2018-09-05

Characterizing the Genomic Heterogeneity of Pediatric Glioblastoma

Hoffman, Mary

UNIVERSITY OF CALGARY

Characterizing the Genomic Heterogeneity of Pediatric Glioblastoma

by

Mary Hoffman

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOCHEMISTRY AND MOLECULAR BIOLOGY

CALGARY, ALBERTA

SEPTEMBER, 2018

Abstract

Brain tumors are currently the most common cause of cancer-related deaths among children. With a 5-year survival rate of 20%, pediatric glioblastoma (pGBM) is a lethal brain tumor with no effective treatment options. Although pathologically indistinct from its adult counterpart, recent work has shown that pGBMs diverge at both the genetic and transcriptional level from the adult malignancy. Genomic analyses have identified a recurrent mutation in *H3F3A*, but this lesion is only present in a fraction of patients and has not contributed to the advancement of effective therapies. Using a longitudinal collection of primary and recurrent pGBMs with matched germlines, I have described a molecularly heterogeneous disease with extreme tumoral evolution. Perhaps my most striking finding is the presence of potentially deleterious structural variants in the patient germlines. Together, these findings suggest a novel hereditary component to tumor etiology which has not been previously described in this malignancy.

Preface

This thesis is original, unpublished, independent work by the author, M. Hoffman. The experiments reported in Chapters 3 and 4 were covered by Ethics Certificate number HREBA.CC-16-0823, issued by the Health Research Ethics Board of Alberta – Cancer Committee for the project "Deriving patient-specific cancer stem cells from brain tumor specimens" on August 22, 2017.

Acknowledgements

I have many people to thank for providing me with overwhelming support and encouragement throughout my Master's degree. First, I would like to thank my supervisor Dr. Marco Gallo. I am incredibly grateful to Marco for teaching me to be an independent scientist and to be confident and proud of my data. Marco's work ethic and passion for research has inspired me these past 2 years and will continue to do so throughout my career. Marco has been an incredible mentor and I am so grateful for the opportunities he has given me. I would also like to thank my co-supervisor, Dr. Jennifer Chan for being such a confident, fearless mentor who has accomplished so much in her career and yet continues to pursue her personal passions. Thank you to everyone in the Gallo Lab that has contributed to my thesis; Dr. Ana Nikolic, Dr. Michael Johnston, and Katrina Ellestad. Ana, thank you for never failing to provide helpful advice and for contributing your intelligent input throughout my project. Michael, thank you for never getting frustrated while I was attempting to learn bioinformatics and for always being patient when I had a never-ending list of questions. I would also like to thank my committee members, Dr. Savraj Grewal and Dr. Jennifer Cobb, for lending their expertise and insight to make my Master's degree a success. Thank you to everyone else who contributed to my thesis, Dr. Donna Senger, Jenny King, Ngoc Ha Dang, Dr. Sorana Morrissy, and Aaron Gillmor. I would also like to thank my family for always believing I am the smartest one in the family and for being my biggest fans. Lastly to my husband Reid, thank you for your unconditional love and support in everything I do. You have pushed me to pursue my passions, have been my biggest support system, and I wouldn't be where I am without you.

Table of Contents

List of Figures

List of Abbreviations

Symbol	Definition
AF	allelic frequency
CNS	central nervous system
CNV	copy number variation
COSMIC	Catalogue of Somatic Mutations in Cancer
del	deletion
DIPG	diffuse intrinsic pontine glioma
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
dsDNA	double stranded deoxyribonucleic acid
FU	fluorescent units
GBM	glioblastoma
gDNA	genomic deoxyribonucleic acid
GEM	gel bead in emulsion
GRCh38	Genome Reference Consortium Human Build 38
HMW	high molecular weight
HS	high sensitivity
Kbp	kilobase pair
Mbp	megabase pair
OMIM	Online Mendelian Inheritance in Man database
PDX	patient-derived xenograft
pGBM	pediatric glioblastoma

RasGAP	Ras GTPase activating protein
RNA	ribonucleic acid
scRNA-seq	single-cell RNA-sequencing
SNV	single nucleotide variant
SV	structural variant
TMZ	temozolomide
tSNE	t-distributed stochastic neighbor embedding
WES	whole-exome sequencing
WGS	whole-genome sequencing

Chapter One: **Introduction**

1.1 Pediatric glioblastoma (pGBM)

1.1.1 Clinical Significance

Brain tumors are the most common solid tumor in children, encompassing 25% of all pediatric cancers with an incidence of 5 cases per 100,000 children per year (Central Brain Tumor Registry of the USA (CBTRUS), 2011) (Fleming & Chi, 2012). There are several different types of brain tumors that are classified histologically and are subsequently labelled as either low grade (grade I and II) or high grade (grade III and IV). A particularly aggressive type of childhood brain tumor is pediatric glioblastoma (pGBM). It is a grade IV malignancy primarily found in supratentorial locations and has a 5-year survival rate of just 20% (CBTRUS, 2011). pGBM is seen in all stages of childhood (0-19 years old) from *in utero*, early infancy, and young adulthood; with the highest incidence occurring in patients 15-19 years old with a median age of approximately 9 years old (Hou et al., 2008; Milano et al., 2009; Şeker & Özek, 2006). Symptoms of pGBM include constant headaches, changes in behaviour, nausea/vomiting, and blurred vision (Fangusaro, 2009; Reddy & Wellons, 2003). These clinical manifestations are due to an increase in intracranial pressure (Fangusaro, 2012). Diagnosis in younger children is often more challenging because the symptoms are non-specific and may go unnoticed or be attributed to another common childhood illness. These symptoms includes lethargy, failure to thrive, vomiting, and macrocephaly (Reddy & Wellons, 2003).

Histopathologically, pGBM is very similar to its adult counterpart and is diagnosed based on the presence of brisk mitotic activity, nuclear pleiomorphism, microvascular proliferation, and

necrosis (Baker, Ellison, & Gutmann, 2016). Treatment includes maximal surgical resection, radiation, and chemotherapy (Cohen et al., 2011; Jalali et al., 2015). This treatment has repeatedly failed to improve the clinical outcome of this disease because the rationale was based on success in adult GBM. Chemotherapeutic drugs shown to provide a survival benefit in adults, such as Temozolomide (TMZ) unfortunately provide no benefit to children with pGBM, with a response rate of 11% (Baker et al., 2016; Cohen et al., 2011; Finlay & Zacharoulis, 2005; Jalali et al., 2015; Warren et al., 2012). Following treatment, tumors inevitably reoccur and cause patients to succumb to this disease. Currently there is no effective treatment for this cancer.

1.1.2 Predisposition Factors

Although the majority of pediatric brain tumors are typically thought of as sporadic diseases, there are 2 factors which increase susceptibility of developing a central nervous system (CNS) tumor, including but not limited to pGBM, in a subset of childhood patients. The first is CNS exposure to radiation therapy, and the second is having a specific genetic syndrome known as familial cancer syndromes. These syndromes/genes that increase the risk of developing CNS tumors include: Neurofibromatosis type 1 (*NF1*), Neurofibromatosis Type 2 (*NF2*), tuberous sclerosis (*TSC1* or *TSC2* genes), Li-Fraumeni (*TP53* or *CHEK2*), Nevoid Basal Cell Carcinoma (*PTCH*), Turcot (*APC, MSH2, MSH6, PMS2, MLH1*), Cowden (*PTEN*), hereditary retinoblastoma (*RB1*), and Rubinstein-Taybi (*CREBBP*) (Bourdeaut et al., 2014; Hegde et al., 2005; Hottinger & Khakoo, 2009; K. J. Johnson et al., 2014; Stefanaki, Alexiou, Stefanaki, & Prodromou, 2013; Yu et al., 2009). Of those syndromes, only Neurofibromatosis type 1, Li-Fraumeni, and Turcot Syndrome have been found to be associated with pGBM in a small subset of patients (A. Johnson

et al., 2017; Salloum et al., 2017). Aside from hereditary cancer syndromes, recent studies have proposed that 6% of all childhood cancer patients, including pGBM, have causative germline variants in known cancer-predisposition genes without harbouring any syndrome (Gröbner et al., 2018; J. Zhang et al., 2015).

Although a rare association between certain genetic syndromes and pGBM has been shown, the majority of tumors are still believed to be sporadic and caused by somatic mutations. Thus, most genomic studies have focused on somatic variants and have rarely analyzed the germline samples. In-depth studies focusing on the germlines of pGBM patients will provide further insight into potentially causative germline variants.

1.2 Somatic Mutations

1.2.1 Histone variant H3.3

A prominent finding which advanced the molecular characterization of pGBM was the discovery of somatic mutations in *H3F3A*, a gene which encodes the histone variant H3.3 (Jeremy Schwartzentruber et al., 2012). Histone H3.3 is deposited in a cell cycle independent fashion at active genes and repetitive regions to regulate gene expression by the HIRA chaperone protein (Goldberg et al., 2010; Szenker, Ray-Gallet, & Almouzni, 2011). H3.3 is additionally deposited by another chaperone complex, ATRX-DAXX, at telomeric and pericentromeric regions (Goldberg et al., 2010; Szenker et al., 2011). H3.3 has also been recently shown to play a role in maintaining genome integrity during mammalian development (Jang, Shibata, Starmer, Yee, & Magnuson, 2015).

Two recurrent mutations in H3F3A which have been well characterized include one that causes an amino acid change from lysine 27 to a methionine (H3.3K27M) and another that causes an amino acid change from glycine 34 to either an arginine (H3.3G34R) or a valine (H3.3G34V) (Jeremy Schwartzentruber et al., 2012). Both mutations interfere with the posttranslational modifications on the histone variant, which subsequently hinders the function of H3.3, but K27M has been described and studied more thoroughly because trimethylation of K27 plays an important role in chromatin compaction and gene repression (Bjerke et al., 2013; Lewis et al., 2013). Tumors that harbor this mutation have been shown to harbor a unique DNA methylation pattern (Sturm et al., 2012).

This lesion was novel because histone genes are highly conserved and no human disorders had previously been associated with genes encoding histones (Jeremy Schwartzentruber et al., 2012). A major caveat with this alteration is that it is only present in ~30% of pGBM cases (Jeremy Schwartzentruber et al., 2012; Wu et al., 2014), while the majority of cases have not revealed any recurrent mutations which can be exploited for therapeutic use. Additionally, this mutation alone is not sufficient to induce gliomagenesis and additional lesions such as *TP53* loss and *PDGFRA* amplification are required to achieve cellular transformation (Funato, Major, Lewis, Allis, & Tabar, 2014; Lewis et al., 2013; Pathania et al., 2017). This suggests that the *H3F3A* mutation may contribute to tumor growth, but alternative mechanisms are required to induce tumor formation in pGBM patients.

1.2.2 Alpha thalassemia/mental retardation syndrome X-linked (ATRX)

Another novel genetic lesion that is found in one-third of pGBM patients and is also involved in chromatin architecture is the somatic mutation in *ATRX*. ATRX forms a chromatin remodeler heterodimer with DAXX, which is responsible for the deposition of H3.3 into heterochromatin and telomeric DNA (He, Mansouri, & Das, 2018). Mutations in *ATRX* or *DAXX* were present in all pGBM cases with H3.3G34R/V mutations but were not shown to significantly co-occur with H3.3K27M tumors (Jeremy Schwartzentruber et al., 2012). Mutations in *ATRX* lead to hypomethylation in CpG rich repetitive sequences, which suggests it may play a role in gene regulation (Gibbons et al., 2000). The *ATRX* mutation in pGBM causes loss of protein expression, and is routinely tested for in the clinic using immunohistochemical analysis to aid in diagnosis (He et al., 2018). Although *ATRX* mutations have been detected in both adult and pGBM, (Gröbner et al., 2018; He et al., 2018; Jeremy Schwartzentruber et al., 2012; Wu et al., 2014) the role of *ATRX* in tumor initiation has yet to be elucidated. Studies have suggested that loss of *ATRX* causes genetic instability in tumors, because of the absence of telomere maintenance and a disruption of DNA repair, which aids in tumor growth to create an aggressive phenotype (Koschmann et al., 2017).

1.3 Molecular Features of pGBM

Although pGBM is phenotypically indistinct from adult GBM, recent studies have shown that pediatric cases are both genetically and transcriptionally distinct from the adult malignancy (Baker et al., 2016; Jeremy Schwartzentruber et al., 2012; Sturm et al., 2012, 2016). Sturm *et al.* showed that pGBM harbors a distinct methylation pattern and can be clustered based on DNA methylation. In these clusters, the pediatric cases did not harbor the classical genetic alterations

commonly found in adult GBM. Alterations such as the *IDH1* mutation, chromosome (chr) 7 gain, chr 10 loss, *CDKN2A* deletions, *EGFR* amplification, and *PDGFRA* amplification are rarely found in pGBM (Sturm et al., 2012). The genetic alterations explored in this paper are characteristic of adult GBM and while it is important to show that these alterations are absent, they do not provide insight into tumor etiology of pGBM. Another genetic lesion which is found in approximately 5% of adult GBM and a subset of diffuse low-grade gliomas is the *BRAF-V600E* mutation (Dahiya et al., 2014). When this mutation was investigated in pGBM, it was only found in 3 of 25 samples (Dahiya et al., 2014). Interestingly, *VEGF* (vascular endothelial growth factor responsible for angiogenesis in adult GBM) was shown to potentially contribute to pGBM pathogenesis (Bodey, Siegel, & Kaiser, 2006) but when treated with a VEGF inhibitor, pGBM did not respond to the treatment and it provided no benefit to the patients, contrary to the adult GBM patients (T. J. MacDonald, Aguilera, & Kramm, 2011). These cases additionally highlight age-related differences in the pathogenesis of adult and pGBM.

Another interesting difference between adult and pediatric GBM is that pGBMs harbor significantly less somatic variants than their adult counterpart (Jones & Baker, 2014), potentially suggesting germline variants and/or impaired developmental processes are playing a larger role in tumor formation.

The assumption that pGBMs have genetically bland genomes and are molecularly homogeneous stems from a few key observations (Jones & Baker, 2014). The first is that there have been few recurrent somatic mutations noted in pGBM, and the second is that those previously described somatic mutations (*H3.3*K27M, *H3.3*G34R/V, *ATRX*) are clonal (Salloum et al., 2017;

Vinci et al., 2018). In the case of *ATRX* mutations (including frameshift insertions/deletions, gains of a stop codon, and missense SNVs) it leads to complete loss of protein expression (J Schwartzentruber et al., 2012) and the H3.3K27M mutation results in global loss of trimethylation of lysine 27 in all tumor cells (Lewis et al., 2013). Traditional sequencing techniques that do not allow for chromosomal level detection also contribute to the notion of a low level of intratumoral heterogeneity in pGBM. Previous techniques may have allowed subclonal variants to go undetected because their low allelic frequencies did not meet the threshold for being called.

Vinci *et al*. recently challenged this dogma by utilizing multi-regional sampling techniques as well as reanalyzing previously published whole exome sequencing (WES) datasets to characterize the presence of subclonal populations that cooperate to maintain an environment conducive to tumor growth. Although their cohort mainly consisted of diffuse intrinsic pontine gliomas (DIPGs), another type of high grade pediatric brain tumor, they were able to show the presence of subclonal populations in pGBM as well (Vinci et al., 2018). This was the first study to show intratumoral heterogeneity in pGBM and further studies with larger cohorts will be valuable in confirming their findings.

My study is unique in the fact that I am characterizing subclonal populations from primary and recurrent tumor samples from the same patient on a genome-wide level. This longitudinal sample collection is incredibly rare and has provided important insight into the level of molecular heterogeneity and the extent of tumoral evolution in pGBM. It has also allowed for a complete investigation into germline variants that could potentially be cooperative to induce tumor formation.

Chapter Two: **Hypothesis and Aims**

Currently, there are no effective treatments options for patients diagnosed with pGBM. Thus, there is a need for a greater understanding of the mechanisms involved in tumor etiology. In this study, I have investigated intra and intertumoral heterogeneity and characterized causative germline variants in pGBM at both the genetic and transcriptional level. ***I hypothesize that pediatric glioblastoma is molecularly heterogeneous and patients with this disease harbor germline variants that contribute to tumor formation.*** This hypothesis was tested in the following aims:

Aim 1: To investigate intrapatient tumor evolution in pGBM

Aim 2: To characterize inter and intratumoral heterogeneity in pGBM

Aim 3: To identify causative mutations in the germlines of pGBM patients

Chapter Three: **Materials and Methods**

3.1 pGBM Sample Collection and Storage

All pGBM tissue samples and matched blood were collected and used for research with appropriate informed consent and with approval by the Health Research Ethics Board of Alberta. Tumor tissue from each patient was preserved in liquid nitrogen and matched blood was preserved at -80°C at the Clark Smith Brain Tumor biobank at the University of Calgary.

3.2 Genomic DNA Extraction

I extracted high molecular weight (HMW) genomic DNA (gDNA) from frozen tumor tissue and frozen whole blood samples with the Qiagen MagAttract® HMW DNA Kit (catalog # 67563) per manufacturers' protocol. I cut and weighed approximately 30 mg of frozen tumor tissue for gDNA extraction from each sample. I used approximately 100-200μL of blood for gDNA extraction from each sample. gDNA fragment size and distribution were quantified with the Agilent Technologies 2200 TapeStation Genomic DNA Assay at the Centre for Health Genomics and Informatics (University of Calgary). I analyzed the results and the samples with gDNA fragment size > 50 kb were used for subsequent library construction.

3.3 Whole-Genome Sequencing (WGS)

I used the 10xGenomics Chromium™ Genome Library Kit & Gel Bead Kit v2, 16 rxns (catalog # PN-120258) for GEM Generation, Barcoding, Post GEM Incubation Cleanup and Library Construction. I diluted the extracted HMW gDNA to 1 ng/μL in Buffer EB and quantified in triplicates using the Qubit dsDNA HS Assay Kit (catalog # Q32851), per manufacturer's protocol. Size and distribution of all sequencing libraries were quantified with the Agilent

Technologies 2200 TapeStation D1000 Assay at the at the Centre for Health Genomics and Informatics (University of Calgary).

I sent the libraries for whole-genome sequencing to The Centre for Applied Genomics at the Hospital for Sick Children (Toronto, ON) and they were sequenced with a HiSeq X Series (Illumina) instrument. Libraries were validated on a Bioanalyzer DNA High Sensitivity chip to check for size and absence of primer dimers and quantified by qPCR using Kapa Library Quantification Illumina/ABI Prism Kit protocol (KAPA Biosystems). Validated libraries were paired-end sequenced on an Illumina HiSeq X platform following Illumina's recommended protocol to generate paired-end reads of 150-bases in length. Each library was loaded on a single lane. All libraries were sequenced at 30x coverage. Post-sequencing, I analyzed the fastq files with the package LongRanger-2.1.6 (10xGenomics) on servers hosted by the Centre for Health Genomics and Informatics (University of Calgary). LongRanger performed alignment of reads to the reference genome GRCh38, assembled linked-reads, generated haplotype blocks, and called single nucleotide variants (SNVs) and structural variants (SVs). LongRanger output files were visualized with Loupe browser 2.1.2 (10xGenomics). I used the LongRanger output files to compare the SNVs and large SVs among and between patients.

3.4 WGS Additional Analysis

Somatic and de novo germline SNVs were called by Michael Johnston, a postdoctoral fellow in the Gallo Lab, using both Mutect2 and Strelka2 (S. Kim et al., 2017). Only SNVs called by both methods were included for downstream analysis. Annotation of SNVs was performed using ANNOVAR (Wang, Li, & Hakonarson, 2010).

SNVs in genes associated with familial cancer syndromes present in both the tumors and the germline control of each patient were called by Ana Nikolic, a PhD candidate in the Gallo Lab. I used the analysis to identify recurrent SNVs in genes associated with familial cancer syndromes.

Phylogenetic trees based on subclonal populations were analyzed and constructed by Aaron Gillmor, a student in Sorana Morrissy's lab, using my WGS data. Aaron also analyzed and constructed the mutational signatures for each subclone using the Catalogue of Somatic Mutations in Cancer (COSMIC) (Forbes et al., 2008).

3.5 Copy Number Assays for *ATRX* Locus

I used custom TaqMan® Copy Number Assay Probes (catalog # 4400294) from Applied Biosystems™ by Thermo Fisher Scientific to detect changes in copy number at the *ATRX* locus in all patient samples. Quantification was performed according to manufacturers' protocols. For each patient I used a control DNA sample, patient germline DNA, primary tumor DNA, and recurrent tumor, when available. I diluted the DNA to 5 ng/μL and quantified in triplicates with the Qubit dsDNA HS Assay Kit (catalog # Q32851). I added the TaqMan® Genotyping Master Mix (10 μL) and TaqMan® Copy Number Assay (1 μL) to each well of a 96-well plate. I then added 20 ng of DNA in triplicate wells of a 96-well plate for each assay condition. I ran the samples on a Bio-Rad CFX Connect™ Real-Time PCR Detection System with the following parameters: hold at 95°C for 10 mins, 40 cycles of 95°C for 15 sec and 60°C for 60 sec. I first normalized the fluorescence signal for the *ATRX* deleted region probe to the fluorescence signal from the control diploid region probe. I additionally normalized the fluorescence signal to the diploid control DNA sample. I designed the TaqMan® Probes which are labeled with FAM at 5' end and MGBNFQ at the 3' end.

For the probe in the *ATRX* deleted region; probe sequence: CACACCCAAATATTGGTAAAAAT, forward primer: TCCAGGACTTAGCAGGATGTGA, reverse primer: ACCCCATCAAGTAGATGGTAAGAAACT. For the probe in the control diploid region; probe sequence: ACACGGGTGTTGAAGACGC, forward primer: AGGCGCTGTGGAAACAATTATAGTA, reverse primer: TGTCCCCGTCAACGATCAC.

3.6 Single-cell RNA Sequencing (scRNA-seq)

pGBM cells were given to Jennifer King and Ngoc Ha Dang in Dr. Donna Senger's Lab where they generated patient derived xenografts (PDXs) using an established lab protocol. I collected the xenograft tumor tissue from the mouse and dissociated them into single cells using the following protocol: I added 10 ml of ACCUTASE™ (catalog # 07920) to the tumor tissue, which I manually dissociated with sterile surgical scissors. I then incubated the sample with ACCUTASE™ and sterile glass beads at 37°C, rotating continuously for 30 min. I removed the glass beads and centrifuged the sample at 1000 rpm for 5 min at 25°C. I then removed and discarded the supernatant and resuspended the cell pellet in neural stem cell media (NeuroCult™ NS-A Proliferation Kit (Human), catalog # 05751).

