Abstract

Ependymomas are childhood brain tumors that occur throughout the central nervous system, but are most common in the hindbrain, also known as the posterior fossa (PF). Current standard therapy comprises maximal safe surgery and radiation, but not cytotoxic chemotherapy as it does not further increase survival. Despite histological similarity, ependymomas from throughout the neuroaxis likely comprise multiple independent entities, each with a distinct molecular pathogenesis. Transcriptional profiling of two large independent cohorts of ependymoma reveals the existence of two demographically, transcriptionally, genetically, and clinically distinct groups of posterior fossa ependymomas, termed Group A and B. Group A patients are younger, have laterally located tumors with a balanced genome, and are much more likely to exhibit recurrence, metastasis at recurrence, and death as compared to Group B patients.

In spite of balanced genomes observed in Group A PF ependymoma, whole-genome and whole-exome sequencing of 47 PF ependymomas reveals an extremely low mutation rate, and zero
significant recurrent somatic single nucleotide variants (SNVs). While devoid of recurrent SNVs and focal copy number aberrations, poor prognosis Group A ependymomas exhibit a CpG island methylator phenotype (CIMP). Transcriptional silencing driven by CpG methylation converges exclusively on targets of the polycomb repressive complex 2 (PRC2) that represses expression of differentiation genes through tri-methylation of H3K27. Group A-CIMP⁺ PF ependymomas are responsive to clinical drugs that target either DNA or H3K27 methylation both in vitro and in vivo. We conclude that epigenetic modifiers are the first rational therapeutic candidates for this deadly malignancy, which is epigenetically de-regulated but genetically bland.
Acknowledgments

I would like to express my sincere appreciation and gratitude to my supervisor and mentor Dr. Michael D. Taylor for teaching me how to think, write, speak, dream, and drink like a research scientist. My accomplishments as a PhD student are tied to the great training environment, leadership, funding, and collaborations that Michael has established in his laboratory. It truly has been an amazing experience, and one simply could not wish for a better supervisor and friend. On the same note, I express my special thanks to the Taylor lab members both past and present for their assistance, advice, and friendship throughout my training. I would also like to thank investigators of the Brain Tumour Research Centre namely Dr. Peter B. Dirks, Dr. James T. Rutka, Dr. Eric Bouffet, and Dr. Abhijit Guha for their support and mentorship throughout my PhD. I would like to thank my committee members Dr. Rod Bremner and Dr. Dwayne Barber for their advice and guidance throughout my training. Collaboration has been such an important aspect of Michael’s lab; I would like to thank the Heidelberg, Germany DKFZ ependymoma team led by Dr. Stefan M. Pfister, Dr. Hendrik Witt, and Dr. Andrey Korshunov for working with us over the past few years, which I think, has led to some great discoveries, and great times along the way. I would like to thank my family and friends for all of their encouragement; I wouldn’t have made it this far without their tireless motivation. I am grateful to my wife, Dr. Kelsey C. Bertrand, for understanding that science takes patience, for encouraging my research goals, and for inspiring me to make life an adventure. Finally, I would like to thank my parents and sister, Karl, Rebecca, and Stephanie Mack, for their unconditional support, and giving me the tools, foundation, and opportunities needed to pursue all of my aspirations.
Table of Contents

ACKNOWLEDGMENTS ........................................................................................................ IV

TABLE OF CONTENTS ....................................................................................................... V

LIST OF ACRONYMS ......................................................................................................... IX

1 CHAPTER 1: INTRODUCTION ......................................................................................... 1
    1.1 THE CLINICAL BASIS OF EPENDYMOMA .............................................................. 1
        1.1.1 Incidence .......................................................................................................... 1
        1.1.2 Histopathology ................................................................................................. 1
        1.1.3 Treatment ......................................................................................................... 3
    1.2 THE GENETIC BASIS OF EPENDYMOMA ........................................................... 5
        1.2.1 Somatic copy number landscape ..................................................................... 5
        1.2.2 Ependymoma gene expression profiling reveals significant inter-tumoral heterogeneity and putative cells of origin .................................................................................. 7
        1.2.3 Cell line and animal models of ependymoma .................................................. 9
    1.3 THE EPIGENETIC BASIS OF EPENDYMOMA .................................................... 10
        1.3.1 Candidate gene approaches identify DNA hypermethylated gene promoters in ependymoma 10
        1.3.2 Genome-wide epigenomic profiling identify pathways targeted by aberrant DNA methylation .................................................................................................................. 11
        1.3.3 Potential applications of epigenetic modifiers for ependymoma treatment .......... 12
        1.3.4 Future steps towards characterizing the ependymoma genome and epigenome .......... 13

2 CHAPTER 2: THESIS RATIONALE AND HYPOTHESES .............................................. 16

3 CHAPTER 3: IDENTIFICATION AND CHARACTERIZATION OF TWO CLINICALLY AND MOLECULARLY DISTINCT SUBGROUPS OF POSTERIOR FOSSA EPENDYMOMA 17
    3.1 CHAPTER 3: SUMMARY .......................................................................................... 17
    3.2 CHAPTER 3: INTRODUCTION .................................................................................. 18
3.3 CHAPTER 3: RESULTS ........................................................................................................... 19
  3.3.1 Ependymoma is comprised of three molecularly distinct diseases .................................. 19
  3.3.2 Group A and B posterior fossa ependymoma are demographically and clinically distinct. 20
  3.3.3 Group A and Group B posterior fossa ependymoma are characterized by distinct copy
number landscapes ................................................................................................................. 21
  3.3.4 Group A and Group B ependymoma are characterized by distinct biological pathways ... 22
  3.3.5 Validation of posterior fossa ependymoma subgroups ......................................................... 24

3.4 CHAPTER 3: DISCUSSION ................................................................................................... 27

3.5 CHAPTER 3: FIGURES AND TABLES .................................................................................... 32

3.6 CHAPTER 3: SUPPLEMENTAL FIGURES .............................................................................. 40

3.7 CHAPTER 3: METHODS ....................................................................................................... 45
  3.7.1 Patients and tumour samples ............................................................................................ 45
  3.7.2 Nucleic acid isolation ....................................................................................................... 46
  3.7.3 Gene expression microarray processing and data filtering .................................................. 47
  3.7.4 Array-based comparative genomic hybridization ............................................................... 48
  3.7.5 Statistical analysis of gene expression based subgroups ..................................................... 49
  3.7.6 Identification of gene expression derived subtypes ............................................................ 50
  3.7.7 Identification of biological pathways distinguishing PFA from PFB ependymoma .......... 51
  3.7.8 Selection of candidate genes and gene signatures ............................................................. 52
  3.7.9 Statistical analysis of clinical parameters ......................................................................... 52
  3.7.10 Construction of ependymoma tissue microarrays .............................................................. 53
  3.7.11 Immunohistochemistry .................................................................................................. 53

4 CHAPTER 4: SOMATIC MUTATIONAL AND EPIGENOMIC LANDSCAPE OF GROUP A
AND GROUP B POSTERIOR FOSSA EPENDYMOMA ................................................................ 54
  4.1 CHAPTER 4: SUMMARY ...................................................................................................... 54
  4.2 CHAPTER 4: INTRODUCTION ............................................................................................. 55
  4.3 CHAPTER 4: RESULTS ........................................................................................................ 56
    4.3.1 The somatic mutational landscape of ependymoma .......................................................... 56
    4.3.2 DNA methylation profiles of ependymoma are regionally specified ............................... 57
    4.3.3 Group A posterior fossa ependymoma demonstrate a CpG island methylator phenotype... 58
    4.3.4 Convergence on PRC2 targets ......................................................................................... 59
    4.3.5 Whole genome DNA methylation profiling of ependymoma ......................................... 60
4.3.6 Group A (CIMP+) ependymoma are highly and specifically inhibited by DNA methylation Inhibitors

4.4 Chapter 4: Discussion

4.5 Chapter 4: Figures

4.6 Chapter 4: Supplemental Figures

4.7 Chapter 4: Methods

4.7.1 Patients and tumour samples

4.7.2 DNA library preparation and Illumina sequencing

4.7.3 DNA sequence data processing

4.7.4 Single nucleotide variant (SNV) detection

4.7.5 Small insertion and deletion (InDel) detection

4.7.6 Computation of recurrently mutated genes

4.7.7 Identification of rearrangements and generating of Circos Plots

4.7.8 Identification of pathways affected by SNVs

4.7.9 Generation of copy number profiles from Illumina 450K methylation data

4.7.10 Methyl-binding domain 2 assisted recovery and sample preparation

4.7.11 Methylation analysis of MBD2-chip data

4.7.12 Illumina Infinium 450K methylation sample preparation and data analysis

4.7.13 Subgroup analysis of gene expression and methylation data

4.7.14 H3K27me3 and EZH2 ChIP-Seq profiling and analysis in PF ependymoma

4.7.15 Sequenom analysis of ependymoma samples

4.7.16 Subgroup stratification of ependymoma samples in a validation cohort

4.7.17 Pathway analysis of DNA methylation data

4.7.18 Whole-genome bisulfite sequencing, DNA preparation, and differentially methylated region analysis

4.7.19 Whole-genome bisulfite sequencing DMR and PMD calling

4.7.20 Ependymoma short-term primary cell culture and in vitro drug treatment

4.7.21 Gene expression profiling of DAC and GSK343 treated cultures

4.7.22 Western blot analysis

4.7.23 Flank injections and in vivo treatments of immunodeficient mice

4.7.24 Cerebellar xenografts and in vivo treatments of immunodeficient mice

4.7.25 Limiting dilution assays (LDAs) of primary ependymoma patient samples or ependymoma xenografts
5 CHAPTER 5: FUTURE DIRECTIONS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>FURTHER EXAMINATION OF TUMOUR HETEROGENEITY IN EPENDYMOMA</td>
<td>105</td>
</tr>
<tr>
<td>5.2</td>
<td>FURTHER INVESTIGATION OF GENOMICALLY BALANCED POSTERIOR FOSSA EPENDYMOMA</td>
<td>106</td>
</tr>
<tr>
<td>5.3</td>
<td>FOUNDATIONS FOR THE PFA EPENDYMOMA CIMP PHENOTYPE</td>
<td>107</td>
</tr>
<tr>
<td>5.4</td>
<td>FURTHER ELUCIDATING THE EPIGENOMIC LANDSCAPE OF POSTERIOR FOSSA EPENDYMOMA</td>
<td>109</td>
</tr>
<tr>
<td>5.5</td>
<td>MOUSE MODELS AND CELLULAR ORIGINS OF POSTERIOR FOSSA EPENDYMOMA</td>
<td>110</td>
</tr>
<tr>
<td>5.6</td>
<td>CLINICAL IMPLICATIONS FOR TREATMENT OF POSTERIOR FOSSA EPENDYMOMA</td>
<td>111</td>
</tr>
</tbody>
</table>

6 ATTRIBUTIONS

7 COPYRIGHT PERMISSIONS

8 REFERENCE LIST
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>aCGH</td>
<td>array comparative genomic hybridization</td>
</tr>
<tr>
<td>ACTB</td>
<td>beta actin</td>
</tr>
<tr>
<td>ALU</td>
<td>arthrobacter luteus</td>
</tr>
<tr>
<td>AQP-1/3/4</td>
<td>aquaporin-1/3/4</td>
</tr>
<tr>
<td>ATRX</td>
<td>ATP-dependent helicase, X-linked helicase II</td>
</tr>
<tr>
<td>BIRC5</td>
<td>baculoviral IAP repeat containing 5</td>
</tr>
<tr>
<td>BLBP</td>
<td>brain lipid binding protein</td>
</tr>
<tr>
<td>ZMYD10</td>
<td>zinc finger, MYND-Type containing 10</td>
</tr>
<tr>
<td>CASP8</td>
<td>caspase 8, apoptosis-related cysteine peptidase</td>
</tr>
<tr>
<td>CDK2/4</td>
<td>cyclin-dependent kinase 2/4</td>
</tr>
<tr>
<td>CDKN-1C/2A/2B/2C</td>
<td>cyclin dependent kinase inhibitor-1C/2A/2B/2C</td>
</tr>
<tr>
<td>cHCL</td>
<td>consensus hierarchical clustering</td>
</tr>
<tr>
<td>CHI3L1</td>
<td>chitinase 3-like 1</td>
</tr>
<tr>
<td>CIMP</td>
<td>CpG island methylator phenotype</td>
</tr>
<tr>
<td>cNMF</td>
<td>consensus non-negative matrix factorization</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>DAC</td>
<td>5'-aza-2'-deoxyazacytidine</td>
</tr>
<tr>
<td>DAPK</td>
<td>death-associated protein kinase 1</td>
</tr>
<tr>
<td>DFKZ</td>
<td>das deutsche krebforchungszentrum</td>
</tr>
<tr>
<td>DLK1</td>
<td>delta-like 1 homolog</td>
</tr>
<tr>
<td>DMR</td>
<td>differentially methylated region</td>
</tr>
<tr>
<td>DNAJC15</td>
<td>DnaJ (Hsp40) homolog, subfamily C, member 15</td>
</tr>
<tr>
<td>DNMT-1/3A/3B</td>
<td>DNA methyltransferase isoform-1/3A/3B</td>
</tr>
<tr>
<td>DZNep</td>
<td>3-deazaneplanocin-A</td>
</tr>
<tr>
<td>EED</td>
<td>embryonic ectoderm development</td>
</tr>
<tr>
<td>EFNA-A3/4</td>
<td>ephrin-A3/A4</td>
</tr>
<tr>
<td>EGFR</td>
<td>epidermal growth factor receptor</td>
</tr>
<tr>
<td>EMA</td>
<td>epithelial membrane antigen</td>
</tr>
<tr>
<td>EPHB-2/3/4</td>
<td>ephrin-B2/B3/B4</td>
</tr>
<tr>
<td>ERRB-2/4</td>
<td>v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2/4</td>
</tr>
<tr>
<td>EZH2</td>
<td>enhancer of zeste homolog 2</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>FHTT</td>
<td>fragile histidine triad</td>
</tr>
<tr>
<td>GAPDH</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GBM</td>
<td>glioblastoma multiforme</td>
</tr>
<tr>
<td>GFAP</td>
<td>glial fibrillary acidic protein</td>
</tr>
<tr>
<td>GO</td>
<td>gene ontology</td>
</tr>
<tr>
<td>GSEA</td>
<td>gene set enrichment analysis</td>
</tr>
<tr>
<td>GSTP1</td>
<td>glutathione S-transferase pi 1</td>
</tr>
<tr>
<td>GTR</td>
<td>gross total resection</td>
</tr>
<tr>
<td>H19</td>
<td>imprinted maternally expressed transcript (non-protein coding)</td>
</tr>
<tr>
<td>H3F3A</td>
<td>H3 histone, family 3A</td>
</tr>
</tbody>
</table>
PTPRN2  | protein tyrosine phosphatase, receptor type, N polypeptide 2
RAB3A  | Ras-related protein Rab-3A
RARB   | retinoic acid receptor, beta
RAS    | rat sarcoma
RASSF1A| ras association domain family member 1
RGC    | radial glial cell
ROI    | region of interest
S100-A6| S100 calcium binding protein-A6
SCNA   | somatic copy number alteration
SINE   | short interspersed nuclear elements
SNV    | single nucleotide variant
SP     | spinal cord
ST     | supratentorial
STAG1  | stromal antigen 1
STR    | sub total resection
SUZ12  | suppressor of zeste 12 protein homolog
TCAG   | the centre for applied genomics
TERT   | telomerase
TET-1/2| tet methylcytosine dioxygenase 1
TGF-Beta| transforming growth factor beta
THAP11 | thap domain containing 11
TIMP3  | TIMP metallopeptidase inhibitor 3
TMA    | tissue microarray
TNC    | tenasin C
TNFRSF-10C/10D| tumour necrosis factor receptor superfamily-10C/10D
TP53   | tumour protein 53
TP73   | tumour protein 73
TRAIL  | TNF-related apoptosis-inducing ligand
TSG    | tumour suppressor gene
VEGF   | vascular endothelial growth factor
WHO    | world health organization
1 Chapter 1: Introduction

1.1 The clinical basis of ependymoma

1.1.1 Incidence

Ependymomas are rare, chemo-resistant, central nervous system tumours, arising in both children and adults. The age-adjusted incidence of ependymoma between 2004-2009 was 0.41 per 100,000\(^1\), and follows a bimodal distribution with peaks at the average age of 5 and 34. Ependymomas can arise along the entire neuro-axis occurring in the supratentorial (ST) brain comprising the cerebral hemispheres, the posterior fossa (PF) comprising the cerebellum and brain stem, and along the entire spinal cord (SP). In children, 90\% of ependymomas arise intracranially, with nearly two-thirds occurring in the PF\(^2\). The incidence of relapse is significantly greater in the pediatric population, and is clinically variable with recurrences observed in some cases 10 – 15 years following treatment of the primary tumour.

