Department of Diagnostics and Intervention, 2025

Germline Genetic Predisposition to Glioma

Adam Rosenbaum

UMEÅ UNIVERSITY

ISBN: 978-91-8070-809-8 (print)
ISBN: 978-91-8070-810-4 (pdf)
ISSN: 0346-6612
Umeå University Medical Dissertations New Series no. 2387
Cover art by Adam Rosenbaum and Sigrid Rosenbaum Velasco
Electronic version available at: http://umu.diva-portal.org/
Printed by: Scandinavian Print Group, Skarpnäck, 2025

Para mi maravillosa esposa Pati y mi querida hija Sigrid, con todo mi corazon.
Gracias por recordarme, día tras día, lo que realmente importa en la vida.

Table of Contents

Abstract

Background: Gliomas are the most common primary malignant tumors of the central nervous system, encompassing a heterogeneous group of subtypes with varying clinical outcomes. Despite advances in the molecular understanding of glioma biology, therapeutic options—especially for higher-grade gliomas—remain limited, and survival outcomes have not been significantly improved. While environmental factors play a minor role in glioma development, germline genetics have emerged as an important contributor toward glioma susceptibility. Nevertheless, the biological mechanisms and complete spectrum of genetic risk factors require further investigation.

Aim: The aim of this thesis was to further contribute to the knowledge of genetic susceptibility to glioma through three studies: a functional study to understand the biological mechanism behind an established risk variant in the *LRIG1* gene (Paper I), a genome-wide association study (GWAS) of common genetic variation investigating risk for pediatric glioma (Paper II), and a study investigating rare germline variants in adult glioma patients from a regional cohort (Paper III).

Methods: In Paper I, CRISPR-Cas9 was used to generate isogenic HEK293 with and without a known risk allele (rs11706832, A/C). Transcriptome profiling and metabolomics were performed to investigate the functional impact of the allele in the cell lines. In Paper II, a meta-GWAS was performed, including over 4,069 pediatric gliomas and 8,778 controls from Sweden, Denmark, and the United States. Germline genotyping was performed using single nucleotide polymorphism (SNP) arrays on blood-derived DNA. Transcriptome-wide association studies (TWAS) and a colocalization analysis were performed using publicly available expression quantitative trait locus data. In Paper III, whole-genome sequencing was performed on blood samples from 113 adult glioma patients from a regional cohort from northern Sweden.

Results: In Paper I, no direct effect from the risk allele on *LRIG1* expression or splicing was found. However, altered expression of mitochondrial genes was observed, possibly due to dysregulation of *SLC25A26*. A global pattern of dysregulated innate immunity was observed and validated using bulk gene-expression data from the Cancer Genome Atlas low-grade gliomas. In Paper II, a genome-wide significant risk locus for childhood pediatric gliomas was identified at 9p21.3 (*CDKN2B-AS1* locus, lead SNP rs573687, odds ratio = 1.27, p = 6.97e-10). TWAS and colocalization results linked this risk to a reduced expression of tumor suppressor *CDKN2B*. In Paper III, 17.6% of the adult glioma patients had pathogenic or likely pathogenic variants in

established predisposition genes, primarily in DNA repair genes. Additionally, gene-wise burden tests suggested enrichment for rare germline variants in *TP53* and in two genes not previously linked to glioma predisposition, *CREBBP* and *DNMT3A*.

Conclusion: Collectively, the results from this thesis demonstrate that genetic predisposition to glioma involves multiple genetic factors and must be studied using a wide range of methodologies, including methods to capture rare germline mutations and common genetic variation. These results highlight the importance of both rare and common germline variants in glioma predisposition, underscoring the potential of integrating germline genetics into future risk stratification and research efforts.

Original Papers

Paper I

Rosenbaum, A., Dahlin, A. M., Andersson, U., Björkblom, B., Wu, W. Y., Hedman, H., Wibom, C., & Melin, B. Low-grade glioma risk SNP rs11706832 is associated with type I interferon response pathway genes in cell lines. Sci Rep 13, 6777 (2023). https://doi.org/10.1038/s41598-023-33923-4

Paper II

Foss-Skiftesvik, J., Li, S., **Rosenbaum, A**., Hagen, C. M., Stoltze, U. K., Ljungqvist, S., Hjalmars, U., Schmiegelow, K., Morimoto, L., de Smith, A. J., Mathiasen, R., Metayer, C., Hougaard, D., Melin, B., Walsh, K. M., Bybjerg-Grauholm, J., Dahlin, A. M., & Wiemels, J. L. (2023). Multi-ancestry genome-wide association study of 4069 children with glioma identifies 9p21.3 risk locus. Neuro-oncology, 25(9), 1709–1720. https://doi.org/10.1093/neuonc/noad042

Paper III

Rosenbaum, A., Wibom, C., Hammermeister Suger, A., Pensch, R., Roy, A., Brännström, T., Rentoft, M., Forsberg-Nilsson, K., Lindblad-Toh, K., Lindström, S., Dahlin A. M., Melin, B. (2025). Rare Germline Variants Contribute to Glioma Predisposition: Whole-Genome Analysis of a regional cohort of glioma patients. Manuscript.

Abbreviations

AF	Allele frequency
BH	Benjamini–Hochberg
CI	Confidence interval
CNS	Central nervous system
CNS5	WHO Classification of CNS Tumours, 5th edition (2021)
CPG	Cancer predisposition gene
CRISPR	Clustered regularly interspaced short palindromic repeats
CT	Computed tomography
DE	Differential expression
DEA	Differential expression analysis
eQTL	Expression quantitative trait locus
FDR	False discovery rate
GC-MS	Gas chromatography-mass spectrometry
GICC	Glioma International Case-Control study
GSEA	Gene set enrichment analysis
HGG	High-grade glioma
gnomAD	Genome Aggregation Database
GTEx	Genotype-Tissue Expression project
GWAS	Genome-wide association study
IDH	Isocitrate dehydrogenase
InDel	Insertion deletion
LD	Linkage disequilibrium
LGG	Low-grade glioma
LoF	Loss of function
MAF	Minor allele frequency
MRI	Magnetic resonance imaging
NGS	Next-generation sequencing
OR	Odds ratio
PCA	Principal component analysis
P/LP	Pathogenic or likely pathogenic
QC	Quality control
RNA-seq	RNA sequencing
SBS	Sequencing by Synthesis
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
s-QTL	Splice quantitative trait locus
SV	Structural variant
SweGen	Swedish Genome Reference Dataset
TCGA	The Cancer Genome Atlas
TTF	Tumor treating fields
TWAS	Transcriptome-wide association study
U-CAN	Uppsala–Umeå Comprehensive Cancer Consortium

UKB	UK Biobank
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
2HG	2-Hydroxyglutarate

Enkel sammanfattning på svenska

Gliom är den vanligaste formen av elakartade hjärntumörer bland både vuxna och barn. Trots att detta är en relativt ovanlig sjukdom står den fortfarande för många cancerorsakade dödsfall, då den ofta är svårbehandlad. Aggressiva former av gliom, såsom glioblastom, har en mycket dålig prognos. Orsakerna till varför gliom uppstår är inte klarlagda och till skillnad från många andra cancerformer så spelar miljöfaktorer en mindre roll. Däremot har man sett att genetiska faktorer kan bidra till en ökad risk.

I den här avhandlingen undersöks flera aspekter av genetisk predisposition för gliom. I en experimentell studie har vi med hjälp av genmodifierade cellinjer undersökt effekten av en specifik riskvariant i *LRIG1*-regionen. Resultaten visade att varianten inte påverkar uttrycket av *LRIG1*-genen, men istället har en möjlig effekt på en angränsad gen och nedströmseffekter på mitokondriellt genuttryck, samt på gener involverade i immunförsvaret. Vi har också genomfört den hittills största genetiska populationsstudien av gliom hos barn där vi identifierar en tydlig riskregion på kromosomarm 9p21.3. Denna region är särskilt kopplad till låggradiga astrocytom och våra resultat pekar mot en påverkan på tumörsuppressorgenen *CDKN2B*. Slutligen har vi analyserat förekomsten av sällsynta potentiellt skadliga genvarianter hos vuxna gliompatienter. Här fann vi att nästan var femte patient bar på en sällsynt variant i någon av de studerade generna, varav ett flertal kända cancergener.

Tillsammans visar resultaten att både vanliga och sällsynta genetiska variationer kan påverka risken att utveckla gliom, sannolikt genom ett flertal biologiska mekanismer. Genom att kombinera funktionella studier med befolkningsbaserade studier och helgenomdata ger denna avhandling en bredare förståelse för sjukdomens komplexa genetiska orsaker, vilket i förlängningen kan bidra till bättre riskbedömningar och eventuellt utvecklingen av nya behandlingar.

Background

Glioma Overview

Definition and General Epidemiology

Gliomas are a heterogeneous group of brain tumors, believed to arise from the central nervous system (CNS)'s supportive glial cells, including astrocytes and oligodendrocytes. Although rare (with an incidence of 5–6 per 100,000), gliomas are the most common malignant primary CNS tumors in adults[1]. In children, approximately half of pediatric CNS tumors are gliomas, predominantly of lower grades. As such, gliomas are a relatively common tumor entity in children[2]. Overall, the term *glioma* represents a broad spectrum of tumor entities with distinct molecular backgrounds and clinical outcomes.

At one end of the spectrum, low-grade gliomas (LGGs), consisting of World Health Organisation (WHO) grades 1–2 classified gliomas, usually develop relatively slowly and are less aggressive. Grade 1 gliomas tend to be well-circumscribed and more amenable to surgical resection. These tumors also have more favorable clinical outcomes than tumors of higher grades. For example, pilocytic astrocytoma, a grade 1 astrocytoma and the most common glioma among pediatric cases[2], can often be treated successfully with surgery alone[3].

On the other end of the spectrum, gliomas of higher grades (WHO grades 3–4) are known for their aggressive progression and poor prognosis. Glioblastomas in particular have a very unfavorable outcome: Patients have a median survival of only 12–15 months, despite multimodal therapy[4].

Gliomas account for roughly 26% of all primary brain tumors. Diffuse gliomas, characterized by infiltrative spread into surrounding tissue, are by far the most common malignant tumor entity among all CNS tumors[1]. High-grade gliomas (HGGs; grades 3–4) are more common within the older population, whereas LGGs (grades 1–2) are more commonly found in children and young adults, as shown in Figure 1.

Efforts to investigate environmental causes of the development of glioma have found that ionizing radiation contributes to glioma development. However, the number of cases attributable to this risk factor is small; thus, occurrences of gliomas do not seem to depend much on environmental causes[5]. Genetic causes are a better-established risk

factor for glioma, and results from a number of studies have contributed to glioma risk now being associated with germline genetics. A familial aggregation of gliomas has been identified[6], and a number of cancer-predisposition syndromes are associated with glioma risk. Genome-wide association studies (GWASs) of common variation in the human genome have shown that several single nucleotide polymorphisms (SNPs) contribute to glioma risk.

Treatment of gliomas still largely relies on regimens developed decades ago[7]. Although new treatments are being studied, the prognosis for patients with HGGs remains poor. An improved understanding of glioma etiology may aid the discovery of new targeted therapies, as has been shown to be successful in other cancers. For example, it has been shown

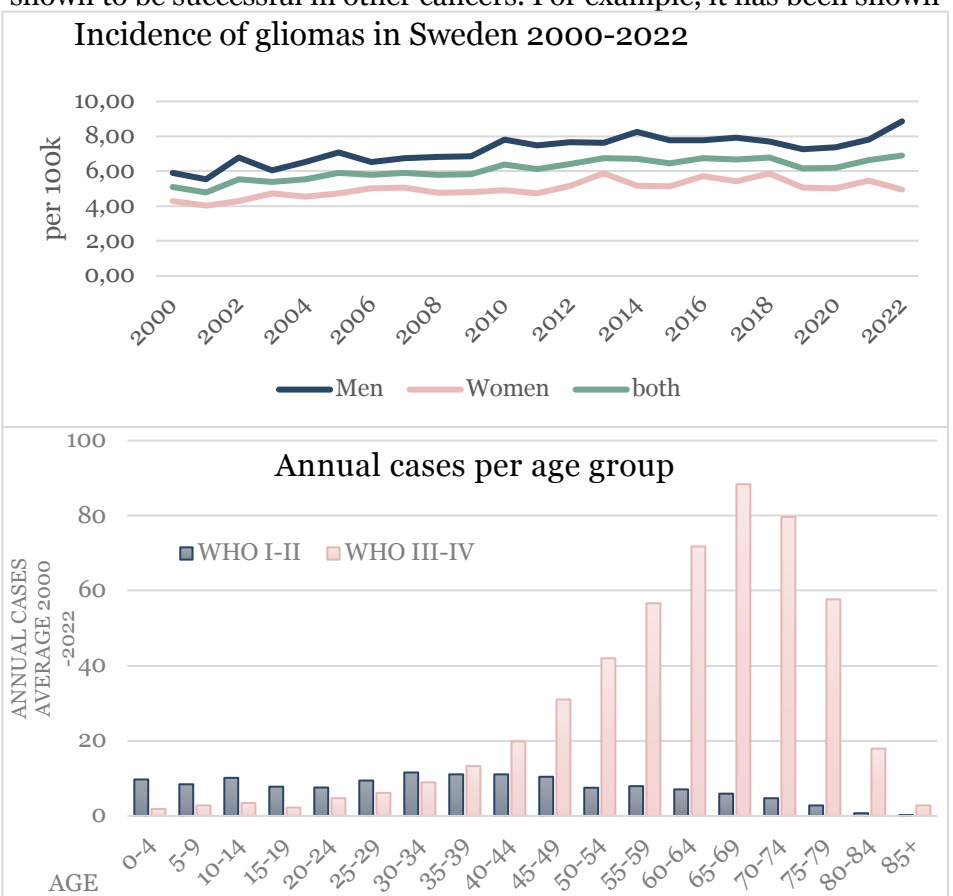

Figure 1: Incidence and annual cases per age group for WHO grades 1–4 gliomas in Sweden. (Source: National Board of Health and Welfare in Sweden statistics database)

that PARP inhibitors are beneficial to breast cancer patients with a *BRCA1/2* germline variant[8].

Treatment Regimens for and Clinical Management of Gliomas

Treatment of gliomas is dictated by tumor grade, location, and patient age and condition. For LGGs, the foundation of therapy is surgery; in some cases, a surgical resection of the tumor alone is potentially curative, especially in pediatric patients, where grade 1 pilocytic astrocytomas are accessible and constrained from healthy brain tissue.[9] When complete resection of the tumor is not possible, adjuvant treatment consisting of radiotherapy and chemotherapy is usually considered.

In contrast, treatment of HGGs, particularly glioblastomas, requires a different approach consisting of several treatment strategies. Although surgery is crucial for these patients as well, a total resection of these infiltrative tumors is often impossible, and adjuvant therapy is always required. The current standard of care for glioblastomas relies on the Stupp regimen[7], which consists of surgery followed by radiotherapy combined with oral administration of temozolomide, an alkylating agent that remains the most widely used and effective drug for HGGs. This protocol, established over two decades ago, offers an overall survival of 27% at 2 years for glioblastoma patients[10].

In recent years, tumor-treating fields (TTF) has emerged as a novel, noninvasive modality in treatment of glioblastomas and can now be considered as a line of therapy alongside surgery, radiotherapy, and chemotherapy. TFF applies low-intensity electric fields to the tumor site. These electrical fields disrupt the mitosis of tumor cells, eventually promoting tumor cell death[11,12]. Adding the treatment modality of TTF to the standard of care for glioblastoma has been shown to increase the median overall survival by around 5 months[13].

Another ongoing field of study for glioma therapy—particularly for glioblastomas—is immunotherapy. This includes immune checkpoint inhibition, adoptive T-cell therapies, tumor vaccines, and oncolytic viral therapies[14]. Although several immunotherapeutic strategies are being investigated in clinical trials for glioblastoma, the results have been variable thus far, with many trials showing limited survival benefits for patients[14]. These results could be attributed to the immunosuppressive tumor microenvironment of gliomas. While gliomas are infiltrated by myeloid cells such as microglia and tumor-associated macrophages, T-cells are infrequent and often functionally exhausted[15,16]. Single-cell

studies have revealed high heterogeneity in the immune landscape of gliomas across subtypes and patient age, showing that immune cells are skewed to immunosuppressive types[17–19].