Following tumor dissociation, I used the Miltenyi Biotec Mouse Cell Depletion Kit (catalog # 130-104-694) and the autoMACS® Pro Separator and columns (catalog # 130-021-101) to remove the mouse cells using the following protocol: I centrifuged the cells at 300xg for 10 min then removed and discarded the supernatant. I resuspended the cell pellet in 80 μL of buffer (0.5% BSA in PBS) and 20 μL of Mouse Cell Depletion Cocktail. I pipette mixed the sample thoroughly and incubated it for 15 min at 4°C. I adjusted the volume to 500 μL using buffer. I ran a flushing

cycle on the autoMACS® Pro Separator, inserted two new columns, and then I ran the column exchange program. I inserted the tube containing the sample and empty tubes for collection into the tube rack in the appropriate locations. I ran the depletion program, and the negative fraction of cells were collected; this fraction represents the enriched human tumor cells. Cells were centrifuged at 1000 rpm for 5 min at room temperature. I resuspended the cell pellet in Dulbecco's Modified Eagle Medium (DMEM) and quantified the viability and concentration of the purified human cells. Cell samples with viability > 80% were used for subsequent scRNA-seq.

The 10xGenomics Chromium™ Single Cell 3' Library & Gel Bead Kit v2, 16 reactions (catalog # PN-120237) was used for GEM Generation, Barcoding, Post GEM-RT Cleanup, cDNA Amplification, and Library Construction. Protocol was followed as per manufacturers' recommendations by Katrina Ellestad, a technician in the Gallo Lab. Size and distribution of all sequencing libraries were quantified with the Agilent Technologies 2200 TapeStation High Sensitivity D1000 Assay at the Centre for Health Genomics and Informatics (University of Calgary).

I sent the libraries for single-cell RNA sequencing at the Centre for Health Genomics and Informatics (University of Calgary). All libraries were sequenced on mid-output 150 cycle NextSeq500 runs, with ~130M reads each. Sequencing (BCL) files were processed with the Cell Ranger 2.1.0 package (10xGenomics). Reads were aligned to the GRCh38 transcriptome. Data standardization and normalization was performed with the Seurat R package (Butler, Hoffman, Smibert, Papalexi, & Satija, 2018).

3.7 Graphing Software and Statistical Analysis

I used GraphPad Prism (Version 7.0; GraphPad Software, Inc., San Diego, CA) for graph generation. I generated the Venn diagrams using the eulerr R package version 4.0.0 (Larsson J. *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*. R package version 4.1.0, https://cran.r-project.org/package=eulerr. I generated the Sankey Diagrams using SankeyMATIC (www.sankeymatic.com). GraphPad Prism (Version 7.0; GraphPad Software, Inc., San Diego, CA) was used for statistical analyses. Results are reported as mean ± standard error. Significance was calculated using the student's t tests with a significance value of 0.05.

Chapter Four: **Results**

4.1 Calgary pGBM Collection

The availability of matched primary and recurrent pGBM patient samples is rare, and the Calgary pGBM cohort was a valuable resource which allowed for new insights into disease etiology and mechanisms of tumoral evolution. This collection included 5 patients with matched non-tumor (germline) DNA; one patient had a diagnostic and second resection of the same tumor, two patients had one tumor recurrence and one patient had three recurrences (Figure 1). The treatment protocols administered to each patient were also available (Figure 2). In this cohort, the ages of the patients range from <1 to 15 years old and the treatment protocols differ vastly with each patient receiving a different chemotherapeutic drug. The variety of treatment protocols given to each patient highlights the lack of understanding and subsequent ineffective therapy options for pGBM patients.

Patient 1 (male, 12 years) received local radiation and no chemotherapy. Patient 2 (female, 9 years) had local radiation and TMZ treatment. TMZ is an alkylating agent that induces apoptosis by triggering the mismatch repair pathway and is often received with concurrent irradiation (Warren et al., 2012). TMZ treatment is the standard treatment for adults diagnosed with GBM but has shown minimal effectiveness in increasing survival of pGBM patients (Stupp, Hegi, Gilbert, & Chakravarti, 2007; Warren et al., 2012). Patient 3 (female, <1 year) received no radiation but did receive carboplatin and etoposide treatment. Although the exact mechanism of action of carboplatin is not known, it is thought to inhibit DNA synthesis by forming interstrand and intrastrand cross-linking of DNA (Go & Adjei, 1999). Etoposide damages DNA by inhibiting

topoisomerase II (Hande, 1998). Patient 4 (male, <1 year) received no treatment, most likely because the patient succumbed to the disease within 2 weeks of diagnosis. Patient 5 (male, 15 years) received cerebrospinal radiation and a chemotherapy treatment consisting of cisplatin, vincristine, and cyclophosphamide. Cisplatin targets rapidly dividing cells by introducing crosslinks between the purine bases on DNA which then interferes with DNA repair mechanisms; causing DNA damage and inducing apoptosis (Dasari & Tchounwou, 2015). Vincristine works by binding to tubulin and stopping the ability of the cell to properly separate its chromosomes during metaphase; subsequently triggering apoptosis (Jordan, 2002). The last chemotherapeutic drug this patient received was cyclophosphamide which, similar to cisplatin, induces interstrand and interstrand crosslinkages in DNA in dividing cells to induce apoptosis (Hall & Tilby, 1992).

4.2 WGS Summary

Rather than using a traditional sequencing approach, I chose to use the 10X Genomics Chromium™ Sequencing Solution. This technology utilizes linked-read technology that allows for sensitive measurement of intratumoral cellular heterogeneity. The product generates linked-reads by tagging short amplicons from a strand of HMW DNA with identical barcodes. During sequencing, the barcodes are then linked together to generate large phased haplotype blocks. The haplotype blocks provide long range information from short next-generation sequencing reads. The linked-read technology also provides chromosomal level resolution which allows for sensitive detection of subclonal events.

For this technique, I first optimized a protocol for the extraction of HMW DNA (>50 kbp) (Figure 3). Starting the library construction with HMW DNA molecules is important for the

Patient #	Primary	Recurrence 1	Recurrence 2	Recurrence 3	Blood
1	✓*	✓*			✓
2	✓	✓			✓
3	✓	✓*			✓
4	✓				✓
5	✓	✓	✓*	✓*	✓

Figure 1. Calgary pGBM collection. Red asterisks indicate that patient-derived xenografts exist for specific tumor samples. Blue asterisks indicate that the recurrence of patient 1 is an exceptional case and is technically a second resection, not a true recurrence.

Patient #	Sex	Age (years)	Radiation	Chemotherapy	Overall survival (days)
1	M	12	local	none	>1434
2	F	9	local	temozolomide	281
3	F	<1	none	carboplatin, etoposide	325
4	M	<1	none	none	16
5	M	15	cerebrospinal, boost	cisplatin, vincristine, cyclophosphamide	192

Figure 2. Therapeutic information for each pGBM patient. Sex, age, radiation, chemotherapy and overall survival is provided.

generation of large phase blocks during downstream sequencing and is also beneficial for haplotype phasing. From my sequencing data of 20 samples, I was able to generate haplotype phase blocks of up to 38 Mbp, which allowed me to detect a total of 476 large structural variants (>30kbp) and 84,907 short deletions (50bp-30kbp) (Figure 4).

4.3 Intrapatient Tumor Evolution

4.3.1 Single nucleotide variants (SNVs)

To address whether the mutational landscape of pGBM evolves throughout disease progression, I used my WGS data to characterize changes in the somatic SNVs from the primary to the recurrent tumors.

Somatic SNVs were called by Michael Johnston, using the patient's germline as the reference genome (Figure 5). I focused specifically on the somatic, non-synonymous coding SNVs for my analysis. The range of SNVs per tumor were 26-126; with an average of 65.6 SNVs per tumor in primary pGBM, and 60 SNVs per tumor in recurrences (Figure 6). Patient 1 had 31 SNVs in the primary and 126 in the second resection. Only 2 mutations in *ARMC9* and *DSE* were conserved in the recurrence (Figure 6A). The SNV in *DSE* is of particular interest because it encodes a dermatan sulfarase epimerase and is expressed in the human brain at the highest levels prenatally and continues to be stably expressed postnatally (BrainSpan Atlas of the Developing Human Brain, 2018). Patient 2 had 109 mutations in the primary and 69 in the recurrence, with mutations in *H3F3A*, *MUC22*, and two in *TP53* shared with the recurrence (Figure 6B). This patient harbored the previously described K27M lesion in *H3F3A*. Patient 3 had 78 mutations in the primary tumor and 62 in the recurrence, with only 1 mutation in *PCNX* shared between both

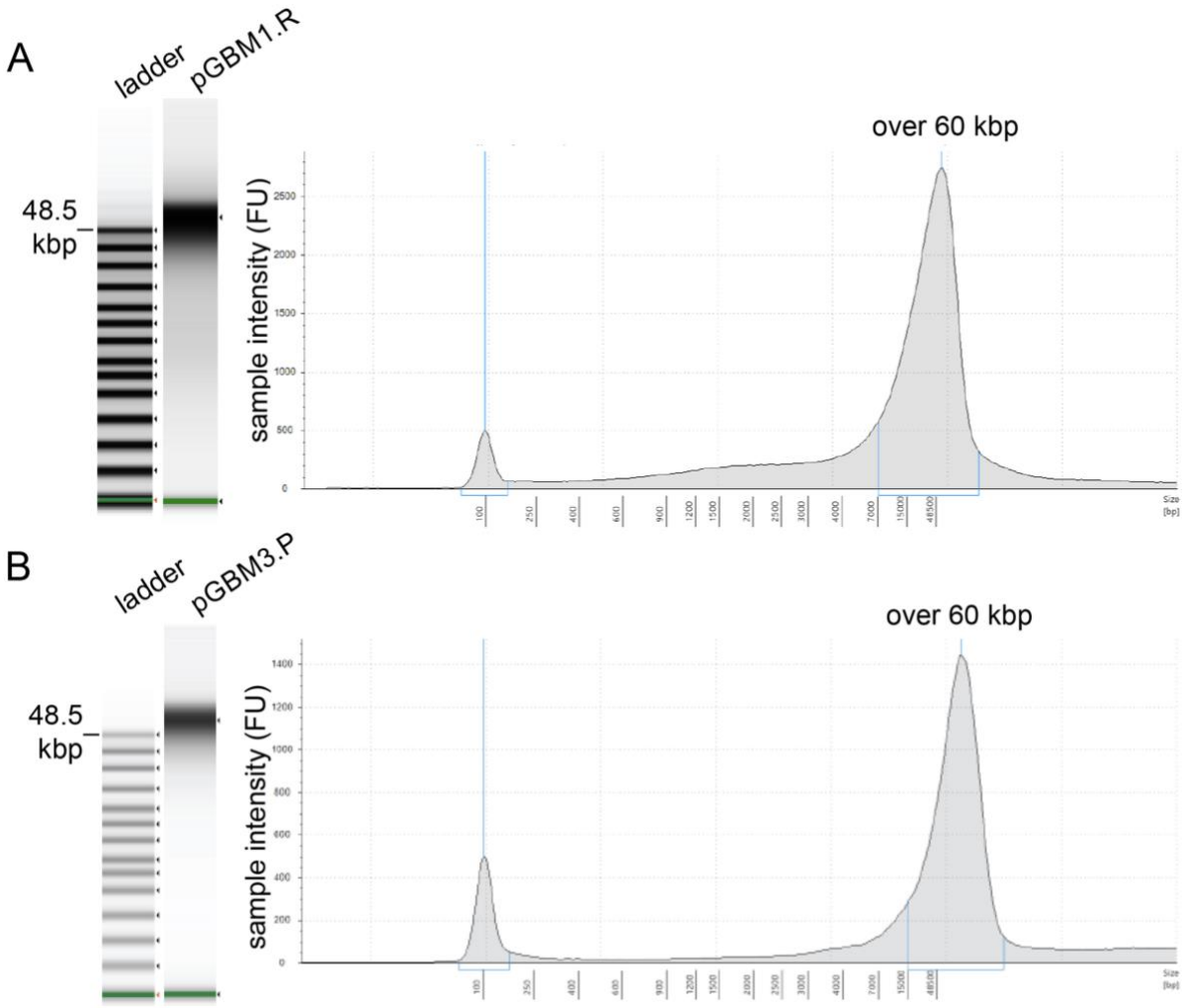


Figure 3. Size analysis of genomic DNA used for whole-genome sequencing. Two representative examples are shown. (A) Tapestation results for genomic DNA extracted from recurrence of patient 1. (B) Tapestation results for genomic DNA extracted from the primary tumor of patient 3.

Patient #	Longest Phase Block (Mbp)	Large SVs (>30kbp)	Short Deletions (50bp-30kbp)
1.Primary	9	49	4239
1.Second resection	1	26	6838
1.Germline	11	20	4153
2.Primary	12	23	4596
2.Recurrence	12	25	4849
2.Germline	21	19	4165
3.Primary	4	15	4481
3.Recurrence	9	20	4367
3.Germline	12	20	4141
4.Primary	12	35	4086
4.Germline	24	31	3977
5.Primary	9	30	3953
5.Recurrence 1	9	23	3945
5.Recurrence 2	11	25	3876
5.Recurrence 3	10	20	3877

Patient #	Longest Phase Block (Mbp)	Large SVs (>30kbp)	Short Deletions (50bp-30kbp)
5.Germline	18	27	4116
1.Parent 1	24	14	3884
1.Parent 2	26	14	3762
3.Parent 1	38	15	3897
3.Parent 2	37	25	3705

Figure 4. Whole genome sequencing summary. Longest phase block, number of large SVs, and the number of short deletions is provided for each sample.

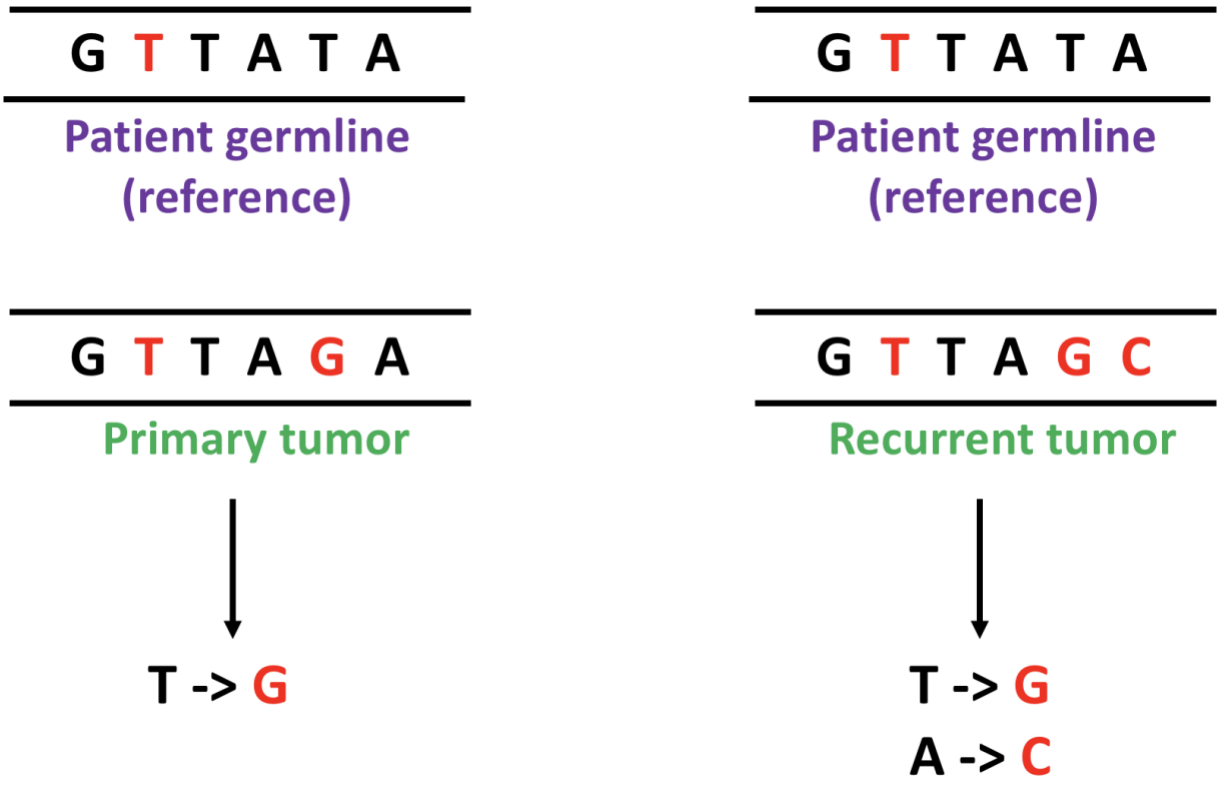


Figure 5. Schematic of single nucleotide variant calling. The patient's germline sample is used as the reference genome when calling somatic single nucleotide variants in the primary and recurrent tumors. Mutations present in both the germline and tumor samples are not called.

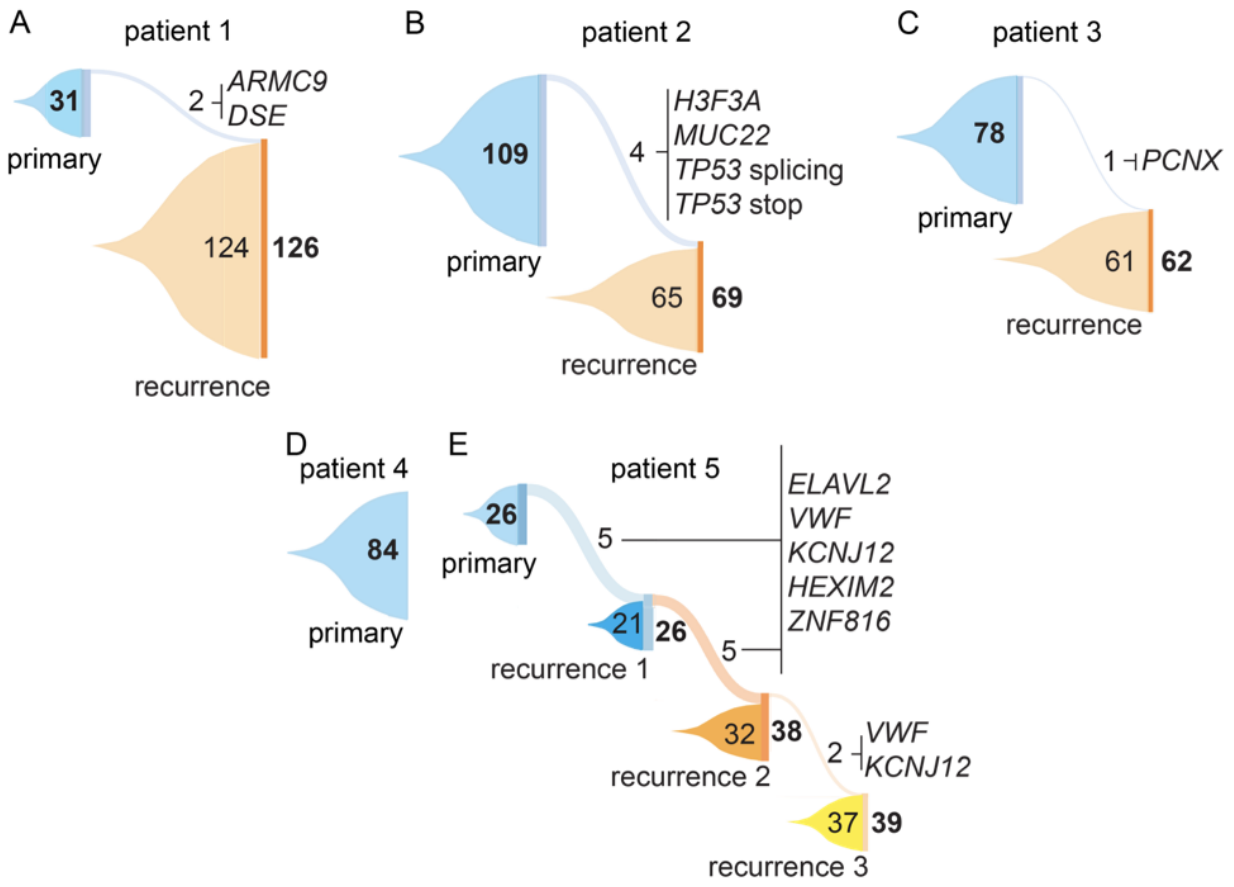


Figure 6. Somatic single nucleotide variants in pGBM samples. Sankey plots illustrating changes in mutational landscapes between primary tumors and recurrences for (A) patient 1 (this patient is an exception because the recurrence is technically a second resection), (B) patient 2, (C) patient 3, (D) patient 4 and (E) patient 5. Bold numbers represent total number of mutations in each profiled tumor. Gene lists represent the mutated genes that are shared between consecutive tumors.

tumors (Figure 6C). Patient 4 had 84 mutations in the primary tumor, recurrence not available (Figure 6D). Patient 5 had 26 SNVs in the primary, 26 in the first recurrence, 38 in the second recurrence, and 39 in the third recurrence. There were 5 SNVs in *ELAVL2*, *VWF*, *KCNJ12*, *HEXIM2*, and *ZNF816* which were shared among the primary, first recurrence, and second recurrence. There were 2 recurrent lesions present in all tumor samples of this patient, in the genes *VWF* and *KCNJ12* (Figure 6E).

One unexpected finding of this analysis was that the overall mutational burden did not drastically increase from primary to recurrence. This was surprising because adult gliomas experience a large number of mutational events following standard of care treatment (radiation and subsequent TMZ) (Johnson *et al.*, 2014), yet patient 2 received radiation and TMZ treatment and the number of SNVs decreased at tumor recurrence. This finding supports the notion that pGBM does not respond to treatment in a similar fashion to adult GBM.