1.1.2 Histopathology

Ependymomas were first recognized as a distinct entity by Cushing and Bailey in their 1926 histological description of brain tumours. To this date much of the same criteria is used in accordance with the 2007 classification of brain tumours as outlined by the World Health
Organization (WHO). The hallmark histological features of ependymoma include: 1) Perivascular pseudorosettes: composed of ependymal cell processes radially arranged towards blood vessels, and 2) True ependymal rosettes: consisting of tumour cells arranged radially surrounding an empty lumen. In terms of immunohistochemical staining, ependymomas are commonly positive for glial fibrillary acidic protein (GFAP), neural cell adhesion molecule protein (NCAM), and epithelial membrane antigen protein (EMA), which allows for further delineation from other histologically similar brain tumours.

The WHO recognizes three grades for ependymomas:

**Grade I:** Comprising subependymomas and myxopapillary ependymomas, which are easily recognizable histological entities associated with a better survival and increased age. (Note: this histological type will not be discussed further in the current thesis).

**Grade II:** Pertains to ependymomas, which lack Grade III features described below.

**Grade III:** Also described as anaplastic, this classification corresponds to ependymomas with “increased cellularity and brisk mitotic activity, often associated with microvascular proliferation and pseudopalisading necrosis”.

While the Grade I criteria for ependymoma are relatively clear, the parameters used to distinguish Grade II and III ependymomas are highly debated as prognostic differences have
been observed in some tumour cohorts but not others\(^3\). This discrepancy occurs even after controlling for potentially confounding factors such as age or tumour location. The lack of reproducibility and reliability of histopathological grading is demonstrated in a systemic review of three independent European trial cohorts by five leading neuropathologists, in which a consensus agreement on the classification of Grade II or III ependymoma was reached in less than half of 221 total cases examined\(^4\). While the prognostic utility of histopathological grading is still debated, immunohistochemistry (IHC) and molecular markers have been proposed as potential solutions e.g. Telomerase (TERT) protein expression\(^5\)\(^-\)\(^7\), V-erb-b2 erythroblastic leukemia viral oncogene homolog 2/4 (ERBB2/4)\(^8\) protein expression, or fluorescence in situ hybridization (FISH) of recurrent chromosomal alterations\(^9\). These IHC and molecular markers remain to be validated in independent and prospective ependymoma cohorts, a challenge necessitating multi-centre collaborative efforts.

1.1.3 Treatment

To this date, treatment for ependymoma remains aggressive surgical intervention and adjuvant radiotherapy. In spite of the histological challenges in identifying high- versus low- risk ependymoma patients, the extent of surgical resection is the most frequently and consistently reported prognostic indicator of ependymoma patient survival. In other pediatric brain tumours, such as medulloblastoma, radiotherapy is typically avoided in children less than 3 years of age due to increased risk of long-term neurological and neuroendocrine sequelae. However, the
aggressive nature of ependymoma in infants and young children combined with a lack of effective chemotherapies has provided rationale for the use of conformal and intensity-modulated radiotherapy in infants. Early evidence in prospective clinical trials has demonstrated that conformal radiation is both effective and associated with minimal short-term neurological side effects in the 5-years following treatment, as evidenced by no detectable verbal or visual learning decline, and relatively stable emotional and behavioral functioning\textsuperscript{10-12}. Whether these outcomes are maintained long-term awaits further evaluation.

Despite improvements in surgical techniques and advances in conformal radiation, recurrence rates remain high in ependymoma patients particularly in the pediatric population\textsuperscript{13}. Although, chemotherapy has been used extensively in the treatment of children with intracranial ependymoma, clinical trial response rates to numerous single agents are less than 12%, with less than 5% of patients experiencing complete responses\textsuperscript{14}. Results from several multi-centre ependymoma clinical trials suggest that there is little evidence that chemotherapy is effective in treatment of this tumour type\textsuperscript{14,15}. As a result, current standard of care for patients with recurrent ependymoma is maximal-safe surgical resection followed by re-irradiation. Despite prolonged overall survival observed from re-irradiation of recurrent ependymoma, the risks of secondary tumours and neurological impairments have yet to be adequately assessed\textsuperscript{13}.

The high recurrence rate of pediatric ependymomas, lack of prognostic histological and molecular markers, and dearth of chemotherapeutic avenues, underscores the importance of
understanding the biological basis of ependymoma such that rationale molecular targets can be identified and rapidly translated in the clinic.

1.2 The genetic basis of ependymoma

1.2.1 Somatic copy number landscape

Efforts to identify driver oncogenes and tumour suppressor genes (TSG) of ependymoma began largely with characterization of these tumours at a DNA copy number level using cytogenetic approaches, DNA-based microarrays and, more recently, whole-genome and whole exome sequencing. Despite higher resolution array technologies, the vast majority of recurrent somatic copy number alterations (SCNA) are broad and involve losses of chromosome: 1p, 3, 6q, 9p, 10q, 13q, 16p, 17, 21 and 22q, and gains of chromosome: 1q, 4q, 5, 7, 8, 9, 12q, and 20\textsuperscript{16}.

The most frequent and focal SCNA in ependymoma is a homozygous deletion encompassing the CDKN2A/Ink4a locus, which is restricted to supratentorial ependymomas (ST). Other non-recurrent events include amplifications of EPHB2, THAP11, PSPH, PCDH cluster, KCNN1, RAB3A, PTPRN2, and NOTCH1, and deletions of PTEN, STAG1, and TNRC6B. As one example, EPHB2 was validated as a bonafide ST ependymoma oncogene and was used to generate the first mouse model of ST ependymoma. In the case of posterior fossa (PF) ependymomas, recurrent and focal copy number alterations pinpointing driver genes have yet to
be discovered, highlighting the difficulty in understanding the biological basis, and identifying novel therapeutic targets for this anatomical subtype of ependymoma.

Chromosome 22 loss has been shown to be the most frequent gross copy number alteration in ependymoma with a frequency ranging from 26% to 71%\textsuperscript{16}. Further, chromosome 22q loss has been observed preferentially in spinal versus intracranial ependymoma, and in adult versus pediatric cases\textsuperscript{16}. The Neurofibromatosis II (NF2) gene is thought to be the candidate TSG of this region, as patients with Neurofibromatosis type II disorder develop a variety of central nervous system tumors including ependymoma, schwannoma, and meningionoma\textsuperscript{16}. However NF2 is mutated exclusively in spinal ependymomas, thus suggesting alternate mechanisms of down-regulation, or another putative chromosome 22q TSG in the case of intracranial ependymoma.

Another broad chromosomal abnormality frequently observed in ependymoma is monosomy 17 with complete or partial loss on both chromosomal p and q arms. Further, several studies, albeit limited to small cohorts, have reported that loss of chromosome 17 occurs more frequently on the p-arm in up to 50% of cases. A clear tumour suppressor gene on 17p has yet to be identified, and the consensus is that TP53 is not the candidate TSG of this region as TP53 mutations are exceedingly rare in both adult and childhood ependymoma.

Chromosome 1q gain has also been consistently reported as a frequent genomic alteration occurring in nearly 22% of cases of intracranial ependymoma. Further, an increased incidence
of 1q gain has been observed in posterior fossa ependymoma and associated with poor clinical outcome\textsuperscript{16-18}. It is thought that the chromosome 1q25 locus harbors a bonafide oncogene involved in the initiation, maintenance, or progression of ependymoma. Efforts have been made to correlate this region of chromosomal gain with gene expression and have identified \textit{CHI3L1} and a family of \textit{S100} genes as up-regulated and potential driver oncogenes, however these candidates remain to be functionally validated.

As seen in numerous cohorts, a large proportion of ependymoma, particularly of posterior fossa localization exhibit balanced genomic profiles with very few genomic abnormalities\textsuperscript{17-20}. This has been observed consistently from low- to high- resolution copy number studies, and associated with young age and poor clinical outcome. Alternative mechanisms of gene regulation such as somatic mutations have been proposed to play a role in ependymoma, perhaps significantly in cases with balanced copy number profiles\textsuperscript{21}. However, whole-exome sequencing across a small series of 8 intracranial ependymomas did not identify a single recurrently mutated gene.

1.2.2 Ependymoma gene expression profiling reveals significant inter-tumoral heterogeneity and putative cells of origin

Using gene expression profiling, and unsupervised clustering, ependymomas have been divided into three principal molecular subgroups, which are separated largely according to anatomical location: 1) Supratentorial (ST), 2) Posterior Fossa (PF), and 3) Spinal cord (SP)\textsuperscript{17}. These three
subgroups have been further divided into molecularly and biologically distinct subtypes of ST and PF ependymoma as defined by distinct clinical features and copy number alterations\textsuperscript{18,20,22}. The genes distinguishing supratentorial, posterior fossa, and spinal ependymomas involve mainly families of genes regulating neural precursor cell proliferation and differentiation\textsuperscript{17}. Supratentorial ependymomas have elevated EPHB-EPHRIN (EPHB2/3/4 and EFNA\textsubscript{A}A3/4), NOTCH (JAGGED\textsubscript{1/2}), and cell cycle related genes (CYCLIN\textsubscript{B2/D1/G2}, CDK2/4, and CDKNIC2C), while posterior fossa ependymomas express many inhibitors of differentiation (ID1/2/4), and the aquaporin family of genes (AQP1/3/4). Spinal ependymomas however are characterized by expression of various homeobox genes including HOXA7/9, HOXB6/7, and HOXC6/10.

While these subgroup gene signatures revealed distinct tumorigenic pathways, Taylor et al. (2005) proposed that these were potential signatures of anatomically distinct cells of origin giving rise to different subgroups of ependymoma\textsuperscript{17}. They further suggested that ependymoma might originate from radial glial cells, a population of primitive neural and multi-potent precursor cells important for neurogenesis and neuronal migration. Indeed, tumour cells isolated from a series of ependymoma primary samples expressed characteristic markers of radial glial cells (RGC), such as BLBP and RC2. These ependymoma cells could be differentiated into neuronal, astrocytic, and oligodenrocytic lineages, thus highlighting the multi-potent capacity and functional similarity to RGCs. Further evidence implicating RGCs as cells of origin of ependymoma was demonstrated by Johnson et al. (2010), in which ex-vivo over-expression of
EPHB2 in p16/INK4A deficient forebrain RGCs led to the formation of the first mouse model of ST ependymoma\textsuperscript{18}.

1.2.3 Cell line and animal models of ependymoma

In the past, lack of clear driver alterations in ependymoma has hampered the ability to generate animal models of this disease. Laboratories have thus relied upon patient-derived ependymoma cultures grown \textit{in vitro} and orthotopic xenograft models generated with limited success particularly in the case of PF ependymoma\textsuperscript{23,24}. Identification of the putative cell-of-origin of ependymoma, and drivers of ST ependymoma, namely EPHB2 amplification and recurrent p16/INK4A deletion, has led to the first animal models of this disease\textsuperscript{8,17}. As novel ependymoma targets are discovered, validation and prioritization of candidates will require accurate pre-clinical models of ependymoma. Atkinson et al., (2011) demonstrate the utility and promise of this approach in Group D ST ependymoma cultures, generated by EPHB2 over expression and CDKN2A/Ink4a deletion in forebrain radial glia\textsuperscript{25}. Here they performed a compound library screen of 7890 compounds, in both tumour and matched normal neural stem cells and identified inhibitors of thymidylate synthase and dehydrogenate reductase, namely 5-fluorouracil as highly and specifically active in mice bearing ST ependymomas. Findings such as these hold therapeutic relevance particularly in ependymoma with currently no chemotherapeutic options available to patients.
1.3 The epigenetic basis of ependymoma

1.3.1 Candidate gene approaches identify DNA hypermethylated gene promoters in ependymoma

Aberrant promoter methylation of CpG di-nucleotides is a well-recognized feature seen in numerous solid and liquid cancers\(^2^6\). As such, early studies have focused on candidate gene promoters, hypermethylated in ependymoma, which have also been reported to be methylated in other cancers. \textit{RASSF1A} has been shown to be the most frequently hypermethylated TSG reported in up to 100\% of ependymomas, and occurring in all clinical and pathological subtypes\(^2^7,2^8\). \textit{HIC1} has also been reported to be commonly methylated, in up to 83\% of ependymomas, with a higher incidence in intracranial tumours\(^2^9\). Furthermore, the \textit{CDKN2A/INK4a} locus, which is focally and recurrently deleted in supratentorial ependymoma\(^1^8\), has been shown to be hypermethylated in 21\% of cases, followed by CDKN2B and p14ARF in 32\% and 33\% of tumors, respectively\(^3^0\). To a lesser extent, putative TSGs found to be hypermethylated in ependymoma include \textit{ZMYD10, GSTP1, DAPK, FHIT, MGMT, DNAJC15, RARB, TIMP3, THBS1, TP73} and the Tumour Necrosis Factor Related Apoptosis-Inducing Ligand (TRAIL) gene family \textit{CASP8, TNFRSF10C} and \textit{TNFRSF10D}\(^3^1-3^4\). Despite the frequency of DNA methylation of these potential TSGs in ependymoma, their role and significance in tumour development remains unclear, requiring validation in independent cohorts, and functional investigation in appropriate ependymoma models. It raises the possibility, however, that other genes and pathways may be epigenetically altered in the
pathogenesis of ependymoma, and suggests that epigenome-wide examinations of DNA methylation in ependymoma are warranted.

1.3.2 Genome-wide epigenomic profiling identify pathways targeted by aberrant DNA methylation

In the last 5 years, the expansion of microarray and next generation sequencing technologies, has allowed for genome-wide investigations of DNA methylation and histone modifications at unprecedented resolution and throughput. Using the Illumina Golden Gate Methylation Cancer Panel 1 (1505 CpG sites), Rogers et al., (2012) profiled a series of 73 primary and 25 recurrent ependymomas. Here they reported that the DNA methylation profiles of ependymoma are distinguished largely according to their location in the central nervous system, supporting the notion that ependymomas arising from different anatomic compartments are molecularly distinct. Furthermore, they demonstrated that ST and SP ependymomas, together, exhibited a larger number of hypermethylated and down-regulated genes in comparison to PF tumours. These changes in DNA methylation were shown to be associated with alterations in gene expression of de novo and maintenance DNA methyltransferases $DNMT1$, $DNMT3A$ and $DNMT3B$. Interestingly, genes involved in immune cell response ($NOD2$, $IRF7$, $IRAK3$, $OSM$ and $PI3$), cell growth and death ($MAPK10$, and $TP73$), and the c-Jun N-terminal kinases (JNK) pathway were found to be hypermethylated. Understanding the contribution of epigenetic alterations in these pathways may reveal mechanisms of ependymoma tumourigenesis, and potential actionable targets for therapeutic intervention.
In contrast to hypermethylation of CpG island promoters in cancer, global hypomethylation is a trend observed in numerous tumour types and is associated with cancer progression. A global decrease in methylation has been observed predominantly at repetitive elements such as LINEs, SINEs and LTRs, which are important for maintaining genomic stability. To elucidate the contribution of DNA methylation alterations at repetitive sequences in ependymoma, Xie et al., (2010) developed a novel genome-wide approach to generate methylation profiles for thousands of Alu elements (The most abundant class of repetitive elements) and their flanking sequences. Here they demonstrated that while the majority of Alu elements and flanking sequences remain unaltered in ependymoma genesis, a small subset of Alu flanking sequences, with low CpG density, exhibited variable methylation patterns. These sequences tended to be hypermethylated in ependymoma at regions proximal to CpG islands and hypomethylated in intergenic regions. Importantly, several of these patterns were shown to be associated with aggressive primary ependymomas and tumour relapse. The impact that these epigenetic alterations have on genomic stability and the pathways and biological processes that they impinge on remain to be elucidated, however point to alterations in the epigenome that may play an important role in the pathogenesis of ependymoma.

1.3.3 Potential applications of epigenetic modifiers for ependymoma treatment

Characterizing the epigenome of ependymoma may hold therapeutic promise, as these marks such as CpG DNA methylation and histone modification are generally reversible by
pharmacologic inhibition. Importantly, inhibitors of DNA methylation (decitabine) and histone deacteylation (Vorinostat) are FDA approved and have shown efficacy in haematological malignancies\textsuperscript{37,38}. This would make them likely rapidly translatable if effective in appropriate ependymoma pre-clinical models. As an example, Milde et al., (2011) established a model reminiscent of high-risk molecular group C supratentorial ependymoma, by \textit{in vivo} transplantation of primary tumour cells in NOD/SCID immunodeficient mice\textsuperscript{23}. Like primary ependymomas, which are highly chemoresistant, cell cultures from this model were refractory to temozolomide, vincristine, and cisplatin. However, these cells were sensitive to a panel of histone deactetylas e inhibitors (HDACi) including Panobinostat, Entinostat, Valproic Acid, and Vorinostat. Importantly, cell cultures treated with Vorinostat at therapeutically achievable doses, demonstrated decreased proliferation, cell cycle arrest, decreased self-renewal capacity, and increased neuronal differentiation. These findings were also supported by Rahman et al., (2010), demonstrating that the ependymoma cell line nEPN2 underwent apoptosis in response to treatment with the HDACi, Trichostatin-A\textsuperscript{39}. Given the rapid development of novel pharmacologic inhibitors of epigenetic marks, it raises the question as to whether these, or at least some, epigenetic modifications are central to ependymoma pathogenesis and whether they may represent novel avenues for therapeutic inhibition.