The patient group of LGG pediatric patients generally have a good long-term survival, with a 5-year overall survival above 90%[20–23]. Nevertheless, many of these patients experience lasting health challenges throughout life, with long-term effects either stemming from the tumor itself or being induced by treatment. Long-term health effects vary significantly depending on tumor location, age at diagnosis, and treatment modalities used[24]. Examples of disease-related long-term effects include motor deficits, ataxia, cranial nerve dysfunction, and seizures; patients may also experience endocrine dysfunction, particularly if treated for hypothalamic tumors. Treatment-related long-term health effects can result from radiotherapy, which is a risk factor for vascular complications, secondary tumors, and neurocognitive decline.[25] Thus, it is notable that, even though the long-term survival of these patients is good, their quality of life may be greatly reduced.

In Sweden, clinical management of glioma patients follows a national standardized care pathway (*standardiserat vårdförlopp*) that exists for many cancers. Individuals with suspected brain tumors, based on symptoms such as neurological deficits, first-time epileptic seizures, cognitive changes, are usually initially evaluated with a computed tomography (CT) scan, followed by magnetic resonance imaging (MRI), which is the preferred modality. A histopathological diagnosis, obtained through biopsy or surgical resection, remains essential for accurate tumor classification. Multidisciplinary tumor conferences further guide the management of treatments, which typically include surgical resection, followed by radiotherapy and chemotherapy according to tumor grade, type, molecular profile, and location[26].

Classification of Gliomas

World Health Organisation Classification

Although little progress has been made in the past decades with therapeutic options for glioma, the classification of gliomas has recently been refined. The WHO provides an international standard on the classification of CNS tumors. The most recent version of this classification—the fifth edition, published in 2021 (CNS5)[27]—emphasizes integration of molecular aspects (building on the 2016 edition), alongside traditional histology and pathology. Classification based on layering histopathological and molecular data to form an

integrated diagnosis is now recommended for all CNS tumors, including glioma[28,29]. The WHO classification system of gliomas is intended to improve diagnostic accuracy and enable the stratification of patients into clinically relevant groups.

One key difference in CNS5 is that diffuse gliomas are now broadly categorized into adult and pediatric types. Adult-type diffuse gliomas are largely defined by *IDH*-mutation status and 1p/19q codeletion, and are categorized into three groups: (1) astrocytoma, *IDH*-mutant; (2) oligodendroglioma, *IDH*-mutant and 1p/19q co-deleted; and (3) glioblastoma, *IDH*-wildtype. According to CNS5, astrocytoma, *IDH*-mutant is regarded as a single type and is given a grade of 2, 3, or 4, depending on histologic and molecular features. Oligodendrogliomas remain graded as CNS WHO 2 or 3, based on mitoses measures and necrosis. In CNS5, glioblastoma is defined as *IDH*-wildtype grade 4 or *IDH*-wildtype diffuse astrocytoma based on certain molecular markers (*TERTp*, *EGFR* amplification, gain of chromosome 7, and loss of chromosome 10), even if the histopathology deviates from the glioblastoma type[28–30]. A graphic overview of the classification of adult-type diffuse gliomas is provided in Figure 2.

Pediatric gliomas have distinct molecular features and clinical behavior. Pediatric diffuse gliomas are subdivided into low- and high-grade groups. Here, the high-grade entities are infiltrative, in contrast to circumscribed low-grade pediatric gliomas such as pilocytic astrocytoma. Diffusively infiltrative pediatric gliomas of high grades, such as diffuse midline glioma, have a poor prognosis.

Key Molecular Markers of Gliomas

Earlier versions of glioma classification relied heavily on the histological features of the tumors. While histology is still a key component in tumor classification, histology alone has been demonstrated to be insufficient in stratifying CNS tumors, as tumors with similar histological features may still have significantly different clinical implications—especially among HGGs. Molecular pathology has increasingly been shown to refine the tumor types in clinically relevant patient groups, and the current classification relies heavily on molecular markers to complement histopathological classification. Below are some key molecular markers that are used to differentiate gliomas.

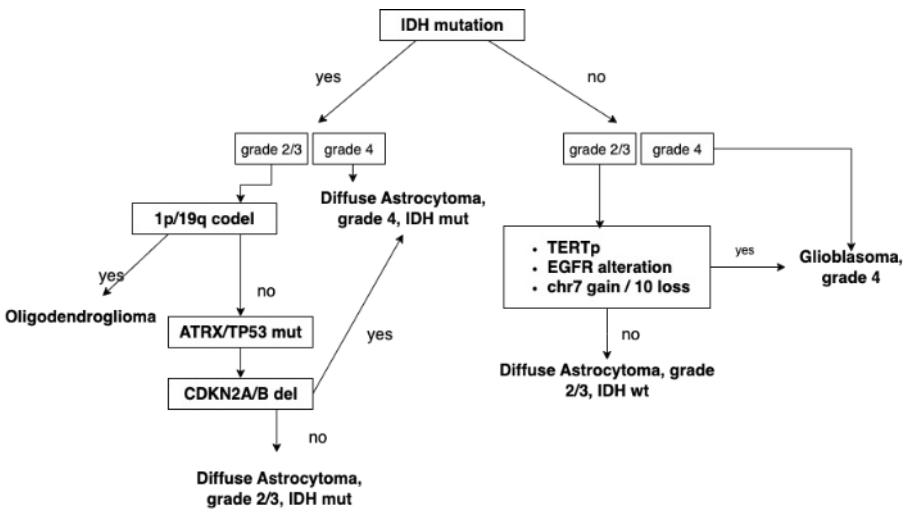

Figure 2: Classification of adult-type diffuse gliomas: a schematic overview of CNS5.

IDH1/*IDH2* mutation

Somatic mutations in genes coding for isocitrate dehydrogenases (*IDH1* or *IDH2*) are one of the most important clinically significant markers. Here, the most common mutation, which results in an amino acid change from arginine 132 to histidine in *IDH1*, distorts the normal enzymatic function and leads to the added ability of catalyzing alpha-ketoglutarate to 2-hydroxyglutarate (2HG)[31]. Accumulation of 2HG leads to the inhibition of demethylases, causing hypermethylation of histones and DNA and blocking cells with this mutation from differentiating[32]. The presence or absence of these mutations in *IDH1/2* alone entails a fundamental division of different tumor entities. This mutation is frequently found in grade 2 and 3 gliomas. Overall, *IDH* mutation is associated with a favorable prognosis and slower disease progression.

TP53 and *ATRX* mutation
Mutations in *TP53* and *ATRX* are common in *IDH*-mutant astrocytomas but are typically not seen in oligodendrogliomas. As such, alterations in these genes can be used to distinguish between ambiguous cases of astrocytomas. Mutations or loss of expression of *ATRX* are frequently seen in *IDH*-mutated grade 2–3 astrocytomas[33] and leads to an alteration of telomere maintenance in astrocytomas[34].

1p/19q co-deletion
Complete deletion of chromosome arms 1p and 19q is required for an oligodendroglioma diagnosis (not seen in astrocytomas), coupled with an *IDH* mutation.

TERT promoter mutation
Mutations in the *TERT* (telomerase reverse transcriptase) promoter (*TERT*p) are one of three molecular criteria that facilitate the classification of grade 4 *IDH*-wildtype astrocytomas to glioblastoma. These mutations activate the promoter, which in turn leads to abnormal telomere maintenance[35]. However, *TERT*p is also commonly seen in oligodendrogliomas; thus, the combination of both *TERT*p and *IDH*-mutation is classified as oligodendroglioma. *TERT*p is seldom seen in *IDH*-mutant astrocytomas.

EGFR alterations
Alterations of the receptor tyrosine kinase *EGFR* (epidermal growth factor receptor) gene, such as amplification or mutation to *EGFRvIII* (caused by the deletion of exons 2–7, leading to a constitutively active protein), are another of the three required molecular markers to upgrade *IDH*-wildtype astrocytomas to glioblastoma. EGFR alterations are seen in approximately 50% of gliomas[36] and promote tumorigenesis through the activation of downstream pathways such as RAS/MAPK/ERK, JAK/STAT, and PI3K/AKT[37].

Chromosome 7 gain/chromosome 10 loss
The simultaneous gain of chromosome 7 and loss of chromosome 10 is the third of the molecular markers required to upgrade *IDH*-wildtype astrocytoma to glioblastoma.

CDKN2A/B deletion
Homozygous deletion of both the *CDKN2A* and *CDKN2B* tumor-suppressor genes is common in HGGs and in *IDH*-mutant astrocytomas, and leads to a strong negative prognosis. As such, the presence of this alteration results in a grade 4 classification.

MGMT promoter methylation

Other than alterations in the tumor genome, gliomas are further defined by epigenetic changes, such as DNA methylation. The most important epigenetic marker is methylation of the *MGMT* (methylguanine–DNA methyltransferase) promoter. Methylation of the *MGMT* promoter silences the expression of *MGMT*. Although not significant for glioma classification, this marker is of great clinical importance, as *MGMT* silencing reduces a tumor's ability to repair DNA damage and thus renders it more susceptible to chemotherapy (and probably also to ionizing radiation) through tumor-alkylating agents such as temozolomide[38].

Methylation Profiling and Epigenetic Signatures

DNA methylation is a molecular process in which a methyl group is added to a nucleotide without further changing the sequence. In humans, this mostly happens at cytosine bases, followed by guanine bases (the so-called CpG site). Methylation has been recognized as an important molecular factor; it has implications for chromatin structure, as well as the suppression of gene expression through the methylation of promoters. The importance of DNA methylation has been further elucidated by recent methylation-based classifiers of CNS tumors, which have been demonstrated to classify tumor entities with high precision based on methylation patterns in the tumor genome alone[39]. Therefore, epigenetic profiling of gliomas has become widely used in subtyping tumor entities.

Other epigenetic signatures, such as histone modifications, often resulting from mutations in *H3F3A*, are frequent in pediatric gliomas but less frequent in adults. Here, the *H3F3Ap*.K27M mutation defines a subgroup of grade 4 gliomas within pediatric patients[40].

Risk Factors and Glioma Predisposition

Environmental Risk

Compared with many other cancers, whose incidence can often be explained by environmental factors (e.g., smoking tobacco products is linked to lung cancer), environmental contributions to glioma are not as pronounced. At most, environmental factors can be said to contribute marginally to glioma incidence[5]. The only well-established environmental risk factor is exposure to ionizing radiation, where high

dosage exposure has been shown to lead to a relative risk of 2–3 of developing glioma[41,42].

Large-scale epidemiological studies, including the Glioma International Case-Control (GICC) study[43], have repeatedly found that a history of allergies, along with asthma and eczema, is associated with a reduced risk of glioma, with about 30% lower risk for individuals with respiratory allergies[44]. These findings are consistent across studies with various study designs and cohorts[45–48]. This reverse association provides an interesting path for further research on gliomas, as it implies that immunological factors are an important contributor to risk[49].

Metabolomic Markers

Although there is little evidence of environmental factors behind glioma development, gliomas have been associated with changes in small molecules—that is, metabolites—in the blood prior to diagnosis. In a nested case-control study on the metabolites in serum from 110 glioblastoma cases and matched controls, elevated serum concentrations of alpha and gamma tocopherols were detected in pre-diagnostic serum samples[50]. Another study identified 43 metabolites with different pre-diagnostic levels in glioma cases years before diagnosis[51]. Moreover, a panel of 15 metabolites in pre-diagnostic plasma were identified in a Swedish cohort as early as 8 years before diagnosis, where the elevated metabolites in future glioma patients were strongly related to oxidative metabolism, among other pathways[52]. Furthermore, a large nested case-control study using matched cases and controls from the Northern Sweden Health and Disease Study (NSHDS) and the European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts showed that changes in plasma metabolite profiles could be detected up to 8 years prior to glioma diagnosis[53]. Here, a panel of 20 metabolites were shown to robustly separate future cases from controls, including metabolites in the TCA cycle and the Warburg effect (in which tumor cells reprogram to favor glycolysis under aerobic conditions)[53].

In all, findings from metabolomic studies further support the idea that systemic changes that contribute to the development of glioma can begin long before clinical onset. This may reflect the early tumor development of gliomas years before diagnosis[54]; it could also point to underlying genetic or environmental risks mediating metabolite changes.

Genetic Predisposition

Several cancer predisposition syndromes are known to increase the risk for gliomas and other CNS tumors. For example, brain tumors are

common among families with Li–Fraumeni syndrome, which is caused by protein-disruptive heterozygous variants in *TP53*[55]. In a study analyzing 475 tumors from 91 families with germline variants in *TP53,* 12% of the tumors were located in the brain[56]. Another syndrome associated with CNS tumors is neurofibromatosis type 1 and 2 (NF1/2)[57]. The CNS tumors most commonly associated with individuals with NF1 are low-grade astrocytomas originating from the optic nerves—a tumor entity more common in pediatric patients[58].

A study conducted in northern Sweden analyzed the first-degree relatives of patients diagnosed between 1985 and 1993 (1,890 relatives and 297 probands) and found a standardized incidence ratio of 3.12 for individuals with first-degree relatives with astrocytoma. The study further found that the increased risk was more pronounced in first-degree relatives of patients with an earlier diagnosis (<58 years)[59]. Another study performed in the United States compared 462 adults with glioma and 443 matched controls and found the odds ratio for having a first-degree relative with a CNS tumor to be 2.3. Less than 1% of the cases had known cancer predisposition syndromes, showing that familial risk is mainly attributable to other heritable factors[60].

With array technologies and the development of next-generation sequencing (NGS), it has become possible to conduct large molecular epidemiological studies, investigating the entire genome for changes that increase the risk of glioma. These studies include GWASs; more recently, several whole-genome sequencing (WGS) or whole-exome sequencing (WES) approaches have also been applied[61–66]. Some of these studies are summarized in Table 1.

While the familial aggregation of gliomas highlights possible genetic predisposition, most gliomas are non-familial and cannot be explained by the inheritance of penetrant variants. Several GWAS[67–70] studies have found common SNPs to be associated with glioma risk. These polymorphisms typically occur in at least one in hundred individuals and generally reside in non-coding regions of the genome, often in proximity to well-known cancer-predisposition genes (CPGs). These SNPs, which have been found in 25 loci in the genome to date, further explain a part of the genetic predisposition to gliomas.

A study utilizing genome-wide complex trait analysis on a total of 3,373 cases and 4,571 controls from three GWAS datasets, while considering the aggregate effects of common SNPs across the genome, estimated that around 25% of the variation in glioma risk can be attributed to common SNPs[71]. Another study investigating glioma risk using polygenic risk

scores on the GICC cohort (10,346 cases and 14,687 controls) modeled the joint effect of more than one million common variants to stratify groups for glioma risk. This model found that the overall lifetime risk reached 1.18% for the highest 95th percentile of the model[72]. Even though a substantial proportion of glioma heritability can be explained through common genetic variation, many of the genetic variants contributing to this heritability remain undefined.

It has further been shown that specific risk loci can be mapped to distinct molecular subtypes and groupings of gliomas. One study, dividing gliomas on three markers – *IDH* mutation, *TERT*p mutation, and 1p/19q codeletion – found the *CCDC26* risk locus shows a specificity for triple-positive oligodendrogliomas (*IDH*- and *TERT*p mutations, and 1p/19q codeletion) and the *TERT*, *CDKN2B–AS1*, *TP53*, and *RTEL1* loci show specificity to *TERT*-only gliomas[73]. A meta-analysis consisting of a total of 5,103 cases and 10,915 controls showed that the 25 germline risk loci are associated with distinct molecular classes of glioma. The main patterns found in the study were as follows: (1) The *TERT* and *TP53* risk loci increased the risk across all glioma subtypes; (2) SNPs in loci around *CDKN2B-AS1*, *EGFR*, and *RTEL1* showed specificity for *IDH*-wildtype gliomas; and (3) variants in loci including *CCDC26*, *LRIG1*, *PHLDB1*, *ETFA*, *MAML2*, and *ZBTB16* were mainly associated with *IDH*-mutated gliomas[74]. Another study found that performing a GWAS stratified for molecular subtypes of glioma uncovered additional loci: SNPs in *D2HGDH* were associated with *IDH*-mutated glioma, and an SNP in *FAM20C* was associated with triple-positive gliomas[75].