My data is consistent with previous reports of low mutational burden in pGBM (Bender et al., 2016; Jeremy Schwartzentruber et al., 2012) and the genomes remain relatively stable throughout tumor progression with respect to overall mutational burden. The exception to this is patient 1 who had a dramatic increase in mutational load upon second tumor resection, and this can potentially be explained by the patient harboring a germline *TP53* deletion of exon 16 passed on from one of his parents (Figure 7). The results of this analysis provide evidence of extreme tumoral evolution in pGBM, shown by the paucity of conserved mutations from the primary to the recurrence in each patient.

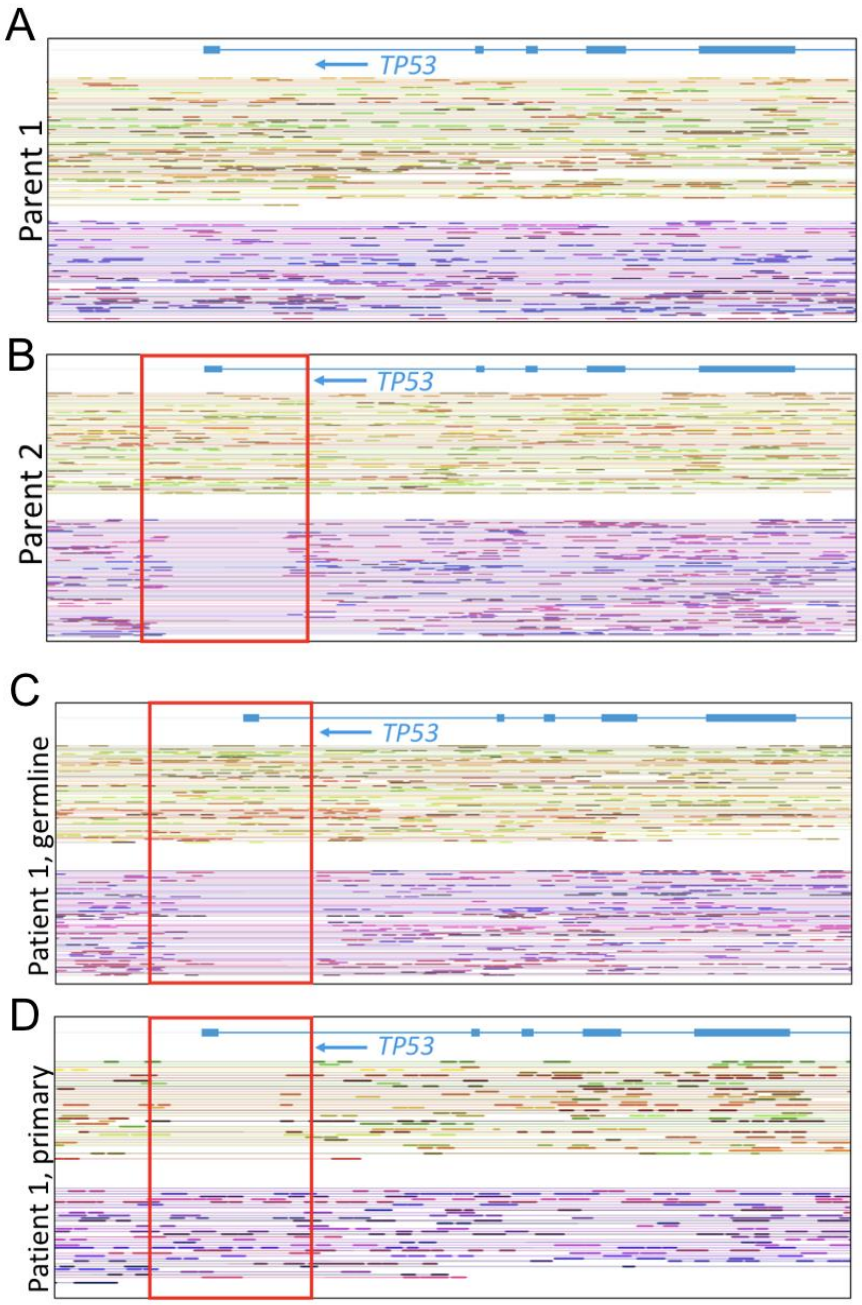


Figure 7. Heterozygous germline *TP53* deletion at exon 16. (A) Parent 1 of Patient 1 (B) Parent 2 of Patient 1 (C) Patient 1, germline (D) Patient 1, primary tumor. The deletion is highlighted by the red outline. Both haplotypes are shown for each patient, green colored reads represent one haplotype and purple-colored reads represent the other haplotype. The haplotypes are not consistent throughout the samples; thus B, C and D represent the same deletion.

4.3.2 Large structural variants (SVs)

I next used my WGS data to characterize changes in the large structural variants (SVs) throughout tumor evolution. Large SVs were defined as affecting >30 kbp of the genome and all samples (tumor and germline) were processed independently. Rather than aligning the tumor samples to the germlines as I had done for the SNVs, I instead aligned them to the GRCh38 reference genome (Figure 8). This approach allowed for detection of large SVs present in both the tumor sample and the germlines, as well as unique SVs in each. The range of large SVs per sample were 19-52; with an average of 29.5 large SVs per tumor and 25.4 large SVs per germline (Figure 9).

For patient 1, the primary had 52 large SVs (31 unique), the recurrence had 27 (7 unique) and the germline had 25 (8 unique). There were 14 large SVs that were shared amongst the primary, recurrent, and the germline (Figure 10A). As previously stated, this patient harbors a germline *TP53* deletion, which potentially accounts for the high number of large SVs present in the primary tumor. I also identified chromosome 6 chromothripsis in the primary tumor of this patient, which accounted for 62% of the large SVs. The apparent loss of chromothripsis the second resected sample provides rationale for the decrease in the number of SVs.

For patient 2, the primary had 25 large SVs (7 unique), the recurrence had 27 large SVs (8 unique), and the germline had 21 large SVs (1 unique) There were 15 large SVs that were shared amongst the primary, recurrent, and the germline (Figure 10B). A novel variant which is present

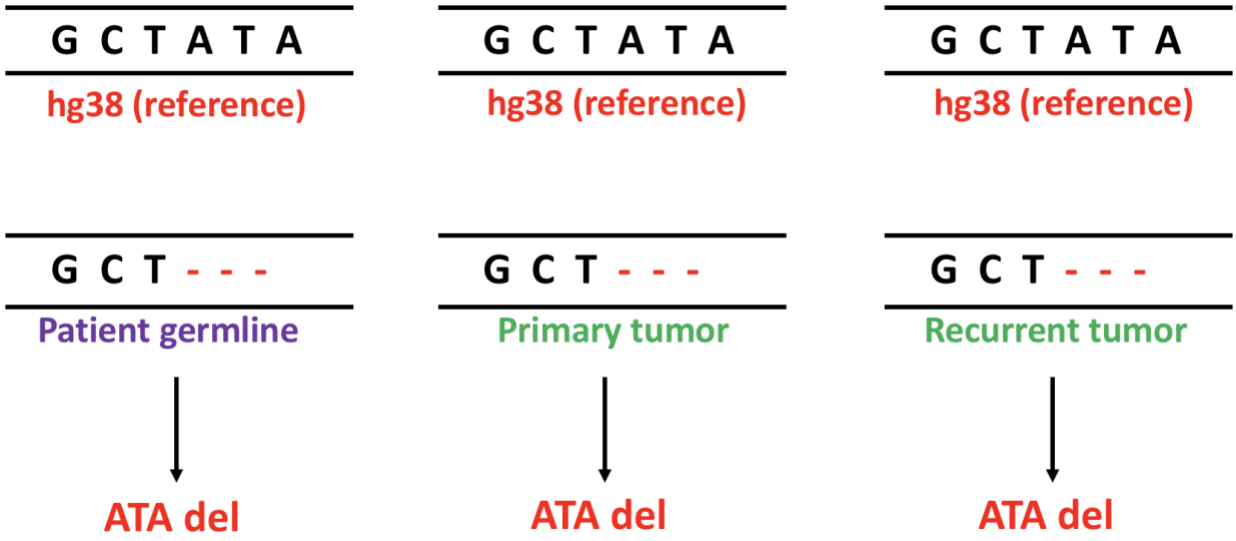


Figure 8. Schematic of large structural variant calling. The GRCh38 human genome was used as the reference genome when calling large structural variants in all patient samples. Germline and tumor samples were analyzed independently.

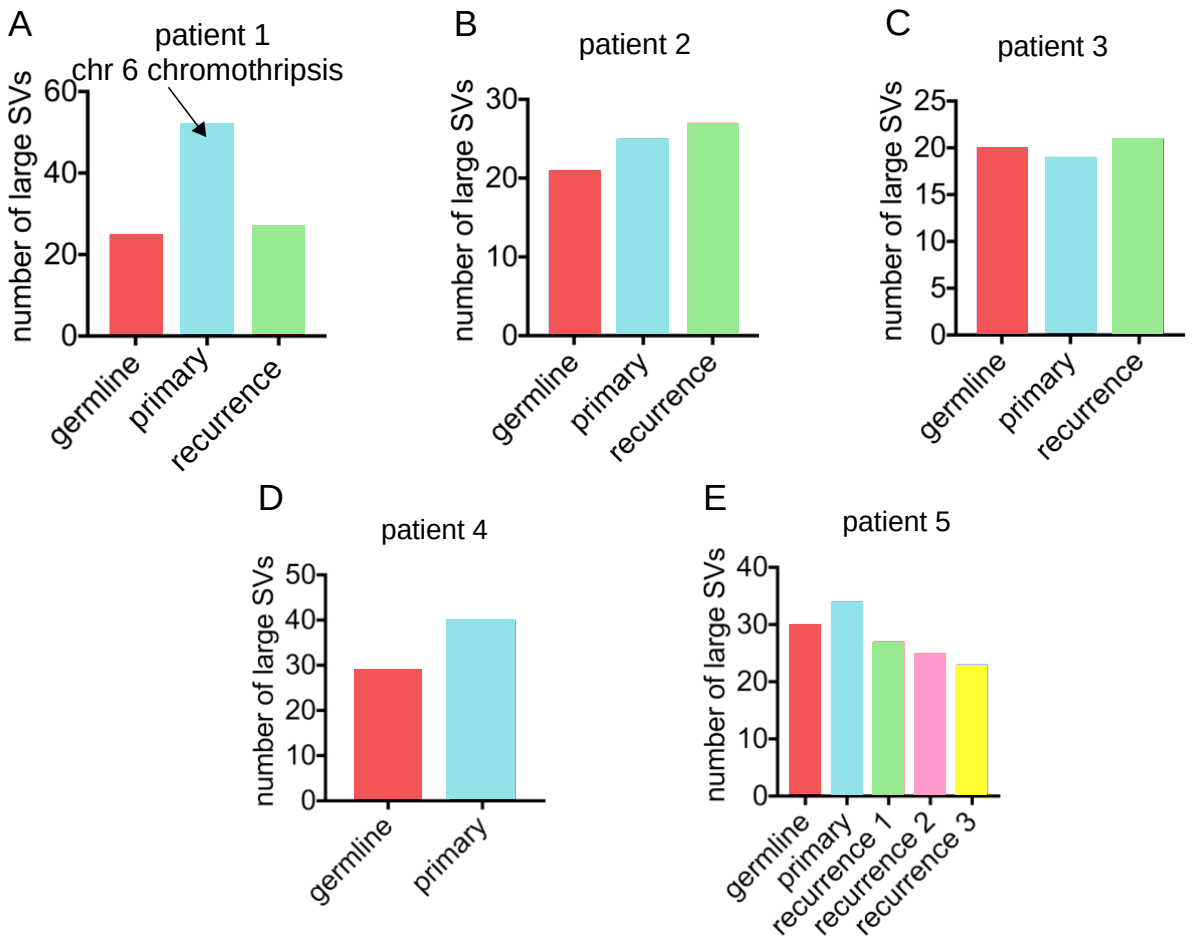


Figure 9. Large structural variants in pGBM samples. Total number of large SVs in the germline, primary and recurrent tumors. (A) Patient 1 (B) Patient 2 (C) Patient 3 (D) Patient 4 and (E) Patient 5. Patient 1 harbored chromosome 6 chromothripsis in the primary tumor.

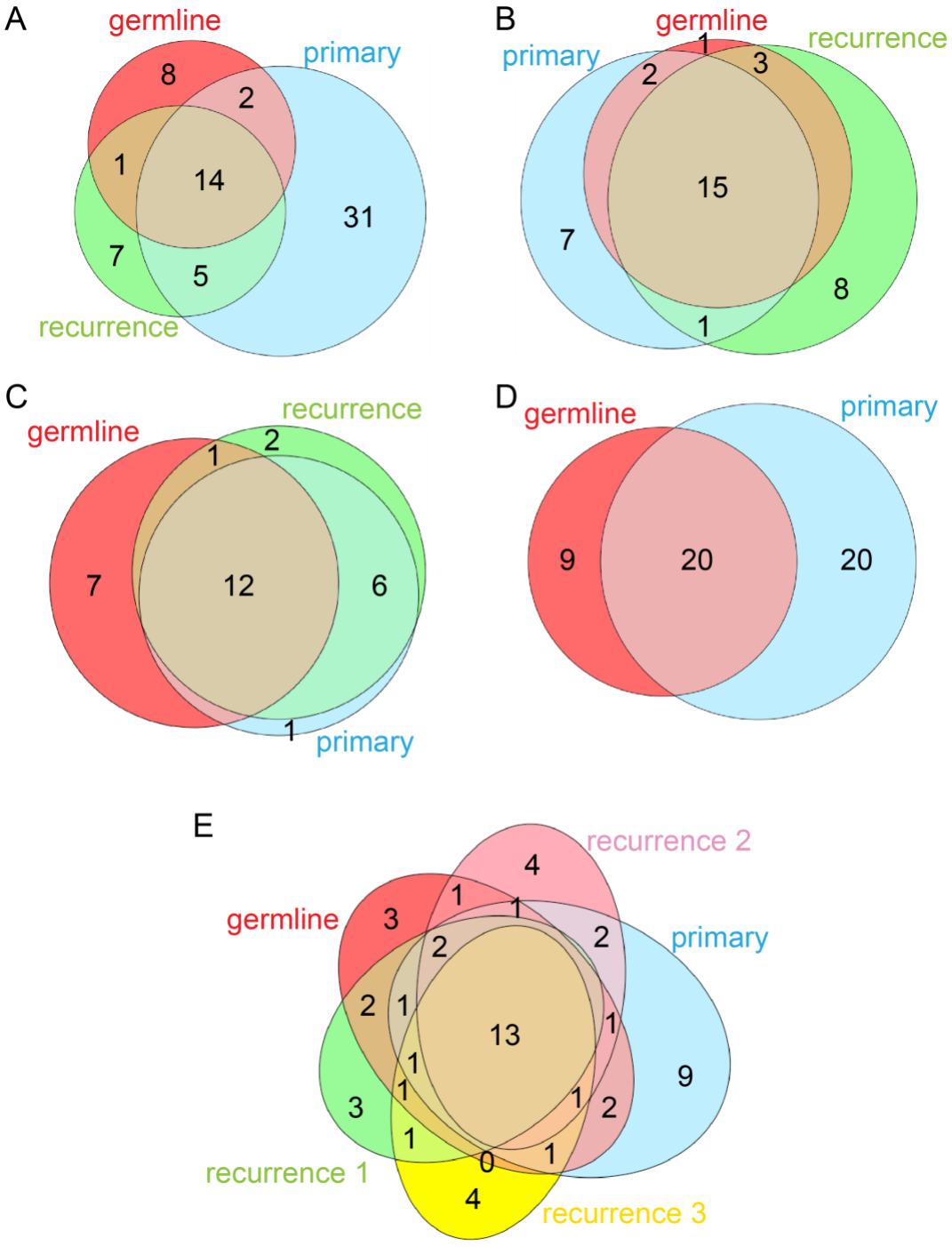


Figure 10. Distribution of large structural variants in each pGBM patient. (A) Patient 1 (B) Patient 2 (C) Patient 3 (D) Patient 4 (E) Patient 5.

in the primary tumor and conserved in the recurrence is the amplification of the *Neurotrophic Receptor Tyrosine Kinase 2 (NTRK2)*. This gene encodes a membrane-bound protein that phosphorylates itself and members of the MAPK pathway. *NTRK2* also suppresses anoikis and overexpression has been shown to contribute to tumor metastases (Douma et al., 2004). The generation of an *NTRK2* fusion gene has been shown in pGBM, but not an amplification as I have shown here (Wu et al., 2014).

For patient 3, the primary had 19 large SVs (1 unique), the recurrence had 21 large SVs (2 unique), and the germline had 20 large SVs (7 unique). There were 12 large SVs that were shared amongst the primary, recurrent, and the germline (Figure 10C). Interestingly, this patient harbors a somatic *ETV6-NTRK3* gene fusion in both the primary and recurrent tumors. This gene fusion has been reported in pGBM before and has been shown to be oncogenic in other cancer types such as acute myeloid leukemia, breast cancer, and papillary thyroid carcinomas (Eguchi et al., 1999; Lannon & Sorensen, 2005; Li et al., 2007; Wu et al., 2014). Wu et al. found the *ETV6-NTRK3* fusion in a 1-month old pGBM patient with only 1 other non-silent alteration in their genome (Wu et al., 2014). This is remarkably similar to our data in that patient 3 from our cohort harbors the least number of SVs and the patient was <1-year-old when they succumbed to their disease. The scarcity of genetic lesions and the young age of these patients suggests that this gene fusion is a robust oncogenic driver that initiates pGBM formation potentially prenatally.

For patient 4, the primary had 40 large SVs (20 unique), and the germline had 29 large SVs (9 unique). There were 20 large SVs that were shared amongst the primary and the germline samples (Figure 10D). This patient harbors a lesion in the *MET* gene in its primary tumor. *MET* is

a proto-oncogene that encodes a receptor tyrosine kinase. Both fusions and amplifications of this gene have been associated with multiple cancer types (Bender et al., 2016; Gherardi, Birchmeier, Birchmeier, & Vande Woude, 2012), and is commonly amplified in adult GBM (Snuderl et al., 2011). This finding is consistent with previous reports of genetic lesions in *MET* occurring in pGBM (Bender et al., 2016; Wu et al., 2014).

For patient 5, the primary had 34 large SVs (9 unique), the first recurrence had 27 (3 unique), the second recurrence had 25 (4 unique), the third recurrence had 23 (4 unique), and the germline had 30 (3 unique). There were 13 large SVs that were shared amongst the primary and recurrent tumors, and the germline (Figure 10E).

The results of this analysis provide further evidence for tumoral evolution in pGBM, shown by the unique large SVs present in each tumor sample that are not shared with the recurrent tumor.

4.3.3 Summary

In this section, I sought to characterize intrapatient tumoral evolution using matched primary and recurrent tumor samples. I looked at the number and composition of both SNVs and large SVs in all tumor samples and have provided evidence of disease progression and drastic changes in the mutational burden from primary to recurrence. Recurrent tumors share very few SNVs with the primary tumor in every patient. There are shared large SVs between primary and recurrences, but there is also a significant number of large SVs that are unique to either the primary or recurrence.

4.4 Inter and Intratumoral Heterogeneity

In order to address my hypothesis of pGBM being molecularly heterogeneous, I compared the SNVs and large SVs between patients (*section 4.4.1*) and I analyzed the presence of subclonal populations within tumors (*section 4.4.2*).

4.4.1 Intertumoral heterogeneity

I began by identifying any recurrent somatic, non-synonymous, exonic SNVs in all primary tumors and all the recurrent tumors. I identified 321 mutations in the primary tumors, 4 of which reoccurred in 2 patients (*EBF4*, *FAM160A2*, *RNF169*, and *VWF)* (Figure 11A). There were 420 lesions in the recurrent tumors, with 6 (*ASXL1*, *HTR1E*, *NBEAL1*, *POC5*, *PRUNE2* and *SZT2)* reoccurring in 2 patients (Figure 11B). There were no recurrent somatic mutations in all tumor samples.

When comparing the 408 large SVs among our patient cohort, I found 16 recurrent large SVs present in at least 2 or more patients. The genes associated with these variants were *ADAM5, ALG1L2, APOBEC3A/B, BTNL8/3, HLADRB1, LCE3C, LSP1, TNNT3, NEGR1, ORC41P, PRH1, RP11748L132, SIRPB1, UGT2B15, UGT2B28, VAMP7,* and *ATRX* (Figure 12). Of those 16, 4 were shared amongst all patients. A possible explanation as to why there were more recurrent large SVs than SNVs is that only somatic and exonic SNVs were analyzed, whereas the large SV data is genome wide and included both somatic and non-somatic variants.

The scarcity of common genetic lesions between patients provides strong evidence of intertumoral heterogeneity amongst pGBM patients, reinforcing the notion that this disease must be treated in a patient-specific manner.

A

genes	number of pGBMs with mutations
EBF4	2
FAM160A2	2
RNF169	2
VWF	2

B

genes	number of pGBMs with mutation
ASXL1	2
HTR1E	2
NBEAL1	2
POC5	2
PRUNE2	2
SZT2	2

Figure 11. Recurrent single nucleotide variants in pGBM samples. (A) Genes mutated in primary pGBM samples and the number of patients that harbor the mutation. (B) Genes mutated in recurrent pGBMs samples and the number of patients that harbor the mutation.

gene	number of patients
ADAM5	3
ALG1L2	3
APOBEC3A/APOBEC3B	2
BTNL8/BTNL3	3
HLADRB1	5
LCE3C	5
LSP1/TNNT3	4
NEGR1/RPL31P12	5
OR4C1P	2
PRH1	4
RP11748L132	3
SIRPB1	4
UGT2B15	4
UGT2B28	4
VAMP7	4
ATRX	5

Figure 12. Recurrent large structural variants in pGBM samples. Genes associated with large structural variants and the number of patients that harbor the variant.

4.4.2 Intratumoral heterogeneity

This next analysis addressed whether pGBM harbours a high degree of intratumoral heterogeneity. My WGS data was used by Aaron Gillmor to infer the number of subclones in each tumor sample and to create phylogenetic trees. The distance between the branches is proportional to the number of genetic differences between subclones (Figure 13). Mutational signatures for each subclone or group of subclones were assigned based on somatic SNVs (Figure 14).

