. Demirkan, H. M. Den Hertog, R. Do, L. A. Donnelly, G. B. Ehret, T. Esko, M. F. Feitosa, T. Ferreira, K. Fischer, P. Fontanillas, R. M. Fraser, D. F. Freitag, D. Gurdasani, K. Heikkila, E. Hypponen, A. Isaacs, A. U. Jackson, A. Johansson, T. Johnson, M. Kaakinen, J. Kettunen, M. E. Kleber, X. Li, J. Luan, L. P. Lyytikainen, P. K. E. Magnusson, M. Mangino, E. Mihailov, M. E. Montasser, M. Muller-Nurasyid, I. M. Nolte, J. R. O'Connell, C. D. Palmer, M. Perola, A. K. Petersen, S. Sanna, R. Saxena, S. K. Service, S. Shah, D. Shungin, C. Sidore, C. Song, R. J. Strawbridge, I. Surakka, T. Tanaka, T. M. Teslovich, G. Thorleifsson, E. G. Van den Herik, B. F. Voight, K. A. Volcik, L. L. Waite, A. Wong, Y. Wu, W. Zhang, D. Absher, G. Asiki, I. Barroso, L. F. Been, J. L. Bolton, L. L. Bonnycastle, P. Brambilla, M. S. Burnett, G. Cesana, M. Dimitriou, A. S. F. Doney, A. Doring, P. Elliott, S. E. Epstein, G. Ingi Eyjolfsson, B. Gigante, M. O. Goodarzi, H. Grallert, M. L. Gravito, C. J. Groves, G. Hallmans, A. L. Hartikainen, C. Hayward, D. Hernandez, A. A. Hicks, H. Holm, Y. J. Hung, T. Illig, M. R. Jones, P. Kaleebu, J. J. P. Kastelein, K. T. Khaw, E. Kim, et al.
Discovery and refinement of loci associated with lipid levels.
Nat Genet, 45(11):1274–1283, 2013.

[386] M. J. Winger, D. R. Macdonald, and J. G. Cairncross.

Supratentorial anaplastic gliomas in adults. the prognostic importance of extent of resection and prior low-grade glioma.
J Neurosurg, 71(4):487–93, 1989.

[387] A. R. Wood, T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, J. Luan, Z. Kutalik, N. Amin, M. L. Buchkovich, D. C. Croteau-Chonka, F. R. Day, Y. Duan, T. Fall, R. Fehrmann, T. Ferreira, A. U. Jackson, J. Karjalainen, K. S. Lo, A. E. Locke, R. Magi, E. Mihailov, E. Porcu, J. C. Randall, A. Scherag, A. A. Vinkhuyzen, H. J. Westra, T. W. Winkler, T. Workalemahu, J. H. Zhao, D. Absher, E. Albrecht, D. Anderson, J. Baron, M. Beekman, A. Demirkan, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, R. M. Fraser, A. Goel, J. Gong, A. E. Justice, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, J. C. Lui, M. Mangino, I. Mateo Leach, C. Medina-Gomez, M. A. Nalls, D. R. Nyholt, C. D. Palmer, D. Pasko, S. Pechlivanis, I. Prokopenko, J. S. Ried, S. Ripke, D. Shungin, A. Stancakova, R. J. Strawbridge, Y. J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, U. Afzal, J. Arnlov, G. M. Arscott, S. Bandinelli, A. Barrett, C. Bellis, A. J. Bennett, C. Berne, M. Bluher, J. L. Bolton, Y. Bottcher, H. A. Boyd, M. Bruinenberg, B. M. Buckley, S. Buyske, I. H. Caspersen, P. S. Chines, R. Clarke, S. Claudi-Boehm, M. Cooper, E. W. Daw, P. A. De Jong, J. Deelen, G. Delgado, et al.
Defining the role of common variation in the genomic and biological architecture of adult human height.
Nat Genet, 46(11):1173–86, 2014.

[388] O. J. Wouters, M. McKee, and J. Luyten.
Estimated research and development investment needed to bring a new medicine to market, 2009-2018.
JAMA, 323(9):844–853, 2020.

[389] M. Wrensch, R. B. Jenkins, J. S. Chang, R. F. Yeh, Y. Xiao, P. A. Decker, K. V. Ballman, M. Berger, J. C. Buckner, S. Chang, C. Giannini, C. Halder, T. M. Kollmeyer, M. L. Kosel, D. H. LaChance, L. McCoy, B. P. O'Neill, J. Patoka, A. R. Pico, M. Prados, C. Quesenberry, T. Rice, A. L. Rynearson, I. Smirnov, T. Tihan, J. Wiemels, P. Yang, and J. K. Wiencke.
Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility.
Nat Genet, 41(8):905–8, 2009.

[390] W. Y.-Y. Wu, G. Johansson, C. Wibom, T. Brannstrom, A. Malmstrom, R. Henriksson, I. Golovleva, M. L. Bondy, U. Andersson, A. M. Dahlin, and B. Melin.
The genetic architecture of gliomagenesis‚Äìgenetic risk variants linked to specific molecular subtypes.
Cancers, 11(12), 2019.

[391] J. Yang, T. Ferreira, A. P. Morris, S. E. Medland, Genetic Investigation of Anthropometric Traits Consortium, Replication Diabetes Genetics and Meta-analysis Consortium, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, T. M. Frayling, M. I. McCarthy, J. N. Hirschhorn, M. E. Goddard, and P. M. Visscher.
Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits.
Nat Genet, 44(4):369–75, S1–3, 2012.