1.3.4 Future steps towards characterizing the ependymoma genome and epigenome
Although genomic and transcriptomic profiling efforts have identified distinct molecular subtypes of ependymoma and revealed potential drivers of the disease, the vast majority of ependymoma tumors are characterized by either large chromosomal alterations, hampering the identification of driver events, or are characterized by very few genomic abnormalities, observed in some of the most aggressive tumours, enriched in PF patients, and occurring in the youngest patient population\textsuperscript{18,20}.

It remains to be seen whether recurrent somatic single nucleotide variants (SNVs) or structural rearrangements (i.e. fusion transcripts) may contribute to the pathogenesis of ependymoma, as reported in several other adult and pediatric CNS neoplasms. Indeed, somatic SNVs identified in CNS cancers have been shown to potentially alter the entire epigenetic landscape during tumor formation, such as the \textit{IDH1} mutations leading to a CpG hypermethylator phenotype in Glioblastoma Multiforme (GBM)\textsuperscript{40}, histone \textit{H3F3A} and \textit{ATRX} mutations in pediatric GBM\textsuperscript{41}, and a collection of chromatin associated genes shown to be mutated in medulloblastoma such as \textit{MLL2}, \textit{MLL3}, and \textit{KDM6A}\textsuperscript{42-45}.

DNA methylation profiling efforts have also been important in the molecular stratification of CNS tumors\textsuperscript{46}, and have also shown promise in distinguishing the principle molecular subgroups of ependymoma and identifying pathways targeted by DNA hypermethylation. As a future step, expanding DNA methylation profiling to platforms with higher CpG coverage surrounding gene promoters might be warranted, and could reveal novel targets, pathways, and mechanisms of epigenetic alteration. Also, given the contributions of aberrant methylation near
repeat elements in the ependymoma epigenome, more global investigations beyond gene promoters might be needed, and could be readily examined with whole-genome bisulfite sequencing.

Ultimately, the contribution of genetic and epigenetic changes in ependymoma pathogenesis may not only improve our understanding of the biology of this disease, but also reveal actionable pathways that could be rapidly translated in the clinic. This is important especially in the case of PF ependymoma, a highly recurrent and chemo-resistant brain tumour, for which the mechanisms and drivers of the disease remain largely unknown.
Chapter 2: Thesis Rationale and Hypotheses

Given that pediatric ependymomas arise most frequently in the posterior fossa, are chemo-resistant, are associated with poor clinical outcome, and exhibit largely balanced genomes, I formulated two hypotheses:

1) Posterior fossa ependymomas may be characterized by significant inter-tumoural heterogeneity, and potential driver oncogene or tumour suppressor genes may be uncovered by delineation of gene expression signatures in large and independent primary ependymoma cohorts.

2) Despite the paucity of focal copy number alterations, posterior fossa ependymomas may harbor recurrent somatic mutations and epigenetic alterations, which may be revealed by whole-genome sequencing, whole exome sequencing, and DNA methylome profiling.

This thesis is divided into two parts addressing the hypotheses above:

1) The first chapter details the identification and characterization of two clinically and molecularly distinct subgroups of posterior fossa ependymoma termed Group A and Group B.

2) The second chapter describes the somatic mutational and epigenomic landscape of Group A and B posterior fossa ependymoma
3 Chapter 3:

Identification and characterization of two clinically and molecularly distinct subgroups of posterior fossa ependymoma

3.1 Chapter 3: Summary

Despite histological similarity, ependymomas from throughout the neuroaxis likely comprise multiple independent entities, each with a distinct molecular pathogenesis. Transcriptional profiling of two large independent cohorts of ependymoma reveals the existence of two demographically, transcriptionally, genetically, and clinically distinct groups of posterior fossa ependymomas (PF), termed Group A and B. Group A patients are younger, have laterally located tumours with a balanced genome, and are much more likely to exhibit recurrence, metastasis at recurrence, and death as compared to Group B patients. Identification and optimization of immunohistochemical markers for PF ependymoma subgroups allowed validation of our findings on a third independent cohort, using a human ependymoma tissue microarray (TMA), and provides a tool for prospective prognostication and stratification of PF ependymoma patients.
3.2 Chapter 3: Introduction

In spite of histopathological similarities, ependymomas are very heterogeneous tumors with disparate mRNA expression profiles, supporting the hypothesis that the histological entity ‘ependymoma’ in fact comprises a group of related diseases\textsuperscript{17,18}. The genetic landscape of ependymoma is also heterogeneous, with subsets of tumors exhibiting frequent gross numerical chromosomal alterations, and others displaying only single focal aberrations, or even a balanced genome\textsuperscript{16}. A recurrent observation in several tumour cohorts is that up to 50\% of PF cases have balanced genomic profiles\textsuperscript{18,19,47-49}. A molecular model for prognostication of pediatric and adult ependymomas was recently published based upon tumors with specific genomic aberrations, or the absence of copy number alterations\textsuperscript{19}. Further addressing the genetic heterogeneity of ependymoma, Johnson et al (2010), reported a comprehensive study cataloguing DNA copy-number alterations in subgroups of ependymoma defined by messenger RNA (mRNA) and microRNA (miRNA) profiles\textsuperscript{18}. Remarkably, they generated a mouse model of supratentorial ependymoma, and present supporting evidence that the cell-of-origin for supratentorial ependymoma resides in the radial glial lineage\textsuperscript{17,18}. Evidence for a molecular ‘driver’ alteration and a cell-of-origin for PF ependymoma are still lacking. We have now applied genomic methodologies to three independent cohorts of human ependymomas in order to uncover the extent of inter-tumoral molecular heterogeneity, and the nature of the clinically relevant subgroups of PF ependymoma.
3.3 Chapter 3: Results

3.3.1 Ependymoma is comprised of three molecularly distinct diseases

In order to delineate the genetic basis of ependymoma we performed gene expression profiling on two non-overlapping primary ependymoma cohorts, which were histologically diagnosed as Grade II or Grade III\(^2\). 102 ependymomas were analyzed in Toronto, Canada using the Affymetrix Exon 1.0 ST gene expression arrays, and 75 ependymomas were analyzed in Heidelberg, Germany on the Agilent two-color 4x44K gene expression arrays. We then examined the subgroup structure within both datasets, applying analytical methods as described by The Cancer Genome Atlas, asking whether the same subgroups were observed using two distinct bioinformatic methods in two independent datasets\(^50\). The 1000 most varying genes in each dataset were used to perform consensus hierarchical clustering (cHCL) from 2 to 10 subgroups (Figure 3.1A-B and Figure S3.1)\(^51\). Despite randomizing the initial number of input genes, we consistently found three stable clusters in both datasets that were generated using distinct microarray technologies (Figure S3.1).

To support the subgroup stratification by cHCL, we performed consensus clustering using non-negative matrix factorization (cNMF)\(^52\), and concluded that both cHCL and cNMF subgroup assignments were highly concordant (Figure 3.1C). We then used an algorithm called SigClust to perform all pairwise comparisons between HCL-defined subgroups and demonstrated that the three subgroups were statistically significant and distinct (Figure 3.1D)\(^53\). Silhouette analysis
demonstrated that 96% of samples in the Toronto and Heidelberg datasets had positive silhouettes, and thus were highly representative of their subgroups (Figure 3.1E)\textsuperscript{54}. This method was also used to identify and remove poorly classified samples, which had negative silhouette values as described previously\textsuperscript{50}. Therefore, using both cHCL and cNMF, we identified 3 principal subgroups of ependymoma divided largely anatomically into ependymomas arising from: 1) the supratentorial region (ST), 2) the posterior fossa (PF), and 3) a third group that included a mixture of both spinal cord and posterior fossa (PF + SP) tumors (Figure S3.1). We next evaluated whether a significant number of the same genes defined the same three subgroups in both datasets. To answer this question, we identified common subtypes across the two datasets by Subclass mapping, and verified that the subgroups identified in both datasets were defined by the same transcriptional signature (Figure 3.1F)\textsuperscript{55}. We conclude that there are three transcriptionally defined principal subgroups of ependymoma, and that a subset of posterior fossa ependymomas is more similar to spinal tumors. The subgroup comprised purely of PF ependymomas was designated Group A (PFA), while PF tumors that clustered with spinal ependymomas in the previous analysis (PF + SP) were designated Group B (PFB, Figure 3.2A-B). We next examined the demographic and clinical differences between Group A and B PF ependymoma.

3.3.2 Group A and B posterior fossa ependymoma are demographically and clinically distinct
In the Toronto and Heidelberg datasets, both independently and combined, Group A tumors developed in younger patients (median age of 2.5 years), while Group B tumors occurred predominantly in older patients (median age of 20, Figure 3.2C). Anatomically, 67% of Group A tumors occurred laterally, invading into surrounding cranial nerves and arteries, while 95% of Group B ependymomas occurred along the midline (Figure 3.2D). Additionally, Group A tumors demonstrated a propensity to invade into the cerebellum (Figure 3.2E). Patients with 5-years of clinical follow-up, harbouring Group A ependymomas, exhibited a significantly increased incidence of tumour relapse (Figure 3.2F), and mortality (Figure 3.2G). 5-year progression-free and overall survival rates were 47% and 69% for Group A tumors (Figure 3.2H - left panel), Figure S3.2), and 79% and 95% for Group B tumors (Figure 3.2H - right panel), respectively. Detailed clinical information for Group A and B tumors of the Toronto and Heidelberg cohort can be found in Witt and Mack et al (2011)\textsuperscript{20}. We conclude that the two subgroups of PF ependymoma identified through transcriptional profiling are both molecularly and clinically distinct, and that their differences are highly clinically significant and relevant.

### 3.3.3 Group A and Group B posterior fossa ependymoma are characterized by distinct copy number landscapes

We analyzed all samples from both cohorts, studied by gene expression profiling, for which sufficient amounts of DNA were available (n=152), by array comparative genomic hybridization (aCGH). Group A and Group B ependymomas were identified as outlined above, yielding 75 tumors where both subgroup assignment and aCGH data were available. Group A
tumors exhibited a largely balanced genomic profile, with an increased occurrence of chromosome 1q gain compared to Group B (Figure 3.3A-B). Normal cell contamination in Group A was a less likely explanation for the balanced genomic landscape observed as the majority of samples were composed of tumor cells (≥80%) by Hematoxylin and Eosin staining (Figure S3.3), and as chromosome 1q gains, when present, were found in the majority of cells as shown by FISH (Figure S3.3). Lastly we compared the copy number intensities derived from the aCGH signals for chromosome 1q gain, 6q loss, and 9q gain, and demonstrated that chromosomal aberrations in Group A had similar intensities as aberrations in Group B (Figure S3.3). This is in line with previous reports that a large proportion of PF ependymoma exhibit a balanced genomic profile 18,19,47-49. In contrast, Group B ependymomas exhibited numerous cytogenetic abnormalities involving whole chromosomes or chromosomal arms, including loss of chromosome 1, 2, 3, 6, 8, 10, 14q, 17q, 22q, and gain of 4, 5q, 7, 9, 11, 12, 15q, 18, 20, 21q (Figure 3.3A-B). Chromosomal aberrations also classified tumors into the recently described cytogenetic risk groups 19. Group A ependymomas included predominantly high-risk groups 2 and 3 (Figure 3.3A), while the vast majority of Group B tumors were classified into the low-risk group 1. We conclude that PF Group A and Group B ependymomas have distinct somatic copy number events in addition to the transcriptional and clinical differences described above.

3.3.4 Group A and Group B ependymoma are characterized by distinct biological processes and signaling pathways
In order to identify the biological basis distinguishing Group A and B we used Gene Set Enrichment analysis\textsuperscript{56}. Gene sets were collected and compiled from the National Cancer Institute (NCI), Kyoto Encyclopedia of Genes and Genomes (KEGG), Protein Families (PFAM), Biocarta, and Gene Ontogeny (GO) pathway databases. In order to visualize significant gene sets as interaction networks we used Cytoscape and the Enrichment Map plug-in (Figure 3.4)\textsuperscript{57}.

Group B ependymomas were characterized by gene sets largely involved in ciliogenesis and microtubule assembly, and to a lesser extent mitochondria and oxidative metabolism. In order to validate this pathway, we selected Kinesin Family Member 27 (KIF27), an established marker of cilia signaling in other brain neoplasms\textsuperscript{58,59}, and demonstrated in an independent TMA cohort, that 91\% of Group B tumors stained positive, while 17\% of Group A tumors showed a pattern of positivity (Figure S3.4).

Group A ependymomas were defined by numerous cancer-related networks namely angiogenesis (HIF-1\(\alpha\) signaling, VEGF signaling, cell migration), PDGF signaling, MAPK signaling, EGFR signaling, TGF-\(\beta\) signaling, integrin signaling, extracellular matrix assembly, tyrosine-receptor kinase signaling, and RAS/Small GTPase signaling. Since extracellular matrix signaling was the most significantly enriched Group A pathway. We therefore selected Tenascin-C (TNC), a type of matrix glycoprotein, which has been reported to be up-regulated in pediatric ependymoma, as a marker of ECM signaling\textsuperscript{60}. 94\% of Group A tumours stained positive for TNC, while only 11\% of Group B tumors showed positivity (Figure S3.4). We
conclude that the marked difference in active signaling pathways between Group A and Group B supports their distinct natures, and suggests possible leads for future subgroup specific targeted therapies.

3.3.5 Validation of posterior fossa ependymoma subgroups

In order to validate the two subtypes of PF ependymoma, an orthogonal technology was applied to an additional and independent cohort of ependymomas. We selected IHC markers up-regulated in each PF subgroup, and stained a TMA comprised of an entirely non-overlapping cohort of 265 PF ependymomas. Using a Goeman’s Global test, subgroup markers were identified from the transcriptional data by their ability to distinguish between the two pre-defined (A vs B) subgroups (Figure 3.5A, Figure S3.4). *Laminin alpha-2 (LAMA2)*, which exhibited an increase in expression in Group A versus B, was selected as a marker of Group A (Figure 3.5B). *Neural Epidermal Growth Factor Like-2 (NELL2)*, which exhibited an increase in expression in Group B versus A, was selected as a marker of Group B (Figure 3.5B). Further, to assure clinical applicability of our markers, we prioritized selection and optimization of commercially available antibodies to LAMA2 and NELL2 as markers of Group A and Group B PF ependymoma, respectively (Figure 3.5C). We then proceeded to stain our TMA validation cohort consisting of 265 PF ependymomas, in which 32% stained positive for LAMA2 and negative for NELL2, 52% stained positive for NELL2 and negative for LAMA2, 8% stained positive for both markers, while 8% stained negative for both markers (Figure 3.5D, Figure
S3.4). Detailed clinical information on tumors staining for LAMA2 or NELL2 in the validation cohort is summarized in Table S3 of Witt and Mack et al., (2011)\(^{20}\).

We also observed numerous clinical disparities between Group A and B patients of the IHC validation cohort consonant with the differences observed in the gene expression cohorts. Patients with Group A ependymomas were significantly younger (median age 4 versus 39 years) (Figure 3.6A), more commonly male (Figure 3.6B), and were more frequently classified as WHO grade III ependymomas (Figure 3.6C). Furthermore, patients with Group A ependymomas had a higher incidence of metastases at the time of recurrence, and had a significantly poorer prognosis as compared to Group B patients (Figure 3.6D-F). Progression-free and overall survival rates were of 24% and 48%, for Group A tumors and 92% and 98% for Group B tumors (Figure 3.6E-F). Interestingly, when cases stained for both or none of the two markers, these tumors had no association with metastasis at time of recurrence, and formed two intermediate patient survival groups, suggesting possible additional inter-tumoral heterogeneity within PF ependymoma. We conclude that Group A tumors are clinically and molecularly distinct from Group B tumors, have a worse prognosis, and can be identified through routine IHC staining of formalin fixed paraffin embedded sections for LAMA2 and NELL2.

The most frequently reported prognostic indicator in ependymomas is the extent of surgical resection\(^{61,62}\). In line with these reports, the extent of resection, in our datasets was an important and independent variable on a Cox proportional hazard model for the entire current cohort of PF tumors when not accounting for subgroups (Table 3.1 and Figure S3.5)\(^{61-68}\). Because Group A
ependymomas are more commonly located laterally in the posterior fossa, we asked whether the observed differences in outcome were secondary to or confounded by a decreased incidence of gross total resection. However, we found that patients with Group B tumors, amenable to complete resection, displayed a 5-year progression-free and overall survival rate of 91% and 100%. Importantly, patients with completely resected Group A tumors, still exhibited worse progression-free and overall survival rates of 18% and 52%, respectively (Figure 3.6G-H). These findings demonstrate that the difference in prognosis between PF ependymoma subgroups is not solely based on an increased incidence of incompletely resected Group A tumors.