Patient 1 had 8 subclonal lineages in the primary sample with subclones 1e and 1f clustering close to the germline, and the other subclones diverging into two major phylogenetic branches, one consisting of subclones 1a, 1b, 1c, 1d and 1g, and the other consisting of 1h (Figure 13A). The mutational signatures for the subclones in patient 1 showed 6 mutational signatures for the subclones 1e and 1f, while the other subclones were mainly composed of signature 1, which is common in most cancer types (Figure 14A). The other patients follow similar patterns, subclones from the primary tumors tend to cluster separately from the subclones in the recurrent tumors (Figure 13B-E). An interesting exception to the pattern is the clustering of recurrent subclones close to the germline. This is most clearly evident in patient 5 where after every recurrence, a subset of subclones clusters near the germline (Figure 13E). This observation raises the intriguing possibility of germline variants contributing to the propagation of tumor recurrences. Taken together, these observations show extensive intratumoral genetic heterogeneity and the presence of subclonal architecture in pGBM.

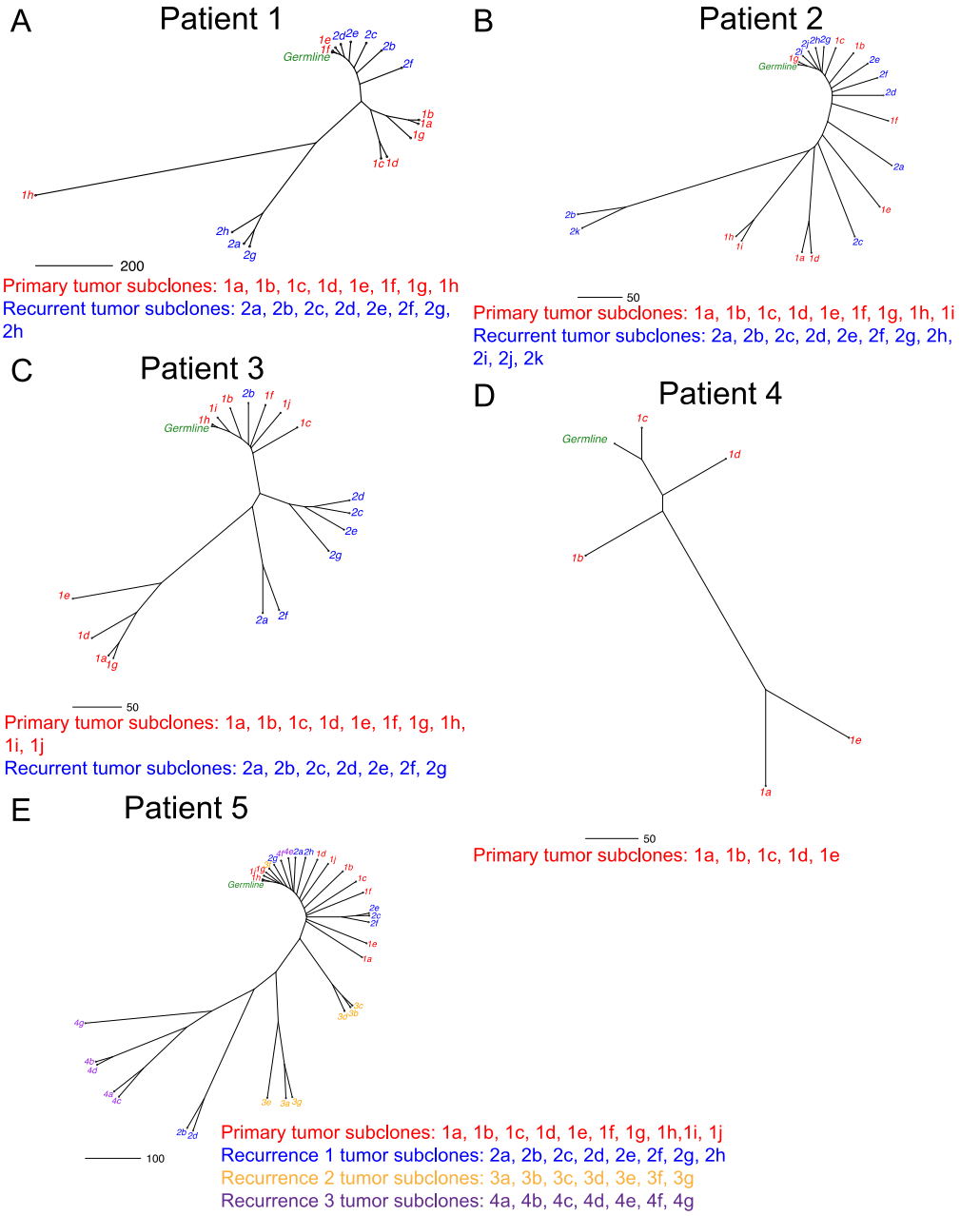
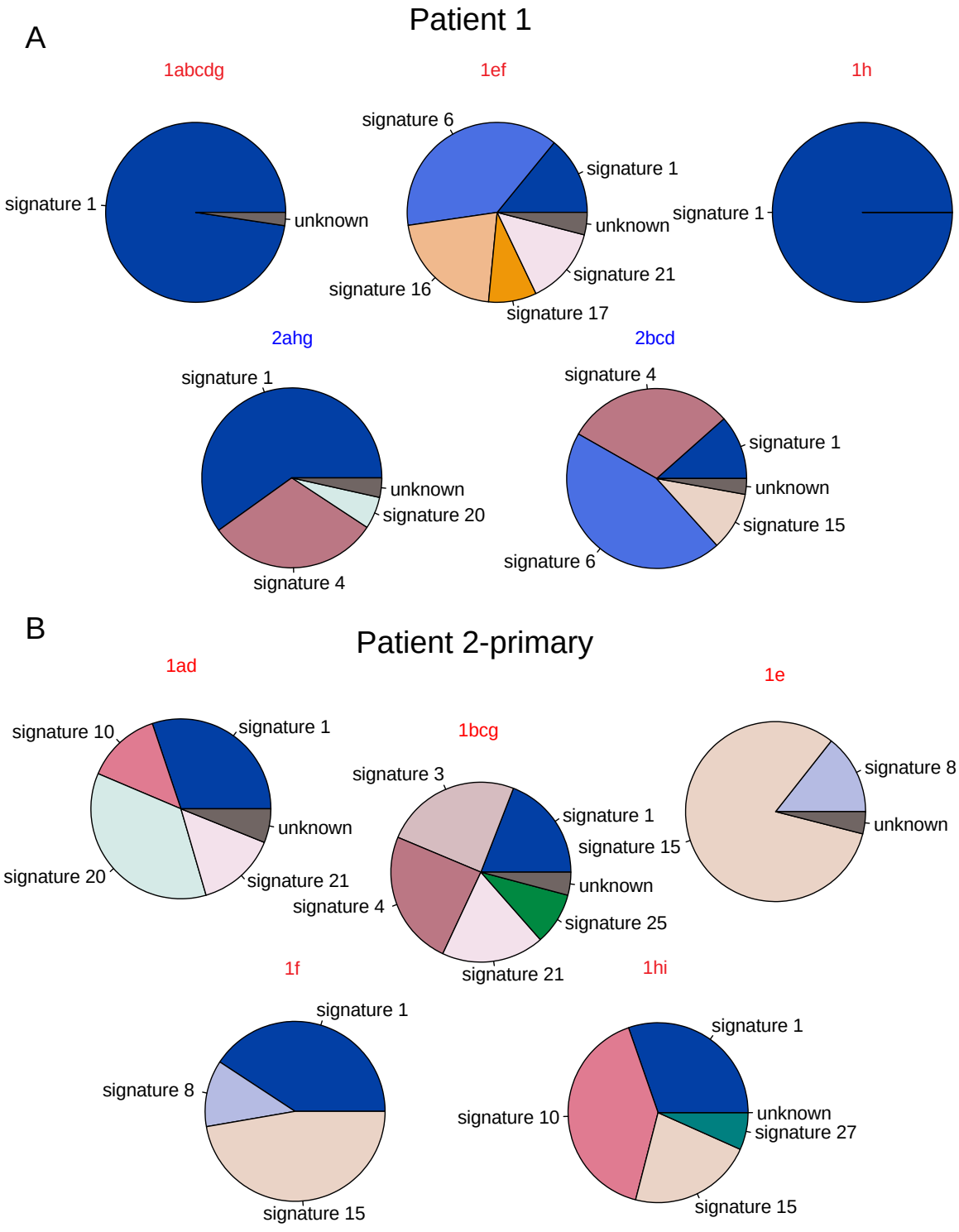
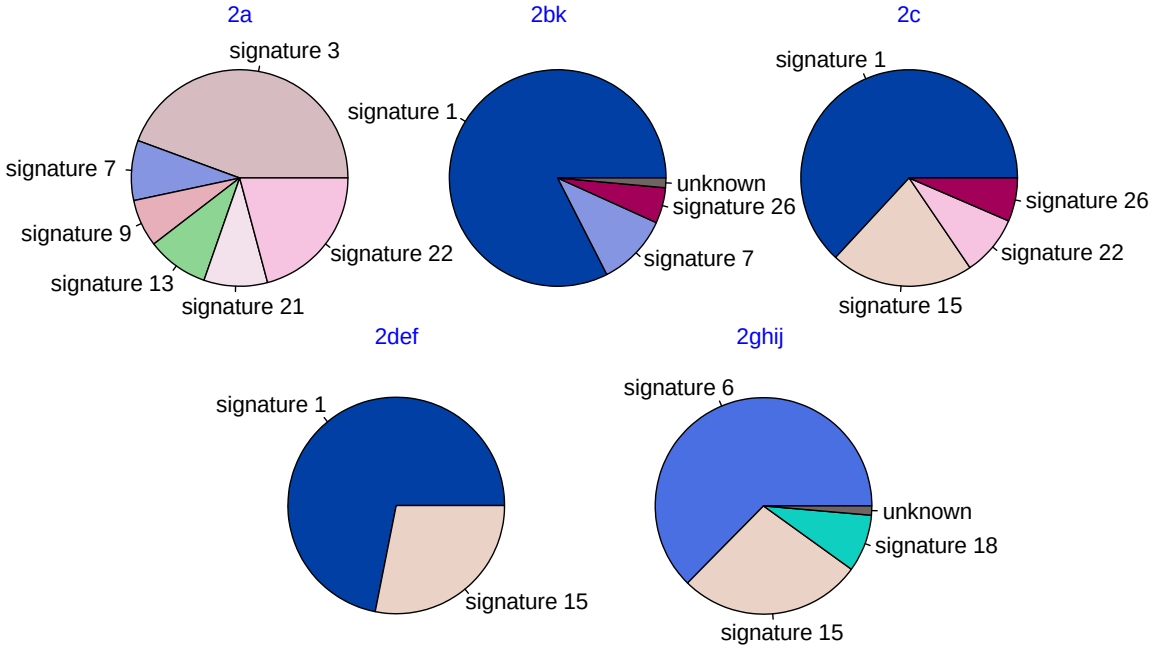


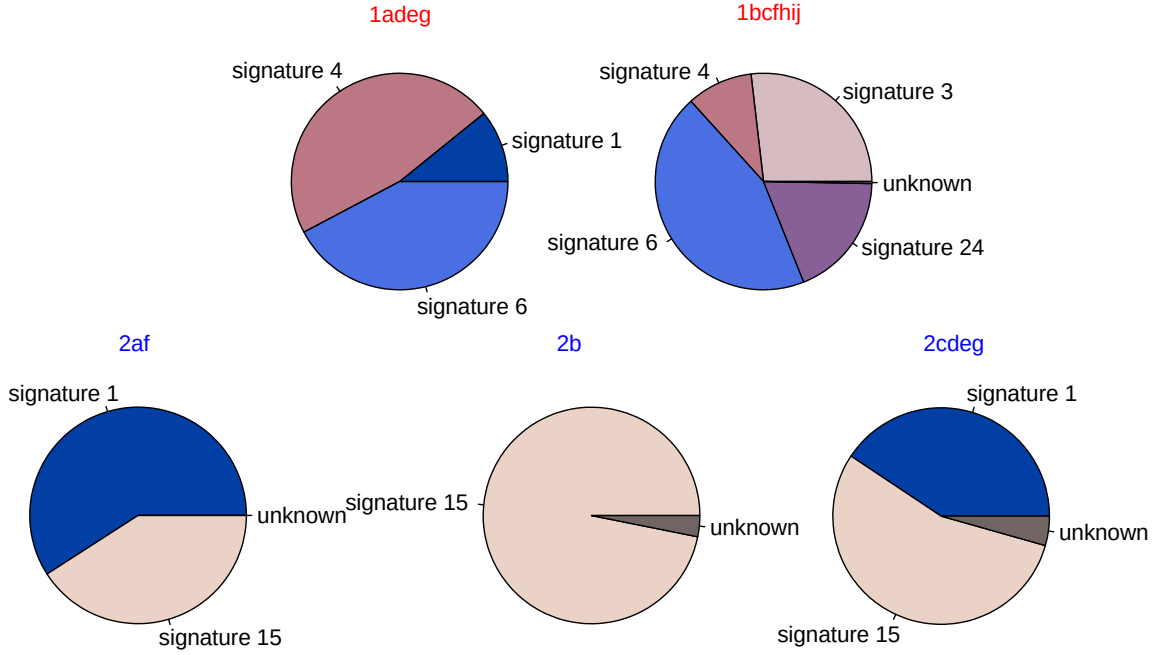
Figure 13. Subclonal architecture in pGBM. (A) Patient 1 (B) Patient 2 (C) Patient 3 (D) Patient 4 (E) Patient 5. Data from whole genome sequencing was used to infer the number of subclones in each tumor sample. Distance between nodes is proportional to the number of genetic differences between subclones. Subclones identified in the same sample are represented with the same color.


A
Patient 1
1abcdg
signature 1
unknown
1ef
signature 6
signature 1
unknown
signature 21
signature 16
signature 17
1h
signature 1
2ahg
signature 1
unknown
signature 20
signature 4
2bcd
signature 4
signature 1
unknown
signature 15
signature 6
B
Patient 2-primary
1ad
signature 10
signature 1
unknown
signature 20
signature 21
1bcg
signature 3
signature 1
signature 15
unknown
signature 4
signature 25
signature 21
1e
signature 8
unknown
1f
signature 1
signature 8
signature 15
1hi
signature 1
signature 10
unknown
signature 27
signature 15

C

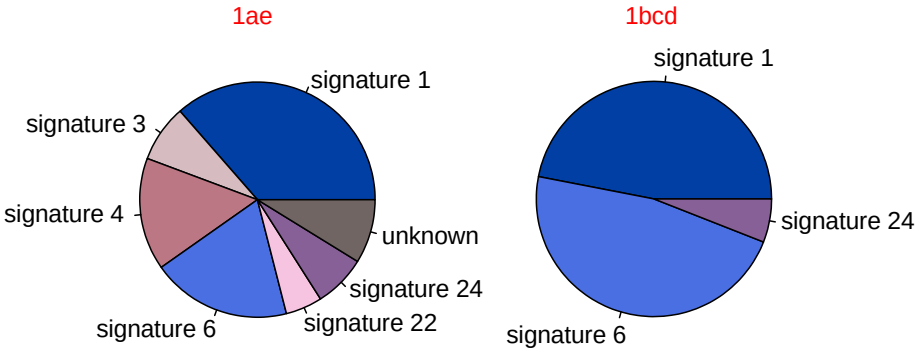

Patient 2-recurrence
2a
signature 3
signature 7
signature 9
signature 13
signature 21
signature 22
2bk
signature 1
unknown
signature 26
signature 7
2c
signature 1
signature 26
signature 22
signature 15
2def
signature 1
signature 15
2ghij
signature 6
unknown
signature 18
signature 15

D


Patient 3
1adeg
signature 4
signature 1
signature 6
1bcfhij
signature 4
signature 3
unknown
signature 6
signature 24
2af
signature 1
unknown
signature 15
2b
signature 15
unknown
2cdeg
signature 1
unknown
signature 15

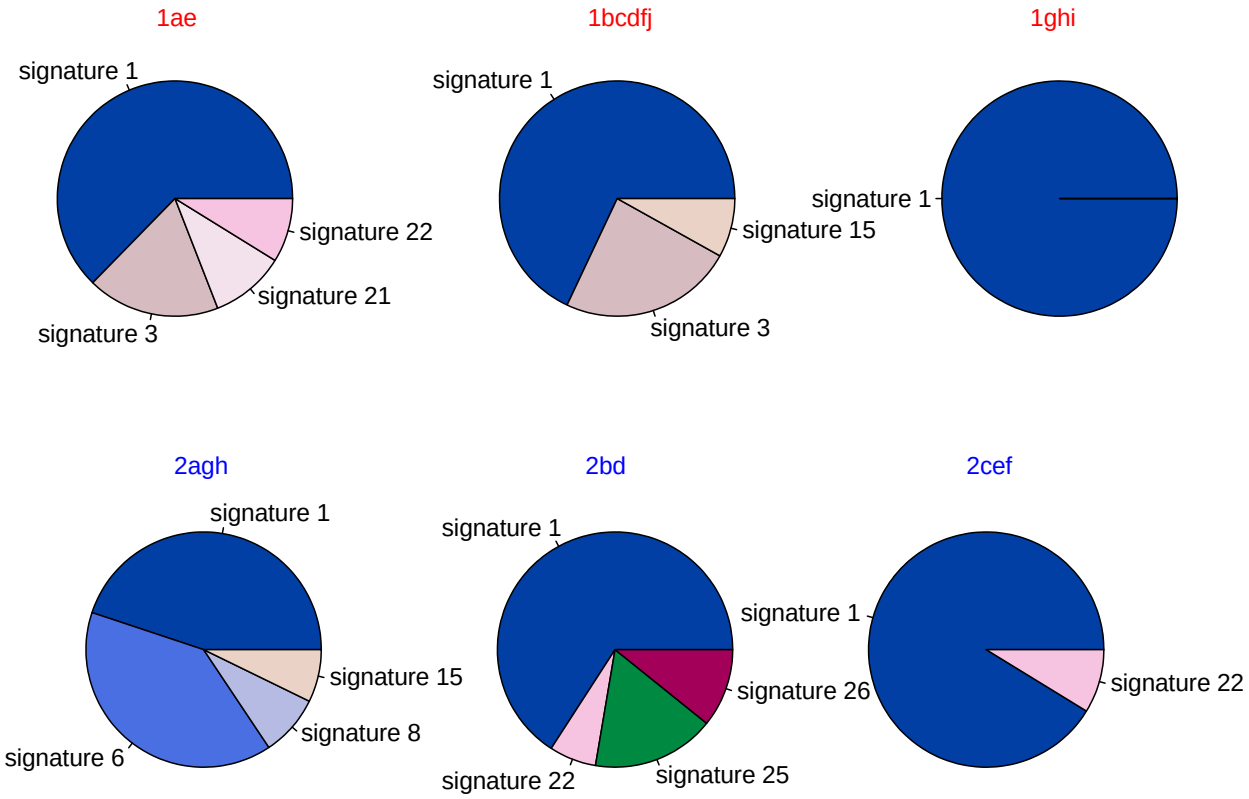
E

Patient 4



F

Patient 5-primary, recurrence 1



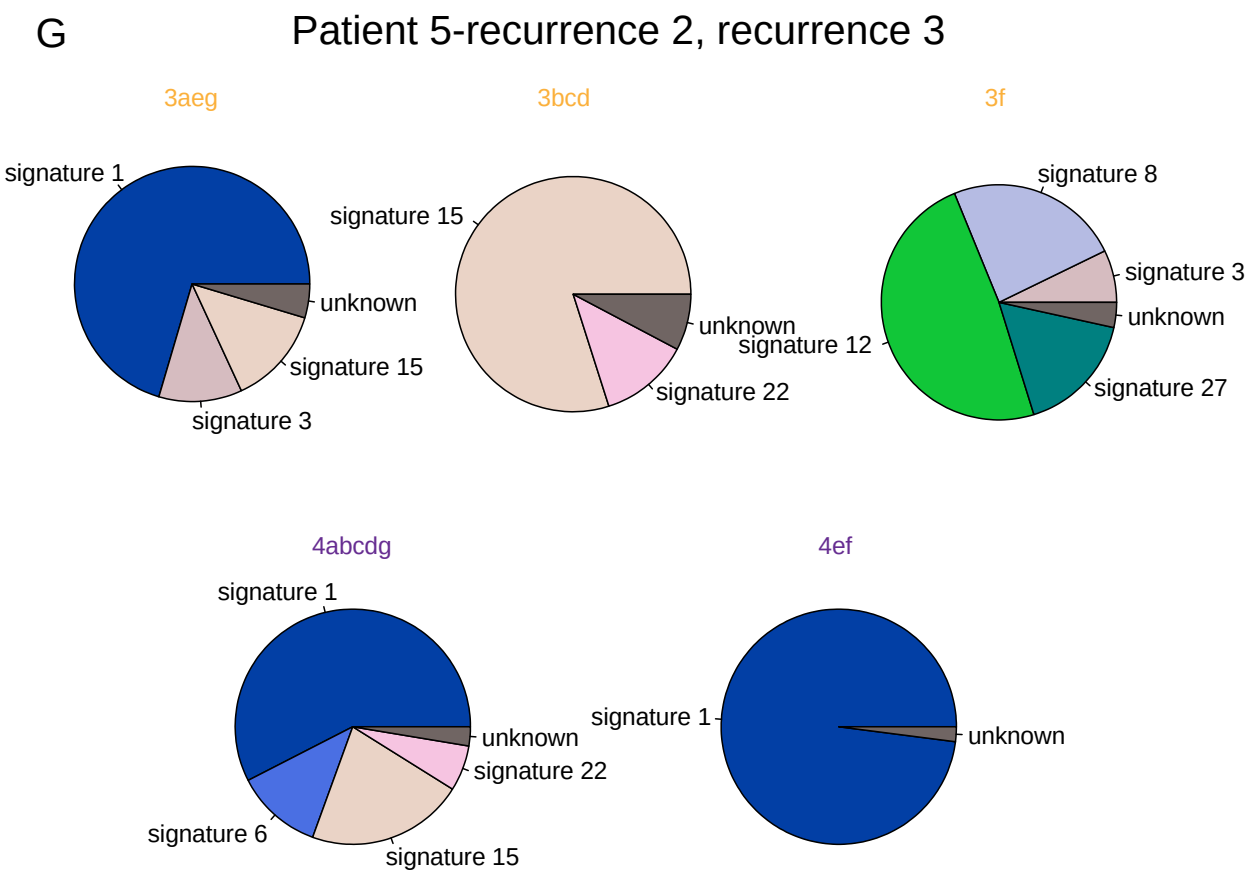


Figure 14. Mutational signatures in pGBM samples. (A) Mutational signatures in the primary (1a-h) and recurrence (2a-g) of patient 1. (B) Mutational signatures in the primary tumor (1a-i) of patient 2. (C) Mutational signatures in the recurrent tumor (2a-k) of patient 2. (D) Mutational signatures in the primary (1a-J) and recurrence (2a-g) of patient 3. (E) Mutational signatures in the primary tumor (1a-e) of patient 4. (F) Mutational signatures in the primary (1a-j) and first recurrence (2a-h) of patient 5. (G) Mutational signatures in the second recurrence (3a-g) and third recurrence (4a-g) of patient 5. COSMIC signatures were assigned based on SNVs for each tumor.