While large-scale GWASs have identified numerous risk loci for glioma, these associations provide little information on which SNPs in a locus are causal and what their downstream effects may be. Hits from GWASs predominately lie in non-coding parts of the genome and are often found in DNase I-hypersensitive sites, where chromatin is more accessible[76]. Furthermore, about 43% of GWAS hits colocalize with known expression quantitative trait loci (eQTLs)[77], indicating that many of these SNPs regulate gene expression. Nevertheless, the risk mechanism for most GWAS hits is not known.

Efforts have been made to study the functionalities of glioma GWAS risk loci. For example, the SNP at the 8q24.2 locus, located in an intron of *CCDC26*, lies within a brain specific enhancer. It has been suggested that this SNP disrupts the binding motif of OCT2/4, affecting its interaction with the promoter of the *MYC* oncogene and thereby leading to increased expression of *MYC*[78]. Another functional study on the 11q23.3 locus (near *PHLDB1*) found several functional SNPs that were in linkage

disequilibrium (LD) to the glioma-associated SNP rs498872. It was found that a particular SNP residing in an enhancer element physically interacts with the promoter of *DDX6*[79]. A study performed on the 20q13.33 locus (near *RTEL1*) identified a functional SNP in LD residing in an enhancer element. Disrupting this SNP altered the expression of several nearby genes, including *STMN3*, *RTEL1*, and *GMEB2*[80]. A recent study mapped the enhancer connectome in glioma cells and found that several glioma-risk SNPs may promote tumor progression through enhancer-mediated regulatory interactions. As an example, the knockout of an enhancer element in an intronic region of *RTEL1* harboring a risk SNP did not affect the expression of *RTEL1* but reduced the expression of the enhancer target *SOX18*[81].

Rare Deleterious Germline Variants

GWASs are fundamentally designed to investigate common genetic variations and their association to disease, given the restrictions of array technologies targeting common SNPs in the human genome. Rare variants are typically not represented by SNP arrays; furthermore, the statistical power to detect the associations of rare variants is lower due to the lower incidence of such variants in the population.

Over 100 CPGs have been identified to date[82]. Among them are well-studied tumor-suppressor genes such as *BRCA1* and *TP53*, in which deleterious mutations lead to an increased risk of cancer. CPGs also encompass genes in which gain-of-function variants lead to an increased risk, such as *RET*, *MET*, *KIT*, and *ALK*[82]. For example, in breast cancer, rare damaging coding variants in genes such as *BRCA1/2*, *CHEK2*, *PALB2*, and *ATM* have been shown to contribute significantly to susceptibility[83]. This finding underscores the importance of investigating rare variants in cancer in order to fully understand genetic susceptibility.

With the advent of NGS technologies, it has become possible to study rare potentially deleterious variants in the human genome and associate them with disease. This approach is guided by a different assumption: While rare deleterious variants occur less frequently, their effects can be assumed to be higher[84]. A number of studies have utilized WGS and WES technologies on glioma patients, and the results from these studies show that considering rare germline variants further helps to explain the inheritable component of glioma that is not captured through GWASs.

WES of familial gliomas found rare variants in *POT1*, a member of the telomere shelterin complex. These variants were found to co-segregate with the disease in two families and were also found in an additional family from a validation cohort[61]. Another study performed WES on

glioma patients from 61 families in Germany, finding that predicted deleterious variants in the DNA proofreading genes *POLE* or *POLD1* occurred in 10/61 (16%) of the families. *POLD1* was found in two patients with oligodendroglioma and *POLE* variants were found in two glioblastoma patients, both with spinal metastasis[65]. A large study performing WGS on 304 families likewise found rare deleterious variants in *POLE* and *POLD1*, as well as in *ATM*, *BRIP1*, *MLH1*, *NBN*, and *PMS2*. The study also found deleterious variants in novel genes not previously associated with glioma or cancer predisposition, with variants in *HERC2* being found to be enriched compared with the controls[63].

Other recent studies have also investigated the role of rare germline variants among pediatric patients with CNS tumors, including glioma. One study examining a cohort of 51 pediatric CNS tumor patients from the Western Balkans (including 15 HGGs and 8 LGGs) found that 27% of the patients carried rare germline variants in the CPGs *CHEK2*, *ELP1*, *MLH1*, *MSH2*, *NF1*, *NF2*, and *TP53*[64]. Another study investigating rare germline variants in CPGs in 280 children with astrocytoma (grades 2–4) diagnosed in California found 33 variants predicted to be deleterious in 11.1% of the patients. These variants were largely found in genes involved in DNA repair, including *FANCA*, *FANCE*, *ATM*, *BLM*, *BRCA1*, *BRCA2*, *NBN*, *POLE*, *PTEN*, *RECQL4*, and *TP53*. A gene variant burden analysis found that rare variants in *TP53* occurred more frequently in pediatric patients diagnosed with glioblastoma than in publicly available controls[62].

Table 1: Overview of studies of genetic predisposition for glioma

Year	Author	Study design	Sample size (cases/controls)	Glioma type	Genes/loci
Common genetic variation (SNPs)					
2009	Shete et al.[68]	GWAS	1,878/3,670	All adult	*TERT, CCDC26, 6, CDKN2A-CDKN2B, RTEL1, PHLDB1*
2009	Wrensch et al.[70]	GWAS	692/3,992	Adult high grade	*CDKN2B, RTEL1*
2011	Sanson et al.[85]	GWAS	2,269/2,500	All adult	*EGFR*
2011	Stacey et al.[69]	Targeted	1,395/45,937	All adult	*TP53*
2012	Rajaraman et al.[86]	GWAS	1,856/4,955	All adult	*RTEL1, TERT, CDKN2B-AS1, EGFR, CCDC26, PHLDB1*
2015	Kinnersley et al.[87]	GWAS	4,147/7,435	GBM, non-GBM adult	*POLR3B, VTI1A, ZBTB16, ETFA*
2017	Melin et al.[67]	GWAS	12,496/18,190	GBM, non-GBM adult	New loci: *RAVER2, MDM4, AKT3, IDH1, LRIG1,OBFC1, MAML2, AKAP6, MPG, LMF1, HEATR3, SLC16A8*
Rare variants					
2014	Bainbridge et al.[61]	WES	55 families	Familial	*POT1*
2020	Muskens et al.[62]	WES	280	Pediatric	*ACD, ATM, BLM, BRCA1, BRCA2, BUB1B, CASP9, CDKN2A, CHEK2, DHCR7, FANCA, FANCE, GJB2, MLH1, NBN, NF1, PMS2, POLE, PTCH1, PTEN, RECQL4, TP53, WT1*
2021	Bu et al.[88]	WES	99/0	Sporadic	*AR, RECQL4, ALK, CFTR, BRCA1, BRCA2, PALB2, ATM, EXT2, MSH3, POLH, PTCH1, SETD2, SLX4, APC, CHEK2, BLM, FANCA, MSH2, MSH6*
2023	Choi et al.[63]	WGS	203/189	Familial	*BRIP1, PMS2, POLE, BAP1, POT1, HERC2, TRPC4AP, KIAA1549, CASC4, CASC5, MYO7A, DCAF8L2, IGSF9, SLC4A7, ZC3H7B, IP6K1, LRRK2, MAST4*
2023	Jovanovic et al.[64]	WES	51/0	Pediatric	*ALK/APC, CHEK2, ELP1, MLH1, MSH2, NF1, NF2, TP53*
2023	Weber et al.[65]	WES	61 families	Familial	*POLE, POLD1*
2023	McDonald et al.[89]	WES	152/0	Sporadic	*BRCA2, MUTYH, CHEK2, BRCA1, NF1, ATM, MSH2, MSH3*
2024	Nurminen et al.[66]	WES	19 families	Familial	*GALNT13, MYO10, AR*
2025	Mateos et al.[90]	WES/WGS	252/0	Pediatric	*PALB2, BRCA2, CHEK2, BRCA1, BRIP1, FANCD2, FANCE, FANCM, MSH2, PMS2, SLX4, LZTR1, FGFR2, HOXB13, SDHA*

Technological Advances

This section highlights some of the technologies that have made it possible to advance in cancer research in general and in glioma research in particular. The key technologies described below have been made widely available to researchers in the past decades and were essential to this thesis.

Next-Generation Sequencing

One of the most ubiquitous innovations in molecular biology in recent decades, NGS consists of a broad range of technologies. Of these, sequencing by synthesis (SBS) is one of the more prevalent technologies, allowing for high-throughput sequencing of the entire human genome and transcriptome. WGS has made it possible for researchers to investigate the whole genome—somatic or germline—in an individual or tumor, which has greatly contributed to the field of cancer biology. Furthermore, RNA sequencing (RNA-seq), which depends on the same underlying technology, has made it possible to study changes in expression patterns in individual tumors. Both WGS and RNA-seq were used in this thesis.

Array technologies

Despite the advances that have been made in sequencing, SNP arrays remain an important tool in molecular epidemiology, allowing researchers to assess the common variations in a population efficiently and more cheaply than NGS technologies allow. As a cornerstone in GWASs, SNP arrays were used in Paper II in this thesis.

Metabolomics

Complementing genomic approaches, high-throughput technologies detecting small molecules and metabolites has given rise to the field of metabolomics. Liquid- or gas-based chromatography, paired with mass spectrometry, has made it possible to study the collective impact of up to thousands of metabolites present in biological materials. While genomic methodologies detect germline, or somatic static DNA variation, metabolomic can capture downstream functional effects which may reflect both genetic and environmental influences. Just as the genomic-based methods, metabolomics has also been used in paper I.

Population-Based Resources and Open Databases

Access to accumulated material from glioma patients, as provided in various biobanks, registries, and collaborative datasets, was a central

component enabling the research in this thesis. Even though gliomas are rare, with a low incidence rate compared with many other cancers, a considerable amount of material has been made available to researchers through the great efforts of national biobanks. Resources and biobanks in Sweden such as the Uppsala–Umeå Comprehensive Cancer Consortium (U-CAN) project[91] played a key role in providing tissue for this thesis.

Aims of This Thesis

Current understanding of the etiology of gliomas remains incomplete, and germline genetic predisposition may only account for a small fraction of gliomas. Nevertheless, identifying and studying the genetic contribution to glioma risk can provide valuable insights into why gliomas start to develop. The three papers that constitute this thesis differ in their methodology and approach, ranging from functional studies to population-based observational studies. By investigating common and rare variants in both adult and pediatric gliomas, this research aims to broaden the current understanding of genetic predisposition to glioma.

The specific aims of this thesis are as follows:

1. **To functionally characterize an LGG risk SNP in the *LRIG1* locus** using CRISPR-edited cell lines, with the goal of understanding the downstream transcriptional and metabolic effects and how these may play a role in gliomagenesis (Paper I);
2. **To identify the contribution of common genetic variation to pediatric glioma risk** by performing the largest meta-GWAS to date on pediatric CNS tumors and thereby uncovering common variants leading to pediatric glioma predisposition (Paper II);
3. **To investigate the contribution of rare germline variants in patients with sporadic glioma**, further expanding the known landscape of germline contribution to glioma risk (Paper III).

Through a variety of methods across the three papers, this thesis contributes to the understanding of glioma predisposition. It investigates three aspects of genetic predisposition through the three studies, including a functional analysis of a known risk SNP, common variation in the population, and rare deleterious variants. Taken together, the findings of this thesis provide a comprehensive perspective of glioma predisposition.

Materials and Methods

This section provides an overview of the key methodologies used in this thesis, together with the detailed applications used in the studies. The different workflows used throughout Papers I–III are visualized in Figure 3.

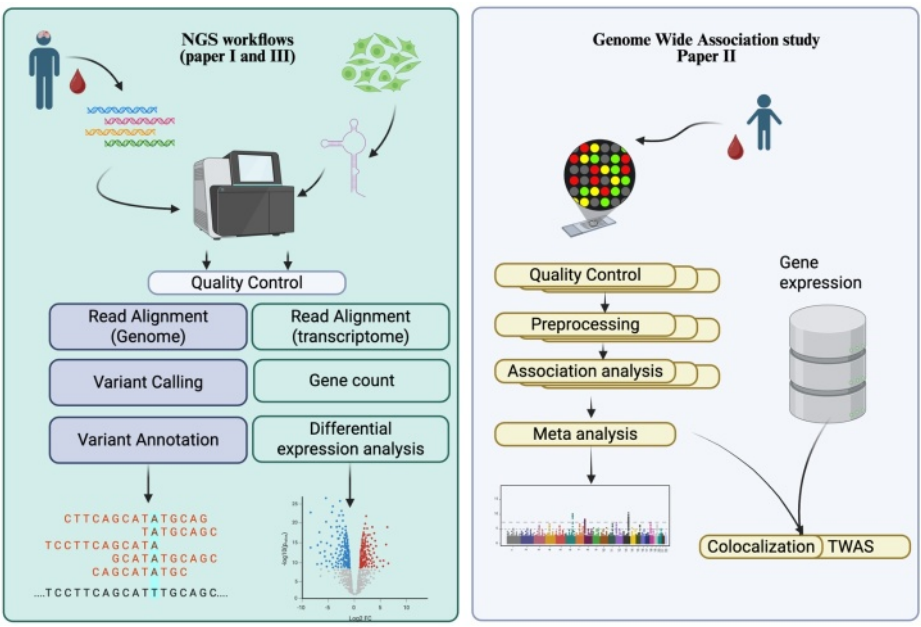

Figure 3: General overview of the main workflows in this thesis, the NGS workflows for RNA and DNA (Papers I and III), and the GWAS workflow (Paper II).

Next-Generation Sequencing (Papers I and III)

NGS is a collective name for the sequencing technologies that make high-throughput sequencing possible. Although the current technologies are more than a decade old, they represent a transformative advancement in molecular biology and genetics and have largely reshaped our understanding of the course of tumor biology, gliomas included. In this thesis, the Illumina sequencing platform was employed in two projects (Papers I and III). Illumina sequencing uses SBS, a process that involves the synthesis of short reads (DNA fragments) by consecutively adding fluorescently labeled nucleotides to a single strand of DNA that is hybridized onto a flow cell. Through the addition of each

nucleotide, a fluorescent read-out is detected, and it is possible to identify the added nucleotide with a high level of accuracy—a process referred to as base calling. This process makes it possible to sequence the entire genome, after DNA extraction and library preparation, resulting in an output of numerous short paired-end reads.

RNA-seq is a process in which mRNA is reversely transcribed into cDNA prior to sequencing. This makes it possible to not only sequence the bases in the transcriptome but also quantify the number of reads stemming from specific genes, thereby enabling measurement of the gene expression of all genes across the human genome. In **Paper I**, RNA-seq was used to assess the transcriptional changes in genetically modified HEK293T cell lines. The total mRNA was extracted from 32 clones of HEK293. Sequencing libraries were prepared from 100 ng of polyA-selected RNA using a TruSeq stranded mRNA library preparation kit. Sequencing was performed using the Illumina NovaSeq platform and an S4 flow cell, yielding approximately 60 million 150 bp read pairs for each sample.

In **paper III,** WGS was used to investigate rare germline variants in glioma patients. DNA was extracted from whole blood (n = 113) and tumors (n = 73) collected from U-CAN. Libraries were prepared from 1 µg DNA using a TruSeq polymerase chain reaction (PCR)-free DNA sample preparation kit with unique dual indexes and an average insert size of 350 bp. For 46 early-onset gliomas, sequencing was performed on the Illumina HiSeqX platform, generating 150 bp paired-end reads at a coverage of 30×. An additional 73 matched tumor-normal pairs were selected and sequenced on NovaSeq 6000, with 30× coverage for blood and 90× for tumor DNA.

All sequencing was performed using the SNP&SEQ Technology Platform in Uppsala.

Genotyping through SNP Arrays (Paper II)

In array technology, fluorescently labeled DNA is hybridized to an array containing pre-designed probes, reflecting hundreds of thousands to millions of known SNPs in the genome in parallel. A fluorescent signal at each probe site in the microarray permits the detection of a specific genotype (e.g., AA, AG, or GG) through software. Companies such as Illumina (e.g., Infinium arrays) and Thermo Fisher (e.g., Axiom arrays) have designed arrays to capture common SNPs based on allele frequencies, as well as structures of LD. Furthermore, the status of

common SNPs not included on the array itself can be imputed by using robust reference haplotypes of the human genome.

In this thesis, genotyping with SNP arrays was performed in Paper II using archived neonatal dried blood spots. The Swedish sample set in this study consisted of 352 glioma cases diagnosed before the age of 15, identified through the Swedish Cancer Registry. DNA was extracted and amplified from blood spots from the Phenylketonuria Screening Registry, and genotyping was conducted with the Illumina HumanOmni2.5Exome. Controls were previously genotyped with OncoArray-500K.