We next evaluated to combination of numerous prognostic factors within Group A and B using a Cox proportional hazard model that included extent of resection, age, gender, WHO grade, use of radiotherapy, NELL2 staining (Group B), and LAMA2 staining (Group A). This analysis revealed that histological markers of Group A and B were the strongest and independent predictors of both overall and progression-free survival (Table 3.2) in patients with PF ependymoma. Indeed, after accounting for PF subgroup, level of resection was no longer prognostic for overall survival and had a lower influence on progression-free survival prediction (Table 3.2), as compared to an analysis that did not include subgroup assignment (Table 3.1). Although these data will need to be validated in the setting of a prospective clinical trial, the markers of Group A and B presented represents an effective and clinically applicable tool for rapid and robust risk stratification of PF ependymoma.
3.4 Chapter 3: Discussion

Our data supports the existence of two molecularly distinct groups of PF ependymoma that although histologically similar, differ in their demographic, transcriptional, clinical, and outcome characteristics. The strength of this hypothesis is supported by our approach using two non-overlapping cohorts of PF ependymomas, studied in two geographic locations, with two distinct expression array technologies, and then subsequent validation using an orthogonal technology, on a third independent cohort of PF ependymomas. Our study of 583 ependymomas represents the largest cohort of ependymomas analyzed to date.

A prior publication studying a smaller cohort of ependymomas (29 total tumors) demonstrated that ependymomas from different regions of the nervous system (supratentorial, posterior fossa, and spine) had regionally specific transcriptional profiles and somatic copy number alterations suggesting that ependymomas from different regions of the nervous system were separate entities\textsuperscript{17}. It is noteworthy that in this publication, all of the supratentorial tumors cluster together, as do all of the spinal ependymomas. However, the spinal tumors cluster between two groups of posterior fossa ependymomas, in keeping with the findings of the current manuscript where some PF ependymomas (Group B) are more similar to spinal ependymomas than they are to the other subgroup of PF ependymomas (Group A).
A recent publication described a mouse model of supratentorial ependymoma, which was generated by transplanting neural stem cells harboring a combination of deletion of \textit{Ink4a/Arf} and/or overexpression of \textit{Ephb2}. These mice developed supratentorial tumors that showed similar patterns of gene expression and cytogenetic events to 1 of 4 subgroups of supratentorial ependymoma identified from studying a total of 83 human ependymomas. The same study also indicated molecular heterogeneity within PF tumors, but this aspect was not a focus of the study, and no correlation with clinical outcome was presented for the PF tumors. Inspection of the subgroups from Johnson et al. (2010) reveals a subgroup that includes both PF and spinal ependymomas, that is genomically unstable, that includes many adults, (Johnson et al. subgroup F) and therefore is similar to PF Group B in the current manuscript. The Johnson et al. study also describes three subgroups comprised almost exclusively of PF ependymomas (Johnson et al. subgroups G, H, and I), which show few cytogenetic aberrations other than gain of chromosome 1q, includes a large number of infants, and are therefore similar to the Group A described in the current manuscript. In line with previous reports, we conclude that a subset of ependymomas, labeled in this study Group A, exhibit a largely balanced chromosomal profile. The high degree of genomic instability in Group B and association with an improved prognosis is a paradox that has been observed in other neoplasms of the breast, stomach, and lung.

In our study, chromosome 22 loss was one of the most frequent genomic alterations, occurring often in Group B PF tumors and rarely in Group A tumors. \textit{Neurofibromin-2 (NF2)}, located at chromosome 22q12.2, is thought to be the candidate tumor suppressor gene of this region.
because patients with \textit{NF2} mutations develop numerous neuro-epithelial neoplasms including spinal ependymoma. However, \textit{NF2} mutations have been found exclusively in ependymomas of the spinal cord, highly suggesting the existence of another chromosome 22q tumor suppressor gene in the case of PF ependymoma\textsuperscript{72,73}.

Examination of molecular pathways characterizing the two PF ependymoma subgroups revealed two diverse patterns of alteration, again suggesting the existence of two biologically distinct classes of posterior fossa ependymoma. In Group B ependymomas only two pathways: ciliogenesis/microtubule assembly and mitochondrial/oxidative metabolism were exclusively deregulated. A more heterogeneous picture of pathway alteration was seen in Group A tumors, including several canonical cancer-associated pathways. Among them were Angiogenesis (HIF-1\textalpha signaling, VEGF pathway), PDGF signaling, MAPK signaling, EGFR signaling, TGF-\beta signaling, Tyrosine-receptor kinase signaling, RAS signaling, and Inteigrin/ECM signaling. Molecular research into the pathways driving Group A and B may yield targets for subgroup specific therapy. Because Group A patients experience poor outcome, and because there are no currently known effective chemotherapeutic regimens in ependymoma, the therapeutic threshold would be very low in this underserved patient population. Also, the highly distinct Group A and B signaling pathways may also be reflective of different cells of origin, in keeping with a mechanism that has recently been shown for supratentorial ependymoma\textsuperscript{18} and for Wnt subgroup medulloblastoma\textsuperscript{74}. The transcriptional profiles of Group A and B presented in this manuscript will serve as a resource to help guide future attempts to pinpoint possible alternative cells of origin giving rise to PF ependymoma.
The most highly differentially expressed genes in Groups A and B, in both the Toronto and Heidelberg datasets, revealed candidate marker genes for distinguishing the two groups; the most striking being up-regulation of *LAMA2* in Group A and *NELL2* in Group B. Other markers of Group A included previously reported biomarkers of poor patient outcome including *CHI3L1, TNC, VEGF, EGFR, ERRB4, BIRC5*, and *S100A6* (Figure S1.4)\(^8,47,75-77\). This evidence, in addition to the lack of prognostic significance for chromosome 1q gain in Group A patients, suggests that some previously reported markers of poor outcome may have been surrogate markers for Group A.

As Group B ependymomas are much less likely to recur, metastasize, or result in the death of the patient, validation of our results in additional cohorts of patients would suggest that Group B patients could be treated less aggressively than Group A patients. Conversely, the poor outcome for Group A patients underlines the need for rapid development of adjuvant therapies for these patients. We anticipate that analysis of additional cohorts of PF ependymoma will further support the existence of at least two divergent molecular variants that are demographically, genetically, transcriptionally, and clinically distinct\(^22\). The antibodies described for *LAMA2* and *NELL2* are both commercially available, and therefore should be widely available across the globe for validation of our results, and eventually for use in prognostication and stratification of PF ependymoma patients. We would also suggest that future clinical trials should prospectively validate IHC staining for *LAMA2* and *NELL2* on formalin-fixed, paraffin-embedded tumor material. Importantly, to further improve our understanding of the molecular biology of these
posterior fossa subgroups, prospective investigations into the cell-of-origin and genetic driver mutations are desperately needed, including modeling of PF ependymoma in the mouse. Finally, the distinct molecular characteristics of these two groups of PF ependymoma suggest that subgroup-specific targeted therapies against subgroup-specific deregulated pathways are needed in future treatments of these tumors.
3.5 Chapter 3: Figures and Tables

Identification of Three Primary Molecular Classes of Ependymoma

(A) Area under empirical cumulative distribution plots (k=2 to k=10), generated from consensus hierarchical clustering of 102 Toronto and 75 Heidelberg samples identifies strongest statistical support for the existence of three primary subgroups of ependymoma. (k denotes the number of clusters) (B) Consensus HCL heatmaps displaying the three robust subgroups of ependymoma defined by gene expression. (C) Consensus Non-Negative Matrix Factorization (NMF) of 102 Toronto and 75 Heidelberg samples at k=3, demonstrates significant concordance with the consensus HCL subgroup classification. Significance of similarity was determined by a Rand index and permutation testing of the Toronto sample labels (see experimental procedures). (D) Significance of HCL subgroup classifications in both datasets determined by pairwise comparisons between all clusters using SigClust. (E) Silhouette analysis identifies "core" samples defined as tumors with positive silhouette values. (F) Submap analysis demonstrates that the HCL-defined clusters identified in the Toronto cohort are nearly identical to the HCL-defined clusters defined in the Heidelberg cohort. Significance of similarity measured by FDR-corrected p-value.
Transcriptome Analysis Distinguishes Two Distinct Subgroups of Posterior Fossa Ependymoma

(A) Graphical illustration of global differences between the transcriptomes of Group A and Group B PF ependymomas determined by principal component analysis. Individual tumor samples are represented as spheres, (red = Group A, blue = Group B) and ellipsoids display two standard deviations around each subgroup. (B) Heatmap of the top 100 most discriminating genes between Group A and Group B ependymomas, in both datasets, as calculated by Goeman’s Global Test statistic. (C) Box plots demonstrating the association of Group A ependymomas with younger age at diagnosis, and Group B ependymomas with older age at diagnosis. Median age is also indicated. p-values were determined by a Mann-Whitney test. (D) - (G) Pie charts demonstrating an association of Heidelberg Group A ependymomas with lateral localization, cerebellar invasion, increased recurrence and death within 5 years from diagnosis as compared with Group B tumors. Statistical significance of PF location, and incidences of recurrence and death was determined by Fisher’s Exact test. (H) Kaplan-Meier survival curves demonstrating a worse progression-free and overall survival in Group A versus Group B ependymomas. Statistical significance was determined by a log-rank test.
Figure 3.3

A

Identification of Subgroup Specific Copy Number Alterations in the Posterior Fossa Ependymoma Genome

(A) Copy number profiling of 75 PF ependymomas using 10K array-CGH identifies disparate genetic landscapes between Group A and Group B tumors. Toronto and Heidelberg copy number datasets have been combined and summarized in a heatmap. The heatmap also displays the association of tumors to cytogenetic risk groups 1, 2 and 3 (Korshunov et al., 2010). Statistically significant chromosomal aberrations (black boxes) are also displayed between both subgroups, calculated by Fisher’s exact test. (B) Median averaged frequencies of DNA copy-number alterations of 45 Group A and 30 PF Group B tumors plotted against their chromosomal position.
Gene Set Enrichment Analysis Delineates Biological Pathways and Processes that Define Two Distinct Variants of Posterior Fossa Ependymoma

Gene Set Enrichment Analysis (GSEA) comparing Group A (red) against Group B (blue) PF ependymoma in the Toronto dataset, illustrating distinct pathways and biological processes between both subgroups. (3.5 % FDR, p=0.01) Cytoscape and Enrichment Map were used for visualization of the GSEA results. Nodes represent enriched gene sets, which are grouped and annotated by their similarity according to related gene sets. Enrichment results were mapped as a network of gene sets (nodes). Node size is proportional to the total number of genes within each gene set. Proportion of shared genes between gene sets is represented as the thickness of the green line between nodes. This network map was manually curated removing general and uninformative sub-networks resulting in a simplified network map shown in Figure 4.
Selection and Optimization of PF Ependymoma Group A and Group B Specific Immunohistochemical Markers

(A) Subgroup-specific expression patterns of selected markers, LAMA2 and NELL2, illustrated by heatmaps in both data sets. Candidate genes were identified using the Goeman’s global test, which assigns a score to each gene based upon its degree of discrimination between defined classes: Group A and Group B. (B) Box plots derived from mRNA expression data displaying overall differences between markers representing Group A (LAMA2) and Group B (NELL2) in the Toronto and Heidelberg cohorts. Comparisons were performed using an unpaired t-test. (C) Representative immunohistochemistry (IHC) staining of LAMA2 and NELL2 on an ependymoma tissue microarray (TMA) composed of 265 PF ependymomas. (D) Pie chart illustrating the distribution of TMA staining for NELL2 and LAMA2. 84% of posterior fossa ependymomas stain positive for a single marker.
Figure 3.6

(A) Box plots for age indicating NELL2-/LAMA2+ tumors represent a significantly younger population than NELL2+/LAMA2- tumors.
(B-C) Bar graphs demonstrating that NELL2-/LAMA2+ tumors are over-represented by males and WHO Grade III tumors.
(D-F) Kaplan-Meier plots demonstrate that NELL2-/LAMA2+ tumors have an earlier time to metastasis, and have a poorer progression-free and overall survival than NELL2+/LAMA2- tumors.

Clinical and Cytogenetic Characteristics Distinguishing IHC Defined Group A and B Subgroups in a Third Non-Overlapping Posterior Fossa Ependymoma Cohort

(A) - (F) Comparing subgroup-specific demographic and clinical information in the validation cohort illustrated by:

(A) Box plots for age indicating NELL2-/LAMA2+ tumors represent a significantly younger population than NELL2+/LAMA2- tumors, (B-C) Bar graphs demonstrating that NELL2-/LAMA2+ tumors are over-represented by males and WHO Grade III tumors, (D-F) Kaplan-Meier plots demonstrate that NELL2-/LAMA2+ tumors have an earlier time to metastasis, and have a poorer progression-free and overall survival than NELL2+/LAMA2- tumors. Statistical significance of age was determined by a Mann-Whitney test; gender and grade by a Fisher’s Exact test, and time to metastasis and survival by a Log-Rank test. 

(G–H) Limiting to gross-totally resected cases, Kaplan-Meier curves demonstrating that NELL2-/LAMA2+ tumors have a significantly poorer survival than NELL2+/LAMA2- tumors. Statistical significance was determined by a Log-rank test.
Table 3.1 Cox Proportional Hazards models of overall survival and progression-free survival examining clinical variables of posterior fossa ependymomas (TMA Validation Cohort)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard Ratio (CI 5%)</th>
<th>Hazard Ratio (CI 95%)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall Survival</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years (4-18 vs &lt; 4)</td>
<td>0.88</td>
<td>0.29</td>
<td>2.66</td>
</tr>
<tr>
<td>Gender (male vs female)</td>
<td>1.74</td>
<td>0.86</td>
<td>3.52</td>
</tr>
<tr>
<td>Age, years (≥ 18 vs &lt; 4)</td>
<td>0.36</td>
<td>0.11</td>
<td>1.22</td>
</tr>
<tr>
<td>Radiotherapy (yes vs no)</td>
<td>0.45</td>
<td>0.17</td>
<td>1.19</td>
</tr>
<tr>
<td>Resection (GTR vs STR)</td>
<td>0.42</td>
<td>0.21</td>
<td>0.86</td>
</tr>
<tr>
<td>Histology (grade 3 vs grade 2)</td>
<td>3.35</td>
<td>1.37</td>
<td>8.18</td>
</tr>
<tr>
<td><strong>Progression-Free Survival</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years (4-18 vs &lt; 4)</td>
<td>0.85</td>
<td>0.39</td>
<td>1.86</td>
</tr>
<tr>
<td>Radiotherapy (yes vs no)</td>
<td>0.72</td>
<td>0.38</td>
<td>1.37</td>
</tr>
<tr>
<td>Gender (male vs female)</td>
<td>1.30</td>
<td>0.83</td>
<td>2.04</td>
</tr>
<tr>
<td>Age, years (≥ 18 vs &lt; 4)</td>
<td>0.42</td>
<td>0.18</td>
<td>1.00</td>
</tr>
<tr>
<td>Histology (grade 3 vs grade 2)</td>
<td>2.74</td>
<td>1.55</td>
<td>4.86</td>
</tr>
<tr>
<td>Resection (GTR vs STR)</td>
<td>0.42</td>
<td>0.26</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Note. No. of patients with fossa posterior ependymoma = 265; CI, Confidence Interval; * Wald Test.
Table 3.2 Cox Proportional Hazards Models for Overall Survival and Progression-Free Survival of Posterior Fossa Ependymoma of TMA Validation Cohort