4.4.3 Summary

I have shown that pGBM is molecularly heterogeneous both between patients and within a single tumor. There is a paucity of shared genetic lesions between patients in the cohort, as evident when analyzing both SNVs and large SVs. The evidence of intratumoral heterogeneity of pGBM is consistent with previous findings (Vinci et al., 2018) and adds additional support for subclonal architecture in both primary and recurrent tumors from the same patient.

4.5 Causative Germline Mutations

Based on my findings of unique variants in the germline samples (Figure 10) and the presence of recurrent subclonal lineages clustering with the germline (Figure 13), I decided to further investigate the germlines of the pGBM patient. To do this, I used the same approach I took for the large SVs, aligning the germline and tumor samples independently to the GRCh38 reference genome (Figure 15). This allowed me to assess whether there were any germline mutations, either SNVs or large SVs which could potentially be contributing to tumorigenesis.

4.5.1 Unique germline variants

The findings of unique large SVs in the germlines of pGBM patients prompted further investigation into this mechanism; "unique" was defined as having breakpoints >2 kbp away from similar SVs in different samples. I focused on the large SVs present in the germline of each patient and by tracking the changes in allelic frequency (AF) throughout tumor progression, I observed dynamic changes. The complete loss of SVs was primarily found in the subclonal germline variants (AF <0.4) (Figure 16) whereas variants with AF >0.4 were relatively stable throughout tumor progression (Figure 17, 18). The observation of subclonal SVs in the germline samples was

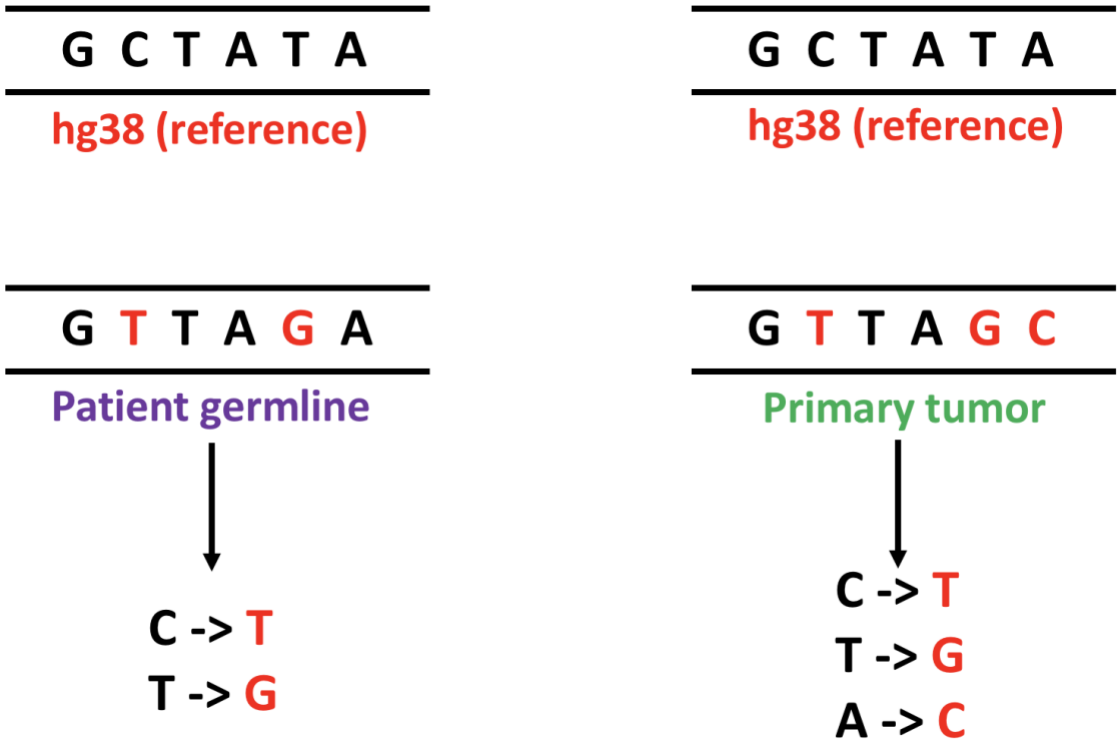


Figure 15. Schematic of germline single nucleotide variant calling. The GRCh38 human genome is used as the reference genome when calling single nucleotide variants in all patient samples. Germline and tumor samples are analyzed independently.

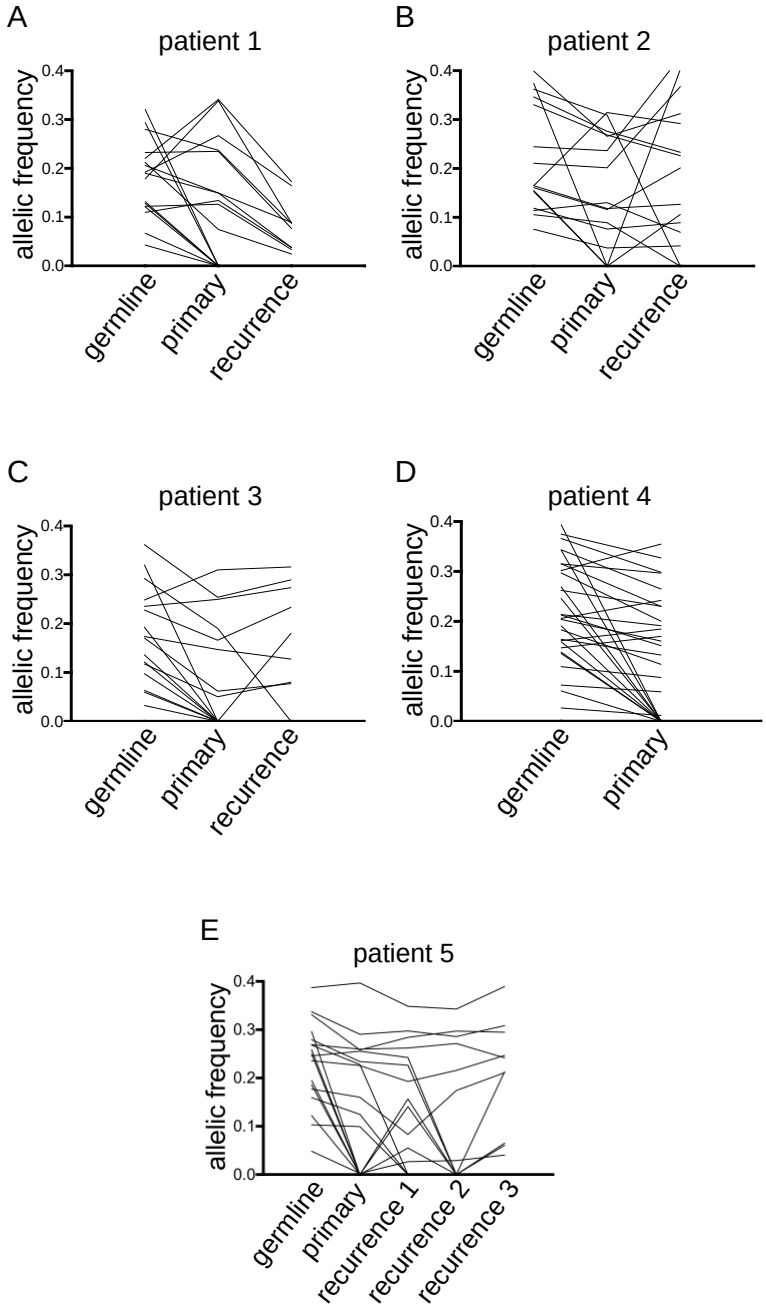


Figure 16. Dynamic contributions of germline structural variants to tumor genomes. Each line represents the dynamics of structural variants that have an allelic frequency < 0.4 in the germline for primary tumors and their recurrences. (A) Patient 1 (B) Patient 2 (C) Patient 3 (D) Patient 4 (E) Patient 5.

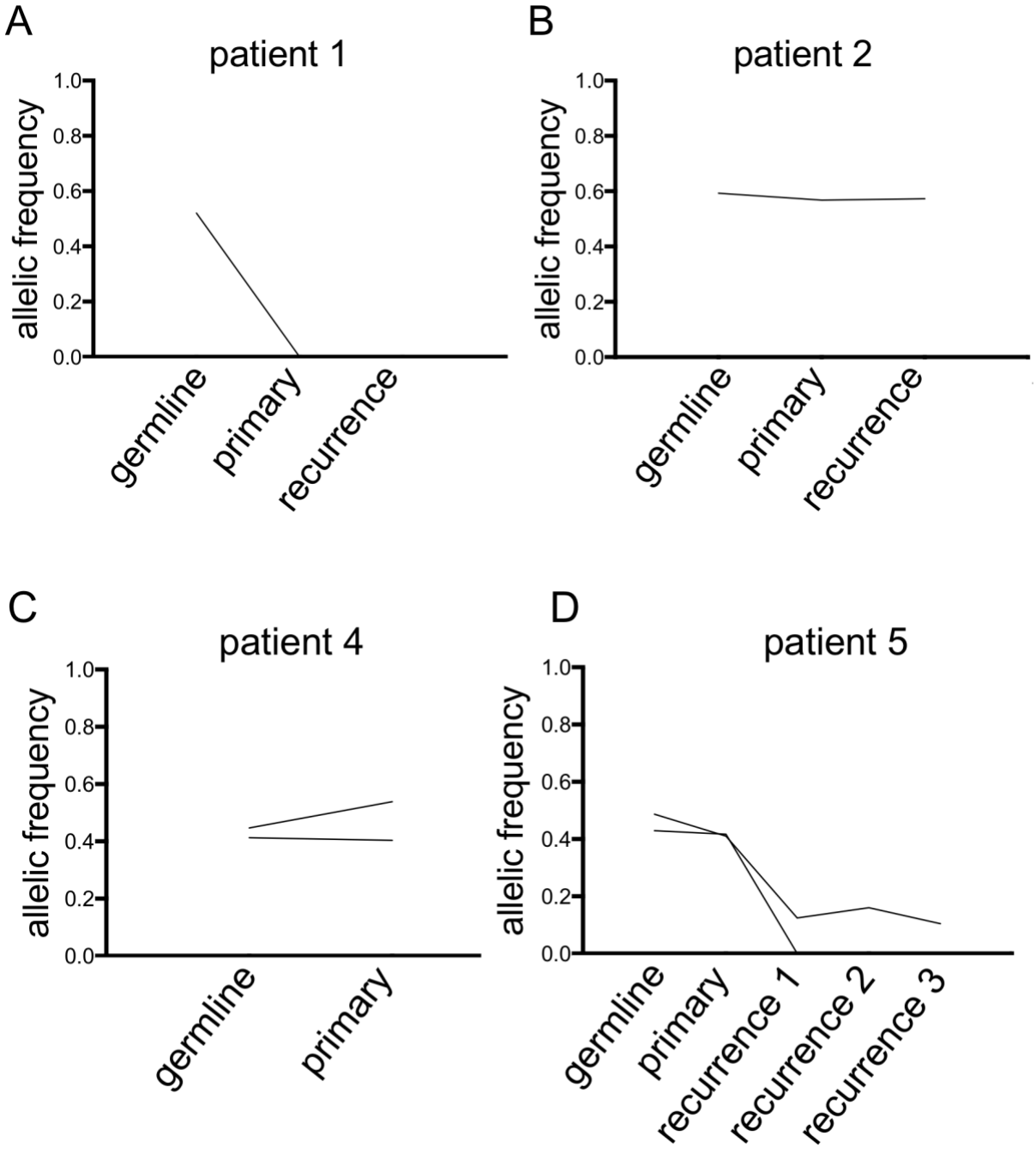


Figure 17. Dynamic contributions of germline structural variants to tumor genomes. Each line represents the dynamics of SVs that have an allelic frequency between 0.4-0.7 in the germline for primary tumors and their recurrences. (A) Patient 1 (B) Patient 2 (C) Patient 4 (D) Patient 5. None for patient 3.

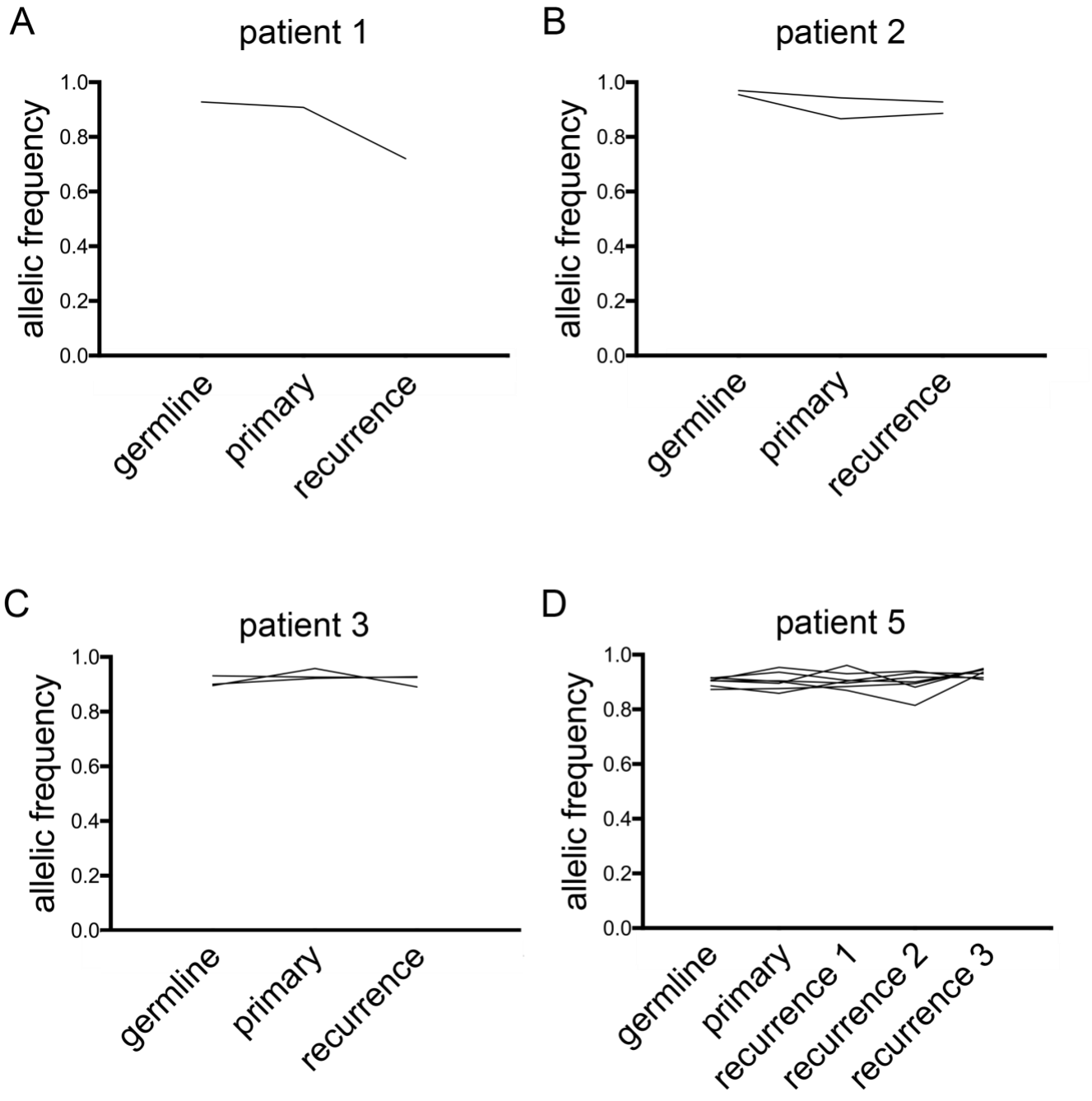


Figure 18. Dynamic contributions of germline structural variants to tumor genomes. Each line represents the dynamics of SVs that have an allelic frequency > 0.7 in the germline for primary tumors and their recurrences. (A) Patient 1 (B) Patient 2 (C) Patient 3 (D) Patient 5. None for patient 4.

surprising and perhaps even more striking was the loss of these variants in the tumor tissue. Fluctuations in the allelic frequency of germline variants suggests they are playing a dynamic role in tumor formation and recurrence.

4.5.2 Mutations in hereditary oncogenes

I next aimed to determine whether pGBM patients were harboring deleterious mutations in genes associated with familial cancer syndromes. Ana Nikolic curated a list of 23 known hereditary oncogenes and performed an analysis to detect germline events in those genes. Even though none of the patients harbored any of the described syndromes, I found 9 hereditary oncogenes genes which were recurrently mutated in 2 or more patients (*APC, BRCA1, BRCA2, MLH1, MSH6, NF1, PMS2, SMARCA4, TP53)* (Figure 19). I checked to see if these mutations were de novo in patient 1 and 3, by using my WGS data from their patient germline samples. I did find the mutations in the parent's blood samples, confirming that they were not de novo germline mutations. I do not have samples from patient 2, 4, or 5's parents, so I was not able to confirm whether they were de novo or not. These data suggest that these mutations might not be pathogenic on their own but could potentially be cooperating with other germline variants to enhance tumor growth and maintenance.

4.5.3 Genes associated with CNS syndromes

While analyzing all the exonic, non-synonymous SNVs, I noticed that a fraction of genes were involved in hereditary syndromes, according to the OMIM database. Although gene ontology analysis did not identify any significantly enriched class of mutated genes, there was a proportion (19.6%) of the genes mutated in the primary tumors which were associated with hereditary

gene	number of patients with mutation
APC	4
BRCA1	3
BRCA2	5
MLH1	2
MSH6	3
NF1	3
PMS2	4
SMARCA4	3
TP53	5

Figure 19. Recurrent germline mutations associated with familial cancer syndromes. Gene associated with a hereditary cancer susceptibility and the number of patients with a mutation in that gene.

syndromes. 41.3% of these genes are associated with syndromes of the central nervous system (CNS) (Figure 20A). Similarly, 20.2% of the 420 genes mutated in all recurrences were associated with hereditary syndromes, 35.2% of which were linked to CNS syndromes (Figure 20B). This data provides additional rationale for pGBM having a hereditary component.

4.5.4 Subclonal germline ATRX deletion

One subclonal germline deletion that stood out was a 211 kbp deletion in *ATRX* (Figure 21A) found in every patient (Figure 12). This deletion was interesting because loss-of-function mutations in *ATRX* have been described in 1/3 of pGBM patients but not a deletion, especially of this magnitude (He et al., 2018). To validate this putative deletion, I performed custom Taqman copy number assays that bind to the deleted locus (Figure 21B). The results of the assay, as shown by a significant decrease in the fluorescent signal in the patients' germline and tumor when compared to a control genome, provides further evidence for the *ATRX* deletion (Figure 21C, 22). Furthermore, PDXs were generated by Jenny King and Ngoc Ha Dang with cells from the recurrence of patient 3, and the second recurrence of patient 5 in order to correlate RNA expression data with my whole-genome data. I used these PDXs to do scRNA-seq and, as predicted by the WGS and Taqman assays, I observed heterogeneous *ATRX* expression in both patients (Figure 21D, E).

4.5.5 Recurrent deletion upstream of neuronal growth regulator 1 (NEGR1)

Another large SV that was detected in all patients' tumor samples and germlines (Figure 12) was a deletion directly upstream of the *NEGR1* gene, potentially at its promoter site. I found evidence of both heterozygous and homozygous deletions in patients (Figure 23A). The protein

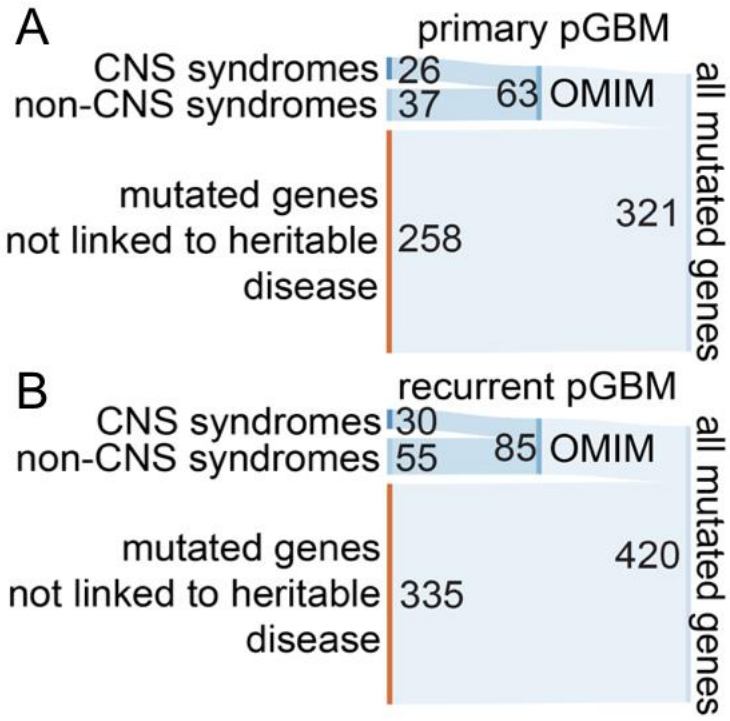


Figure 20. Single nucleotide variants in genes associated with syndromes. (A) Analysis of all exonic, non-synonymous SNVs in all primary tumors reveals a fraction of genes involved in hereditary syndromes, according to the OMIM database. Out of the 63 genes with links to hereditary disease, 26 were specifically linked to central nervous system (CNS) syndromes. (B) Analysis of all exonic, non-synonymous SNVs in all recurrent tumors Out of the 85 genes with links to hereditary disease, 30 were specifically linked to CNS syndromes.