CRISPR-Cas9 Genome Editing (Paper I)

CRISPR-CAS9 genome editing is another transformative methodology in molecular biology, one that allows for high-precision targeted genomic editing. This technology is adapted from the bacterial adaptive immune system, which uses RNA to direct the Cas9 nuclease to a specific region of DNA. Cas9 induces a double-strand break at the target that can then be repaired through homology-directed repair using a template. In this methodology, the template is often utilized to introduce specific genetic changes in the repaired sequence[92].

CRISPR-Cas9 technology was used in **Paper I** to introduce a risk allele, that has been found to predispose to LGGs in adults[67], at a specific genomic locus in HEK293T cells (derived from embryonic kidney cells). Starting from a HEK293T cell line with a single allele of *LRIG1* (generated by the CRISPR-mediated deletion of one copy), editing at rs11706832 was achieved by co-transfection with a plasmid encoding Cas9 and a site-specific gRNA, along with a single-stranded oligonucleotide with the desired nucleotide conversion (C → A). Edited clones were identified by means of Sanger sequencing, where four clones with the A allele and four with the C allele were selected. The genome editing procedure and clone generation were performed by GenScript (Piscataway, NJ, USA). Four replicates of each clone were isolated and cultivated 11 days prior to harvesting. This resulted in 32 samples that were subsequently subjected to RNA-seq and gas chromatography-mass spectrometry (GC-MS).

Metabolomics and GC-MS Analysis (Paper I)

Gas chromatography-mass spectrometry (GC-MS) is a widely used technique in metabolomics. With this technique it is possible to detect small, volatile metabolites in a biological sample. In GC-MS, compounds

are first vaporized and separated while in gas phase, based on their volatility, and interaction with a chromatography column. They are then identified and quantified based on their mass-to-charge ratio using mass spectrometry. In paper I, intracellular metabolites were extracted from snap-frozen HEK293T cells and analyzed using GC-MS with an electron ionization step (GC-EI-MS).

Bioinformatics and Statistics

This section contains an overview of key methods of analysis, bioinformatics, and statistics utilized in this thesis.

NGS Preprocessing (Papers I and III)

Many bioinformatic pipelines using NGS start with raw sequencing format FASTQ files for paired-end reads, as was done in this thesis in **Paper I** (RNA-seq) and **Paper III** (WGS). Two FASTQ files are given per sample, where the DNA fragment is read from both directions. After ensuring that the reads are good quality, the raw reads are often subjected to alignment to a reference genome (WGS reads, **Paper III**) or transcriptome (RNA-seq reads, **Paper I**). The read alignment differs between RNA-seq reads and WGS: The former must use aligner software that is splice-sensitive, taking into account that any read stemming from a gene might encompass different exon–exon junctions, depending on which transcript it comes from. After alignment, the results are typically returned in a BAM format, where each read is marked with the genomic position and other details on the alignment.

RNA-Seq and Differential Expression Analysis (Paper I)

Starting from the aligned sequencing reads in BAM format, the expression of each gene can be quantified by counting the number of times a read pair is mapped to a gene. In **Paper I**, the salmon[93] software was used to quantify the transcript abundance directly from the raw FASTQ files. The reads were also mapped to the reference transcriptome using STAR[94] for a detailed alignment of the short reads. In addition, the aligned reads were used to obtain more detailed information on the splicing of the transcriptome, where alignment is necessary to study the specific occurrences of exon–exon junctions for each gene.

Once the gene- or transcript-level expression counts are available, normalization of the raw count data is essential before further analysis. Normalization of the reads from an RNA-seq experiment must take into account factors such as library size, RNA composition, and sample-

specific biases. Without normalization, comparisons of the expression between samples can easily lead to false positives.

In **paper I**, the quantified transcript counts were aggregated to the gene-level expression. Differential expression analysis (DEA) was then carried out using DESeq2[95], applying standard normalization and variance modeling. Genes with expression levels of <10 mapped reads in more than half the samples were excluded. Replicates for each clone were aggregated, and statistical testing was done using Wald's test comparing cell lines carrying the C or A allele at the risk SNP rs11706832. When performing the DEA, it was targeted within 1 Mb of *LRIG1* and globally, including all available reference genes. For the global analysis, multiple testing was accounted for using the Benjamini–Hochberg (BH) method; genes with an adjusted p value of <0.05 and an absolute log2 fold change of >0.5 were considered to be significantly differentially expressed. Validation on the *IDH*-mutated LGG gene-count data from TCGA was performed using the same preprocessing, but here the expression difference was tested with likelihood-ratio testing across all possible genotypes (A/A, A/C, C/C), in addition to Wald's test.

For the analysis of alternative splicing, aligned reads were quantified using stringtie[96], quantifying the read mapping to exon–exon junctions. The expression of specific exon–exon junctions was calculated as the ratio to the total gene expression for genes within 1 Mb of *LRIG1*. The difference in exon–exon junctions was then investigated between cell lines with A and C alleles.

WGS Analysis, Variant Calling, Annotation, and Burden Testing (Paper III)

In a typical WGS pipeline, once reads from WGS are mapped to the reference genome, they are further marked for duplicate reads (identifying PCR duplicates) and undergo base quality recalibration. Accurate preprocessing is essential, as sequencing artifacts can introduce biases in the downstream analysis.

In **Paper III**, WGS data from northern Sweden glioma cases were aligned to the GRCh38 reference genome using bwa[97], and variant calling for single nucleotide variants (SNVs) and indels was performed with the Genome Analysis Toolkit (GATK) HaplotypeCaller, following the GATK best practices workflow[98]. Joint genotyping was used across all samples, and variants were filtered using GATK Variant Quality Score Recalibration. Only the variants that passed the quality filters were retained for downstream assessment.

Following variant calling, variant annotation is a crucial step in order to interpret the biological—or even clinical—implications of each variant. Any individual will differ from the reference genome at millions of positions, with the large majority of these variants being normal genetic variations. To identify variants with a possible clinical impact, variants are often annotated according to their frequency of appearance in the general population (as common variants are unlikely to be deleterious for an individual). Variants are then further annotated based on whether they are coding or not and on the predicted consequence of a coding variant (e.g., synonymous, missense, frameshift, etc.). Many tools and databases are now available for performing *in-silico* prediction of deleteriousness on variants, so this information can also be used in the annotation step.

In **Paper III**, all SNVs and indels were annotated using the Ensembl Variant Effect Predictor (VEP) tool[99], supplementing the annotation with publicly available data from AlphaMissense[100], Zoonomia[101] constraint scores, SpliceAI[102] predictions, and Swedish population allele frequencies from SweGen[103], Combined Annotation-Dependent Depletion (CADD)[104], and ClinVar[105]. In addition to calling SNVs and indels, larger structural variants (SVs) were called using Delly[106] and manta[107]. Only the SVs identified by both callers were retained; these were then further harmonized and genotyped across all samples using SURVIVOR[108]. The SVs were annotated using VEP, and only those predicted to have a high impact on any of the analyzed genes were considered for downstream analysis. For the tumor-normal WGS pairs, somatic SNVs and indels were called using GATK Mutect2 and Strelka2[109], with only consensus variants between the two callers retained. Somatic copy number alterations were called with CNVkit[110] and structural variants with Manta.

Following variant calling and annotation, germline variants were filtered based on their rarity in the Swedish population (allele frequency (AF) < 0.001 in SweGen) and whether their CADD score was >20 (1% top deleterious SNVs in the genome). Variants were further filtered using four different strategies: (1) all coding variants; (2) ClinVar-classified pathogenic or likely pathogenic (P/LP) variants; (3) predicted deleterious variants (based on loss of function (LoF) prediction, CADD score > 30, or a positive spliceAI prediction); and (4) putatively regulatory variants, defined here as rare variants that overlap predicted regulatory elements.

Gene-level burden tests were then performed using the one-sided Fisher's exact test, comparing rare variant frequency (using one of the

four filtering strategies) with the frequency in healthy individuals from Västerbotten County, denoted herein as "ACpop." Genes with $p < 0.05$ were considered significant hits, and genes with $p < 0.1$ were still retained for validation. Validation of the findings from gene-wise burden tests in the northern Sweden glioma patients was performed by downloading processed WES calls for individuals with glioma in TCGA[111]. Burden tests were performed using the same filtering strategies. Further validation was then performed using a matched case-control burden test on individuals from the UK Biobank WES data, using REGENIE[112] and adjusting for age, sex, and 10 principal components based on the genetic data.

Genome-Wide Association Study and Meta-Analysis (Paper II)

A standard GWAS workflow typically begins using data from SNP arrays (as in **Paper II**). Once raw intensity data are generated, the genotype status on hundreds of thousands to millions of SNPs is inferred through software. This step is followed by quality-control (QC) procedures, at both the individual level and the SNP level. QC steps involve removing samples with excessive incompletion, sex mismatches, and/or evidence of cryptic relatedness between individuals. SNPs may then be filtered based on the call rate, Hardy–Weinberg equilibrium, and minor allele frequencies. Following QC, imputation can be done to impute the genotypical status of SNPs not included in the arrays used. This is possible if robust haplotype information for the study population is available. In this way, the status of additional millions of SNPs can be inferred for each individual and used for downstream analysis.

Association testing is then carried out, typically under an additive genetic model, testing each SNP against the disease status (e.g., pediatric glioma in **Paper II**). For binary outcomes, such as diseased versus healthy, testing is done using logistic regressions adjusted for possible confounders such as sex and population stratification, often inferred by means of principal components analysis (PCA). This gives the effect size and p value for each tested SNP. Often, SNPs with a small genetic distance from each other—that is, in the same haplogroup—have similar effect sizes, as these are often inherited together in LD; this makes it difficult to conclude which SNP at a given genomic locus is functional or underlying for the tested disease.

In collaborative efforts between different universities and countries, as in **Paper II**, raw genotype data often cannot be shared directly for ethical and legal reasons. Nevertheless, increasing the statistical power by including data from several institutions can still be done by utilizing

the summary statistics from several independently performed GWASs in a meta study. A meta-analysis can be performed under a fixed-effect model assuming similar effect sizes for SNPs between studies or a random-effect model assuming different effects of the SNPs. In **Paper II**, we used both a fixed- and a random-effect model on the summary statistics from each GWAS.

In **Paper II**, QC of the genotype data was uniformly performed across all three cohorts using PLINK 1.9[113]. Individuals with more than 2% missing genotypes, sex mismatches, excessive heterozygosity, and cryptic relatedness were filtered out. Outliers were identified and filtered using an iterative process of principal components. The SNP-level QC excluded variants with a minor allele frequency (MAF) < 0.01, over 2% missingness, and deviations from the Hardy–Weinberg equilibrium. Phasing and imputation were performed separately for U.S. and Scandinavian cohorts, due to regulatory differences. The Swedish and Danish cohorts were phased using SHAPEIT4[114] and IMPUTE5[115], using the haplotype reference consortium panel[116]. Only genotypes with MAF > 0.01 and imputation scores > 0.6 were retained.

Association testing was done separately for each cohort using PLINK2[117], adjusting for sex and significant principal components for each analyzed tumor subtype. A meta-analysis was performed on the results from all three cohorts, in two settings: (1) including only cohorts from European descent and (2) including all cases and controls of all descents combined. The meta-analysis was conducted both under a cohort fixed-effect assumption using METAL[118] and assuming random-effect differences across the cohorts using metafor[119].

Colocalization Analysis Using Quantitative Traits (Paper II)

Understanding the biological mechanisms behind GWAS hits remains a major challenge. Typically, a GWAS may result in several loci that are associated with a trait, but no further biological context is given; due to LD, it is often not possible to determine which SNP in a locus is causative to the trait. One way to complement the results from a GWAS is to utilize quantitative traits. Data on the association between a given SNP and the quantitative traits of nearby genes are now publicly available for various human tissues. One example of such a quantitative trait is the expression of nearby genes associated with the SNP status (eQTL). This data can be utilized to add one layer of interpretation to the GWAS results, by investigating whether the SNPs associated with disease also colocalize with the SNPs that regulate gene expression. In doing so, the probability of a shared causal SNP between the disease risk and gene regulation can be inferred.

In **paper II**, a colocalization analysis at the *CDKN2B-AS1* risk locus obtained from the meta-analysis was performed. Associations between SNP expression for the genes *CDKN2B* and *CDKN2A* were downloaded for 13 brain tissues (12 from GTEx and one from the CommonMind[120] dorsolateral prefrontal cortex), as well as for blood (from the eQTL catalogue)[121]. Splice-QTL (sQTL) data were also downloaded from GTEx v8 to investigate alternative splicing events at nearby genes in the locus. Bayesian inference using the coloc package[122] was applied to calculate the probability of colocalization between traits. A posterior probability of ≥0.8 was considered significant. We also compared our results from the meta-analysis with the summary statistics from adult glioma by performing an association analysis on cases and controls from the GICC study. The association results from the GICC cohort were then compared with the summary statistics from the cross-ancestry meta-analysis for pediatric astrocytoma.

Transcriptome-Wide Association Study (Paper II)

Another way to utilize quantitative traits such as eQTLs is to test whether genetically inferred gene expression can be associated with a disease. This approach is known as a transcriptome-wide association study (TWAS). Even though gene expression is not directly measured in the study individuals, it can be estimated with publicly available eQTL data. One way to conceptualize a TWAS is to think of it as imputing the expression levels, just as SNPs not directly measured can be imputed from public haplotype data.

In **paper II**, we applied a TWAS to investigate whether predicted gene expression in brain tissues can be associated with pediatric glioma risk. Here, we downloaded and used pre-trained SNP-expression models based on 13 brain tissues from GTEx v8[123], available via PredictDB, and conducted the TWAS on summary statistics from the meta-GWAS using S-PrediXcan[124] individually on all 13 brain tissues. The results from the individual-run TWAS were integrated into a single model using S-MultiXcan[125] to account for shared expression similarities across the tested brain tissues. Genes with adjusted p values < 0.05 were considered to be significantly associated with pediatric glioma risk.

Accounting for the Testing of Multiple Hypotheses (Papers I–III)

All the papers in this thesis were explorative to some extent. In Paper I, we investigated all genes in the genome for differential expression, as is typical for a DEA; in Paper II, we agnostically tested the entire genome for SNPs associated with glioma; finally, in Paper III, we attempted to

find novel associations of rare variants across hundreds of genes that could be associated with glioma. In short, thousands to millions of statistical tests were performed in each study. Having a *p* value cutoff—that is, only considering tests with $p < 0.05$—meant that, out of 100 statistical tests, we could expect five to be "significant" purely by chance (i.e., false positives). This false discovery rate (FDR) is accounted for differently in each paper.

In **Paper I**, we adjusted the raw *p* values from the DEA using the BH method, in which each *p* value is directly recalculated in an attempt to control the FDR among the significant results. In **Paper II**, we set a static cutoff for genome-wide significance ($p < 5e\text{-}8$). Considering the large number of tests carried out in a GWAS, this is a sensible approach, since *p*-value Bonferroni corrections are typically too conservative, not accounting for the LD between closely located SNPs[126]. Applying such correction methods thus implies a high risk of filtering out true associations. In addition, the *p* values were further controlled in **Paper II** using methods such as quantile-quantile plots (qq plots) analyzing the *p*-value inflation to obtain a general overview of the risk of false positives. **Paper III** utilized qq plots and *p*-value inflation; however, no further correction on the *p* values was applied.

Gene Set Enrichment Analysis (Papers I and II)

Another problem with explorative analyses—where a number of genes may be associated with a trait, for example—is the biological interpretability of these results. As stated before, some of these traits will even be false positives. One way to overcome this issue and place the aggregated results of multiple significant genes within a biological context is to perform a gene set enrichment analysis (GSEA). Genes involved in well-known biological processes and ontologies have been curated over the years, and the key genes in many biological processes are known. A GSEA identifies which genes known to be involved in certain biological contexts are over-represented in a gene set (e.g., from a DEA). In this way, a GSEA can shed some light on the biological relevance of the findings from an explorative study. GSEAs were performed in **Papers I** and **II**.

Study Subjects and Cohorts

The studies reported in **Papers II** and **III** relied heavily on real patient samples and data derived from biobanks, registries, and consortia. This section provides an overview of the most important resources used in this thesis.