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard Ratio</th>
<th>95% CI</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall Survival</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (male vs female)</td>
<td>1.42</td>
<td>0.67 - 3.04</td>
<td>0.3591</td>
</tr>
<tr>
<td>Resection (GTR vs STR)</td>
<td>0.66</td>
<td>0.32 - 1.34</td>
<td>0.2470</td>
</tr>
<tr>
<td>Age, years (4-18 vs &lt; 4)</td>
<td>1.81</td>
<td>0.77 - 4.25</td>
<td>0.1746</td>
</tr>
<tr>
<td>Histology (grade 3 vs grade 2)</td>
<td>1.92</td>
<td>0.79 - 4.68</td>
<td>0.1486</td>
</tr>
<tr>
<td>Age, years (&gt; 18 vs &lt; 4)</td>
<td>0.44</td>
<td>0.16 - 1.24</td>
<td>0.1202</td>
</tr>
<tr>
<td>Radiotherapy (yes vs no)</td>
<td>0.29</td>
<td>0.11 - 0.77</td>
<td>0.0135</td>
</tr>
<tr>
<td>NELL2 (positive vs negative)</td>
<td>0.12</td>
<td>0.03 - 0.44</td>
<td>0.0012</td>
</tr>
<tr>
<td>LAMA2 (positive vs negative)</td>
<td>10.55</td>
<td>2.81 - 39.60</td>
<td>0.0005</td>
</tr>
<tr>
<td><strong>Progression-Free Survival</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (male vs female)</td>
<td>1.17</td>
<td>0.74 - 1.85</td>
<td>0.5037</td>
</tr>
<tr>
<td>Radiotherapy (yes vs no)</td>
<td>0.80</td>
<td>0.44 - 1.47</td>
<td>0.4744</td>
</tr>
<tr>
<td>Age, years (4-18 vs &lt; 4)</td>
<td>0.68</td>
<td>0.34 - 1.35</td>
<td>0.2686</td>
</tr>
<tr>
<td>Histology (grade 3 vs grade 2)</td>
<td>1.42</td>
<td>0.79 - 2.57</td>
<td>0.2423</td>
</tr>
<tr>
<td>Age, years (&gt; 18 vs &lt; 4)</td>
<td>1.80</td>
<td>0.99 - 3.27</td>
<td>0.0535</td>
</tr>
<tr>
<td>Resection (GTR vs STR)</td>
<td>0.53</td>
<td>0.34 - 0.83</td>
<td>0.0061</td>
</tr>
<tr>
<td>NELL2 (positive vs negative)</td>
<td>0.32</td>
<td>0.17 - 0.61</td>
<td>0.0005</td>
</tr>
<tr>
<td>LAMA2 (positive vs negative)</td>
<td>8.45</td>
<td>4.08 - 17.49</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
3.6 Chapter 3: Supplemental Figures

Figure S3.1

(A) Empirical cumulative distribution function (ECDF) plots illustrating that subgroup stability is achieved at a maximal number of three subgroups in both Toronto and Heidelberg datasets. (B) Heatmaps from k=2 to k=4 depicting the three principal subgroups from consensus hierarchical clustering in both Toronto and Heidelberg datasets. Consensus hierarchical clustering was performed with the top 1000 genes exhibiting the greatest median absolute deviation. (k denotes number of clusters) (C) Area under the ECDF plots demonstrating that the principal number of subgroups in both datasets is three despite varying the number of input genes from 800 to 2000. (D) Kaplan-Meier curves comparing the progression-free and overall survival for the three gene-expression defined subgroups identified in the Heidelberg data set. (E) Box plots illustrating the anatomical distribution of ependymomas (ST, PF, SP) for each gene-expression defined subgroup. While two groups were composed of largely ST or PF ependymomas, the third group was comprised of a mixture of PF and SP ependymomas.
Figure S3.2

(A) Kaplan-Meier curves of Heidelberg mRNA-defined Group B tumors, demonstrating no prognostic significance between patients <18 years versus >18 years for both PFS and OS. Statistical significance for all Kaplan-Meier curves in Figure S2 was calculated using a log-rank test. (B) Kaplan-Meier curves of Heidelberg mRNA-defined Group B tumors, demonstrating no prognostic significance between patients <18 years versus >18 years for both PFS and OS. Statistical significance was observed in the case of OS. (C) Kaplan-Meier curves of Heidelberg mRNA-defined Group B tumors, demonstrating no prognostic significance in PFS and OS between ependymomas with sub-total and gross-total resection. (D) Kaplan-Meier curves of Heidelberg mRNA-defined Group A tumors, demonstrating no prognostic significance between patients <4 years versus >4 years for both progression-free survival (PFS) and overall survival (OS). (E) Kaplan-Meier curves of Heidelberg mRNA-defined Group A tumors, demonstrating that Group A patients who received radiotherapy had an improved PFS compared to those who did not receive radiotherapy. No prognostic significance was observed in the case of overall survival. (F) Kaplan-Meier curves of Heidelberg mRNA-defined Group A tumors, demonstrating that sub-total resection is associated with poor PFS and OS compared to gross-total resected tumors. (G) Kaplan-Meier curves of Heidelberg mRNA-defined Group A tumors, demonstrating an improved PFS in patients with midline occurring ependymomas, as compared to lateral tumors. No prognostic significance was observed in the case of OS.
Figure S3.3

(A) – (C) Assessing the presence of normal cell contamination in a series of Group A tumors. (A) Haematoxylin and Eosin staining demonstrating ≥ 80% tumor cell composition as determined by neuropathological assessment. (B) Representative fluorescent in situ hybridization images demonstrating that tumors with 1q gain, as determined by array comparative genomic hybridization (aCGH), harbored 1q gain in a majority of tumor cells. (C) aCGH heatmaps comparing the intensities of copy number alterations between Group A and Group B, and demonstrating a similar signal intensity in aberrations present in both groups.
Representative IHC staining for KIF27 and TNC illustrating the clear difference between positive and negative staining cases

In comparison, only 11% of Group-B tumors showed activation of this pathway. Statistical significance of pathway markers was determined using a Fisher’s Exact Test.

Pathway involvement in Group B tumors could be demonstrated by positive staining of 91% Group B tumors by a Ciliogenesis marker KIF27, in contrast 17% of Group-A tumors showed pattern of positivity. Statistical significance was determined by the Goeman’s global test.

Also shown are the ranks for each gene, which were determined by genes exhibiting the highest ‘Influence Score’ in both datasets as reported to be associated with poor patient outcome in ependymoma. Statistical significance was determined using an unpaired t-test.

Pie graphs demonstrating a greater than 94% concordance between gene expression (microarray) and protein expression (IHC) for NELL2 and LAMA2. Statistical significance was determined using a Fisher’s Exact Test.

(A) Box plots of selected Group A markers as defined by gene expression in both datasets. Markers selected have been previously reported to be associated with poor patient outcome in ependymoma. Statistical significance was determined using an unpaired t-test. Also shown are the ranks for each gene, which were determined by genes exhibiting the highest ‘Influence Score’ in both datasets as determined by the Goeman’s global test. (B) Pie graphs demonstrating a greater than 94% concordance between gene expression (microarray) and protein expression (IHC) for NELL2 and LAMA2. Statistical significance was determined using a Fisher’s Exact Test. (C) Ciliogenesis involvement in Group B tumors could be demonstrated by positive staining of 91% Group B tumors by a Ciliogenesis marker KIF27, in contrast 17% of Group-A tumors showed pattern of positivity. Statistical significance of pathway markers was determined using a Fisher’s Exact Test. (D) 94% of Group-A tumors showed ECM/Integrin pathway activation as determined positive expression of Tenascin C, a representative candidate ECM/Integrin protein. In comparison, only 11% of Group-B tumors showed activation of this pathway. (E) Representative IHC staining for KIF27 and TNC illustrating the clear difference between positive and negative staining cases.
Figure S3.5

(A) Kaplan-Meier curves of NELL2-/LAMA2+ tumors demonstrating no prognostic difference in PFS and OS between tumors with or without chromosome 1q gain. Chromosome 1q gain was determined by fluorescent in situ hybridization. Statistical significance for all Kaplan-Meier curves in Figure S6 was calculated using a Log-Rank test. (B) Kaplan-Meier curves of the TMA tumor cohort demonstrating that PF ependymoma patients with sub-total resection have a statistically poorer outcome than patients with gross-total resection. (C) Kaplan-Meier curves of NELL2-/LAMA2+ tumors demonstrating that level of resection is prognostically significant only in the case of PFS, and not OS. (D) Kaplan-Meier curves of NELL2+/LAMA2- tumors demonstrating that level of resection is not prognostically significant in both PFS and OS. (E) Kaplan-Meier curves of NELL2-/LAMA2+ tumors demonstrating no prognostic difference in PFS and OS between patients <4 years and >4 years. (F) Kaplan-Meier curves of NELL2+/LAMA2- tumors demonstrating no prognostic difference in PFS and OS between patients <18 years and >18 years. (G) Kaplan-Meier curves of NELL2-/LAMA2+ tumors demonstrating no prognostic difference in PFS and OS between patients who received or did not receive radiotherapy. (H) Kaplan-Meier curves of NELL2+/LAMA2- tumors demonstrating no prognostic difference in PFS and OS between patients who received or did not receive radiotherapy.
3.7 Chapter 3: Methods

3.7.1 Patients and tumour samples

Clinical samples and data were utilized in accordance with research ethics board approval from both The Hospital of Sick Children (Toronto, Ontario) and DKFZ (Heidelberg, Germany). Informed consent was obtained from all patients. Patient clinical details can be found in Publication Table S1\textsuperscript{20} of expression profiling data sets, and in Publication Table S3\textsuperscript{20} for the TMA validation cohort. Only patients with WHO II and III grade ependymomas were analyzed in this study, while all other variants including subependymomas, ependymoblastomas, and myxopapillary ependymomas were excluded. Heidelberg sample cohort: Snap-frozen primary tumor samples (n=75) for gene expression analysis, and paraffin-embedded samples (n=406) for a tissue microarray analysis were collected from a single-center study at the Burdenko Neurosurgical Institute in Moscow, Russia. All tumors were banked at the time of primary diagnosis between 1993 and 2007, and diagnosed as M0 at the time of surgical resection. At least 80\% of tumor cell content was estimated in all Heidelberg and TMA samples by staining cryosections (~5um thick) of each sample with hematoxylin and eosin. Diagnoses were confirmed by histopathologic assessment by at least two neuropathologists, including a central pathology review that utilized the 2007 WHO classification for CNS tumors. The treatment strategy is described below in brief: Patients with grade 2 tumors after complete tumor resection received no adjuvant postoperative treatment and were left under observation. Patients aged 4 years or older, who had grade 3 tumors or grade 2 tumors after incomplete resection received
radiotherapy. Evaluation of the extent of surgical resection was based on neurosurgical assessment, and postoperative contrast magnetic resonance imaging within 48 hours of tumor resection. Tumor removal was evaluated as complete with no visible tumor tissue remaining (i.e. Gross-Totally Resected), or incomplete (i.e. Sub-Totally Resected). All remaining patients younger than 4 years received adjuvant chemotherapy. The anatomical location of tumors within the posterior fossa was divided into those occurring along the midline (i.e. protruding into the 4th ventricle) and those that occurred laterally growing along the cerebellopontine angle. Median and lateral classification was determined by neuro-radiological observation, in which only final reports were used. Tumors in which a significant portion of the tumor (>33%) had protruded through the Foramen of Luschka into the cerebellopontine angle were identified as lateral ependymomas. A tumor with a large medial component, and a small tongue of tumor protruding out of the Foramen of Luschka would be classified as a medial tumor. Toronto sample cohort: The tumor cohort analyzed in Toronto was composed of 102 snap-frozen samples, which were received from numerous centers.

3.7.2 Nucleic acid isolation

Extraction of high molecular weight DNA from frozen tumor samples of Heidelberg cohort was carried out as previously described. Briefly, genomic DNA from peripheral blood mononuclear cells of healthy donors (pool of ten male and female donors, respectively, age 25-40 years) used as a control was isolated by use of the Qiagen DNA Blood Midi-Kit (Hilden, Germany). Regarding the Heidelberg cohort, RNA was extracted from biopsy samples after
milling of the frozen sample in a Micro-Dismembrator S (B. Braun, Melsungen, Germany). Frozen cell powder was immediately disintegrated in Trizol (Invitrogen, Carlsbad, USA) solution and extracted by RNeasy mini columns (Qiagen, Hilden, Germany) following the manufacturer’s instructions. Extraction of high-molecular weight DNA and high-quality RNA from frozen tumor samples of the Toronto cohort was carried out as previously described79.

3.7.3 Gene expression microarray processing and data filtering

Toronto ependymoma samples were analyzed on the Affymetrix Human Exon 1.0 ST Gene Chip at the London Regional Genomics Centre (London, Ontario). Sample library preparation, hybridization, and quality control were performed according to Affymetrix recommended protocols. CEL files were imported into Affymetrix Expression Console (Version 1.1) and gene level analysis (CORE content) was performed. Arrays were Quantile normalized (sketch) and summarized using PLIER with PM-GCBG background correction. Probesets were annotated according to the human genome build HG19 (GRCh37).

Processing and hybridization of RNA isolated from 75 ependymomas of the Heidelberg cohort was processed and hybridized to the 4x44K feature Agilent Whole Human Genome Oligo Microarray according to the manufacturer’s instructions. All hybridization experiments were investigated by labeling tumor sample against reference probes normal brain pool, and scanned in a two-color Agilent Scanner G25505B (Agilent, Santa Clara, USA) according to the manufacturer’s specification. Array raw data were generated from scanned images using Feature
Extraction 9.1 Software (Agilent, Santa Clara, USA). Pre-processing of the data, and quality control were conducted with our in-house developed ChipYard framework for microarray data analysis (http://www.dkfz.de/genetics/ChipYard/) using R\textsuperscript{80} and Bioconductor\textsuperscript{81} packages. Feature signals had to fulfill the following criteria to be considered for further analysis: a minimal signal to background ratio ≥ 1.5 in at least one channel; a minimal raw signal of 200; a mean to median spot intensity ≤ the 75\% Quantile + 3 times the interquartile range of all features on the array; and a feature replicate standard deviation ≤ 0.25 per array. Normalization of raw signals was performed using vsn\textsuperscript{82}. Based on BLASTing the probes sequence information against the genome, biological annotations were retrieved from EnsEMBL (version 54, NCBI Build 36 of the human genome reference sequence). Gene expression data for both Toronto (GSE27279) and Heidelberg (GSE27287) datasets can be found at the NCBI Gene Expression Omnibus repository.

3.7.4 Array-based comparative genomic hybridization

Array- (or Matrix-) CGH\textsuperscript{83} at an average probe spacing of 0.4 Mb was carried out as previously described\textsuperscript{19} for the Heidelberg and Toronto data sets. Microarray data analysis was performed as previously reported\textsuperscript{84}. The extracted raw data was preprocessed and normalized with the R software (R Development Core Team, 2010) using the microarray data analysis framework ChipYard. To eliminate low-quality spots, three different quality criteria were applied, described elsewhere\textsuperscript{19}. A k-state hidden Markov model was fitted to the normalized data to identify genomic breakpoints. The median absolute deviation (MAD) of balanced regions was used to
define the sample experimental variation (SDa) per array. Genomic alterations were defined as log2-ratios larger than 2 times SDa (gains) or smaller than 2 times SDa (losses). Clones with incomplete mapping information, missing data in more than 20% of patient samples or known to harbor copy-number polymorphisms according to the Database of Genomic Variants (February 2007, http://projects.tcag.ca/variation/project.html) were excluded from further analyses. This resulted in a total of 10471 evaluable clones per each tumor sample. Normalized log2-ratios and raw data sets are available at the NCBI Gene Expression Omnibus (GSE27287) for the Toronto data set and the DKFZ Heidelberg data set.

3.7.5 Statistical analysis of gene expression based subgroups

Missing values in the Agilent gene expression dataset were imputed using k-nearest neighbor algorithm. In both datasets, duplicate probesets were filtered to the highest variant probesets. For HCL consensus clustering, datasets were reduced to 1000 probesets exhibiting the largest median absolute deviation (MAD). For NMF consensus clustering, datasets were reduced to 5000 probesets exhibiting the largest MAD scores. Since the Heidelberg dataset contained negative values, probesets were further preprocessed as previously described.
3.7.6 Identification of gene expression derived subtypes

To detect robust sample clusters, we used hierarchical clustering with agglomerative average linkage as our method for consensus clustering. (R package: ConsensusClusterPlus; \textsuperscript{88}. Datasets were median-centered and the distance measure was computed as 1 minus the Pearson’s correlation coefficient. Clustering was performed over 1000 iterations at a sub-sampling ratio of 0.8. SigClust was used to compute the significance of each of the identified clusters in a pairwise fashion (R package: SigClust)\textsuperscript{53}. To establish whether samples were representative of their subtype, silhouette analysis was performed to identify “core” samples (R package: cluster)\textsuperscript{30}. Only members with a positive silhouette value were retained for further analysis as representative samples of their subgroup assignment. Subclass mapping using the Submap module (version 3), within GenePattern software was performed to assess the commonality of the subtypes identified in the two data sets\textsuperscript{55,89}. Non-negative matrix factorization (NMF) was used to assess the sample memberships at a 3-subgroup classification. NMF (R package: NMF version 0.5.02) was performed on each dataset for 1000 re-sampling iterations using the parameters described previously\textsuperscript{52}. Sample memberships were then compared against the HCL analysis using a Rand index. The significance of the Rand index was assessed by permutation of the Toronto sample labels and calculating the Rand Index over 10,000 iterations in order to generate a null distribution of Rand index values.

Principal components analysis was performed within Partek Genomics Suite (Partek Inc.) to compare Group A and Group B posterior fossa subtypes with the same 1000 genes used for
HCL consensus clustering. Both groups were highlighted by their HCL cluster membership, and encircled by ellipsoids measuring 2 standard deviations around the center of each subgroup.