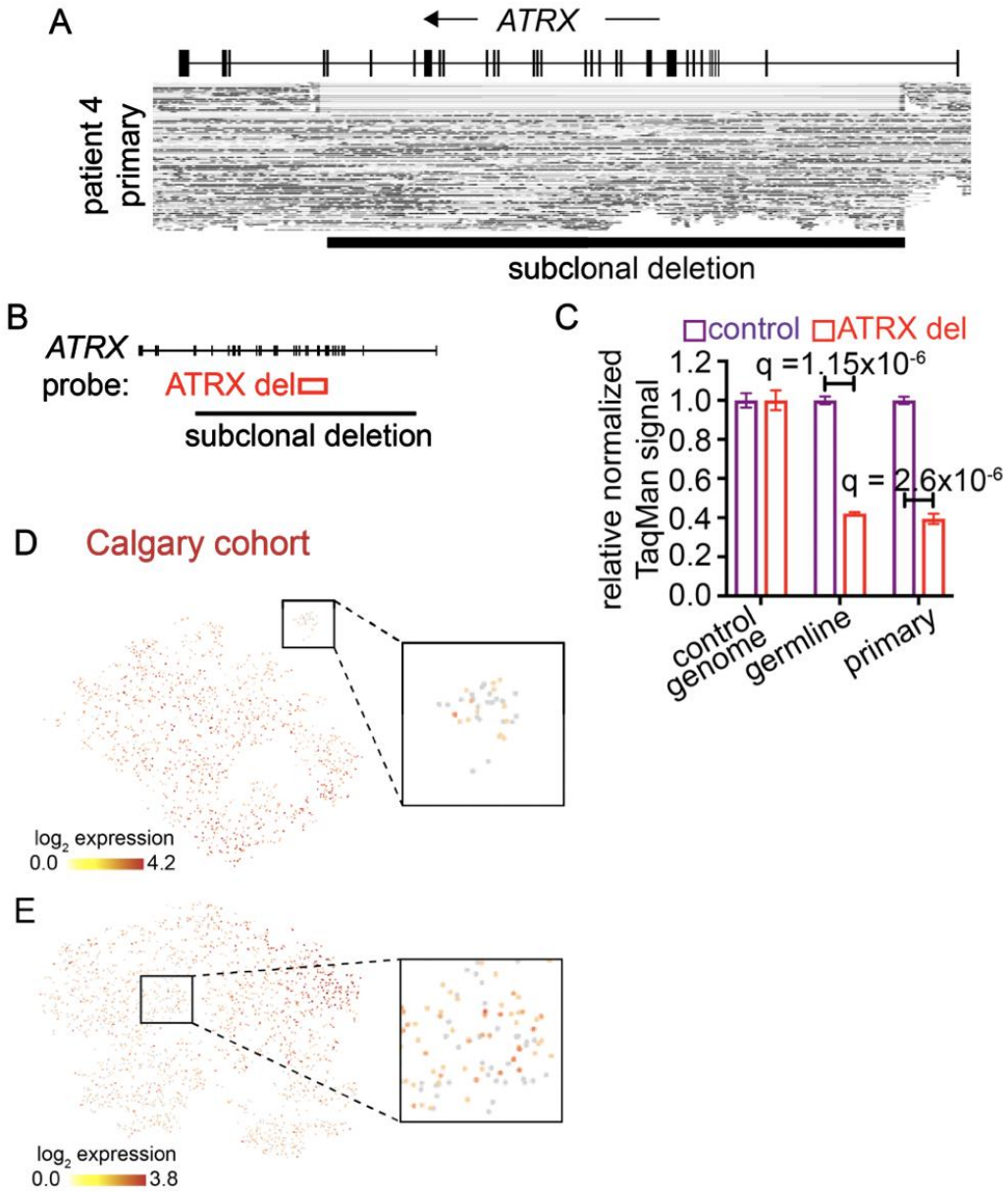


Figure 21. Subclonal ATRX deletions in the germline and tumors of pGBM patients. (A) Putative subclonal deletions at the *ATRX* locus in pGBM patient. (B) Diagram describing the binding locations of the TaqMan probe (ATRX del) designed to detect copy number variations. (C) TaqMan assays to detect copy number variation at the *ATRX* locus on samples for patient 4. (D) Single-cell RNA-seq data derived from patient's 3 recurrence. *ATRX* expression levels are represented in this tSNE plot. (E) Single-cell RNA-seq data for recurrence 2 of patient 5. This tSNE plot displays expression levels for ATRX.

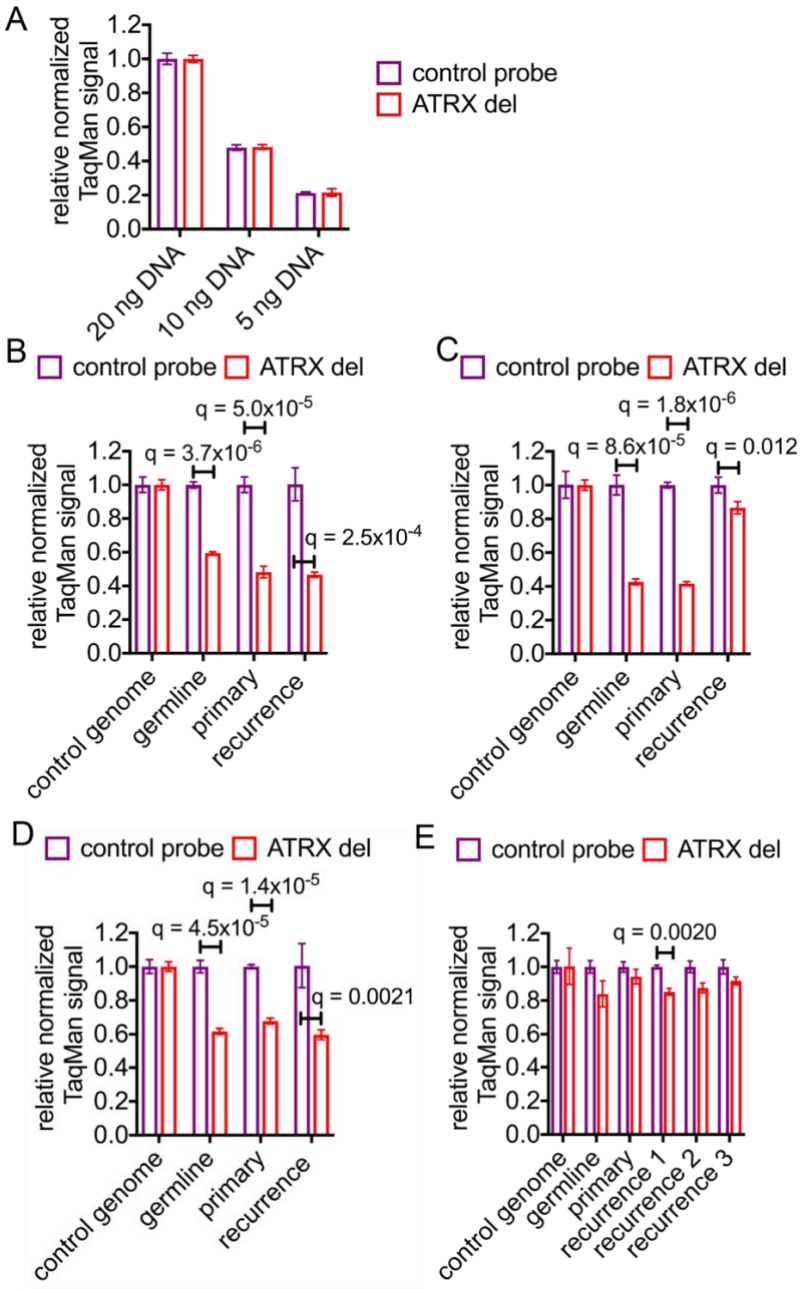


Figure 22. Copy number assays using TaqMan probes confirm the deletion at the *ATRX* locus. (A) TaqMan probes were assessed for their sensitivity with decreasing amount of genomic DNA as a template. All probes responded comparably to changes in amounts of DNA. (B) Copy number TaqMan assay for patient 1. (C) Copy number TaqMan assay for patient 2. (D) Copy number TaqMan assay for patient 3. (E) Copy number TaqMan assay for patient 5. For all assays, multiple t test statistics was used to derive q values.

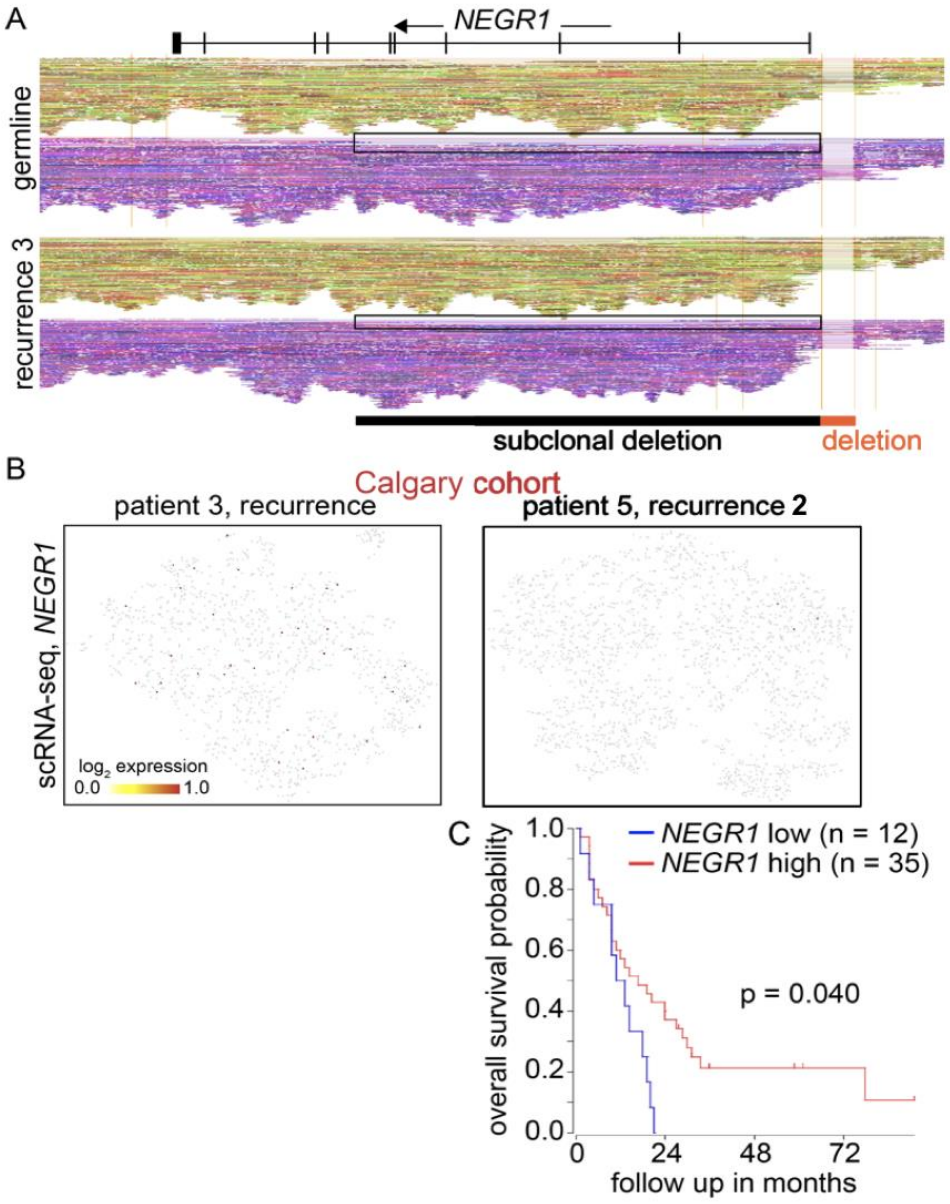


Figure 23. Germline deletions at the *NEGR1* locus. (A) Coverage of linked-reads at the *NEGR1* locus. Linked-reads are grouped based on haplotypes (red/green hues for one haplotype, purple/violet for the second). A deep deletion is observed in the region immediately upstream of the gene for the germline and the third recurrence for a pGBM patient. A subclonal deletion that extends into the gene body is also visible. (B) Single-cell RNA seq for a pGBM xenograft derived from the recurrence of patient 3. Expression levels of *NEGR1* are shown. (C) Kaplan Meier curve for pediatric GBM patients in the Paugh dataset, stratified by expression of *NEGR1*. Log-rank p = 0.040.

that is encoded by this gene is normally expressed in brain (Figure 24), but my scRNA-seq data from the PDXs shows loss of expression (Figure 23B). This deletion is significant because loss of NEGR1 has been detected in many cancer types and has been shown to contribute to an increase in cell migration and invasion (H. Kim, Hwang, Lee, Hong, & Lee, 2014). Another important aspect of this genetic lesion is that by analyzing a previously published pGBM cohort (Paugh et al., 2010), overall survival of patients could be stratified based on *NEGR1* expression, with patients having lower expression experiencing worse prognosis (Figure 23C).

4.5.6 Alterations in immune stimulatory genes

Two more recurrent alterations that I found in both the tumor samples and germlines of pGBM patients were deletions in the genes butyrophilin-like protein 8/3 (*BTNL8/BTNL3)* (Figure 25) and *apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A/3B* (*APOBEC3A/APOBEC3B)* (Figure 26). Data from the scRNA-seq show loss of expression of all 4 genes (Figure 25C, D and Figure 26C, D).

4.5.7 De novo germline mutations

In order to detect potentially causative de novo germline mutations, I performed WGS on blood samples from the parents of patient 1 and patient 3 (Figure 4). I identified all the unique SNVs that were present in the patient's blood and absent in the parent's blood. Patient 1 had 10 de novo, exonic, non-synonymous mutations (Figure 27A) and patient 3 had 4 de novo, exonic, non-synonymous mutations (Figure 27B).

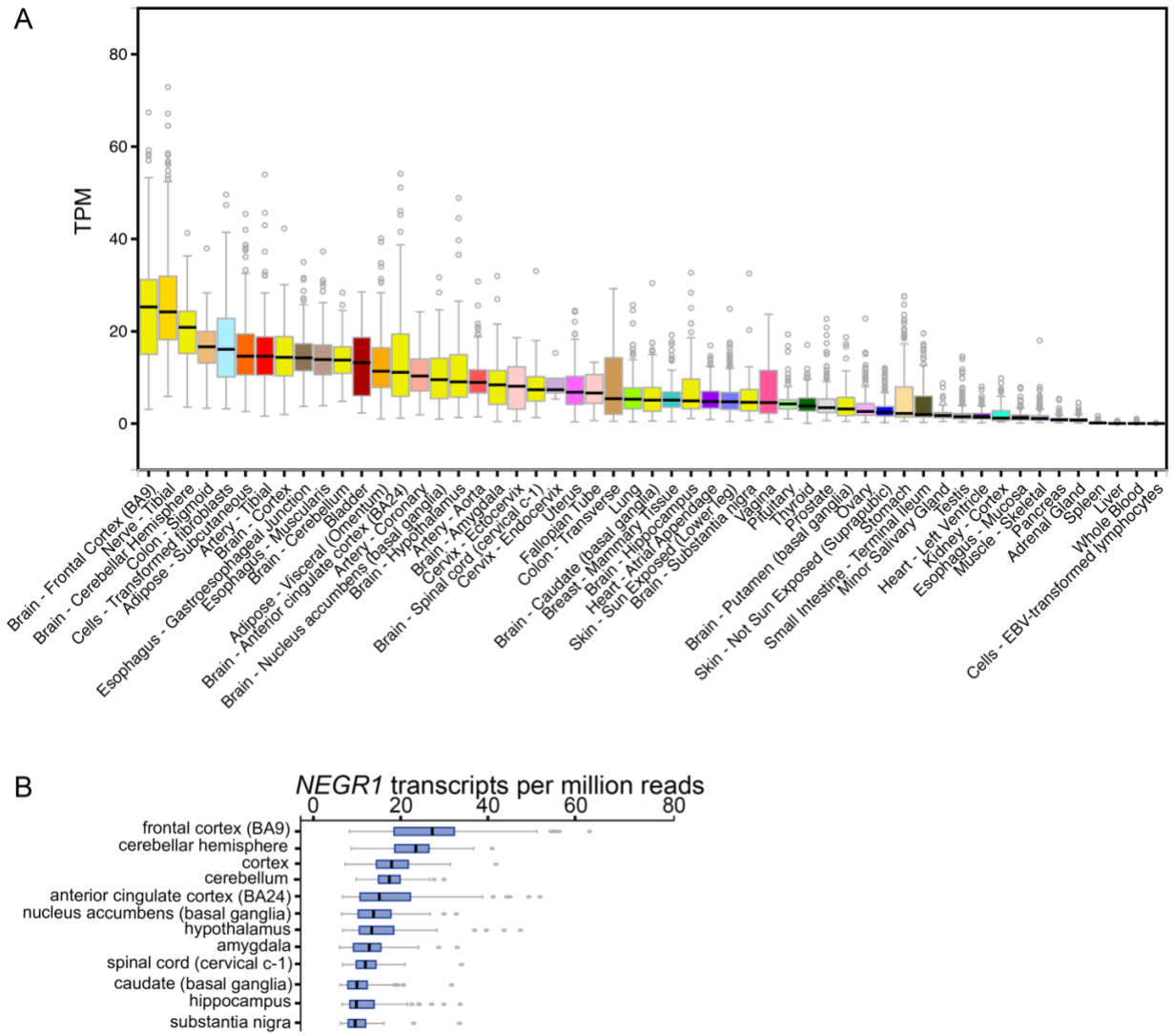


Figure 24. Expression of *NEGR1* in human tissues. (A) Expression of *NEGR1* was plotted for different human tissues. Source: GTex database. (B) Expression of *NEGR1* is brain tissues, based on data from the GTex database.

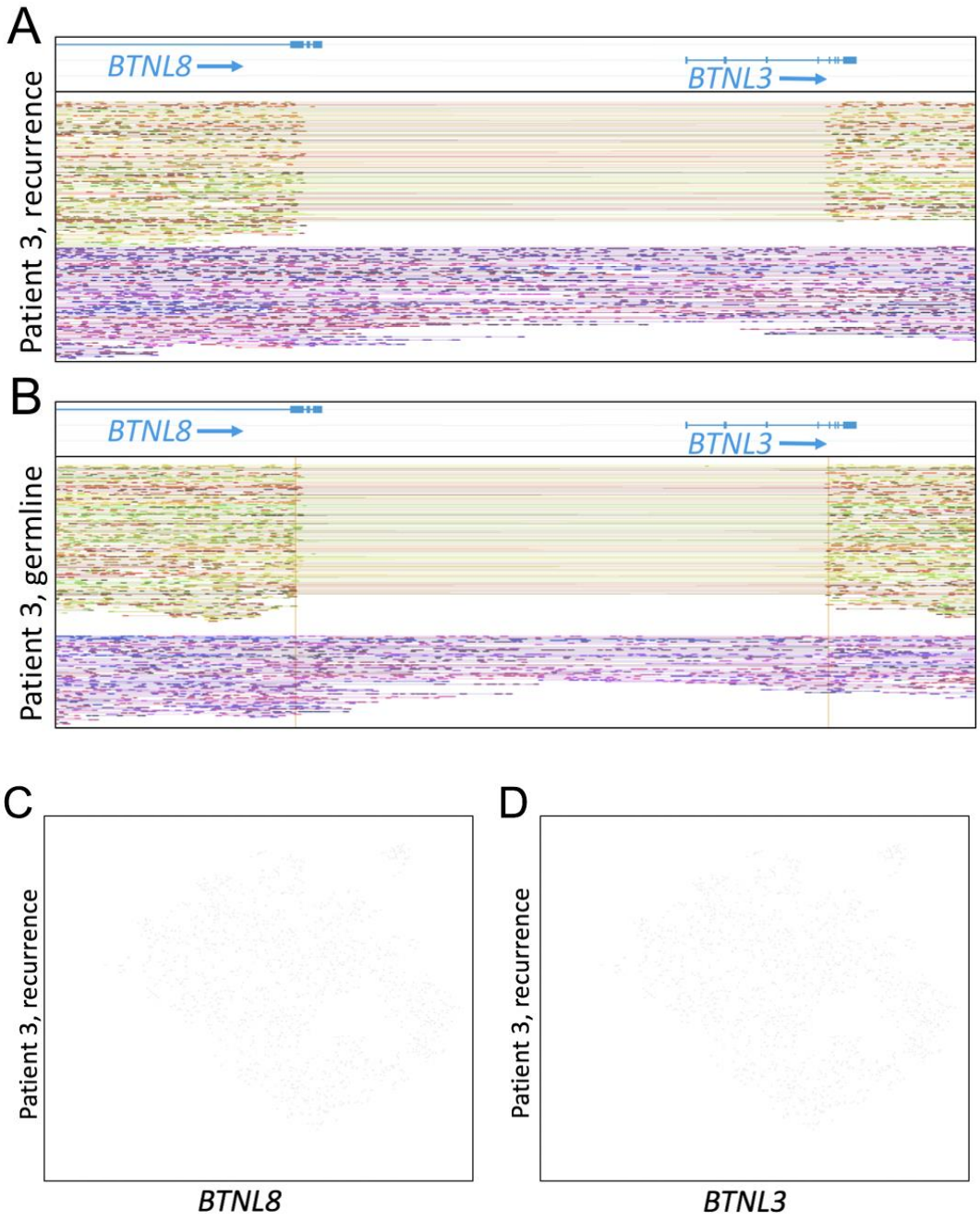


Figure 25. Heterozygous germline deletion between *BTNL8* and *BTNL3.* (A) Patient 3, recurrence (B) Patient 3, germline (C) Single-cell RNA seq data derived from the recurrence of patient 3. Expression levels of *BTNL8* are shown in the tSNE plot (D) *BTNL3* are shown in the tSNE plot.