The Swedish Pediatric Glioma Cohort (Paper II)

For the Swedish part of the meta-GWAS in **Paper II**, we used a national cohort of children diagnosed with glioma between 1976 and 2004, identified through the Swedish Cancer Registry. Dried neonatal blood spot samples were collected from the Swedish Phenylketonuria Screening Registry, and histopathological diagnoses were gathered from the Swedish Pediatric Brain Tumor Quality Registry. A total of 352 cases diagnosed within ≤15 years were included in **Paper II**. A total of 874 Swedish control subjects previously used in the GWAS of adult gliomas were used as the controls for this study[67].

To validate our findings in **Paper II**, we used WGS data derived from blood from the Children's Brain Tumor Network (CBTN)[127] and the Gabriella Miller Kids First study (BASIC3).

Uppsala–Umeå Comprehensive Cancer Consortium (Paper III)

Many of the adult glioma samples used in **Paper III** were obtained through U-CAN[91], a Swedish project for the collection of biological material (e.g., blood and tumor samples) from patients. Many of the samples from northern Sweden were collected as part of the U-CAN study from patients admitted through Umeå University Hospital. We also used data from several patients admitted through Uppsala University Hospital. Within U-CAN, blood and tumor samples from glioma patients have been gathered since 2010; at Umeå University Hospital, local samples have been gathered since 2005.

Two main and partly overlapping patient groups were retrieved from U-CAN for **Paper III**: an early-onset set of patients (n = 46) diagnosed between the ages of 18 and 40, and a matched tumor-blood set (n = 73), consisting of patients for which sufficient tumor material was available for tumor WGS. Six patients were included in both groups, and their material was sequenced on two occasions.

In glioma cases with tumor tissue available, classification following the 2016 WHO CNS tumor guidelines was performed using histopathology and then harmonized with the 2021 CNS5 standard, where possible, when the *IDH* status was available. For instance, three cases initially classified as *IDH*-mutant glioblastoma under CNS4 were reclassified as astrocytoma grade 4. For validation, we also obtained preprocessed germline WGS data for an additional series of 105 glioblastoma patients from Uppsala that were part of U-CAN.

External Public Data and Databases

This section describes the most notable sources of publicly available data used in this thesis.

The Cancer Genome Atlas (Papers I and III)

TCGA is a resource for cancer researchers that contains molecular and clinical data for multiple cancer types[128]. For **paper I**, gene count data from RNA-seq for 325 IDH-mutated LGGs were downloaded, along with germline genotypes of the same individuals. For **paper III**, WES-derived variant calls[111] for 806 glioma patients were downloaded.

The Haplotype Reference Consortium (Paper II)

The Haplotype Reference Consortium[116] contains haplotypes of 64,976 individuals of predominantly European genetic ancestry. In **Paper II**, a subset of 11,227 individuals was used for SNP imputation for the Swedish individuals.

The GTEx Consortium (Papers I and II)

The Genotype-Tissue Expression (GTEx) consortium[129] aims to map expression and splicing throughout the entire human transcriptome. Thus far, expression data is available for 49 different human tissues, sampled from 838 individuals. GTEx has also mapped the expression of genes and common SNPs, providing millions of tissue-specific SNP-expression and SNP-splicing associations.

Information on eQTL for the risk SNP in **Paper I** was gathered through GTEx. The eQTL data used for both the colocalization analysis and TWAS in **Paper II** originated from GTEx brain tissue samples.

The Genome Aggregation Database (Paper III)

The Genome Aggregation Database (gnomAD) is an initiative from the Broad Institute that currently contains data on allele frequencies aggregated from hundreds of thousands of individuals subjected to WGS or WES. It is a ubiquitous resource for both genetic research and clinical decisions. In this thesis, gnomAD was used in **Paper III** as a control reference for TCGA variant calls.

SweGen (Paper III)

SweGen[103] contains WGS-derived allele frequencies for a cross-section of 1000 Swedish individuals, making aggregated allele frequencies

available for researchers. In **Paper III**, SweGen was used to filter common variants in the overall Swedish population.

ACpop (Paper III)

ACpop[130] contains aggregated allele frequencies for 300 individuals from Västerbotten County in northern Sweden, recruited from the Västerbotten Intervention Program (VIP). The individuals recruited into ACpop met the criteria of having reached the age of 80 without a cancer diagnosis, making ACpop a valuable resource for studying potential predisposition variants for cancer. In this thesis, ACpop was used in **Paper III** as a healthy reference population in contrast with our northern Sweden glioma cohort.

The UK Biobank (Paper III)

The UK Biobank[131] is a prospective cohort containing disease, health, and genetic data for around 500,000 individuals from the UK, recruited within the age span of 40–69. This large resource allows researchers to perform large-scale epidemiological and genetic studies of disease.

In this thesis, the UK Biobank was used in **Paper III** to perform a matched case-control study, investigating rare variants for individuals diagnosed with glioma. More specifically, WES data for 833 individuals of European Ancestry with a glioma diagnosis were obtained from the UK Biobank, and 8,315 age-matched cancer-free individuals from the Biobank were used as controls.

Results and Discussion

Paper I: Functional Investigation of the *LRIG1* Risk SNP

Background and Study Rationale

To date, several risk SNPs have been found to increase the risk of developing glioma in adults. While GWASs can give some indication of the genes involved for the trait studied, they tell little about the mechanism of the given SNPs. For glioma in adults, the SNP rs11706832 in the second intron of *LRIG1* was of particular interest, as this SNP leads to an increased risk of developing LGG in adults (OR = 1.15)[67]. *LRIG1* is a negative regulator of receptor tyrosine kinases and can act as a tumor suppressor for various tumor entities through this mechanism[132–135]. However, publicly available data shows that the SNP in *LRIG1* is not an eQTL of this risk SNP; rather, the LGG risk SNP rs11706832 affects the expression of the neighboring gene *SLC25A26*. In this study, we aimed to clarify the functional role of rs11706832 by constructing isogenic cell lines with CRISPR-CAS9 and investigating the effects on the transcriptome and metabolome in these cell lines.

Results

To investigate the functional effect of the LGG risk SNP rs11706832, we employed CRISPR-Cas9 genome editing. gRNAs targeting the genomic region around rs11706832 were designed. HEK293T cells were transiently transfected with plasmids carrying the gRNAs, cas9, and a single-stranded oligodeoxynucleotide carrying the genetic modification C to A (HEK293T cells carry the minor risk allele C). The transfected cells were cloned, and the status at the *LRIG1* risk loci was determined through Sanger sequencing. Four successful knock-in clones (non-risk allele A), and four negative knock-ins (risk allele C) were isolated and used in the study. Four replicates of each cell line were cultivated separately, resulting in a total of 32 samples used in the downstream analysis.

A first check of differential expression between the HEK293T cell lines was done using real-time quantitative polymerase chain reaction (RT-qPCR) on *LRIG1* and the neighboring *SLC25A26* gene, revealing no differential expression of either gene. These results were consistent with the RNA-seq data, where none of 12 genes in a region of 5×10^6 bp around the SNP were shown to be differentially expressed. Furthermore,

no splicing effect was seen from the rs11706832 risk allele C carriers in *LRIG1* or any of the genes in proximity. It was thus observed that there is no direct gene regulation effect on nearby genes resulting in differential expression and no splicing stemming from the risk SNP.

Although differential expression for *SLC25A26* was not significant (p = 0.19), we carried on with the hypothesis that possible downstream effects from a regulation of *SLC25A26* could be noted in our cell lines. As SLC25A26 is a transporter protein carrying the methyl-group-donor S-adenosylmethionine into the mitochondria, we hypothesized that regulation of *SLC25A26* could affect the methylation of mitochondrial DNA and be noted in the expression of mitochondrial mRNA. Out of the 13 mitochondrial genes detected in our data, 12 were shown to be differentially expressed between our cell lines, all with increased expression for the cell lines with the risk allele (Figure 4a). It was also noted that the expression of mitochondrial genes was reversely

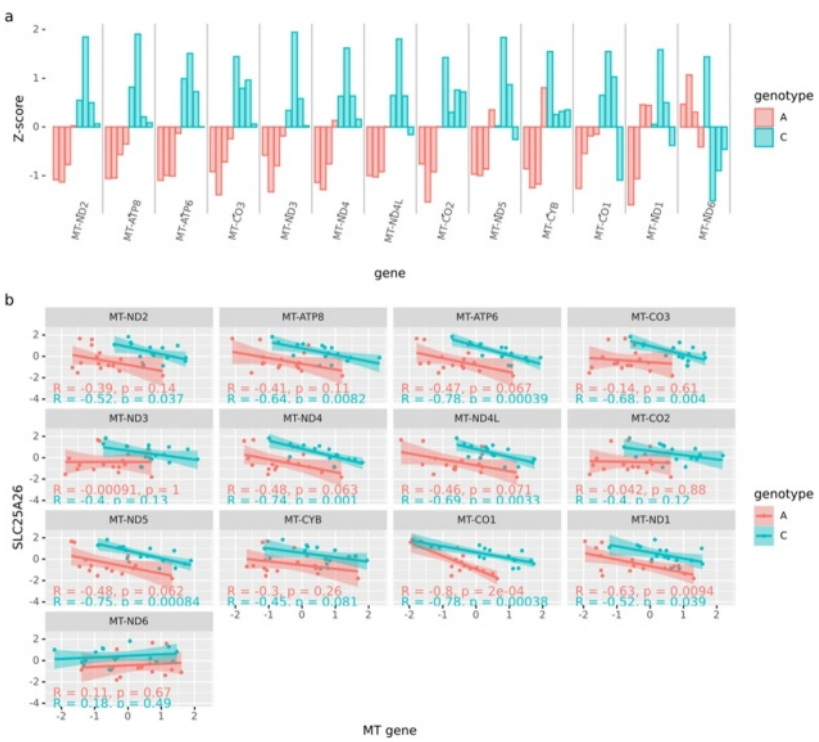

Figure 4: (a) Differential expression of mitochondrial genes between risk (C) and non-risk allele (A) in cell lines. (b) Correlation between SLC25A26 *and the expression of mitochondrial RNA. Figure reproduced from Rosenbaum et al., 2023, scientific reports, CC BY 4.0. No changes made*

associated with the expression of *SLC25A26*, which was in line with the hypothesis that *SLC25A26* regulates the mitochondrial genome through DNA methylation[136,137] (Figure 4b).

A genome-wide gene-expression analysis revealed 74 differentially expressed genes between the cell lines. Gene set enrichment on these genes showed that they were involved in a number of biological processes related to the innate immune response (*TRIM17, TLR3, OAS1, CXCL16, SP100, TRIM34, TRIM14, B2M,* and *IFITM1*) and type I interferon response (*IFITM1, SP100, OAS1, IFI35, IFIT3, IFIT5,* and *IFIT2)* being differentially expressed in the modified HEK293T cells.

To validate our findings from the cell lines, we downloaded data from 325 *IDH1*-mutated LGGs from TCGA. In this data, individuals homozygous to the risk allele (C/C) had a higher expression of *SLC25A26,* as seen in GTEx, while no differential expression of *LRIG1* was noted. A genome-wide DEA of the TCGA samples, conditioned on the status of the *LRIG1* risk SNP for these individuals, revealed 40 differentially expressed genes. Among these genes, *OAS1* was also found

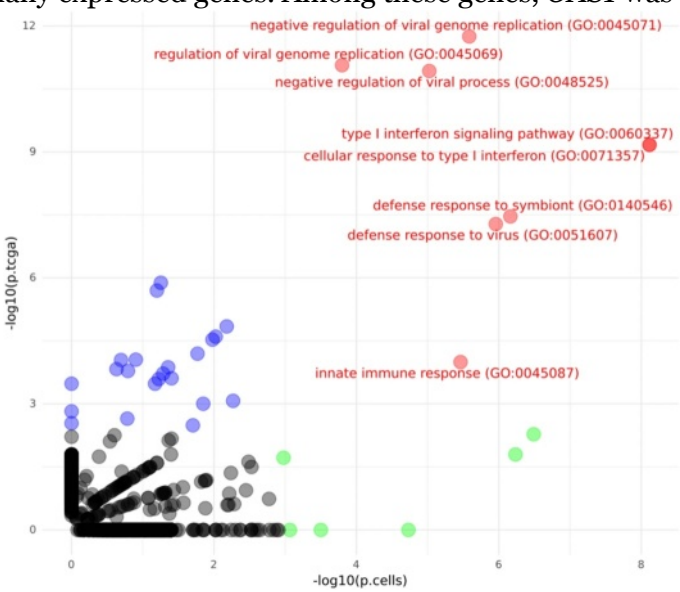

Figure 5: Gene set enrichment p *values for differentially expressed genes in TCGA* (y*-axis) and our cell lines* (x*-axis). Ontology terms in red are significant in both the cell lines and TCGA data, demonstrating that the same biological pathways are conditioned by the risk SNP in both the cell lines and real tumors. Figure reproduced from Rosenbaum et al., 2023, scientific reports, CC BY 4.0. No changes made*

to be differentially expressed in our cell lines. Gene set enrichment of TCGA significant genes showed that the same terms of biological processes related to the innate immune response were enriched (Figure 5). Furthermore, an analysis of 53 of the genes involved in interferon type I response, as seen in the cell lines, revealed that 13 of these genes were significantly different after correcting for multiple testing in TCGA. These genes were expressed at lower levels for individuals with the C/C alleles in the risk loci, compared with A/A individuals.

Using GC-MS on the cell lines to characterize soluble intracellular metabolites, we assessed the levels of 111 metabolites. None of these metabolites were found at different levels in our cell lines after correction for multiple testing.

Discussion

Haploinsufficiency of *LRIG1* has been shown to be a potential driver for LGGs by affecting tyrosine kinase inhibition capabilities[133,135,138]. When this locus —which is located in the second intron of *LRIG1*— appeared as a genome-wide significant GWAS locus, it was reasonable to hypothesize that the functionality of this locus involved dysregulation of *LRIG1*. However, the study showed no differential regulation of *LRIG1* in either expression or splicing in HEK293T cell lines. This finding was further confirmed upon examining the public data available from GTEx and TCGA. Thus, this study showed that the increased risk of developing LGGs stemming from this locus is not due to *LRIG1* dysregulation.

The solute carrier family-25 member-26 gene *SLC25A26*, which is adjacent to *LRIG1*, was found by GTEx to be an eQTL of rs11706832, with the expression of *SLC25A26* being positively associated with the risk allele C. Although *SLC25A26* expression was not significantly different in our modified HEK293T cell lines, the observed trend was consistent with the direction suggested by GTEx. An analysis of the splicing of *SLC25A26* in the cell lines showed no effect on the splicing of *SLC25A26*. Taking into account the biological function of *SLC25A26*, the expression of mitochondrial genes was indeed differentially expressed in our cell lines conditioned on the risk allele of rs11706832. We found that the expression of several mitochondrial genes was negatively correlated with the expression of *SLC25A26*, particularly in cell lines with the risk allele, which aligns with previous reports of *SLC25A26* involvement in mitochondrial methylation[136,137].

A genome-wide DEA on the cell lines, conditioned on having or not having the risk allele, highlighted genes involved in the innate immune response, such as the interferon type I induced proteins *IFIT2*, *IFIT3*,

and *IFIT5*, and the 2'-5'-oligoadenylate synthetase 1 protein coded by *OAS1*. Validation in *IDH1*-mutated LGGs in TCGA showed that the only overlapping differentially expressed gene was *OAS1*. However, gene set enrichment highlighted the exact same terms of biological pathways as in our cell lines, suggesting that the risk allele affects pathways related to the innate immune response (Figure 5). While these findings imply that the downregulation of interferon response and potentially a less active innate immune system may be the downstream effects of the SNP, the link between the risk SNP and the innate immune system remains unclear. One hypothesis derived from the results of this study is that changes in mitochondrial expression, mediated through *SLC25A26*, could be related to immune signaling. Supporting this hypothesis, studies have shown that mitochondrial activity is associated with innate immunity[139–141].

A major limitation of this study is its use of a non-glial cell line, HEK293T, as a model. This cell line was used for technical reasons, as we had access to a haploid subclone at the *LRIG1* locus[133]. It is possible that the effects observed in this study are cell-line specific and not representative of the tumor cells in LGGs. However, we observed a strong concordance between GSEA results from our genetically edited HEK293T cell lines and bulk RNA expression data from TCGA (Figure 5). Thus, it is possible that the observed effects may be present *in vivo*, possibly within the tumor microenvironment rather than in the tumor cells themselves.