3.7.7 Identification of biological pathways distinguishing group A from group B posterior fossa ependymoma

Gene Set Enrichment Analysis\textsuperscript{56}, as visualized in Cytoscape (version 2.7.0)\textsuperscript{90}, and the Enrichment Map software\textsuperscript{57}, were used to identify the biological processes discriminating Group A from Group B ependymomas. Gene sets were compiled from NCI\textsuperscript{91}, KEGG\textsuperscript{92}, PFAM\textsuperscript{93}, Biocarta: Pathway Resource (www.biocarta.com), and Gene Ontology (http://www.geneontology.org) databases. Using the Toronto dataset, GSEA analysis was performed using gene-set permutations with a FDR cutoff of 0.035 and p-value cutoff of 0.01. The network map was manually curated removing general and uninformative sub-networks and nodes resulting in a simplified network map shown in Figure 4. The complete network map for the Toronto dataset can be found in Witt and Mack et al., Figure S4A\textsuperscript{20}. Due to differences in microarray platforms, different GSEA parameters were needed in order to generate the network map comparing Group A and Group B in the dataset from Heidelberg (Publication supplemental Figure S4B; FDR<0.10, p-value<0.01)\textsuperscript{20}. GSEA results for both Toronto and Heidelberg datasets can be found in Publication Supplemental Table S2\textsuperscript{20}. 
3.7.8 Selection of candidate genes and gene signatures

We computed the test statistic of Goeman’s Global test for logistic regression to derive the contributions made by each of the genes on discriminating between the two groups\(^9^4\). Individual transcripts were then ranked by their influence. Finally, we selected 100 candidate genes with highest ranks in both datasets, as representative subgroup markers for Group A and Group B.

3.7.9 Statistical analysis of clinical parameters

Estimation of survival time distribution was performed by the method of Kaplan and Meier. For comparisons of two or more survival curves, the log-rank test was used. Comparisons of binary and categorical patient characteristics between age groups were performed by use of a two-sided Fisher’s exact test. To evaluate the status of recurrence, secondary metastasis and mortality 5 years after diagnosis, we selected patients with a 5-year follow-up within the expression and validation data set. Wilcoxon rank sum test was used to perform two-sample location tests for at least ordinal covariates. Multivariate Cox regression analysis was used to evaluate the impact of subgroup markers together with prognostically relevant clinical and histopathological factors. The result of a test was always judged as statistically significant when the corresponding two-sided p-value was smaller than 5%. The prognostic value of clinical and molecular factors was assessed by their estimated hazard ratios including 95% confidence intervals. All statistical computations were performed with the statistical software environment R, version 2.9.0\(^8^0\) using
the R packages Design, version 2.2-0, and aCGH, version 1.18.0, from the Bioconductor project.

3.7.10 Construction of ependymoma tissue microarrays

Hematoxylin and eosin (HE) stained sections from all 406 paraffin blocks were prepared to define representative tumor regions. In addition, 10 samples of non-tumor brain tissues were included as a control. All tissue specimens were arrayed into a recipient block as previously described95.

3.7.11 Immunohistochemistry

Antibodies against the following antigens were used: LAMA2 (Abnova, Taipei City, Taiwan; H00003908-M0; dilution 1:2500), NELL2 (Abcam, Cambridge, MA; ab80885; dilution 1:1000), KIF27 (Altas, Stockholm, Sweden; HPA022246; dilution 1:500) and TNC (Novus Biologicals, Cambridge, UK; NB110-68136; dilution: 1:1000). TMA staining was performed, evaluated and scored as published96.
4 Chapter 4:

Somatic mutational and epigenomic landscape of Group A and B posterior fossa ependymoma

4.1 Chapter 4: Summary

Ependymomas are common childhood brain tumors that occur throughout the nervous system, but are most common in the pediatric hindbrain. Current standard therapy comprises surgery and radiation, but not cytotoxic chemotherapy as it does not further increase survival. Whole-genome and whole-exome sequencing of 47 hindbrain ependymomas reveals an extremely low mutation rate, and zero significant recurrent somatic SNVs. While devoid of recurrent SNVs and focal copy number aberrations, poor prognosis hindbrain ependymomas exhibit a CpG island methylator phenotype (CIMP). Transcriptional silencing driven by CpG methylation converges exclusively on targets of the polycomb repressor complex 2 (PRC2) that represses expression of differentiation genes through tri-methylation of H3K27. CIMP-positive (CIMP+) hindbrain ependymomas are responsive to clinical drugs that target either DNA or H3K27 methylation both in vitro and in vivo. We conclude that epigenetic modifiers are the first rational therapeutic candidates for this deadly malignancy, which is epigenetically de-regulated but genetically bland.
4.2 Chapter 4: Introduction

Ependymomas are malignancies that occur throughout the nervous system, but are more common in the hindbrain in children, as opposed to supratentorial and spinal cord tumors, which are more frequently diagnosed in adulthood. Despite being histologically identical, ependymomas from different regions of the nervous system are biologically and clinically distinct\(^1,2\). Current therapy for all ependymoma patients consists of maximal safe surgical resection, followed by radiation therapy\(^3\). While adjuvant chemotherapy is routine for most children with malignant brain tumors, it is not part of the current standard of care for ependymoma patients as multiple clinical trials have failed to show any survival benefit after cytotoxic chemotherapies\(^4,5\). Even at the time of disease recurrence, chemotherapy has not been shown to be effective for ependymomas; therefore, many children with recurrent ependymoma undergo a full but palliative second course of cranial irradiation\(^6-8\). Indeed, while treatment protocols for many other childhood malignancies have changed and improved in the past two decades, ependymoma therapy remains stagnant. The mechanisms underlying the chemoresistance of ependymoma are not known.

Within each anatomic compartment (supratentorial/hindbrain/spinal), there is additional inter-tumoral heterogeneity in the form of well-documented molecular subtypes of ependymoma\(^9-11\). Ependymoma subtypes are clinically and functionally relevant, as rational therapies may only be effective in a single subtype of the disease\(^12\). Ependymomas are thought to arise from the regionally distinct radial glial cells (RGCs). Differences between those RGC populations are likely carried forward in the neoplasm, and may account for a portion of the observed heterogeneity\(^1,9\). Hindbrain ependymomas occur within the posterior fossa of the skull, and are clinically referred to as ‘posterior fossa’ (PF) ependymomas. There are two clear and distinct
subtypes of PF ependymoma; one that occurs in older children and adults with very good prognosis (Posterior Fossa Group B, or PFB), and another found predominantly in infants, which is associated with poor prognosis in spite of maximally aggressive therapy (Posterior Fossa Group A, or PFA)\textsuperscript{10,11}.

4.3 Chapter 4: Results

4.3.1 The somatic mutational landscape of ependymoma

To uncover the biology of PF ependymomas, we undertook whole genome sequencing of tumor and matching germline DNA from five ependymomas (3 PFA, 2 PFB), and whole exome sequencing of an additional 42 PF ependymomas and their matching germline DNA (24 PFA, and 18 PFB) (Figure 4.1A-B, Supplemental Figure S4.1, Publication Supplemental Tables S1-2)\textsuperscript{97}. Unlike some other childhood malignancies, the rate of somatic SNVs did not correlate significantly with the age at diagnosis (Supplemental Figure S4.1)\textsuperscript{13}. Further, the rate of somatic SNVs was extremely low in PF ependymomas, with an average of 5.0 somatic non-synonymous SNVs per exome across the entire cohort (Figure 4.1B), and low in both PFA (4.6 SNVs per tumor) and PFB ependymomas (5.6 SNVs per tumor, Publication Supplemental Table S3)\textsuperscript{97}. Perhaps the most surprising result was that there were zero significant recurrent mutations across the cohort of 47 PF ependymomas as detected by two different algorithms MUTSIG\textsuperscript{14} and MUSIC\textsuperscript{15} (Figure 4.1C, Publication Supplemental Tables S3-6)\textsuperscript{97}. Despite the absence of significant recurrent SNVs, PFB harbored frequent and recurrent large-scale copy number alterations (CNAs) indicative of chromosomal aneuploidy (Supplemental Figure S4.2).
Compared to other malignancies, PF ependymomas have a very low rate of SNVs per megabase, and the lowest number of recurrent significant SNVs, making PF ependymoma the first malignancy for which genome sequencing across a broad cohort (n = 47) has failed to identify any significantly and recurrently mutated genes (Figure 4.1C, Publication Supplemental Tables S3-7)\(^9^7\).

### 4.3.2 DNA methylation profiles of ependymoma are regionally specified

A number of other childhood nervous system malignancies, including medulloblastoma, retinoblastoma, glioblastoma, atypical teratoid/rhabdoid tumor, and neuroblastoma, have recently been demonstrated to harbor a paucity of recurrent mutations, with a significant proportion of the recurrent events converging on epigenetic mechanisms\(^13^{16-26}\). Due to the absence of recurrent and significant SNVs and CNAs, we hypothesized that PFA ependymomas could be driven by epigenetic mechanisms. We studied DNA methylation patterns in a discovery cohort of 79 ependymomas using methyl-binding domain-2 (MBD2) protein recovery followed by hybridization to Nimblegen 385K CpG island promoter plus microarrays (MBD2-chip). Unsupervised consensus clustering of CpG methylation profiles yielded three distinct subgroups, composed of supratentorial, posterior fossa, and mixed spinal/posterior fossa tumors respectively, in a pattern highly similar to that yielded by unsupervised clustering of gene expression profiles (Figure 4.2A, Supplemental Figure S4.3)\(^10\). The group of pure posterior fossa tumors corresponds to PFA ependymomas, while the PFB ependymomas cluster with the spinal ependymomas. We validated our discovery cohort findings through study of a non-overlapping cohort of 48 PF ependymomas using an orthogonal technology (Illumina Infinium 450k methylation arrays). In these validation experiments, the DNA methylome of PFA
ependymomas was very distinct from PFB tumors (Figure 4.2B, Supplemental Figure S4.4). Unsupervised clustering of CpG methylation signatures was very robust, supporting two major molecular subtypes, even after applying a number of distinct bioinformatic and biostatistical techniques (Supplemental Figure S4.4). We conclude that PFA and PFB ependymomas have very distinct methylomes, and that epigenetic biomarkers could be used to develop a clinically relevant molecular classification of PF ependymomas. To this end, we identified three genes that exhibited increased CpG methylation in most PFA tumors, but not in PFB tumors (Supplemental Figure S4.5). We determined the presence of CpG hypermethylation representing PFA tumors using a mass spectrometry based technology (Sequenom) on our training cohort (Supplemental Figure S4.5-4.6, and Publication Supplemental Table S8). We were able to validate our Sequenom based biomarker panel on an independent cohort of ependymomas using FFPE tissues to predict both progression free and overall survival (Supplemental Figure S4.5-4.6 and Publication Supplemental Table S8)\textsuperscript{97}. We conclude that division of PFA from PFB ependymomas using a mass spectrometry based biomarker should be feasible in a clinical setting.

4.3.3 Group A posterior fossa ependymoma demonstrate a CpG island methylator phenotype

We next compared the extent of promoter CpG methylation in PFA ependymomas to that of PFB ependymomas and found that PFA tumors have a much higher extent of CpG island methylation (Figure 4.2C-E, Supplemental Figure S4.7-4.8, and Publication Tables S9-14)\textsuperscript{97}. In comparison to PFB ependymomas, PFA tumors have more methylated CpG sites (Figure 4.2C), more genes with significant CpG methylation (Figure 4.2D), and more genes that are
transcriptionally silenced by CpG hypermethylation (Figure 4.2E). We conclude that PFA ependymomas exhibit a ‘CpG Island Methylator’ or ‘CIMP’ phenotype, and suggest that PFA ependymomas be referred to as PFA CIMP positive (PFA-CIMP+) ependymomas, and PFB as PFB CIMP negative (PFB-CIMP-) (Supplemental Figure S4.9-4.10, Publication Supplemental Table S15). To determine the mechanism by which CpG hypermethylation driving transcriptional silencing promotes the pathogenesis of PFA ependymoma, we performed a pathway analysis in our discovery cohort of PFA and PFB ependymomas studied by MBD2-chip (Figure 4.3A, Publication Supplemental Table S16). Although olfactory signaling was the only significant pathway enriched in PFB ependymomas, genes CpG methylated in PFA ependymoma showed a remarkable convergence on genes documented as silenced in embryonic stem (ES) cells by the polycomb repressive complex 2 (PRC2). In our non-overlapping, independent validation dataset studied by Illumina Infinium 450K arrays, we observed no significant pathways in the PFB ependymomas, while the PFA tumors exhibited the same convergence on gene targets that are silenced by PRC2 in ES cells (Figure 4.3B, Publication Supplemental Table S17). The PRC2 complex contains the histone methylase EZH2, which tri-methylates H3K27, thereby driving gene silencing. Genes known to be required for differentiation, and which are silenced by PRC2 have been documented to frequently undergo cancer specific CpG methylation, and it is described that both DNA and histone methylation contribute to ongoing gene silencing in these cancers.

4.3.4 Convergence on PRC2 targets

We next sought to validate these pathway findings by performing H3K27me3 ChIP-seq in 11 primary PF ependymomas. Our findings demonstrate distinct H3K27me3 signatures in PFA-
CIMP+ versus PFB-CIMP- ependymomas (Figure 4.3C, Supplemental Figure S4.11-4.12, Publication Supplemental Table S18). Furthermore, the gene expression of H3K27me3 target genes can robustly stratify PFA-CIMP+ from PFB-CIMP- tumours, thus highlighting the distinct epigenetic differences between these subgroups (Supplemental Figure S4.11). Examination of differential H3K27me3 targets demonstrated a convergence and significant overlap with PRC2 targets in ES cells observed exclusively in PFA-CIMP+ tumours (Figure 4.3D). Further, a significant proportion of shared PFA-CIMP+ and ES cell H3K27me3 targets were CpG hypermethylated exclusively in PFA (CIMP+) tumours, a pattern that was not detected in PFB-CIMP- ependymoma (Figure 4.3E-F, Publication Supplemental Tables S18-19). We hypothesize, therefore, that hyperactivity of the PRC2 complex leading to tumor suppressor gene silencing with subsequent gene silencing by DNA CpG hypermethylation contributes to the pathogenesis of PFA-CIMP+ ependymoma.

4.3.5 Whole genome DNA methylation profiling of ependymoma

We next sought to expand our global analysis of CpG methylation by performing whole-genome bisulfite sequencing in 3 PFAs, 3 PFBs, 3 fetal normal brains, and 3 adult normal brains (Publication Supplemental Tables S20-22). Here we observed the same patterns of increased CpG methylation at CpG islands occurring specifically in PFA-CIMP+ tumours consistent with a CpG island methylator phenotype (Figure 4.4A,C, Supplemental Figure S4.13). In line with other solid tumours we identified additional cancer-specific epigenome patterns including hypomethylation of repetitive elements (LINEs, SINEs, and LTRs) restricted to PFA-CIMP+ ependymoma and subgroup specific partially methylated domains (Figure 4.4B,D, Supplemental Figure S4.13-4.14) These findings illustrate genome-wide DNA methylation alterations in PFA-
CIMP+ ependymoma, concurrent with a silent genome exhibiting few CNAs and no significant and recurrent somatic SNVs.

4.3.6 *Group A (CIMP+) ependymoma are highly and specifically inhibited by DNA methylation Inhibitors*

While our genomic and epigenomic data suggest that over-activity of the PRC2 complex, and/or subsequent promoter CpG hypermethylation may be involved in driving the pathogenesis of PFA-CIMP+ ependymomas, they do not address whether or not these mechanisms continue to be necessary for tumor maintenance, and would therefore constitute an effective target for therapy. Functional assessment of CpG and histone methylation in ependymoma is harshly limited by the complete lack of established ependymoma cell lines, xenografts, or transgenic mouse models\(^{1,9,10}\). To this end we established four short-term, patient-derived primary ependymoma cultures from two PFA-CIMP+ tumors, and two childhood ST ependymomas. We were unable to grow any PFB-CIMP- ependymomas in vitro. Treatment of PFA-CIMP+ cultures with the DNA de-methylating agents 5-aza-2'-deoxycytidine (decitabine) resulted in marked de-repression of genesets enriched in EZH2 targets and known DNA hypermethylated genes in other solid cancers (Figure 4.4E-F) Furthermore, compared to ST ependymoma primary cultures in vitro, Decitabine demonstrated significant anti-neoplastic effect on both PFA-CIMP+ tumors at low dose nanomolar levels (Figure 4.4G). In order to model the effects of Decitabine on PFA-CIMP+ cultures as early, and representative of the patient tumour as possible, we derived a passage zero ex-vivo culture from a PFA-CIMP+ metastasis. In this ex-vivo culture we demonstrate significant impairment of neurosphere colony formation upon DNA methylation blockade (Figure 4.4H, Supplemental Figure S4.15) Because decitabine is
FDA approved for the treatment of hematopoietic malignancies, we propose that it could be rapidly re-purposed in a clinical trial for children with PFA-CIMP+ ependymoma\textsuperscript{29}. In addition, we observed additive effects on combining decitabine and an FDA approved HDAC inhibitor (suberoylanilide hydroxamic acid, SAHA) against PFA-CIMP+ ependymoma (Supplemental Figure S4.16).