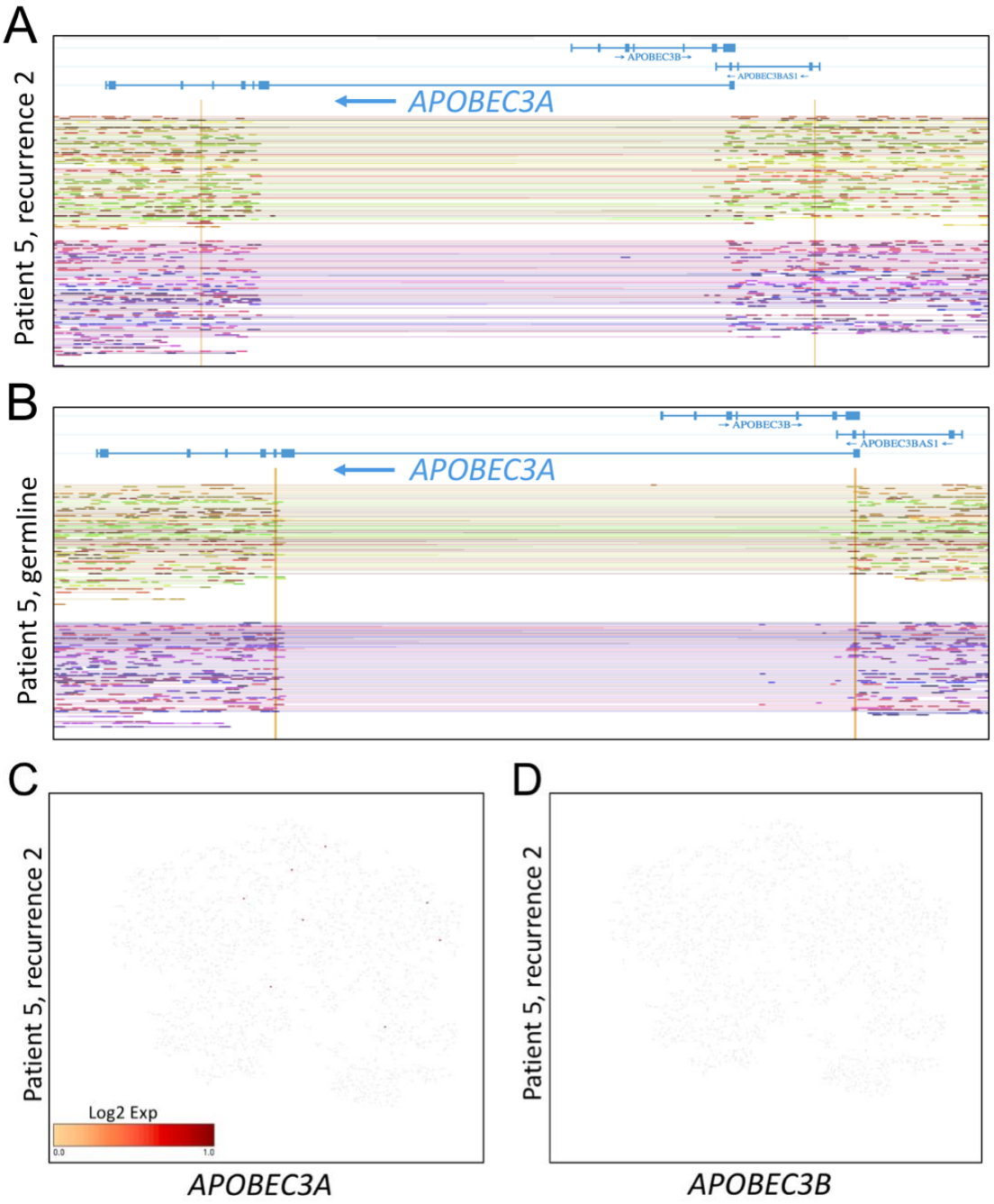


Figure 26. Homozygous germline deletion of *APOBEC3A* and *APOBEC3B*. (A) Patient 5, recurrence 2 (B) Patient 5, germline (C) Single-cell RNA seq data derived from the second recurrence of patient 5. Expression levels of *APOBEC3A* are shown in the tSNE plot (D) *APOBEC3B* are shown in the tSNE plot.

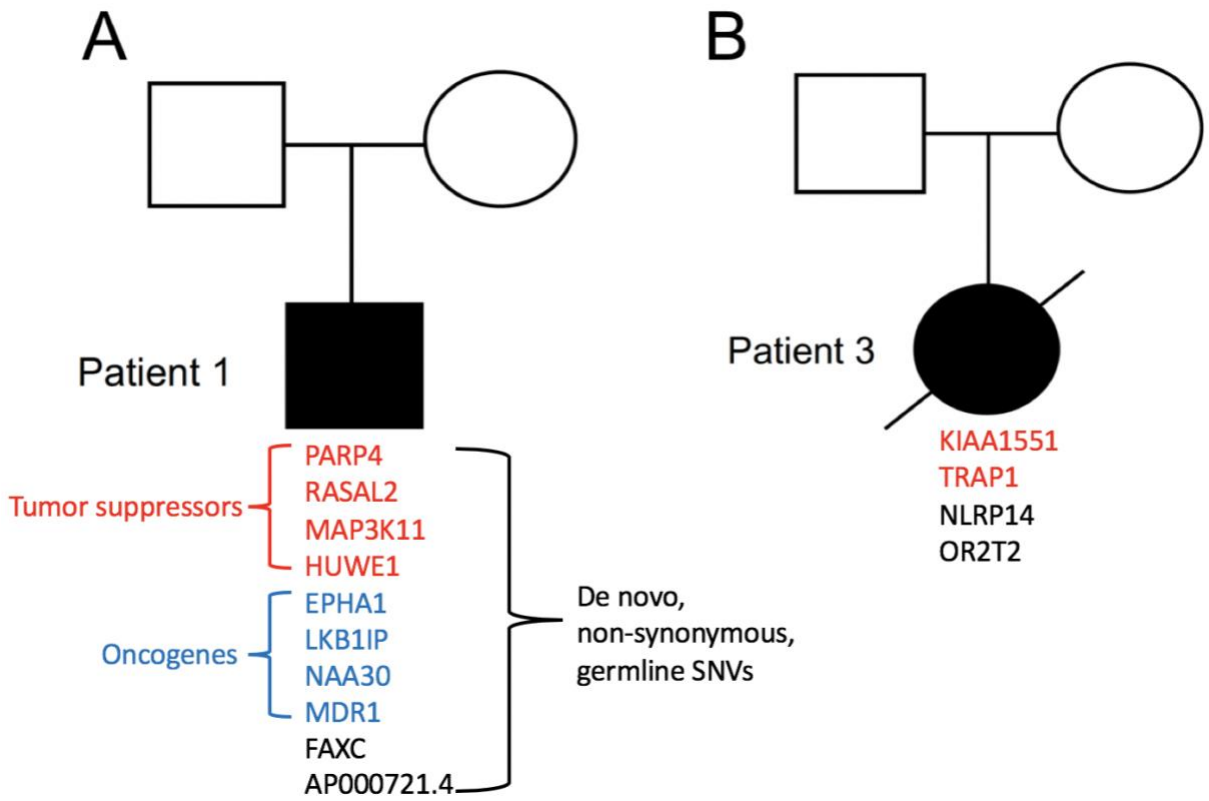


Figure 27. De novo single nucleotide variants in the pGBM patient's germline. (A) De novo non-synonymous SNVs in 10 genes that are present in the germline of patient 1 but not in either of the parent's germline are shown. (B) De novo non-synonymous SNVs in 4 genes that are present in the germline of patient 3 but not in either of the parent's germline are shown.

4.5.8 Summary

In this section, I aimed to identify causative mutations present in the germline of pGBM patients. I found recurrent large SVs that are potentially cooperating in tumor maintenance including a subclonal *ATRX* deletion, a deletion upstream of *NEGR1*, and recurrent deletions in immune stimulatory genes. I have also identified de novo, exonic, non-synonymous SNVs in 2 pGBM patients. Taken together, these data suggest a novel role for the germline and its contribution to carcinogenesis of pGBM.

Chapter Five: **Discussion**

pGBM is a disease which has been overlooked due to the assumption that it is akin to adult GBM, which has been studied relentlessly. With recent insights showing that these malignancies diverge at both the genetic and transcriptional level, comes the need for independent characterization of pGBM. By utilizing a rare collection of matched primary and recurrent tumor samples, I have characterized evolutionary trajectories of pGBM, intratumoral heterogeneity, and a novel germline predisposition factor.

5.1 Extreme Tumoral Evolution Upon Recurrence

In my first aim, I sought to investigate the extent of mutational evolution in tumors when they reoccur following treatment. Using a longitudinal collection of primary and recurrent tumors, I addressed this aim by performing WGS on both primary and recurrent samples and comparing the mutational landscape between the malignancies.

Initially, I described differences and similarities between the SNVs in the matched primary and recurrent tumor samples. This analysis showed very few (<5) SNVs being propagated throughout disease progression in each case. This finding of extreme tumoral evolution upon recurrence was surprising as I expected the recurrence to share many of the same genetic lesions with the primary tumor. I also expected evident driver mutations being shared amongst samples, but this was not observed. Rather, I observed an almost completely separate entity emerge upon recurrence, which exemplifies the need for recurrent tumor sampling and genetic testing following tumor recurrence.

Interestingly, I also observed the number of SNVs remained relatively stable upon recurrence, with the exception of patient 1. This was surprising because these patients were treated with chemotherapy and radiation, which typically induces SNVs. One possible mechanism that explains this observation is that a population of slow cycling cells remained following treatment and thus were immune to the accumulation of mutations during treatment and subsequently repopulated the tumor and acquired new mutations. Another possible explanation is that the majority of cells were depleted by treatment and allowed rare subclonal populations to repopulate the recurrence. The mechanism of tumoral evolution needs to be elucidated further by conducting follow up experiments. Experiments which utilize the established pGBM PDX protocol and assess how the primary tumor responds to various treatments *in vivo* will be beneficial in characterizing how these tumors evolve and progress. Growing these cells *in vivo* will also allow for longitudinal experiments that examine the rate at which pGBM evolves when either left untreated or treated by collecting samples from various timepoints throughout disease progression under both conditions. These future experiments be beneficial in further characterizing tumoral evolution.

5.2 Interpatient Heterogeneity

I next investigated whether pGBMs are genetically heterogeneous between patients. To address this aim, I used WGS data to compare and contrast genetic lesions. Although I have highlighted a few previously described oncogenic mechanisms (*NTRK2* amplification, *ETV6-NTRK3* fusion, *MET* lesion) found in individual patients, the lack of consistency demonstrates the presence of alternative mechanisms driving this disease. In terms of SNVs, there were no recurrent somatic mutations shared amongst all tumor samples. While looking at the large SVs, I found 16

out of 408 which reoccurred in at least 2 or more patients and 4 large SVs present in all pGBM patients, in both germline and tumor samples. A reason for the small discrepancy between shared SNVs and large SVs could be caused by the fact that I analyzed SNVs in coding regions, but the large SVs were analyzed on a genome wide level. Regardless, these observations highlight a high degree of interpatient heterogeneity as observed by the paucity of recurrent SNVs and large SVs between patients. One possible explanation for this observation may be that there is no single genetic driver of pGBM but rather a collection of variants which contribute to tumor formation and maintenance. Another explanation could be that a genetic lesion in the intergenic or non-coding regions is driving tumorigenesis. Further projects which utilize my WGS data and focus specifically on analyzing the SNVs in the intergenic and non-coding regions of pGBM would further elucidate mechanisms of tumor initiation. Another interesting follow up experiment to this aim would be to introduce the 4 recurrent large SVs in neural stem cells in vitro and observe whether the cells acquire tumorigenic properties.

5.3 Intratumoral Heterogeneity

The next analysis I did addressed my hypothesis that pGBM is molecularly heterogeneous, specifically within individual tumors. Using copy number variations (CNVs) and large SVs, we were able to generate phylogenetic trees whose branches represent major subclones present within the tumors. Mutational signatures were also assigned to each major subclone, where we found stark differences in the mutational landscape amongst subclones present within the same tumor. As this analysis was being done, another group described the presence of cooperative subclones in DIPG and pGBM using previously published whole exome sequencing (WES) datasets (Vinci et

al., 2018). Although my data are unique in that sequencing was performed on a whole genome level, the data complements their findings of intratumoral heterogeneity in pGBM.

The presence of subclonal architecture within pGBMs is an important observation because it contradicts the current dogma in the field that pGBMs are homogeneous with relatively bland genomes (Jones & Baker, 2014). This is especially important when exploring new treatment options because targeting the major cell populations may leave behind a population of subclonal cells capable of repopulating the tumor. Follow up experiments for this aim would be to perform multi-regional sampling on pGBM tumors to further confirm our findings and to determine the extent of subclonal populations within pGBM. Currently, it is difficult to acquire multi-regional samples from the same tumor without close collaboration with the surgeon. The current protocol divides the tumor pieces amongst many different departments for various reasons (pathology for diagnosis, research labs, tumor bank, etc.). I believe that when multi-regional sampling for research use is incorporated into the protocol, we will be able to detect and target rare subclonal populations that play a role in tumor formation and maintenance.

5.4 Germline Variants

The next set of analyses were conducted to address my hypothesis of germline variants contributing to tumor formation in pGBM. I approached this aim by sequencing the germline and the tumor samples separately and aligning them to the GRCh38 reference genome rather than aligning the tumor sample to the germline control. Traditional sequencing methods use a germline control, when available, as the reference genome in order to call all somatic mutations. This approach eliminates the ability to identify interesting genetic lesions present in the germline,

because they are removed when they are detected in the germline and assumed to be non-pathogenic.

One advantage of having a relatively small cohort is that I was able to manually compare large SVs present in the germline with tumor samples of all the patients. An unexpected finding was the presence of unique large SVs in the germlines and not the tumors of these children. Additionally, in patient 3 and patient 5, the number of large SVs actually decreased in the tumor samples compared to the germline. I investigated the germline variants further by tracking the allelic frequency of each germline variant throughout tumor progression. I found evidence of subclonal variants (AF<0.4) in the germline. I also found that a proportion of germline subclonal variants are lost in the tumor samples. One possible explanation is that the tumor is more capable of selecting against the subclonal populations with low AFs to the point of non-detection in the tumor samples. There were also large SVs which increased AF upon tumor initiation. These dynamic changes demonstrate a role for the germline in pGBM.

5.4.1 Germline deletion of ATRX

One of the recurrent large SVs that was intriguing was a subclonal deletion in the *ATRX* gene that encompasses the majority of the coding region. This deletion was interesting because it was confirmed in the germline and tumor samples of 4/5 patients. While somatic point mutations in *ATRX* have been described in pGBM, there have been no studies which have detected this large-scale deletion. There are a number of reasons why this deletion may have gone undetected until now; the first is that the deletion is present in both the germline and tumor samples so traditional

sequencing techniques would have removed it from analysis because it is not somatic. Another reason is that it is subclonal and may have gone undetected because of a low AF.

Currently, loss of ATRX expression is tested for in diagnostic labs in pGBM cases because it is a well described feature of pGBM. It is unlikely that the point mutation in *ATRX* is causing the complete loss of protein expression. A more plausible explanation is that the loss of protein is being caused by the germline deletion and may or may not be acting with the SNV to reduce expression. This is seen in my data where I did not detect any coding, non-synonymous *ATRX* mutations in our patients and yet I observed heterogeneous *ATRX* expression.

To test whether this deletion is contributing to tumor formation or maintenance, follow up experiments that induce the deletion in neural stem cells and then monitor the cells for tumorigenic properties will address the impact of this genetic lesion. Another beneficial experiment would be to overexpress *ATRX* in pGBM cells to test whether the tumorigenic properties can be dampened.

5.4.2 Germline deletion upstream of NEGR1

The recurrent deletion directly upstream of *NEGR1* in pGBM was present in the germlines and tumor samples of all the pGBM patients. I detected both heterozygous and homozygous genotypes amongst the patients. This deletion has not been described in pGBM before but loss of *NEGR1* expression has been found in other cancers and has been shown to contribute to an increase in cell migration and invasion (H. Kim et al., 2014). I also detected loss of *NEGR1* expression in at least 2 pGBM patients, most likely because the gene's promoter is being deleted. This deletion has most likely not been detected in pGBM previously because it is not somatic, and it is also

present in an intergenic region, so it cannot be detected using techniques that are not genome-wide, such as WES.

This deletion was interesting because *NEGR1* has recently been proposed to play a role in tumor suppression (H. Kim et al., 2014). *NEGR1* primarily localizes to the cell membrane, where it interacts with other cells to control growth and prevent malignant transformation. Overexpression of this protein in an ovarian cancer cell line significantly decreased the proliferative and colony forming capacity of the cells (H. Kim et al., 2014). Follow up experiments that overexpress *NEGR1* in pGBM will be key in characterizing the role this variant plays in tumor etiology.

5.4.3 Cooperation of germline variants

I have highlighted two key recurrent germline variants found in pGBM, a subclonal deletion in *ATRX* and a clonal deletion directly upstream of *NEGR1*. While I propose that there is no single driver of pGBM, I believe these germline variants may be acting synergistically to either predispose developing embryos to become cancerous or be contributing to tumor maintenance by providing an environment conducive to malignant transformation.

Additional examples of germlines variants which are potentially contributing to a predisposition of pGBM are the recurrent clonal deletions between the genes *BTNL8/BTNL3* and *APOBEC3A/APOBEC3B.* These lesions were not present in every sample, the *BTNL8/BTNL3* deletion was present in 3 out of 5 patients and the *APOBEC3A/APOBEC3B* deletion was detected in 2 out of the 5 patients. I also observed no expression of *APOBEC3A*, *APOBEC3B, BTNL8, or BTNL3* in the PDX cells. Neither of these deletions have been described in pGBM previously but

I wanted to investigate them further because they both play a role in immune activation. *BTNL8* has not been studied extensively and its exact role has been a topic of debate but recent work has shown that one of its roles is to stimulate the immune system by priming naïve T cells (Chapoval et al., 2013). *APOBEC3* genes encode cytidine deaminases that protect cells against viral infection (Harris & Liddament, 2004). Deregulation of other *APOBEC* genes have been involved in promoting carcinogenesis but the functions of *APOBEC3A* and *APOBEC3B* have yet to be elucidated. Since these proteins reside in the nucleus and are known to bind DNA, they could potentially have alternative roles such as binding to DNA or RNA and activating an innate immune response (Harris & Liddament, 2004; Henderson & Fenton, 2015). What is known in the literature is that the deletion between *APOBEC3A/3B* has been previously characterized and has been shown to generate a fusion transcript shown to be associated with breast cancer risk (T. Zhang et al., 2013).

These deletions may be causing subsequent suppression or alterations of the immune system, allowing tumor cells to evade immune detection and promote tumor propagation. The downregulation of *BTNL8* and *APOBEC3A* and *APOBEC3B* could potentially be cooperating with previously described germline variants to maintain a suitable tumor microenvironment.

The remaining 2 germline variants that were present in every patient were a duplication in *HLA-DRB1* and a deletion of the last exon of *late cornified envelope protein 3C* (*LCE3C*). The deletion in *LCE3C* is commonly found in the population (37%), according to the Database of Genomic Variants (J. R. MacDonald, Ziman, Yuen, Feuk, & Scherer, 2014), and has not been previously associated with any cancer types. Although the duplication in *HLA-DRB1* could

potentially deregulate the immune system as well because it encodes the HLA class II histocompatibility antigen DRB1 beta chain and has been associated in inflammatory diseases such as rheumatoid arthritis (Källberg et al., 2007), I did not detect any gene expression in the PDXs. The HLA locus is also difficult to align to because of a high degree of polymorphism in the region (Segawa, Kukita, & Kato, 2017), so I was not confident in this variant call. For these reasons, I chose not to investigate these 2 variants further.

Additional analysis to detect the frequency of the above germline variants and how commonly they co-occur will provide more insight into their importance in terms of tumor initiation. The tumor microenvironment of pGBM is also a subject which has yet to be elucidated but is important because it would enable one to predict whether immunotherapy would be an optimal treatment option for this disease.

5.4.4 SNVs in patient germlines

Following the characterization of recurrent large SVs in the germline of pGBM patients, I decided to investigate whether there were any potentially deleterious SNVs in the germlines as well. I began by using the previously described familial cancer syndromes and detected germline SNVs in those genes (analysis done by Ana Nikolic). I found 9 genes which were recurrently mutated in 3 or more patients (*APC, BRCA1, BRCA2, MLH1, MSH6, NF1, PMS2, SMARCA4, TP53).* The significance of these mutations in these pGBM patients is not clear because none of the patients harbour the syndromes associated with these gene mutations and the parents of patient 1 and patient 3 also harboured all of the same mutations. This was surprising because I did not expect exonic, non-synonymous mutations in these genes to be present in a healthy individual.

This could potentially be explained by the mutations not being deleterious or they are deleterious when present in combination with other germline variants.

I next investigated whether there were any de novo SNVs present in the germline of these patients and fortunately, I had the resources to address this aim. I identified all the unique SNVs that were present in the patient's blood and absent in the parent's blood. Patient 1 had 10 de novo, exonic, non-synonymous mutations and patient 3 had 4 de novo, exonic, non-synonymous mutations. Of the 10 genes affected by the SNVs in patient 1, 8 of them are associated with cancer related genes such as tumor suppressors and proto-oncogenes. *PARP4* encodes a member of the poly(ADP)-ribose family, and has been shown to be mutated in the germlines of patients with thyroid and breast cancer (Ikeda et al., 2016). The *RASAL2* gene is a tumor suppressor that encodes a RasGAP (McLaughlin et al., 2013), *MAP3K11* is also a tumor suppressor that encodes a protein which regulates mitogen- and stress-activated kinase signaling (Knackmuss et al., 2016), and *HUWE1* encodes an E3 ubiquitin ligase that was shown to be a tumor suppressor in mouse models of skin cancer and colorectal cancer (Inoue et al., 2013; Myant et al., 2017). The tyrosine kinase *EPHA1*, is amplified in carcinoma cell lines and has been shown to play a role in colorectal cancer (Herath, Doecke, Spanevello, Leggett, & Boyd, 2009). The protein encoded by *LKBP1P* associates with p53 to regulate apoptotic pathways (Karuman et al., 2001). *NAA30* is upregulated in GBM-initiating cells and knockdown of this gene reduced cell proliferation (Mughal et al., 2015). Lastly, induction of *MDR1* causes breast and ovarian cancer to confer resistance to certain chemotherapies (Vaidyanathan et al., 2016).