Paper II: Multi-Ancestry Meta-GWAS of Pediatric Gliomas

Background and Study Rationale

Genetic susceptibility to gliomas stemming from common genetic variations (i.e., SNPs) has been extensively studied in adults. However, adult and pediatric gliomas differ significantly in their molecular composition, origin, and clinical implications[142]. Even though low-grade astrocytoma is one of the most common tumor entities in children, pediatric gliomas are still uncommon; thus, GWAS studies have previously been limited to small sample sizes[143]. We performed a meta-GWAS on data from the United States, Sweden, and Denmark, amounting to a total of 4,069 children with glioma and 8,778 controls.

Results

In this study, we combined GWAS summary statistics from three independent pediatric glioma and control cohorts from the United States, Sweden, and Denmark. The samples and selection criteria from each cohort for the locally performed GWAS were as follows. The U.S. (California) study included 3,150 cases and 3,154 matched controls; DNA extraction was done on neonatal blood spots, genotyped using the Affymetrix Axiom Precision Medicine Diversity Array. The Sweden cohort included 352 pediatric glioma cases and 874 controls (where these controls were previously used in an adult glioma GWAS)[67] (Table 2); DNA was extracted from neonatal blood spots and genotyped using Illumina arrays. The Denmark cohort included 567 glioma cases and 4,750 controls. The cases from all countries were diagnosed with glioma up to 15 years of age.

Regression analyses were performed separately for each country, using a similar methodology. The resulting summary statistics from each GWAS were combined in a meta-analysis. The strongest finding from the meta-analysis involved the 9p21.3 locus within the gene *CDKN2B-AS1*. At this locus, 36 SNPs reached the genome-wide significant threshold ($p < 5\times10^{-8}$) with the strongest association for rs573687, with an odds ratio (OR) of 1.27 (95% confidence interval [CI]: 1.18–1.37, p $= 6.97\times10^{-10}$). The results at this locus also included the known risk SNPs for adult glioma, rs634537 and rs2157719. The SNP mapping to the locus exhibited unidirectional effects for astrocytoma risk across all

Table 2: Overview of patients and controls across all three cohorts used for the meta-GWAS

Sweden cohort	
Cases	352
Controls	874
Denmark cohort	
Cases	567
Controls	4,750
U.S. cohort	
Cases	3,150
Controls	3,154
Total	
Cases	4,069
Controls	8,778

genetic ancestries included in the meta-study, showing a shared cross-ancestry risk effect stemming from this locus (Figure 6).

When the dataset was stratified by WHO grade and tumor type, the *CDKN2B-AS1* association was particularly seen in pediatric LGGs, driven by low-grade astrocytoma. When the dataset was stratified for HGGs and glioblastomas, no SNP at the locus reached genome-wide significance. When the dataset was stratified for patients of European decent (European-Americans from the U.S. cohort, together with the Swedish and Danish cohort), the *CDKN2B-AS1* association retained its significance, and the minor allele frequencies at the locus were

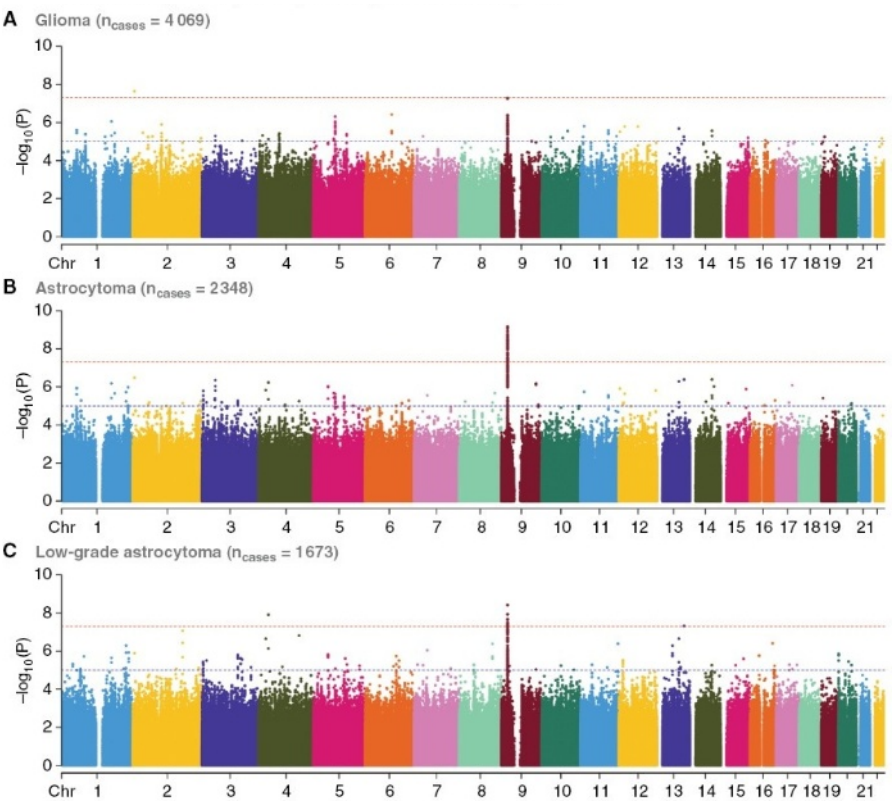

Figure 6: Manhattan plots with chromosome and chromosome position at the x*-axis and the* p *value from the meta-analysis for the tested SNP on the* y*-axis. (a) All gliomas, (b) astrocytomas only, and (c) low-grade astrocytomas. Figure adapted from Foss-Skiftesvik, Li, Rosenbaum et al., 2023, Neuro-Oncology, CC BY-NC 4.0. Changes: cropped*

significantly higher in the European sub-cohorts than in patients of non-European descent from the U.S. cohort.

The *CDKN2B-AS1* association was replicated in a separate case control cohort (270 cases and 2,080 controls), where the associations of seven SNPs at the locus reached $p < 1e\text{-}5$.

To evaluate the potential effects of gene regulation, we conducted a TWAS on the results from the meta-analyses, using SNP-expression models on 13 publicly available human brain tissues. This analysis identified a significant association in only one of 18,628 genes tested—namely, the tumor suppressor *CDKN2B*, localized in the same locus as *CDKN2B-AS1*. More specifically, the resulting TWAS showed that a decreased expression of *CDKN2B* was associated with childhood astrocytoma.

To further investigate the potential functional mechanism underlying the genome-wide significant SNPs at *CDKN2B-AS1,* we performed a colocalization analysis between the SNPs and the eQTLs and sQTLs. The colocalization analysis was performed on SNP-expression data available

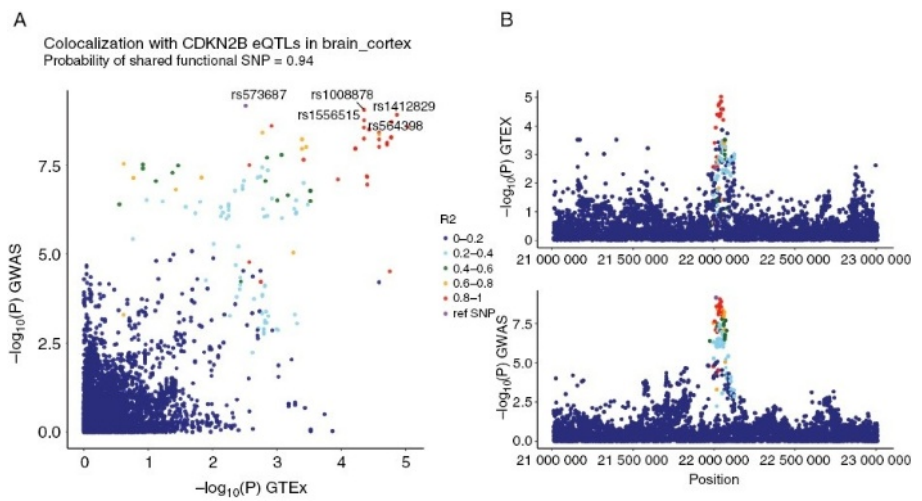

Figure 7: (A) Colocalization of pediatric astrocytoma associations with associations for the expression of CDKN2B. *(B) Regional plots showing the* p *values of eQTL associations (top) and* p *values from the meta-analysis (bottom), with colors indicating linkage disequilibrium to the lead SNP. Figure reproduced from Foss-Skiftesvik, Li, Rosenbaum et al., 2023, Neuro-Oncology, CC BY-NC 4.0.*

for human brain tissues through GTEx. The results showed that the SNPs from our meta-GWAS largely colocalized with SNPs having an effect on the expression of *CDKN2B*, most notably in eQTLs from the cerebral cortex (colocalization probability: 0.94) (Figure 7). No colocalization of our results with *CDKN2A* was found in any brain tissue. We further investigated the colocalization of our results and the splicing of *CDKN2B-AS1* using sQTL data from GTEx, determining a probability of 0.77 of our results colocalizing with the sQTLs for *CDKN2B-AS1* in sQTL data from the pituitary gland.

Discussion

Although only one risk locus was found to be associated with the development of pediatric gliomas (more specifically, low-grade astrocytoma), the results of this paper represent an important advance in our understanding of genetic predisposition to pediatric glioma, highlighting the 9p21.3, *CDKNB2-AS1* locus as the first genome-wide significant risk locus of astrocytoma in children.

From a functional perspective, we observed an association between downregulation of the nearby tumor-suppressor gene *CDKN2B* and the risk of developing low-grade pediatric glioma, as shown through both a TWAS analysis and a colocalization analysis using publicly available eQTL data. This finding suggests that *CDKN2B* is the main affected gene in this locus in terms of overall expression. An additional colocalization analysis of the *CDKN2B-AS1* gene itself suggested that the risk loci also colocalize with splicing of *CDKN2B-AS1*. Interestingly, our results in the risk locus colocalize with GWAS data from adult gliomas, with the distinction that this locus has a stronger association with HGGs, such as glioblastoma, in adult gliomas, although the association is also seen in LGGs[67]. In contrast, the risk locus is not associated with HGGs in children. The *CDKN2B-AS1* gene itself has been implicated in various cancers[144], in which it is suggested to act through diverse mechanisms such as cis-regulation of *CDKN2B* and *CDKN2A*[145] or binding to microRNA[146].

This study further highlights the difference in predisposition between adult and pediatric glioma, where the former has been associated with several common risk loci in the genome, whereas this study only found one associated locus for pediatric glioma. One possible source of this discrepancy is the rarity of pediatric glioma in contrast to adult glioma, resulting in higher sample sizes and statistical power for adult glioma. Still, the overabundance of common risk SNPs for adult glioma compared with pediatric glioma may be due in part to the difference in etiology between the two, where the molecular heterogeneity of adult

gliomas may reflect a broader and more diverse etiology compared with pediatric gliomas.

Paper III: Rare Germline Variants Contributing to Glioma Predisposition

Background and Study Rationale

The etiology underlying glioma development remains largely unexplained. Several studies have investigated genetic glioma predisposition, ranging from GWASs to familial studies of rare variants. This study aimed to investigate the occurrences of rare potentially pathogenic germline variants among a local cohort of 113 individuals, utilizing WGS.

Results

The study subjects included 113 glioma patients recruited through the U-CAN project at the Umeå University Hospital in northern Sweden. The samples comprised two partly overlapping patient groups: patients diagnosed between 18 and 40 years of age (n = 46) and cases with available matched tumor-blood samples (n = 73). The most common diagnosis was glioblastoma (30% in <40 years, and 53% in >40 years). All patients had been subjected to WGS on blood-derived DNA.

The genomic scope of this study was focused to include 651 genes, including well-known CPGs, genes known for glioma predisposition, and commonly somatically mutated genes across all cancers. Variants were called in the 651 genes for all 113 individuals. After an initial filtering, a total of 20 variants in unique individuals (17.6%) in 19 genes had previous records in the ClinVar database describing them as P/LP. The identified variants predominately resided in genes related to DNA repair, including *BRCA1*, *BARD1*, *FANCA*, *RECQL4*, *ERCC2*, and *CHEK2*. Two of the variants also resided in *TP53*. All P/LP variants are shown in Table 3. As a comparison, we used the same filtering on variants from the ACpop project, consisting of WGS-derived allele frequencies from 300 healthy individuals from Västerbotten County in northern Sweden. In ACpop, we found 22 variants classified as P/LP. Comparing the number of alleles with P/LP variants between our patients and ACpop, we found that P/LP variants were more common among our patient group ($p = 4.44 \times 10^{-5}$, Fisher's exact test).

We consolidated our findings in a second Swedish cohort consisting of 105 individuals diagnosed with glioblastoma, where variant calls from

WGS were available from our collaborators in Uppsala. Just as in our cohort from northern Sweden, we found several P/LP variants in genes related to DNA repair, including *FANCA*, *RECQL4*, *ERCC2*, *CHEK2*, and *TP53*.

Table 3: Overview of P/LP variants found in northern Sweden glioma patients

Position (protein)	ClinVar Entry	Gene	Variant type	Diagnosis
chr1:43338634:G:C (p.Arg102Pro)	P	*MPL*	Missense	Glioblastoma (grade 4)
chr2:214781360:C:CT (p.Asp172ArgfsTer10)	P	*BARD1*	Frameshift	Glioblastoma (grade 4)
chr3:14148560:A:G (.)	LP	*XPC*	Splice donor	Oligodendroglioma (grade 2)
chr3:48572138:G:A (p.Arg2338Ter)	P	*COL7A1*	Stop gained	Astrocytoma (grade 4)
chr5:224432:C:T (p.Arg75Ter)	P/LP	*SDHA*	Stop gained	Astrocytoma (grade 2)
chr7:142751919:C:T (p.Arg116Cys)	P/LP	*PRSS1*	Missense	Astrocytoma (grade 2)
chr8:144513412:G:A (p.Gln757Ter)	P	*RECQL4*	Stop gained	Glioblastoma (grade 4)
chr8:144515891:C:T (.)	LP	*RECQL4*	Splice acceptor	Astrocytoma (grade 2)
chr9:97675579:G:A (p.Arg228Ter)	P	*XPA*	Stop gained	Oligodendroglioma (grade 2)
chr10:70598772:C:T (p.Gly317Arg)	P	*PRF1*	Missense	Glioblastoma (grade 4)
chr11:77156022:T:A (p.Ile134Asn)	LP	*MYO7A*	Missense	Astrocytoma (grade 3)
chr15:38351477:CAGAG:C (p.Gly385IlefsTer20)	P/LP	*SPRED1*	Frameshift	Pilocytic astrocytoma (grade 1)
chr16:89815966:T:A (p.Lys34Ter)	P	*FANCA*	Stop gained	Astrocytoma (grade 2)
chr17:7674972:C:T (.)	P	*TP53*	Splice acceptor	Glioblastoma (grade 4)
chr17:7675139:C:T (p.Arg158His)	P/LP	*TP53*	Missense	Glioblastoma (grade 4)
chr17:43093514:C:A (p.Glu673Ter)	P	*BRCA1*	Stop gained	Glioblastoma (grade 4)
chr17:58709926:GT:G (p.Thr259LeufsTer4)	P/LP	*RAD51C*	Frameshift	Glioblastoma (grade 4)
chr19:45352801:C:G (p.Arg616Pro)	P	*ERCC2*	Missense	Astrocytoma (grade 2)
chr22:20993977:G:A (p.Trp469Ter)	P	*LZTR1*	Stop gained	Glioblastoma (grade 4)
chr22:28725338:T:C (p.Arg117Gly)	P/LP	*CHEK2*	Missense_	Glioblastoma (grade 4)

In addition to directly filtering rare variants for known pathogenicity against ClinVar, we wanted to find potentially novel genes whose rare variants could be associated with glioma predisposition. Here, we employed gene-wise variant burden tests, comparing—per gene—the number of rare coding variants found in our patient cohort with the variants found in ACpop under the same filtering criteria. As a first step,

we filtered the variants to include only rare coding variants and found that six genes out of the 651 tested (i.e., *TP53*, *ALK*, *CREBBP*, *GAS7*, *KMT2A*, and *POU5F1*) had significantly more rare coding variants compared with the reference population. When we applied stricter filtering criteria, only including variants with potentially high impact, we identified significant enrichment for *SMAD4*. We further found the genes *PREX2*, *CACNA1D*, *DNMT3A*, *PRSS1*, and *TP53* to have some enrichment of high impact variants, albeit not significant ($0.05 < p < 0.1$).