The tool compound 3-Deazaneplanocin-A (DZNep) is known to target the PRC2 complex and result in diminished tri-methylation of H3K27 through degradation of PRC2 complex proteins\textsuperscript{30}. Treatment of PFA-CIMP+ ependymoma, but not ST ependymoma with DZNep is highly effective in the nanomolar range \textit{in vitro} (IC50 for E517 = 95 nm, E520 = 262 nm) (Figure 4.5A). We also observed additive effects between DZNep and SAHA, and DZNep and Decitabine (Supplemental Figure S4.16). Treatment of PFA-CIMP+ ependymomas with DZNep, compared to controls, results in decreased expression of EZH2, decreased tri-methylation of H3K27, and increased cleavage of PARP (Figure 4.5B). \textit{In vivo} treatment of established xenografts of human PFA-CIMP+ ependymoma with DZNep using either a flank model (Figure 4.5C), or an orthotopic intra-cerebellar xenograft model (Figure 4.5D) results in decreased tumor volume and improved survival. Furthermore, PFA-CIMP+ ependymoma cells isolated from tumor xenografts treated \textit{in vivo} with DZNep have a markedly reduced colony forming ability compared to controls, suggesting that the compound targeted ependymoma cells with clonogenic or tumor initiating potential (Supplemental Figure S4.15). In addition, treatment with a recently published, and extremely potent, highly selective, S-adenosylmethionine competitive small molecular inhibitor of EZH2 (GSK-343) results in significant de-repression of gene expression in PFA-CIMP+ ependymoma including genes which are known targets of PRC2 in ES cells (Figure 4.5E-F)\textsuperscript{31}. We found that treatment with GSK343 but not an inactive compound with the same molecular backbone (GSK669) resulted in diminished levels
of H3K27me3, and had a potent anti-neoplastic effect against PFA-CIMP+ ependymoma (Figure 4.5G, Supplemental Figure S4.15). These findings are further supported in a passage zero PFA-CIMP+ ex-vivo culture treated immediately with GSK343, which significantly impaired neurosphere colony formation (Figure 4.5H). We therefore hypothesize that ongoing hypermethylation of promoter CpG islands and H3K27 contribute to the maintenance of PFA-CIMP+ ependymoma, and that targeting these epigenetic mechanisms represents the first identified rational targets for this chemotherapy resistant, epigenetically dysregulated, and genetically bland childhood cancer.

4.4 Chapter 4: Discussion

Our findings demonstrate that Group A posterior fossa ependymoma, which exhibit no recurrent focal copy number alterations, are also characterized by a stable mutational landscape. While a number of other pediatric malignancies have a very low incidence of recurrent somatic mutations, we are unaware of any other malignancies with zero significant recurrently mutated genes. The silent genomic architecture of posterior fossa ependymoma was validated in a study led by Parker et al., (2014), in which whole-genome sequencing and RNA-sequencing of 62 PF ependymomas, failed to identify a recurrently mutated gene or fusion transcript. The possibility of somatic alterations in non-coding space still remains, and may be adequately explored in the future as the cost of high-coverage whole-genome sequencing decreases. To this date, the most notable and frequently reported example is TERT promoter mutation as shown in numerous solid cancers. TERT promoter mutations, however, are non-recurrent and highly infrequent in ependymoma.
A number of recent seminal publications have demonstrated that childhood nervous system tumors harbor very few recurrent genetic events, and that many of the recurrent events converge on genes important in epigenetic processes such as CpG hypermethylation, post-translational modification of histones, and even mutation of the actual histone genes\textsuperscript{13,16-20,22-24}. While mutation of \textit{IDH1}, \textit{IDH2}, \textit{TET1}, \textit{TET2}, and/or \textit{DNMT3a} have been documented in other types of cancer with a CIMP phenotype, we did not observe any such mutations in PF ependymomas\textsuperscript{32-34} (Publication Supplemental Table S23)\textsuperscript{97}.

Subgroups of patients with a CIMP phenotype have a better prognosis for some cancer histologies\textsuperscript{35}, but not others\textsuperscript{2,27,36-40}, suggesting that CIMP positive tumors represent distinct subgroups of disease, but that the CIMP phenotype itself is not intrinsically benign or responsive to therapy. PFA-CIMP+ ependymomas have a nearly normal genetic code and a very poor prognosis in comparison to aneuploid PFB (CIMP-) ependymomas where 5-year overall survival exceeds 95\% of patients. Many of the cytotoxic chemotherapeutics currently used clinically function through promoting damage to the genomic DNA, which subsequently induces cancer cells with deranged and disorganized genomes to undergo apoptosis. In light of the nearly normal genetic code found in PFA-CIMP+ ependymomas, perhaps it is not surprising that a therapy based on DNA damage has not shown efficacy in clinical trials.

A recurring theme in several solid cancers studied to date, is a convergence of gene expression and DNA methylation networks upon genes known to be silenced by the PRC2 complex in ES cells, particularly in poorly differentiated cancers\textsuperscript{102}. Our data demonstrate hyperactivity of DNA CpG methylation and disparate PRC2 H3K27me3 signatures in poor prognosis PF ependymomas that may be necessary for tumor maintenance. The cause and effect of this ES cell signature remains unclear, however Kim et al., (2010) suggest that the similarities between
cancer and ES cells may be attributed to an active MYC transcriptional network. Indeed, MYC gene expression has been shown to be up-regulated preferentially in PFA-CIMP+ ependymoma as compared to PFB-CIMP- tumours (Data unpublished). We hypothesize that PF ependymomas may arise from primitive neural precursors, arising in the radial glial cell lineage, which may fail to undergo terminal differentiation during neurogenesis. Evaluating this effect may involve over-expression of active DNA and/or H3K27 methyltransferases in hindbrain radial glial cells during murine development.

Given that aggressive infant ependymomas have no chemotherapeutic options available, the threshold is low for rapid translation of a rationale, novel, and effective drug therapy. Our findings demonstrate that inhibitors that target DNA CpG methylation, PRC2/EZH2, and/or HDAC inhibitors represent the first rational strategies for therapy of this untreatable disease, and should be considered for testing in future clinical trials for children with PFA-CIMP+ ependymoma.
Somatic single nucleotide variants are rare in the posterior fossa ependymoma genome

(A) Summary of clinical and genomic details of posterior fossa ependymomas stratified according to Group-A and Group-B ependymoma (Wilcoxon rank sum test). (B) Bar graphs summarizing the numbers and frequencies of SNVs detected by whole-genome and whole-exome sequencing of posterior fossa ependymomas. (C) Comparison of numbers of significant and recurrently mutated genes, and mutation rates, in several whole-genome and whole-exome sequencing studies of adult and pediatric cancers (FDR < 0.1).
Figure 4.2

DNA methylation profiles suggest that Group A ependymoma demonstrate a CpG island methylator phenotype

(A) Unsupervised hierarchical clustering of 79 ependymoma DNA methylation profiles. (B) Heatmap of 48 PF ependymoma DNA methylation profiles. Group A and Group B clinical differences were assessed using a two-sided Fisher’s Exact Test. (C) Volcano plot comparing the number of significant methylated CpG sites between Group A and Group B (p<0.05, Wilcoxon-test, FDR-corrected). Differences in number of methylated genes (D) and methylated/silenced genes (E) in Group A versus Group B (p<0.0001, binomial-distribution-test).
**Figure 4.3**

Group A (CIMP+) and Group B (CIMP-) ependymoma are distinguished by CpG hypermethylated and H3K27 tri-methylated genes related to PRC2 occupancy in ES cells

(A-B) CpG methylated pathways in Group A(CIMP+) and Group B(CIMP-) ependymomas in a discovery and validation cohort. (C) Differential H3K27me3 binding sites distinguishing Group A and Group B (p<0.01(MACSv2.0), p<0.05(R:DiffBind)). Venn diagrams comparing: (D) Group A and B H3K27me3 genes with ES cell PRC2 genes (E) Group A H3K27me3 and DNA hypermethylated genes with ES cell H3K27me3 genes. (F) Group B H3K27me3 and DNA hypermethylated genes with ES cell H3K27me3 genes (binomial-distribution-test).
Figure 4.4

Whole-genome bisulphite sequencing validates a CpG island methylator phenotype in Group A ependymoma

(A-B) Heatmap of DNA methylation at CpG islands and repetitive regions in Group-A(CIMP+) versus Group-B(CIMP-).

(C-D) Proportion of hyper-methylated versus hypo-methylated regions at CpG islands and repetitive elements in Group-A and Group-B(p<2.2e-16, binomial-distribution-test). Top 10 pathways up-regulated (E) and differentially expressed (F) upon 5-aza-2'-deoxycytidine (DAC) treatment of E517-PF2 and E520-PF1(p<0.0001, binomial-distribution-test). (G) Survival analysis of E479-ST2, E479-ST1, E517-PF2 and E520-PF1 cells treated 7d with DAC(p=0.05, two-sided-t-test, error-bars=s.e.m., technical:n=6 over biological:n=2). (H) Limiting dilution assay of zero-passage Group-A cells treated 2wks with DAC(p=4.18e-10, chi-square-test).
Modulation of H3K27 methylation has anti-neoplastic effects against Group A ependymoma

(A) Survival of E479-ST1, E478-ST2, E520-PF1 and E517-PF2 cells treated 7d with 3-Deazaneplanocin-A (DZNep, \( p < 0.0001 \), two-sided t-test, error-bars=s.e.m., biological:\( n=3 \)). (B) EZH2, H3K27me3, H3K4me3, and cleaved-PARP protein expression in E520-PF1 and E479-ST1 cells treated 7d with DZNep (500 nM). (C) E520-PF1 flank tumor volumes following treatment with DZNep (\( p = 0.0087 \), Wilcoxon-test, error-bars=s.d.) (D) Survival of E520-PF1 posterior fossa tumour bearing mice treated with DZNep (\( p = 0.033 \), log-rank-test). Top 10 pathways up-regulated (E) and differentially expressed (F) upon treatment of E517 and E520 with GSK343. (G) Cell proliferation of E520-PF1 and E479-ST1 cells treated 11d with GSK343 (active-inhibitor) or GSK669 (inactive-inhibitor) (\( p = 0.0022 \), two-sided t-test, error-bars=s.e.m., technical:\( n=9 \) over biological:\( n=3 \)). (H) Limiting dilution assay of passage-zero Group-A cells treated 2wks with GSK343 (\( p = 2.83\times 10^{-5} \), chi-square-test).
4.6 Chapter 4: Supplemental figures

Figure S4.1

(A) Circos plots depicting whole genome sequencing profiles in three Group A and two Group B ependymomas. From outer track to inner track depicts the following: 1) Chromosomal location, 2) Copy number alterations, and 3) Chromosomal translocations and focal insertions, deletions, and duplications. Inter-chromosomal translocations are shown in green, intra-chromosomal translocations show in purple, focal deletions in blue, and focal duplications/insertions in red.

(B) Dot plot demonstrating the lack of correlation of non-silent somatic single nucleotide variants with age. An R² and was computed using a Pearson product moment correlation, which follows a t-distribution for p-value calculation.
Figure S4.2

(A) Unsupervised consensus non-negative matrix factorization of a discovery cohort consisting of 79 primary PF ependymoma copy number profiles generated by 10K array comparative genomic hybridization (Agilent). Shown are subgroup profiles at various subgroup classification (k) ranks. (B) Comparison of the numbers of chromosomal arm alterations in Group A versus Group B PF ependymoma in a discovery cohort. Statistical significance was assessed using a Wilcoxon-rank sum test. (C) Comparison of the numbers of chromosomal alterations in Group A versus Group B in a validation cohort of 48 PF ependymoma, in which copy number profiles were generated by Illumina Infinium 450K methylation arrays. Statistical significance was assessed using a Wilcoxon-rank sum test. (D-E) Copy number profile of a representative Group A (UHS40) and Group B (GER-1E31) PF ependymoma. Green highlight indicates a significant copy number change above noise. The red indicates a copy neutral ratio of zero.
Figure S4.3

(A) Unsupervised consensus hierarchical clustering (HCL) of 47 primary ependymoma gene expression profiles (Affymetrix Exon 1.0ST), which also had matched DNA methylation profiles illustrated in Fig. 2.2A. Clustering was performed by median centering probes, using the average linkage method, and with a subset of 1000 genes exhibiting the greatest standard deviation. (B-C) Change in the cumulative distribution function plots (ΔCDF) of DNA methylation (MBD2-chip) and gene expression (Affymetrix Exon 1.0ST) clustering results demonstrating the different levels of statistical confidence for differing number of subgroups. k denotes the number of subgroups. (D) Overlap analysis between DNA methylation and gene expression based subgrouping. The degree of overlap was analyzed using the Rand index with 10,000 sample permutations of sample labels to measure statistical significance. (E-F) Silhouette analysis identifies “core” samples defined as samples with positive silhouette widths for both DNA methylation (MBD2-chip) and gene expression (Affymetrix 1.0ST) based consensus HCL.
Figure S4.4

(A) Principal components analysis of gene expression and DNA methylation profiles between Group A and Group B PF ependymomas. DNA methylation profiles were generated on the validation cohort using the Illumina 450K methylation platform. (C) Subgroup analysis distinguishing PFA versus PFB ependymoma by DNA methylation profiles (Illumina) using K-means, NMF, HCL and SOM consensus clustering. (D) Change in the cumulative distribution function plots (dCDF) of DNA methylation clustering results (C) demonstrating the different levels of statistical confidence for differing number of subgroups. k denotes the number of subgroups. (E) Unsupervised k-means consensus clustering of 48 PF ependymoma DNA methylation profiles generated by Illumina Infinium 450K methylation arrays. Shown are heatmaps from rank k2 to k4. (F) Lorenz curves for DNA methylation cluster analysis performed from k2 to k10. (G) Change in Gini values for cluster analysis demonstrating two principle subgroups of PF ependymoma defined by DNA methylation.
Figure S4.5

(A) DNA methylation dot plots for the top 3 methylated genes (CRIP1, CYP26C1, and PKP1) for which a Sequenom assay could be designed and optimized for robust stratification. (B) Calculation of misclassification errors based upon a value of threshold and number of genes selected for the prediction analysis of microarrays (PAM) algorithm. (C) Calculation of misclassification errors based upon a value of threshold and number of genes selected for the prediction analysis of microarrays (PAM) algorithm relative to subgroup. (D-E) Kaplan-Meier curves are shown for training samples classified into Group A (CIMP+) and Group B (CIMP-) for both progression-free and overall survival. Number of cases in each subgroup is shown below the graph. Statistical significance for all Kaplan-Meier curves was measured using a log-rank test. (F-G) Using the markers identified in (A), class prediction was performed on a non-overlapping cohort of 62 PF ependymomas (DNA isolated from frozen or FFPE). Samples of the predicted dataset were provided by The Hospital for Sick Children, MD Anderson Cancer Centre, Children’s Hospital Boston and the University of Michigan. Depicted are Kaplan-Meier curves for progression-free and overall survival, with significance measured using a Log-rank test. Number of cases in each subgroup is shown below the graph.
Figure S4.6

(A) Age distribution of Group A (CIMP+) versus Group B (CIMP-) PF ependymomas in a training cohort stratified by the Sequenom 3 gene biomarker panel. (B) Age distribution of Group A (CIMP+) versus Group B (CIMP-) PF ependymomas in a validation cohort stratified by the Sequenom 3 gene biomarker panel. (C) Progression-free survival for Group A (CIMP+) versus Group B (CIMP-) PF ependymoma for gross-totally resected tumours stratified by the 3 gene biomarker panel. (D) Overall survival for Group A (CIMP+) versus Group B (CIMP-) PF ependymoma for gross-totally resected tumours stratified by the 3 gene biomarker panel. (E) Multivariate analysis of posterior fossa ependymoma progression free survival including radiation, grade, level of resection and 3-marker subgroup stratification as potential prognostic factors.
Clinical associations regarding tumour location, gender, and age are shown in the top panel. (B) Volcano plot demonstrating an over-representation of significantly methylated regions of interest (ROI) in Group A (CIMP+) versus Group B (CIMP-). Significance of methylation differences was determined using a Wilcoxon-rank sum test with FDR correction using the Benjamini-Hochberg method. (C) Bar chart demonstrating a significantly greater number of methylated ROIs in Group A (CIMP+) versus Group B (CIMP-) as measured by a binomial distribution test. (D) Biological processes identified in Group A (CIMP+) and Group B (CIMP-) as determined by a pathway association test using genesets C2 and C5 provided by MSigDB (Broad Institute).
Figure S4.8

(A) Group A (CIMP+) DNA methylation profiles plotted against Group B (CIMP-) methylation profiles. Red dotted line denotes equivalent levels of DNA methylation between Group A and B. (B) Starburst plot comparing significant differentially methylated genes versus significant differentially expressed genes. Significant genes or methylated probes were identified using a Wilcoxon-Rank sum test with FDR correction of p-values using the BH method. (C) Bar graph comparing the differences in number of significantly methylated CpG sites. (D-E) Dot plots demonstrating the differences between methylated and silenced CpG sites and genes, within a given tumour, between Group A (CIMP+) and Group B (CIMP-) ependymoma. Significant differences were calculated using a Wilcoxon-rank sum test with FDR (BH correction).
Figure S4.9

(A) Unsupervised NMF consensus clustering of CpG probes exhibiting a standard deviation greater than 0.2 is shown at subgroup rank 3 for normal brain, Proneural-GBM, and CIMP+ GBMs tumours. (B) Heatmap of CpG probes demonstrating increased methylation of CpG sites in GBM-CIMP and Proneural-GBM versus Normal Brain. (C) Boxplot of CpG probes demonstrating increased methylation of CpG sites in GBM-CIMP and Proneural-GBM versus Normal Brain. (D) Unsupervised NMF consensus clustering of CpG probes exhibiting a standard deviation greater than 0.2 is shown at subgroup rank 3 for normal brain, Group A (CIMP+), and Group B (CIMP-) PF ependymomas. (E) Heatmap of CpG probes demonstrating increased methylation of CpG sites in Group A (CIMP+) PF ependymoma versus Group B (CIMP-) and Normal Brain. (F) Heatmap of CpG probes demonstrating increased methylation of CpG sites in Group A (CIMP+) PF ependymoma versus Group B (CIMP-) and Normal Brain.
**Figure S4.10**

(A) Overlap of genes significantly DNA hypermethylated in Breast, Group A Ependymoma, and GBM CIMP (FDR < 0.05). Significance of overlap was calculated by comparing observed versus expected proportions using the Binomial distribution test.