Of the 4 genes in patient 3, 2 tumor suppressors were identified, *KIAA1551* and *TRAP1*. *KIAA1551* was recently identified as a candidate tumor suppressor and is frequently deleted in different cancer types (Cheng et al., 2017). *TRAP1* has been shown to regulate cancer cell metabolism and is downregulated in cancers such as ovarian, bladder, and renal (Matassa, Agliarulo, Avolio, Landriscina, & Esposito, 2018).

The de novo SNVs play tumorigenic roles in other cancer types but have not been described in pGBM. They provide further evidence of cooperative germline component to this disease and previously published cohorts should be reanalyzed to assess whether these germline variants are present in other pGBM cases.

5.5 Germline Predisposition Component

The detection of recurrent mutations in hereditary oncogenes and germline SNVs associated with CNS syndromes suggests a role for genetic predisposition in this disease. Specifically, the deletions in *ATRX* and *NEGR1* may play cooperative roles with other germline variants associated with CNS development and function in predisposing developing embryos to cancer susceptibility. The SNVs in genes linked to hereditary syndromes that disrupt multiple unrelated networks of genes are compelling when considering early-life onset of this malignancy and suggests these may also contribute to tumor initiation. This notion of a heritable susceptibility to pGBM will advance our understanding of this disease while challenging the dogma that this is a sporadic disease. Further investigation into the germline of pGBM patients using larger cohorts will further elucidate mechanisms of tumor initiation and etiology.

To continue this work and to solidify the importance of the germline variants in pGBM, focusing on combinations of SVs and SNVs in these tumors and the frequency within established cohorts would contribute insight into the current data. The data I have generated are an invaluable resource for further projects not only in pGBM, but also in the field of pediatric cancer biology.

5.6 Caveats

No project is without limitations and caveats. Thus, although the WGS method chosen for this project was innovative and had many advantages, there were also some challenges that accompanied using newfound technology. The lack of available public datasets with linked-read data proved to be challenging when attempting to compare my dataset with normal samples. Pioneering a new technology also has downsides when it comes to data interpretation and overall lack of experience in the field with this type of data. I validated the findings with additional experiments, such as Taqman assays, and correlated my observations with scRNA-seq as a means to overcome these obstacles and ensure my data was reliable.

Another caveat to my observations is that it was difficult to determine if the recurrent genetic lesions I observed were significant or not because of the small sample size. I determined the frequency of the recurrent large SVs in the general population using the Database of Genomic Variants and they were all <40% (J. R. MacDonald et al., 2014). That being said, just because a genetic lesion is found in the general population does not mean it is not significant. It could be acting in a synergistic manner with other genetic lesions to propagate tumor maintenance. Further studies that determine the frequency of combinations of these recurrent lesions in the general population will further elucidate the importance of each variant.

References

Baker, S. J., Ellison, D. W., & Gutmann, D. H. (2016). Pediatric Gliomas as Neurodevelopmental Disorders. *Glia*, *64*(6), 879–895.

Bender, S., Gronych, J., Warnatz, H. J., Hutter, B., Gröbner, S., Ryzhova, M., … Jones, D. T. W. (2016). Recurrent MET fusion genes represent a drug target in pediatric glioblastoma. *Nature Medicine*, *22*(11), 1314–1320.

Bjerke, L., Mackay, A., Nandhabalan, M., Burford, A., Jury, A., Popov, S., … Jones, C. (2013). Histone H3.3 mutations drive pediatric glioblastoma through upregulation of MYCN. *Cancer Discovery*, *3*(5), 512–519.

Bodey, B., Siegel, S. E., & Kaiser, H. E. (2006). Up-regulation of VEGF expression and related neo-angiogenesis in childhood high-grade gliomas: Implications for anti-angiogenic anti-neoplastic therapy. *In Vivo*, *20*(4), 511–518.

Bourdeaut, F., Miquel, C., Richer, W., Grill, J., Zerah, M., Grison, C., … Delattre, O. (2014). Rubinstein–Taybi Syndrome Predisposing to Non-WNT, Non-SHH, Group 3 Medulloblastoma. *Pediatric Blood Cancer*, *61*(2), 383–386.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, *36*(5), 411–420.

Chapoval, A. I., Smithson, G., Brunick, L., Mesri, M., Boldog, F. L., Andrew, D., … Mezes, P. S. (2013). BTNL8, a butyrophilin-like molecule that costimulates the primary immune response. *Molecular Immunology*, *56*(4), 819–828.

Cheng, J., Demeulemeester, J., Wedge, D. C., Vollan, H. K. M., Pitt, J. J., Russnes, H. G., … Van Loo, P. (2017). Pan-cancer analysis of homozygous deletions in primary tumours uncovers rare tumour suppressors. *Nature Communications*, *8*(1221), 1–14.

Cohen, K. J., Pollack, I. F., Zhou, T., Buxton, A., Holmes, E. J., Burger, P. C., … Heideman, R. L. (2011). Temozolomide in the treatment of high-grade gliomas in children: a report from the Children's Oncology Group. *Neuro-Oncology*, *13*(3), 317–323.

Dahiya, S., Emnett, R. J., Haydon, D. H., Leonard, J. R., Phillips, J. J., Perry, A., & Gutmann, D. H. (2014). BRAF-V600E mutation in pediatric and adult glioblastoma. *Neuro-Oncology*, *16*(2), 318-319.

Dasari, S., & Tchounwou, P. B. (2015). Cisplatin in cancer therapy : molecular mechanisms of action. *European Journal of Pharmacology*, *740*, 364–378.

Douma, S., Laar, T. Van, Zevenhoven, J., Meuwissen, R., Van Garderen, E., & Peeper, D. S. (2004). Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB. *Nature*, *430*(7003), 1034–1039.

Eguchi, B. M., Eguchi-ishimae, M., Tojo, A., Morishita, K., Suzuki, K., Sato, Y., … Kamada, N. (1999). Fusion of ETV6 to Neurotrophin-3 Receptor TRKC in Acute Myeloid Leukemia With t(12;15)(p13;q25). *Blood*, *93*(4), 1355–1364.

Fangusaro, J. (2009). Pediatric high-grade gliomas and diffuse intrinsic pontine gliomas. *Journal of Child Neurology*, *24*(11), 1409–1417.

Fangusaro, J. (2012). Pediatric High Grade Glioma: a Review and Update on Tumor Clinical

Characteristics and Biology. *Frontiers in Oncology*, *2*(105), 1–10.
Finlay, J. L., & Zacharoulis, S. (2005). The treatment of high grade gliomas and diffuse intrinsic pontine tumors of childhood and adolescence: A historical - and futuristic - perspective. *Journal of Neuro-Oncology*, *75*(3), 253–266.
Fleming, A. J., & Chi, S. N. (2012). Brain tumors in children. *Current Problems in Pediatric and Adolescent Health Care*, *42*(4), 80–103.
Forbes, S. A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., & Menzies, A. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). *Current Protocols in Human Genetics* (Unit-10.11).
Funato, K., Major, T., Lewis, P. W., Allis, C. D., & Tabar, V. (2014). Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation. *Science*, *346*(6216), 1529–1534.
Gherardi, E., Birchmeier, W., Birchmeier, C., & Vande Woude, G. (2012). Targeting MET in cancer: rationale and progress. *Nature Reviews*, *12*(2), 89–103.
Gibbons, R. J., Mcdowell, T. L., Raman, S., Rourke, D. M. O., Garrick, D., Ayyub, H., & Higgs, D. R. (2000). Mutations in ATRX , encoding a SWI / SNF-like protein , cause diverse changes in the pattern of DNA methylation. *Nature Genetics*, *24*(4), 368–371.
Go, R. S., & Adjei, A. A. (1999). Review of the Comparative Pharmacology and Clinical Activity of Cisplatin and Carboplatin. *Journal of Clinical Oncology*, *17*(1), 409–422.
Goldberg, A. D., Banaszynski, L. A., Noh, K. M., Lewis, P. W., Elsaesser, S. J., Stadler, S., … Allis, C. D. (2010). Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell*, *140*(5), 678–691.
Gröbner, S. N., Worst, B. C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V. A., … Pfister, S. M. (2018). The landscape of genomic alterations across childhood cancers. *Nature*, *555*(7696), 321–327.
Hall, A. G., & Tilby, M. J. (1992). Mechanisms of action of, and modes of resistance to, alkylating agents used in the treatment of haematological malignancies. *Blood Reviews*, *6*(3), 163–173.
Hande, K. R. (1998). Etoposide: Four decades of development of a topoisomerase II inhibitor. *European Journal of Cancer*, *34*(10), 1514–1521.
Harris, R. S., & Liddament, M. T. (2004). Retroviral restriction by APOBEC proteins. *Nature Reviews Immunology*, *4*(11), 868–877.
He, J., Mansouri, A., & Das, S. (2018). Alpha Thalassemia / Mental Retardation Syndrome X-Linked , the Alternative Lengthening of Telomere Phenotype , and Gliomagenesis : Current Understandings and Future Potential. *Frontiers in Oncology*, *7*(322), 1–6.
Hegde, M. R., Chong, B., Blazo, M. E., Chin, L. H. E., Ward, P. A., Chintagumpala, M. M., … Richards, C. S. (2005). A Homozygous Mutation in MSH6 Causes Turcot Syndrome. *Human Cancer Biology*, *11*(13), 4689–4694.
Henderson, S., & Fenton, T. (2015). APOBEC3 genes: Retroviral restriction factors to cancer drivers. *Trends in Molecular Medicine*, *21*(5), 274–284.
Herath, N. I., Doecke, J., Spanevello, M. D., Leggett, B. A., & Boyd, A. W. (2009). Epigenetic silencing of EphA1 expression in colorectal cancer is correlated with poor survival. *British Journal of Cancer*, *100*(7), 1095–1102.

Hottinger, A. F., & Khakoo, Y. (2009). Neurooncology of Familial Cancer Syndromes. *Journal of Child Neurology*, *24*(12), 1526–1535.

Hou, L. C., Bababeygy, S. R., Sarkissian, V., Fisher, P. G., Vogel, H., Barnes, P., & Huhn, S. L. (2008). Congenital glioblastoma multiforme: Case report and review of the literature. *Pediatric Neurosurgery*, *44*(4), 304–312.

Ikeda, Y., Kiyotani, K., Yew, P. Y., Kato, T., Tamura, K., Yap, K.-L., … Grogan, R. H. (2016). Germline PARP4 mutations in patients with primary thyroid and breast cancers. *Endocrine-Related Cancer*, *23*(3), 171–179.

Inoue, S., Hao, Z., Elia, A. J., Cescon, D., Zhou, L., Silvester, J., … Mak, T. W. (2013). Mule/Huwe1/Arf-BP1 suppresses Ras-driven tumorigenesis by preventing c-Myc/Miz1-mediated down-regulation of p21 and p15. *Genes and Development*, *27*(10), 1101–1114.

Jalali, R., Rishi, A., Goda, J. S., Sridhar, E., Gurav, M., Sharma, P., … Gupta, T. (2015). Clinical outcome and molecular characterization of pediatric glioblastoma treated with postoperative radiotherapy with concurrent and adjuvant temozolomide: A single institutional study of 66 children. *Neuro-Oncology Practice*, *3*(1), 39–47.

Jang, C., Shibata, Y., Starmer, J., Yee, D., & Magnuson, T. (2015). Histone H3.3 maintains genome integrity during mammalian development. *Genes & Development*, *29*(13), 1377–1392.

Johnson, A., Severson, E., Gay, L., Vergilio, J., Elvin, J., Suh, J., … Ramkissoon, S. H. (2017). Comprehensive Genomic Profiling of 282 Pediatric Low- and High-Grade Gliomas Reveals Genomic Drivers, Tumor Mutational Burden, and Hypermutation Signatures. *The Oncologist*, *22*(12), 1478–1490.

Johnson, B. E., Mazor, T., Hong, C., Barnes, M., Aihara, K., McLean, C. Y., … Costello, J. F. (2014). Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science*, *343*(6167), 189–194.

Johnson, K. J., Cullen, J., Barnholtz-Sloan, J. S., Ostrom, Q. T., Langer, C. E., Turner, M. C., … Scheurer, M. E. (2014). Childhood Brain Tumor Epidemiology: A Brain Tumor Epidemiology Consortium Review. *Cancer Epidemiology, Biomarkers & Prevention*, *23*(12), 2716–2736.

Jones, C., & Baker, S. J. (2014). Unique genetic and epigenetic mechanisms driving paediatric diffuse high-grade glioma. *Nature Reviews Cancer*, *14*, 651–661.

Jordan, M. (2002). Mechanism of Action of Antitumor Drugs that Interact with Microtubules and Tubulin. *Current Medicinal Chemistry-Anti-Cancer Agents*, *2*(1), 1–17.

Källberg, H., Padyukov, L., Plenge, R. M., Rönnelid, J., Gregersen, P. K., van der Helm-van Mil, A. H. M., … Alfredsson, L. (2007). Gene-Gene and Gene-Environment Interactions Involving HLA-DRB1, PTPN22, and Smoking in Two Subsets of Rheumatoid Arthritis. *The American Journal of Human Genetics*, *80*(5), 867–875.

Karuman, P., Gozani, O., Odze, R. D., Zhou, X. C., Zhu, H., Shaw, R., … Wimalasena, J. (2001). The Peutz-Jegher gene product LKB1 is a mediator of p53-dependent cell death. *Molecular Cell*, *7*(6), 1307–19.

Kim, H., Hwang, J.-S., Lee, B., Hong, J., & Lee, S. (2014). Newly Identified Cancer-Associated Role of Human Neuronal Growth Regulator 1 (NEGR1). *Journal of Cancer*, *5*(7), 598–608.

Kim, S., Scheffler, K., Halpern, A. L., Bekritsky, M. A., Noh, E., Källberg, M., … Saunders, C. T. (2017). Strelka2: Fast and accurate variant calling for clinical sequencing applications. *BioRxiv*.

Knackmuss, U., Lindner, S. E., Aneichyk, T., Kotkamp, B., Knust, Z., Villunger, A., & Herzog, S. (2016). MAP3K11 is a tumor suppressor targeted by the oncomiR miR-125b in early B cells. *Cell Death and Differentiation*, *23*(2), 242–252.

Koschmann, C., Calinescu, A., Nunez, F. J., Mackay, A., Thomas, D., Mendez, F., … Castro, M. G. (2017). ATRX Loss Promotes Tumor Growth and Impairs Non- Homologous End Joining DNA Repair in Glioma. *Science Translational Medicine*, *8*(328), 1–12.

Lannon, C. L., & Sorensen, P. H. B. (2005). ETV6 – NTRK3 : a chimeric protein tyrosine kinase with transformation activity in multiple cell lineages. *Seminars in Cancer Biology*, *15*(3), 215–223.

Lewis, P. W., Muller, M. M., Koletsky, M. S., Cordero, F., Lin, S., Banaszynski, L. A., … Allis, C. D. (2013). Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science*, *340*(6134), 857–862.

Li, Z., Tognon, C. E., Godinho, F. J., Yasaitis, L., Hock, H., Herschkowitz, J. I., … Orkin, S. H. (2007). ETV6-NTRK3 Fusion Oncogene Initiates Breast Cancer from Committed Mammary Progenitors via Activation of AP1 Complex. *Cancer Cell*, *12*(6), 542–558.

MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L., & Scherer, S. W. (2014). The Database of Genomic Variants: A curated collection of structural variation in the human genome. *Nucleic Acids Research*, *42*(D1), 986–992.

MacDonald, T. J., Aguilera, D., & Kramm, C. . (2011). Treatment of high-grade glioma in children and adolescents. *Neuro-Oncology*, *13*(10), 1049–1058.

Matassa, D. S., Agliarulo, I., Avolio, R., Landriscina, M., & Esposito, F. (2018). Trap1 regulation of cancer metabolism: Dual role as oncogene or tumor suppressor. *Genes*, *9*(4), 1–12.

McLaughlin, S. K., Olsen, S. N., Dake, B., De Raedt, T., Lim, E., Bronson, R. T., … Cichowski, K. (2013). The RasGAP Gene, RASAL2, Is a Tumor and Metastasis Suppressor. *Cancer Cell*, *24*(3), 365–378.

Milano, G. M., Cerri, C., Ferruzzi, V., Capolsini, I., Mastrodicasa, E., Genitori, L., & Aversa, F. (2009). Congenital Glioblastoma. *Pediatric Blood Cancer*, *53*(1), 124–126.

Mughal, A. A., Grieg, Z., Skjellegrind, H., Fayzullin, A., Lamkhannat, M., Joel, M., … Stangeland, B. (2015). Knockdown of NAT12/NAA30 reduces tumorigenic features of glioblastoma-initiating cells. *Molecular Cancer*, *14*(1), 1–18.

Myant, K. B., Cammareri, P., Hodder, M. C., Wills, J., Von Kriegsheim, A., Győrffy, B., … Sansom, O. J. (2017). *HUWE1* is a critical colonic tumour suppressor gene that prevents MYC signalling, DNA damage accumulation and tumour initiation. *EMBO Molecular Medicine*, *9*(2), 181–197.

Pathania, M., De Jay, N., Maestro, N., Harutyunyan, A. S., Nitarska, J., Pahlavan, P., … Salomoni, P. (2017). H3.3K27M Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas. *Cancer Cell*, *32*(5), 684–

700.
Paugh, B. S., Qu, C., Jones, C., Liu, Z., Adamowicz-Brice, M., Zhang, J., … Baker, S. J. (2010). Integrated Molecular Genetic Profiling of Pediatric High-Grade Gliomas Reveals Key Differences With the Adult Disease. *Journal of Clinical Oncology*, *28*(18), 3061–3068.
Reddy, A. T., & Wellons, J. C. (2003). Pediatric High-Grade Gliomas. *The Cancer Journal*, *9*(2), 107–112.
Salloum, R., McConechy, M. K., Mikael, L. G., Fuller, C., Drissi, R., DeWire, M., … Jabado, N. (2017). Characterizing temporal genomic heterogeneity in pediatric high-grade gliomas. *Acta Neuropathologica Communications*, *5*(78), 1–12.
Schwartzentruber, J., Korshunov, A., Liu, X. Y., Jones, D. T., Pfaff, E., Jacob, K., … Jabado, N. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*, *482*(7384), 226–231.
Schwartzentruber, J., Korshunov, A., Liu, X. Y., Jones, D. T. W., Pfaff, E., Jacob, K., … Jabado, N. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*, *482*(7384), 226–231.
Segawa, H., Kukita, Y., & Kato, K. (2017). HLA genotyping by next-generation sequencing of complementary DNA. *BMC Genomics*, *18*(1), 1–12.
Şeker, A., & Özek, M. M. (2006). Congenital glioblastoma multiforme. *Journal of Neurosurgery: Pediatrics*, *105*(6), 473–479.
Snuderl, M., Fazlollahi, L., Le, L. P., Nitta, M., Zhelyazkova, B. H., Davidson, C. J., … Iafrate, A. J. (2011). Mosaic Amplification of Multiple Receptor Tyrosine Kinase Genes in Glioblastoma. *Cancer Cell*, *20*(6), 810–817.
Stefanaki, K., Alexiou, G. A., Stefanaki, C., & Prodromou, N. (2013). Tumors of central and peripheral nervous system associated with inherited genetic syndromes. *Pediatric Neurosurgery*, *48*(5), 271–285.
Stupp, R., Hegi, M. E., Gilbert, M. R., & Chakravarti, A. (2007). Chemoradiotherapy in malignant glioma: Standard of care and future directions. *Journal of Clinical Oncology*, *25*(26), 4127–4136.
Sturm, D., Orr, B. A., Toprak, U. H., Hovestadt, V., Jones, D. T. W., Capper, D., … Ellison, D. W., Korshunov, A., & Kool, M. (2016). New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. *Cell*, *164*(5), 1060–1072.
Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D., Jones, D. T. W., Konermann, C., … Plass, C., Jabado, N., & Pfister, S. M. (2012). Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell*, *22*(4), 425–437.
Szenker, E., Ray-Gallet, D., & Almouzni, G. (2011). The double face of the histone variant H3.3. *Cell Research*, *21*(3), 421–434.
Vaidyanathan, A., Sawers, L., Gannon, A. L., Chakravarty, P., Scott, A. L., Bray, S. E., … Smith, G. (2016). ABCB1 (MDR1) induction defines a common resistance mechanism in paclitaxel- and olaparib-resistant ovarian cancer cells. *British Journal of Cancer*, *115*(4), 431–441.
Vinci, M., Burford, A., Molinari, V., Kessler, K., Popov, S., Clarke, M., … Jones, C. (2018). Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. *Nature Medicine*, *24*, 1204-1215.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), 1–7.

Warren, K. E., Gururangan, S., Geyer, J. R., McLendon, R. E., Poussaint, T. Y., Wallace, D., … Kun, L. E. (2012). A phase II study of O6-benzylguanine and temozolomide in pediatric patients with recurrent or progressive high-grade gliomas and brainstem gliomas: A Pediatric Brain Tumor Consortium study. *Journal of Neuro-Oncology*, *106*(3), 643–649.

Wu, G., Diaz, A. K., Paugh, B. S., Rankin, S. L., Ju, B., Li, Y., … Baker, S. J. (2014). The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nature Genetics*, *46*(5), 444–450.

Yu, C.-L., Tucker, M. A., Abramson, D. H., Furukawa, K., Seddon, J. M., Stovall, M., … Kleinerman, R. A. (2009). Cause-Specific Mortality in Long-Term Survivors of Retinoblastoma. *Journal of the National Cancer Institute*, *101*(8), 581–591.

Zhang, J., Walsh, M. F., Wu, G., Edmonson, M. N., Gruber, T. A., Easton, J., … Downing, J. R. (2015). Germline Mutations in Predisposition Genes in Pediatric Cancer. *New England Journal of Medicine*, *373*(24), 2336–2346.

Zhang, T., Cai, J., Chang, J., Yu, D., Wu, C., Yan, T., … Lin, D. (2013). Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma. *Human Molecular Genetics*, *22*(6), 1262–1269.