To validate the findings from the burden analysis, we downloaded variant calls from WES data from TCGA (n = 806 gliomas) and performed the same gene-wise burden tests against the allele frequencies in gnomAD. We found a significant enrichment of rare coding variants in *TP53*, *CREBBP*, *GAS7*, *NBN*, *STRN*, and *DNMT3A*. To further consolidate our findings, we performed a large-scale association analysis on 833 glioma cases from the UK Biobank on matched controls. The only significant association from the UK Biobank was for *TP53*.

Discussion

By applying WGS to a regional cohort of patients with glioma of sporadic origin, this study found that the occurrences of rare variants further explain the genetic predisposition of glioma in adults. More specifically, we demonstrated how rare variants were the probable cause of a significant portion of the gliomas in the patients included in this study, with 17.6% having P/LP variants. Although several other studies have highlighted the association of rare deleterious variants with gliomas, many of these have mainly focused on gliomas of known familial origin[61,63,65,66]. This study further expands the current understanding of genetic predisposition by focusing on sporadic gliomas, showing that the presence of rare variants in our cohort greatly exceeded the number of rare variants in a healthy reference population.

It is notable that many of the P/LP variants were found to reside in genes involved in DNA repair, including genes such as *BRCA1*, *BARD1*, *FANCA*, *RECQL4*, and *CHEK2*—a pattern that was validated in another cohort of unrelated Swedish patients: In an other set of 105 glioblastoma patients from Uppsala, we found that 13 individuals had P/LP variants in any of the tested genes, including variants in SDHA, RECQL4, FANCA, TP53, ERCC2, CHEK2, and MPL. This finding suggests that a high proportion of sporadic gliomas can be attributed to rare deleterious variants in genes related to DNA repair.

This study also provided a more exploratory analysis looking at the gene-wise burden of rare variants. The analysis took into account variants that were not previously recorded as P/LP but had the potential to affect the gene in which they reside based on rarity and other *in silico* measures of deleteriousness. A higher frequency of such variants was found in a total of five genes, including *TP53*. When we validated these findings in two independent cohorts, TCGA and the UK Biobank, only *TP53* showed consistent enrichment across all three datasets. *DNMT3A* and *CREBBP* showed an increased burden of rare variants in TCGA, but this was not replicated in the UK Biobank. Of these two validation datasets, UK Biobank provides the most robust data, given that it is harmonized between the cases and controls, and furthermore relies on a larger number of cases and better matched controls. Since these two genes do not validate in UK Biobank, their role in glioma predisposition remains unclear.

No apparent difference in somatic mutations spectra was seen between carriers and non-carriers of rare germline P/LP variants in any CPGs. This result suggests that the mechanism of predisposition does not further dictate the molecular state of the tumors.

Conclusion

The principal aim of this thesis was to broaden the current understanding of genetic predisposition to glioma. This was done through a multifaceted approach across the three papers presented in this thesis, highlighting that the genetic predisposition of a heterogeneous disease such as glioma requires a wide range of studies. Glioma development in adults and children can still be considered largely sporadic, despite the findings in this thesis and those of earlier studies. This work, however, rests on the premise that even modest germline predisposition can illuminate fundamental mechanisms of tumor development when carefully examined, broadening our understanding of glioma etiology and hopefully providing a basis for future clinical insights.

To recapitulate, the specific aims of this thesis were as follows:

1. **To functionally characterize an LGG risk SNP in the LRIG1 locus (Paper I);**

2. **To identify the contribution of common genetic variation to pediatric glioma risk (Paper II);**

3. **To investigate the contribution of rare germline variants in patients with sporadic glioma (Paper III).**

Although the three aims were investigated through distinct methodological approaches, together they contribute to provide a greater picture of glioma predisposition.

Paper I

This study fulfilled the first aim of the thesis by providing a functional hypothesis for how a common noncoding SNP might influence predisposition to one subtype of glioma.

Paper II

This study fulfilled the second aim of the thesis by providing the first GWAS association with pediatric glioma.

Paper III

This study fulfilled the third aim of the thesis by demonstrating that rare germline variants play a substantial role in glioma susceptibility and can be uncovered even in small sample sizes by utilizing multiple cohorts and integrating publicly available data.

A Unifying Perspective on Glioma Predisposition

Taken together, the findings of this thesis reveal that the genetic predisposition of glioma is polygenic and multifactorial. **Paper I** linked a common noncoding SNP to downstream immunological effects through a functional experiment; **Paper II** identified a risk locus in *CDKN2B-AS1*, colocalizing with lower expression of the tumor suppressor *CDKN2B,* as a driver of pediatric glioma risk; and **paper III** expanded the records of rare variation occurring in adult glioma risk.

This thesis illustrates the methodological diversity required to study cancer predisposition, ranging from functional experiments to exploratory population studies. Each study in this thesis provides a unique layer of information that builds to a broader understanding. While no single genetic variant can largely explain the occurrence of glioma, the results from this work contribute to uncovering the complex issue of glioma predisposition.

Future Directions

Building on previous work, the results from this thesis show that genetic predisposition to glioma is a complex area. This thesis has described the contribution of both rare and common variants and how they contribute to the risk of developing glioma. Looking forward, several promising research directions emerge to further deepen the understanding gained herein, which may guide clinical decisions regarding glioma patients in the future.

One of the clearest paths for future work is the need to refine the functional implications of the risk loci found in Paper II, as well as the implications of the candidate genes found in Paper III. As shown in Paper I, the risk association of germline genetics may not exert effects directly in terms of regulating the nearest gene, for example. The *CDKN2B-AS1* locus that was found to be associated with a higher risk of pediatric glioma requires further investigation. This study implies that the increased risk may come as a result of *CDKN2B* downregulation. However, the analysis was done on publicly available eQTL data from deceased adults (GTEx), and further functional studies are required. Functional experiments like those performed in **Paper I** could be one way to unveil the mechanism of this locus and provide a deeper understanding of the etiology behind pediatric astrocytoma. Likewise, the possible connections of *CREBBP* and *DNMT3A* to glioma risk must be further confirmed with functional studies—possibly through animal models in which these genes are knocked out.

As both common and rare variants have been shown to contribute to the overall risk of developing glioma, the prospect of screening for future cases of glioma holds promise. Although no single association found in this thesis is sufficiently strong to make population-wide genomic screening feasible or beneficial, future approaches may make targeted risk assessment possible for high-risk individuals. Such efforts could integrate common variation, rare variants, and other molecular markers associated with glioma, such as blood metabolite levels. For this to be possible, large-scale population studies using sophisticated machine learning methods may be applied on translational data.

Another research path is how germline factors such as rare variants affect individual prognostics and whether this can be used to guide future treatment options for patients. This path will require well-planned clinical studies. For example, it is known that individuals with a mismatch repair deficiency are less likely to respond to temozolomide but may benefit from newer immunological treatments[147]. It could be

hypothesized that individuals with germline variants in genes involved in the DNA repair pathway may respond differently to treatment. The accumulation of more WGS data in projects such as the UK Biobank, coupled with clinical data, will make it possible to answer such questions.

Ultimately, the extent to which the findings of this thesis—and similar research in the field—can be translated into real clinical benefit for patients will depend on the integration of these accumulated insights on genetic predisposition with other layers of data. Incorporating additional risk markers, such as those from metabolomics, may help to create a more complete picture of glioma susceptibility. Such an integrated approach might be used for risk assessment and prognostics, and could potentially guide more individualized treatment strategies for glioma patients in the future.

Acknowledgements

This work was supported by grants from **Umeå University Hospital**, **the Swedish Research Council**, **the Swedish Cancer Society**, **the Swedish Childhood Cancer Fund**, **and the Cancer Research Foundation in Northern Sweden**.

To all the study participants and patients who generously contributed samples and data to this work—whether in Sweden, Denmark, the United States, or through international collaborations—my deepest gratitude. **To all of you I have not met** but who have nonetheless shaped and enabled every page of this thesis—I am truly humbled by your contribution.

To my supervisor, **Anna Dahlin**: You are a wonderful person and researcher; I could not have asked for a better mentor on this journey. You always seem to have the answers to everything. Thank you.

To my co-supervisor, **Beatrice Melin**: Your encouragement and generosity have been—and will continue to be—a great source of inspiration to me.

To my co-supervisor, **Carl Wibom**: Thank you for your insightful discussions, guidance, and support.

To all the members of my research group: **Barbro**, **Benny**, **Bishnupriya**, **Ida Maria**, **Matilda**, **Mahnaz**, **Mikael**, **Minh**, **Sebastian**, and **Wendy**. You are a group of brilliant people; thank you for all you have taught me.

To, my fellow PhD students, and everyone at the Oncology lab, and oncology journal club: Thank you.

To **Irina**, **Jessica**, and **Maria**, for helping me review this thesis, thank you.

To **Tufve**, **Monica**, **Idah**, **Maria**, and everyone else who has supported me during this PhD: Thank you.

To my colleagues **Elin**, **Eva**, **Katja**, **Linda**, **Magnus**, **Moa**, **Niklas**, **Pär**, **Soumyadeep**, **Tomas,** and to all of Klinisk Genetik at Norrlands Universitetssjukhus: I truly enjoy working with you.

To all my colleagues, collaborators, and co-authors: Thank you for your invaluable contributions, ideas, and support.

To my wife, **Patricia Velasco**: Thank you for your unwavering support, kindness, and love. You have been my anchor and greatest source of encouragement. This thesis could not have been written without you, and I dedicate every step of this journey to you.

To my daughter, **Sigrid Rosenbaum Velasco**: If you stumble upon a dusty copy of this thesis in the future, I don't expect you to read it. I only hope you read this: I love you more than words can say.

To my father, **Peter Nyberg**, and his wife, **Ingrid Larsson**: Thank you for your love, guidance, and support—and for always being there when I needed you.

To my mother, **Mari Rosenbaum**: Thank you for your love and support, which have meant more than I can express.

To my big brother, **William Rosenbaum,** my sister-in-law, **Emilie Thelander,** and my nieces **Noel** and **Viggo**: Thank you for your love and support, and for always bringing joy and perspective into my life.

To my little brother, **Fabian Rosenbaum**: I look up to you immensely, even though you're the youngest.

To my grandmother, **Astrid Rosenbaum**: Your support has been a constant presence in my life.

A mis suegros**, Eugenia Gutiérrez y Patricio Velasco**: Gracias por recibirme siempre con los brazos abiertos y por hacerme sentir parte de la familia desde el primer día.

A mi cuñada, **Graciela Velasco**: Tu generosidad y cercania me hacen sentir que la vida me regaló una hermana.

A mi sobrino, **Francisco Velasco**: Para mi no eres solo un sobrino, eres también uno de mis mejores amigos.

A mi cuñado, **Gonzalo Velasco** y a su familia, **Josefa**, **Teodoro** y **Fhai**: Gracias por su energia y sentido del humor y las risas compartidas que siempre me han hecho sentir como en casa.

And to all my colleagues, family, and friends not mentioned here: Thank you.

References

1. Ostrom QT, Price M, Neff C, et al. CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2016-2020. *Neuro Oncol*. 2023;25(12 Suppl 2):iv1-iv99.

2. Ostrom QT, de Blank PM, Kruchko C, et al. Alex's Lemonade Stand Foundation Infant and Childhood Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2007-2011. *Neuro Oncol*. 2015;16 Suppl 10(Suppl 10):x1-x36.

3. Pollack IF, Agnihotri S, Broniscer A. Childhood brain tumors: current management, biological insights, and future directions. *J Neurosurg Pediatr*. 2019;23(3):261-273.

4. Ohgaki H, Kleihues P. Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *J Neuropathol Exp Neurol*. 2005;64(6):479-489.

5. Vienne-Jumeau A, Tafani C, Ricard D. Environmental risk factors of primary brain tumors: A review. *Rev Neurol (Paris)*. 2019;175(10):664-678.

6. Scheurer ME, Etzel CJ, Liu M, et al. Familial aggregation of glioma: a pooled analysis. *Am J Epidemiol*. 2010;172(10):1099-1107.

7. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352(10):987-996.

8. Tutt ANJ, Garber JE, Kaufman B, et al. Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer. *N Engl J Med*. 2021;384(25):2394-2405.

9. Collins KL, Pollack IF. Pediatric low-grade gliomas. *Cancers (Basel)*. 2020;12(5):1152.

10. Stupp R, Hegi ME, Mason WP, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol*. 2009;10(5):459-466.

11. Guo X, Yang X, Wu J, et al. Tumor-treating fields in glioblastomas: Past, present, and future. *Cancers (Basel)*. 2022;14(15):3669.

12. Voloshin T, Schneiderman RS, Volodin A, et al. Tumor Treating Fields (TTFields) hinder cancer cell motility through regulation of microtubule and acting dynamics. *Cancers (Basel)*. 2020;12(10):3016.

13. Ballo MT, Conlon P, Lavy-Shahaf G, Kinzel A, Vymazal J, Rulseh AM. Association of Tumor Treating Fields (TTFields) therapy with survival in newly diagnosed glioblastoma: a systematic review and meta-analysis. *J Neurooncol*. 2023;164(1):1-9.

14. Liu Y, Zhou F, Ali H, Lathia JD, Chen P. Immunotherapy for glioblastoma: current state, challenges, and future perspectives. *Cell Mol Immunol*. 2024;21(12):1354-1375.

15. Miller TE, El Farran CA, Couturier CP, et al. Programs, origins and immunomodulatory functions of myeloid cells in glioma. *Nature*. 2025;640(8060):1072-1082.

16. Sharma P, Aaroe A, Liang J, Puduvalli VK. Tumor microenvironment in glioblastoma: Current and emerging concepts. *Neurooncol Adv*. 2023;5(1):vdad009.

17. Abdelfattah N, Kumar P, Wang C, et al. Single-cell analysis of human glioma and immune cells identifies S100A4 as an immunotherapy target. *Nat Commun*. 2022;13(1):767.

18. Kirschenbaum D, Xie K, Ingelfinger F, et al. Time-resolved single-cell transcriptomics defines immune trajectories in glioblastoma. *Cell*. 2024;187(1):149-165.e23.

19. Darmanis S, Sloan SA, Croote D, et al. Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep*. 2017;21(5):1399-1410.

20. Roka K, Scheinemann K, Avula S, et al. European standard clinical practice recommendations for primary pediatric low-grade gliomas. *EJC Paediatr Oncol*. 2024;4(100169):100169.

21. Gnekow AK, Walker DA, Kandels D, et al. A European randomised controlled trial of the addition of etoposide to standard vincristine and carboplatin induction as part of an 18-month treatment programme for childhood (≤16 years) low grade glioma - A final report. *Eur J Cancer*. 2017;81:206-225.

22. Stokland T, Liu JF, Ironside JW, et al. A multivariate analysis of factors determining tumor progression in childhood low-grade glioma: a population-based cohort study (CCLG CNS9702). *Neuro Oncol*. 2010;12(12):1257-1268.

23. Wisoff JH, Sanford RA, Heier LA, et al. Primary neurosurgery for pediatric low-grade gliomas: a prospective multi-institutional study from the Children's Oncology Group. *Neurosurgery*. 2011;68(6):1548-1554; discussion 1554-5.

24. Armstrong GT, Conklin HM, Huang S, et al. Survival and long-term health and cognitive outcomes after low-grade glioma. *Neuro Oncol*. 2011;13(2):223-234.

25. de Bont JM, Schouten-van Meeteren AYN. Long-term quality of survival after pediatric low-grade glioma. *Childs Nerv Syst*. 2024;40(10):3341-3355.

26. Nationellt vårdprogram tumörer i hjärna, ryggmärg och dess hinnor. Accessed August 15, 2025. https://kunskapsbanken.cancercentrum.se/diagnoser/hjarna/vardprogram/

27. International Agency for Research on Cancer. *WHO Classification of Tumours Editorial Board. World Health Organization Classification of Tumours of the Central Nervous System. 5th Ed.Lyon*. International Agency for Reasearch on Cancer; 2021.

28. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol*. 2021;23(8):1231-1251.

29. Torp SH, Solheim O, Skjulsvik AJ. The WHO 2021 Classification of Central Nervous System tumours: a practical update on what neurosurgeons need to know-a minireview. *Acta Neurochir (Wien)*. 2022;164(9):2453-2464.

30. Gritsch S, Batchelor TT, Gonzalez Castro LN. Diagnostic, therapeutic, and prognostic implications of the 2021 World Health Organization classification of tumors of the central nervous system. *Cancer*. 2022;128(1):47-58.

31. Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*. 2009;462(7274):739-744.