[392] J. J. Yang, C. Cheng, W. Yang, D. Pei, X. Cao, Y. Fan, S. B. Pounds, G. Neale, L. R. Trevino, D. French, D. Campana, J. R. Downing, W. E. Evans, C.-H. Pui, M. Devidas, W. P. Bowman, B. M. Camitta, C. L. Willman, S. M. Davies, M. J. Borowitz, W. L. Carroll, S. P. Hunger, and M. V. Relling.
Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia.
JAMA, 301(4):393–403, 2009.

[393] C. Yao, G. Chen, C. Song, J. Keefe, M. Mendelson, T. Huan, B. B. Sun, A. Laser, J. C. Maranville, H. Wu, J. E. Ho, P. Courchesne, A. Lyass, M. G. Larson, C. Gieger, J. Graumann, A. D. Johnson, J. Danesh, H. Runz, S. J. Hwang, C. Liu, A. S. Butterworth, K. Suhre, and D. Levy.
Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease.
Nat Commun, 9(1):3268, 2018.

[394] J. Yarmolinsky, C. J. Bull, E. E. Vincent, J. Robinson, A. Walther, G. D. Smith, S. J. Lewis, C. L. Relton, and R. M. Martin.
Association between genetically proxied inhibition of HMG-CoA reductase and epithelial ovarian cancer.
JAMA, 323(7):646–655, 2020.

[395] S. A. Yeh, J. T. Ho, C. C. Lui, Y. J. Huang, C. Y. Hsiung, and E. Y. Huang.
Treatment outcomes and prognostic factors in patients with supratentorial low-grade gliomas.
Br J Radiol, 78(927):230–5, 2005.

[396] W. K. Yung, R. E. Albright, J. Olson, R. Fredericks, K. Fink, M. D. Prados, M. Brada, A. Spence, R. J. Hohl, W. Shapiro, M. Glantz, H. Greenberg, R. G. Selker, N. A. Vick, R. Rampling, H. Friedman, P. Phillips, J. Bruner, N. Yue, D. Osoba, S. Zaknoen, and V. A. Levin.
A phase II study of temozolomide vs. procarbazine in patients with glioblastoma multiforme at first relapse.

Br J Cancer, 83(5):588–93, 2000.

[397] W. K. Yung, J. J. Vredenburgh, T. F. Cloughesy, P. Nghiemphu, B. Klencke, M. R. Gilbert, D. A. Reardon, and M. D. Prados.
Safety and efficacy of erlotinib in first-relapse glioblastoma: a phase II open-label study.
Neuro Oncol, 12(10):1061–70, 2010.

[398] C. Zappa and S. A. Mousa.
Non-small cell lung cancer: current treatment and future advances.
Transl Lung Cancer Res, 5(3), 2016.

[399] H. Zhao, W. Cai, S. Su, D. Zhi, J. Lu, and S. Liu.
Allergic conditions reduce the risk of glioma: a meta-analysis based on 128,936 subjects.
Tumour Biol, 35(4):3875–80, 2014.

[400] J. Zheng, V. Haberland, D. Baird, V. Walker, P. C. Haycock, M. R. Hurle, A. Gutteridge, P. Erola, Y. Liu, S. Luo, J. Robinson, T. G. Richardson, J. R. Staley, B. Elsworth, S. Burgess, B. B. Sun, J. Danesh, H. Runz, J. C. Maranville, H. M. Martin, J. Yarmolinsky, C. Laurin, M. V. Holmes, J. Z. Liu, K. Estrada, R. Santos, L. McCarthy, D. Waterworth, M. R. Nelson, G. D. Smith, A. S. Butterworth, G. Hemani, R. A. Scott, and T. R. Gaunt.
Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases.
Nat Genet, 2020.

[401] Z. Zheng, D. Huang, J. Wang, K. Zhao, Y. Zhou, Z. Guo, S. Zhai, H. Xu, H. Cui, H. Yao, Z. Wang, X. Yi, S. Zhang, P. C. Sham, and M. J. Li.
QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes.
Nucleic Acids Res, 48(D1):D983–D991, 2020.

[402] D. Zhou, Y. Jiang, X. Zhong, N. J. Cox, C. Liu, and E. R. Gamazon.
A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis.
Nat Genet, 52(11):1239–1246, 2020.

[403] W. Zhou, P. Mukherjee, M. A. Kiebish, W. T. Markis, J. G. Mantis, and T. N. Seyfried.
The calorically restricted ketogenic diet, an effective alternative therapy for malignant brain cancer.
Nutr Metab (Lond), 4:5, 2007.

[404] Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R. Wray, P. M. Visscher, and J. Yang.

Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets.
Nat Genet, 48(5):481–487, 2016.

[405] D. S. Ziegler, M. Wong, C. Mayoh, A. Kumar, M. Tsoli, E. Mould, V. Tyrrell, D.-A. Khuong-Quang, M. Pinese, V. Gayevskiy, R. J. Cohn, L. M. S. Lau, M. Reynolds, M. C. Cox, A. Gifford, M. Rodriguez, M. J. Cowley, P. G. Ekert, G. M. Marshall, and M. Haber.
Brief report: Potent clinical and radiological response to larotrectinib in TRK fusion-driven high-grade glioma.
Br J Cancer, 119(6):693–696, 2018.

[406] G. Zuccoli, N. Marcello, A. Pisanello, F. Servadei, S. Vaccaro, P. Mukherjee, and T. N. Seyfried.
Metabolic management of glioblastoma multiforme using standard therapy together with a restricted ketogenic diet: Case report.
Nutr Metab (Lond), 7:33, 2010.