32. Lu C, Ward PS, Kapoor GS, et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*. 2012;483(7390):474-478.

33. Liu XY, Gerges N, Korshunov A, et al. Frequent ATRX mutations and loss of expression in adult diffuse astrocytic tumors carrying IDH1/IDH2 and TP53 mutations. *Acta Neuropathol*. 2012;124(5):615-625.

34. Abedalthagafi M, Phillips JJ, Kim GE, et al. The alternative lengthening of telomere phenotype is significantly associated with loss of ATRX expression in high-grade pediatric and adult astrocytomas: a multi-institutional study of 214 astrocytomas. *Mod Pathol*. 2013;26(11):1425-1432.

35. Bell RJA, Rube HT, Kreig A, et al. Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science*. 2015;348(6238):1036-1039.

36. Brennan CW, Verhaak RGW, McKenna A, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013;155(2):462-477.

37. An Z, Aksoy O, Zheng T, Fan QW, Weiss WA. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. *Oncogene*. 2018;37(12):1561-1575.

38. Hegi ME, Diserens AC, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med*. 2005;352(10):997-1003.

39. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. *Nature*. 2018;555(7697):469-474.

40. Karremann M, Gielen GH, Hoffmann M, et al. Diffuse high-grade gliomas with H3 K27M mutations carry a dismal prognosis independent of tumor location. *Neuro Oncol*. 2018;20(1):123-131.

41. Ron E, Modan B, Boice JD Jr, et al. Tumors of the brain and nervous system after radiotherapy in childhood. *N Engl J Med*. 1988;319(16):1033-1039.

42. Braganza MZ, Kitahara CM, Berrington de González A, Inskip PD, Johnson KJ, Rajaraman P. Ionizing radiation and the risk of brain and central nervous system tumors: a systematic review. *Neuro Oncol*. 2012;14(11):1316-1324.

43. Amirian ES, Armstrong GN, Zhou R, et al. The Glioma International Case-Control study: A report from the Genetic Epidemiology of Glioma International Consortium. *Am J Epidemiol*. 2016;183(2):85-91.

44. Amirian ES, Zhou R, Wrensch MR, et al. 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*. 2016;25(2):282-290.

45. Chen C, Xu T, Chen J, et al. Allergy and risk of glioma: a meta-analysis: Allergy and risk of glioma. *Eur J Neurol*. 2011;18(3):387-395.

46. Linos E, Raine T, Alonso A, Michaud D. Atopy and risk of brain tumors: a meta-analysis. *J Natl Cancer Inst*. 2007;99(20):1544-1550.

47. Schwartzbaum J, Jonsson F, Ahlbom A, et al. Cohort studies of association between self-reported allergic conditions, immune-related diagnoses and glioma and meningioma risk: Allergic Conditions and Glioma and Meninglioma Risk. *Int J Cancer*. 2003;106(3):423-428.

48. Zhao H, Cai W, Su S, Zhi D, Lu J, Liu S. Allergic conditions reduce the risk of glioma: a meta-analysis based on 128,936 subjects. *Tumour Biol*. 2014;35(4):3875-3880.

49. Ostrom QT, Edelson J, Byun J, et al. Partitioned glioma heritability shows subtype-specific enrichment in immune cells. *Neuro Oncol*. 2021;23(8):1304-1314.

50. Björkblom B, Wibom C, Jonsson P, et al. Metabolomic screening of pre-diagnostic serum samples identifies association between α- and γ-tocopherols and glioblastoma risk. *Oncotarget*. 2016;7(24):37043-37053.

51. Huang J, Weinstein SJ, Kitahara CM, Karoly ED, Sampson JN, Albanes D. A prospective study of serum metabolites and glioma risk. *Oncotarget*. 2017;8(41):70366-70377.

52. Jonsson P, Antti H, Späth F, Melin B, Björkblom B. Identification of pre-diagnostic metabolic patterns for glioma using Subset analysis of Matched Repeated Time points. *Cancers (Basel)*. 2020;12(11):3349.

53. Löding S, Andersson U, Kaaks R, et al. Altered plasma metabolite levels can be detected years before a glioma diagnosis. *JCI Insight*. 2023;8(19):e171225.

54. Körber V, Yang J, Barah P, et al. Evolutionary trajectories of IDHWT glioblastomas reveal a common path of early tumorigenesis instigated years ahead of initial diagnosis. *Cancer Cell*. 2019;35(4):692-704.e12.

55. Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science*. 1990;250(4985):1233-1238.

56. Kleihues P, Schäuble B, zur Hausen A, Estève J, Ohgaki H. Tumors associated with p53 germline mutations: a synopsis of 91 families. *Am J Pathol*. 1997;150(1):1-13.

57. McClatchey AI. Neurofibromatosis. *Annu Rev Pathol*. 2007;2(1):191-216.

58. Lobbous M, Bernstock JD, Coffee E, et al. An update on neurofibromatosis type 1-associated gliomas. *Cancers (Basel)*. 2020;12(1):114.

59. Malmer B, Grönberg H, Bergenheim AT, Lenner P, Henriksson R. Familial aggregation of astrocytoma in northern Sweden: an epidemiological cohort study. *Int J Cancer*. 1999;81(3):366-370.

60. Wrensch M, Lee M, Miike R, et al. Familial and personal medical history of cancer and nervous system conditions among adults with glioma and controls. *Am J Epidemiol*. 1997;145(7):581-593.

61. Bainbridge MN, Armstrong GN, Gramatges MM, et al. Germline mutations in shelterin complex genes are associated with familial glioma. *J Natl Cancer Inst*. 2015;107(1):384.

62. Muskens IS, de Smith AJ, Zhang C, et al. Germline cancer predisposition variants and pediatric glioma: a population-based study in California. *Neuro Oncol*. 2020;22(6):864-874.

63. Choi DJ, Armstrong G, Lozzi B, et al. The genomic landscape of familial glioma. *Sci Adv*. 2023;9(17):eade2675.

64. Jovanović A, Tošić N, Marjanović I, et al. Germline variants in cancer predisposition genes in pediatric patients with central nervous system tumors. *Int J Mol Sci*. 2023;24(24):17387.

65. Weber CAM, Krönke N, Volk V, et al. Rare germline variants in POLE and POLD1 encoding the catalytic subunits of DNA polymerases ε and δ in glioma families. *Acta Neuropathol Commun*. 2023;11(1):184.

66. Nurminen R, Afyounian E, Paunu N, et al. Previously reported CCDC26 risk variant and novel germline variants in GALNT13, AR, and MYO10 associated with familial glioma in Finland. *Sci Rep*. 2024;14(1):11562.

67. Melin BS, Barnholtz-Sloan JS, Wrensch MR, et al. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat Genet*. 2017;49(5):789-794.

68. Shete S, Hosking FJ, Robertson LB, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet*. 2009;41(8):899-904.

69. Stacey SN, Sulem P, Jonasdottir A, et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet*. 2011;43(11):1098-1103.

70. Wrensch M, Jenkins RB, Chang JS, et al. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat Genet*. 2009;41(8):905-908.

71. Kinnersley B, Mitchell JS, Gousias K, et al. Quantifying the heritability of glioma using genome-wide complex trait analysis. *Sci Rep*. 2015;5(1):17267.

72. Nakase T, Guerra GA, Ostrom QT, et al. Genome-wide polygenic risk scores predict risk of glioma and molecular subtypes. *Neuro Oncol*. 2024;26(10):1933-1944.

73. Labreche K, Kinnersley B, Berzero G, et al. Diffuse gliomas classified by 1p/19q co-deletion, TERT promoter and IDH mutation status are associated with specific genetic risk loci. *Acta Neuropathol*. 2018;135(5):743-755.

74. Wu WYY, Johansson G, Wibom C, et al. The genetic architecture of gliomagenesis-genetic risk variants linked to specific molecular subtypes. *Cancers (Basel)*. 2019;11(12):2001.

75. Eckel-Passow JE, Drucker KL, Kollmeyer TM, et al. Adult diffuse glioma GWAS by molecular subtype identifies variants in D2HGDH and FAM20C. *Neuro Oncol*. 2020;22(11):1602-1613.

76. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190-1195.

77. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-1330.

78. Yanchus C, Drucker KL, Kollmeyer TM, et al. A noncoding single-nucleotide polymorphism at 8q24 drives IDH1-mutant glioma formation. *Science*. 2022;378(6615):68-78.

79. Baskin R, Woods NT, Mendoza-Fandiño G, Forsyth P, Egan KM, Monteiro ANA. Functional analysis of the 11q23.3 glioma susceptibility locus implicates PHLDB1 and DDX6 in glioma susceptibility. *Sci Rep*. 2015;5(1):17367.

80. Ali MW, Patro CPK, Zhu JJ, et al. A functional variant on 20q13.33 related to glioma risk alters enhancer activity and modulates expression of multiple genes. *Hum Mutat*. 2021;42(1):77-88.

81. Bi J, Mo W, Liu M, et al. Systematic decoding of functional enhancer connectomes and risk variants in human glioma. *Nat Cell Biol*. Published online August 19, 2025:1-10.

82. Rahman N. Realizing the promise of cancer predisposition genes. *Nature*. 2014;505(7483):302-308.

83. Wilcox N, Dumont M, González-Neira A, et al. Exome sequencing identifies breast cancer susceptibility genes and defines the contribution of coding variants to breast cancer risk. *Nat Genet*. 2023;55(9):1435-1439.

84. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-753.

85. Sanson M, Hosking FJ, Shete S, et al. Chromosome 7p11.2 (EGFR) variation influences glioma risk. *Hum Mol Genet*. 2011;20(14):2897-2904.

86. Rajaraman P, Melin BS, Wang Z, et al. Genome-wide association study of glioma and meta-analysis. *Hum Genet*. 2012;131(12):1877-1888.

87. Kinnersley B, Labussière M, Holroyd A, et al. Genome-wide association study identifies multiple susceptibility loci for glioma. *Nat Commun*. 2015;6:8559.

88. Bu L, Hameed NUF, Luo C, et al. Germline ALK variations are associated with a poor prognosis in glioma and IDH-wildtype glioblastoma. *J Neurooncol*. 2021;152(1):27-36.

89. McDonald MF, Prather LL, Helfer CR, et al. Prevalence of pathogenic germline variants in adult-type diffuse glioma. *Neurooncol Pract*. 2023;10(5):482-490.

90. Mateos MK, Ajuyah P, Fuentes-Bolanos N, et al. Germline analysis of an international cohort of pediatric diffuse midline glioma patients. *Neuro Oncol*. Published online March 12, 2025:noaf061.

91. Glimelius B, Melin B, Enblad G, et al. U-CAN: a prospective longitudinal collection of biomaterials and clinical information from adult cancer patients in Sweden. *Acta Oncol*. 2018;57(2):187-194.

92. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014;346(6213):1258096.

93. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-419.

94. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.

95. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.

96. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290-295.

97. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bioGN]*. Published online March 16, 2013. Accessed February 2, 2025. http://arxiv.org/abs/1303.3997

98. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498.

99. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.

100. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381(6664):eadg7492.

101. Christmas MJ, Kaplow IM, Genereux DP, et al. Evolutionary constraint and innovation across hundreds of placental mammals. *Science*. 2023;380(6643):eabn3943.

102. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535-548.e24.

103. Ameur A, Dahlberg J, Olason P, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. *Eur J Hum Genet*. 2017;25(11):1253-1260.

104. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894.

105. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067.

106. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28(18):i333-i339.

107. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220-1222.

108. Jeffares DC, Jolly C, Hoti M, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun*. 2017;8(1):14061.

109. Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591-594.

110. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016;12(4):e1004873.

111. Huang KL, Mashl RJ, Wu Y, et al. Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell*. 2018;173(2):355-370.e14.

112. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021;53(7):1097-1103.

113. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.

114. Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. *Nat Commun*. 2019;10(1):5436.

115. Rubinacci S, Delaneau O, Marchini J. Genotype imputation using the Positional Burrows Wheeler Transform. *PLoS Genet*. 2020;16(11):e1009049.

116. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283.

117. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7.

118. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.

119. Viechtbauer W. Conducting Meta-Analyses inRwith themetaforPackage. *J Stat Softw*. 2010;36(3):1-48.

120. Hoffman GE, Bendl J, Voloudakis G, et al. CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci Data*. 2019;6(1):180.

121. Kerimov N, Hayhurst JD, Peikova K, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet*. 2021;53(9):1290-1299.

122. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383.

123. Barbeira AN, Bonazzola R, Gamazon ER, et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol*. 2021;22(1):49.

124. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun*. 2018;9(1):1825.

125. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet*. 2019;15(1):e1007889.

126. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet*. 2016;24(8):1202-1205.

127. Lilly JV, Rokita JL, Mason JL, et al. The children's brain tumor network (CBTN) - Accelerating research in pediatric central nervous system tumors through collaboration and open science. *Neoplasia*. 2023;35(100846):100846.

128. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113-1120.

129. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)-Analysis Working Group, Statistical Methods groups-Analysis Working Group, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213.

130. Rentoft M, Svensson D, Sjödin A, et al. A geographically matched control population efficiently limits the number of candidate disease-causing variants in an unbiased whole-genome analysis. *PLoS One*. 2019;14(3):e0213350.

131. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.

132. Billing O, Holmgren Y, Nosek D, Hedman H, Hemmingsson O. LRIG1 is a conserved EGFR regulator involved in melanoma

development, survival and treatment resistance. *Oncogene*. 2021;40(21):3707-3718.

133. Faraz M, Herdenberg C, Holmlund C, Henriksson R, Hedman H. A protein interaction network centered on leucine-rich repeats and immunoglobulin-like domains 1 (LRIG1) regulates growth factor receptors. *J Biol Chem*. 2018;293(9):3421-3435.

134. An Y, Zhao Z, Ou P, Wang G. Expression of LRIG1 is associated with good prognosis for human non-small cell lung cancer. *Medicine (Baltimore)*. 2015;94(47):e2081.

135. Mao F, Holmlund C, Faraz M, et al. Lrig1 is a haploinsufficient tumor suppressor gene in malignant glioma. *Oncogenesis*. 2018;7(2):13.

136. Menga A, Palmieri EM, Cianciulli A, et al. SLC25A26 overexpression impairs cell function via mtDNA hypermethylation and rewiring of methyl metabolism. *FEBS J*. 2017;284(6):967-984.

137. Kishita Y, Pajak A, Bolar NA, et al. Intra-mitochondrial methylation deficiency due to mutations in SLC25A26. *Am J Hum Genet*. 2015;97(5):761-768.

138. Gur G, Rubin C, Katz M, et al. LRIG1 restricts growth factor signaling by enhancing receptor ubiquitylation and degradation. *EMBO J*. 2004;23(16):3270-3281.

139. West AP, Shadel GS. Mitochondrial DNA in innate immune responses and inflammatory pathology. *Nat Rev Immunol*. 2017;17(6):363-375.

140. West AP, Shadel GS, Ghosh S. Mitochondria in innate immune responses. *Nat Rev Immunol*. 2011;11(6):389-402.

141. Song Y, Zhou Y, Zhou X. The role of mitophagy in innate immune responses triggered by mitochondrial stress. *Cell Commun Signal*. 2020;18(1):186.

142. Sturm D, Pfister SM, Jones DTW. Pediatric gliomas: Current concepts on diagnosis, biology, and clinical management. *J Clin Oncol*. 2017;35(21):2370-2377.

143. Dahlin AM, Wibom C, Andersson U, et al. Genetic variants in the 9p21.3 locus associated with glioma risk in children, adolescents,

and young adults: A case-control study. *Cancer Epidemiol Biomarkers Prev*. 2019;28(7):1252-1258.

144. Hjazi A, Ghaffar E, Asghar W, et al. CDKN2B-AS1 as a novel therapeutic target in cancer: Mechanism and clinical perspective. *Biochem Pharmacol*. 2023;213:115627.

145. Kotake Y, Nakagawa T, Kitagawa K, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene*. 2011;30(16):1956-1962.

146. Ma ML, Zhang HY, Zhang SY, Yi XL. LncRNA CDKN2B-AS1 sponges miR-28-5p to regulate proliferation and inhibit apoptosis in colorectal cancer. *Oncol Rep*. 2021;46(4). doi:10.3892/or.2021.8164

147. Negm L, Chung J, Nobre L, et al. The landscape of primary mismatch repair deficient gliomas in children, adolescents, and young adults: a multi-cohort study. *Lancet Oncol*. 2025;26(1):123-135.