(B) Overlap of genes significantly DNA hypermethylated and silenced in Breast, Group A Ependymoma, and GBM CIMP (FDR < 0.05).

(C) Overlap of ES cell PRC2 regulated genes significantly DNA hypermethylated in Breast, Group A Ependymoma, and GBM CIMP (FDR < 0.05).

(D) Overlap of genes DNA hypermethylated or specifically marked by H3K27me3 in Group A or B against the COSMIC cancer gene consensus list.
Figure S4.11

(A) Group A (top panel) and Group B (lower panel) specific EZH2 target profiles generated in 11 PF primary ependymomas. Heatmaps are shown of -400bp to +400bp around EZH2 peaks identified using MACS2 (p < 0.01). Differential peaks were identified using the DIFFBind package using a p < 0.05 cutoff. (B) Pair-wise sample – sample Pearson correlations (top panel) of differential Group A and B EZH2 ChIP-seq peaks demonstrating robust differences between Group A and Group B EZH2 binding profiles. Heatmap of differential EZH2 peaks in Group A versus B is shown in the lower panel. (C) Unsupervised hierarchical clustering of Group A and B EZH2 target genes demonstrating robust segregation of Group A and B based upon a core set of EZH2 bound genes. (D–E) Heat map of differentially expressed EZH2 target genes between Group A and B identified using SAM analysis FDR < 0.05. The top 10 differentially expressed EZH2 targets in Group A and B are also shown. (F) Overlap of Group A (CIMP+) specific DNA hypermethylated genes and Group A specific EZH2 targets. (G) Overlap of Group B (CIMP-) specific DNA hypermethylated genes and Group B specific EZH2 targets.
Figure S4.12

(A) Group A (top panel) and Group B (lower panel) specific H3K27me3 target profiles generated in 11 PF primary epidernomas. Heatmaps are shown of -400bp to +400bp around H3K27me3 peaks identified using by MACS2 (p < 0.01). Differential peaks were identified using the DIABIND software using a p < 0.05 cutoff. (B) Unsupervised consensus hierarchical clustering of 51 Group A and B primary epidernomas gene expression signatures (Affymetrix Exon 1.0ST) limited to H3K27me3 predicted target genes. Clustering was performed by median centering probes, using the average linkage method, and with a subset of 10,017 genes exhibiting the greatest standard deviation. (C) Examination of H3K27me3 binding profile at the CRIP1 locus in Group A, Group B, and H1 human embryonic stem cells. DIABIND detected significant binding regions (p < 0.05) are depicted in the lower panels (yellow-PFAS, blue-PFB) along with the location of Group A hypermethylated region (red) and CpG island location (green). Heatmap (right panel) also demonstrates CpG hypermethylation (CpG sites on horizontal axis) occurring preferentially in Group A (CIMP+) PF epidernomas (Tumour samples on vertical axis). (D-E) Heat map of differentially expressed H3K27me3 target genes between Group A and B identified using SAM analysis FDR < 0.05. The top 10 differentially expressed EZH2 target genes in Group A and B are also shown. (F) Overlap of Group A (CIMP+) specific DNA hypermethylated genes and Group A specific EZH2 bound plus H3K27me3 marked genes. (G) Overlap of Group B (CIMP-) specific DNA hypermethylated genes and Group B specific EZH2 bound plus H3K27me3 marked genes.
A islands. Statistical significance was assessed using a Mann-Whitney transcription by a purple arrow. Direct comparison of DMRs between Group A (CIMP+) and Group B (CIMP-) at repetitive regions (LINES, SINES, and TRs). Statistical significance was assessed using a binomial distribution test of proportions. Figure S4.13

(A) From outer to inner track, chromosomal location, then circos plots depicting whole genome DNA methylation profiles in 3 Group A (CIMP+), 3 Group B (CIMP−), 3 fetal normal brains, and 3 adult normal brains. DNA methylation was averaged in 10kb windows and shown as a heatmap with hypomethylation (pale white) and hypermethylation (dark blue). (B) Example of Group A (CIMP+) specific DNA methylation accumulating at the promoter of ESRRG compared to Group B (CIMP−). Fetal normal brain, and adult normal brain. Location of CpG island is shown by a green bar, and direction of transcription by a purple arrow. (C) Comparing the DNA methylation levels detected in DMRs between Group A (CIMP+) and Group B (CIMP−) at CpG islands. Statistical significance was assessed using a Mann-Whitney Test. (D) Comparing the DNA methylation levels detected in DMRs between Group A (CIMP+) and Group B (CIMP−) at repetitive regions (LINES, SINES, and LTRs). Statistical significance was assessed using a Mann-Whitney Test. (E) Direct comparison of DMRs between Group A (CIMP+) and Group B (CIMP−) at various genetic loci, generated by WGBS, validating the CIMP identified in Figure 2. Statistical significance was determined using the binomial distribution test. (F) Comparison of the proportion of differentially methylated versus hypomethylated CpG island differentially methylated regions (DMRs) in Group A versus Fetal Normal Brain, Group B versus Adult Normal Brain, and Group A versus Group B. Statistical significance was determined using the binomial distribution test. (G) Comparison of the proportion of hypermethylated versus hypomethylated repeat region DMRs in Group A versus Fetal Normal Brain, Group B versus Adult Normal Brain, and Group A versus Group B. Statistical significance was assessed using a binomial distribution test of proportions.
Figure S4.14

(A) Example at chromosome 18 demonstrating partially methylated domains identified specifically in Group A, which are associated with areas of closed chromatin identified by FAIRE-seq in PF ependymomas. (B-C) Variance of gene expression for genes identified in Group A (CIMP+) and Group B (CIMP-) partially methylated domains. Statistical significance of variation difference was assessed using a Mann-Whitney test. (D-E) Ranked list of most genes found within partially methylated domains with the most variable gene expression. Statistical significance of variant gene expression was detected using an F-test with FDR correction using the BH method.
Figure S4.15

(A) Venn diagrams comparing the overlap of genes up-regulated in Group A (CIMP+) cultures (E517, E520) compared to Group A specific DNA hypermethylated genes or Group A specific H3K27me3 marked genes. (B) Comparison of genes up-regulated in E517 or E520 by GSK343 or 5-aza-2'-deoxycytidine with Group A specific H3K27me3 marked genes. (C) Limiting dilution colony forming assay for ependymoma cells isolated from E520-PF1 flank tumours, which were treated in vivo with DZNep (3mg/kg/d) or vehicle. A statistical comparison was made between DMSO (black), and DZNep (orange) using a chi-square test. (D) Western blot analysis of H3K27me3 protein levels following treatment of E520-PF1 cells with GSK343 (500nm) over 9 days. Alpha-Tubulin protein expression was used to assess equal protein loading. (E) Confidence intervals for stem cell frequencies associated with LDAs shown in Figure 2.5i and 2.5j. (F) Goodness of fit test for LDAs shown in Fig 2.4h and 2.6h.
Figure S4.16

Combinations of epigenetic modifiers and chemotherapeutic compounds, used to treat a panel of ependymoma primary cultures over a 7 day course. Experiments were performed in 96-well plates and cell survival was assessed using an MTS assay.
4.7 Chapter 4: Methods

4.7.1 Patients and tumour samples

Tumour samples, clinical information, and animal studies were processed in approval with local ethics board from both the institutions as described previously\textsuperscript{12}. From all patients informed consent was obtained, as described previously\textsuperscript{12}. No patient underwent chemotherapy or radiotherapy prior to the surgical removal of the primary tumor. This study included only primary samples for analysis, and further excluded WHO grade I histological variants of ependymoma. Detailed clinical description of patient characteristics is shown for the sequencing cohort in Publication Supplemental Table S2\textsuperscript{97} and for the methylation cohort in Supplementary Table S8. Tumour subgrouping was based on gene expression profiling or immunohistochemical analysis as described previously\textsuperscript{12}. At least 80\% of tumor cell content was estimated in all tumour samples of the sequencing cohort by staining cryosections (~5um thick) of each sample with hematoxylin and eosin. Diagnoses were confirmed by histopathologic assessment by at least two neuropathologists, including a central pathology review that utilized the 2007 WHO classification for CNS tumors\textsuperscript{2}.

4.7.2 DNA library preparation and Illumina sequencing

Tumor and control samples were individually processed, in every case thorough histological examination proved that each tumor consisted of >80\% tumor cells, in most cases it was >95\%. DNA from tumor and control samples (blood) were prepared and sequenced individually. The Agilent SureSelect Human All Exon 50 Mb target enrichment kit (v3 initially switched to v4
subsequently) was used to capture all human exons for deep sequencing, using the vendor’s protocol v2.0.1. The SureSelect Human All Exon Kit targets regions of 50 Mb in total size, which is approximately 1.7% of the human genome. Briefly, 3µg of genomic DNA were sheared with a Covaris S2 to a mean size of 150bp. 500ng of library were hybridized for 24h at 65°C with the SureSelect baits. The captured fragments of the tumor samples and controls were sequenced in 105bp single end mode on an Illumina HiSeq2000 deep sequencing instrument (based on Illumina, Inc., v3 sequencing chemistry). Median coverage of whole-exome sequenced tumor samples was 157-fold (range 43-469×) and for control samples (blood DNA) 146-fold (range 80-222×). In addition, whole genome libraries (before the exome hybridization step) were sequenced three lanes each in paired end 105bp mode on the HiSeq2000, as described earlier by Jones et al\textsuperscript{15}.

To increase the coverage of the samples for whole exome-sequencing we used the following strategy: Exome capture was initially carried out with Agilent SureSelect (Human All Exon 50 Mb) in-solution reagents using the default Illumina adapters (without barcode). To introduce Illumina Multiplex barcodes into the existing libraries at a later stage, 15ng final exome-enriched library (without barcode) were used as a template in a 50ul PCR reaction. The Herculase II Fusion enzyme (Agilent) was used together with the NEBNext Universal PCR primer for Illumina and NEBNext Index primer (NEB #E7335S) with the following conditions: The initial denaturation step for 2min at 98°C was followed by 4 cycles of 30sec 98°C, 30sec 57°C, 1min 72°C, and a final 10min at 72°C step. 6-7 barcoded samples were then sequenced on the Hiseq2000 in 2x100bp paired end mode.
### 4.7.3 DNA sequence data processing

For each sequencing lane, read pairs were mapped to the human reference genome (hg19, NCBI build 37.1, downloaded from the UCSC genome browser at http://genome.ucsc.edu/) using BWA\(^{43}\) version 0.5.9-r16 with default parameters and maximum insert size set to 1 kb. We used SAMtools\(^{44}\) to generate a chromosomal coordinate-sorted BAM file. Post-processing of the aligned reads included merging of lane-level data and removal of duplicate read pairs per sequencing library using Picard tools (version picard-1.48, http://picard.sourceforge.net). Lane, library and sample information was captured in the read group tag in the header of the merged final BAM file. Only uniquely aligned reads (minimum mapping quality of 1) were considered for downstream mutation analysis. Coverage calculations following duplicate removal considered all informative bases of the reference genome (excluding Ns). A mean Phred-scaled base quality of at least 25 across the length of the read was required. For target capture sequencing, only bases of reads overlapping the targets +/-100 bases were considered for coverage calculations. Sequencing statistics are given in Publication Supplemental Table 1 and 2\(^97\).

### 4.7.4 Single nucleotide variant (SNV) detection

Our analysis pipeline for single nucleotide variant (SNV) detection integrates publicly available tools with custom in-house software and applies several filtering and annotation steps. SNV calling is based on SAMtools mpileup\(^{44}\) and bcftools\(^{44}\) (version 0.1.17), using parameter adjustments to allow calling of somatic variants. Default settings of bcftools are designed for diploid samples, but due to tumour heterogeneity, polyploidy and normal cell contamination,
tumour genomes often have a significantly lower mutant allele frequency than that seen in normal diploid genomes. Therefore, somatic SNVs are often not called by standard tools designed for detection of single nucleotide polymorphisms, e.g. in population studies such as the 1000 Genomes Project (http://www.1000genomes.org/). Initial SNV candidates were identified by using SAMtools mpileup for each tumour sample, considering only reads with a minimum mapping quality of 30 and bases with a minimum base quality of 13, after application of the extended base alignment quality (BAQ) model. BAQ is the Phred-scaled probability of a read base being misaligned⁴⁵, and it is designed to reduce false SNV calls caused by misalignments. After the pile up of high quality bases at each position of the input BAM file, bcftools applies the prior and performs the actual SNV calling. We changed the default probability of calling a variant if P(ref|D) < 0.5 to 1.0, which results in all positions containing at least one high quality non-reference base to be reported as a variant. Therefore, this initial set of SNV candidates contains a high fraction of false positive calls, but ensures that true somatic mutations with low allele frequency (well below the expected 50% allele frequency) are reported. This initial SNV call set is then subjected to various filters. SNVs covered by fewer than three reads in the tumour and control sample, with somatic allele frequency <10%, or with only one read supporting the variant were excluded. Additionally, a minimum of 10 high quality reads available at the corresponding position in the control sample were required, in order to be able to distinguish somatic from germline variants. Local sequence context can lead to incorrect base calls, but typically involves reads sequenced from a single strand only. Thus, if the variant call was supported by reads from only one strand, the +/-10 bases around the SNV were automatically screened for Illumina specific error profiles⁴⁶ and excluded if a profile was matched. For all tumour SNV calls the pipeline generates a pileup of the bases in the normal sample considering only uniquely mapping reads. SNV calls were categorized as germline or somatic according to whether there was evidence for the same event at the same locus in the
BAM file of the tumour-matched control sample. Filtered calls were annotated with RefSeq gene annotations, dbSNP build 135 and variants from the 1000 Genomes project. Calls matching the position of known dbSNP (up to version 131) or known 1000 Genome variants were excluded from the high-confidence somatic call set (calls matching the position of dbSNP version >131 but not the position of 1000 Genome variants were retained because cancer relevant somatic mutations, such as several TP53 mutations, have been included in more recent dbSNP versions). Additionally, we filtered out SNVs that were found in at least 1% of the control samples or at least 1% of a set of 162 unrelated controls from other studies, because they constitute likely unannotated, naturally occurring SNPs and/or false positives, e.g. Artifacts related to sequencing and mapping. The pipeline integrates Annovar (http://www.openbioinformatics.org/annovar/) to determine whether the observed amino acid change has synonymous, non-synonymous, nonsense, or splice-site changing properties on the encoded protein. Variants were further annotated with genes listed in the Cancer Gene Census (http://www.sanger.ac.uk/genetics/CGP/Census/) and entries from the Catalogue of Somatic Mutations version 57 (COSMIC, http://www.sanger.ac.uk/genetics/CGP/cosmic/), in addition to the full RefSeq gene summary, full gene name, and genomic size. A subset of sequence variants and Indels were validated by capillary sequencing by Sanger using purified PCR products. Primer sequences are available upon request.

4.7.5 Small insertion and deletion (InDel) detection

Small insertions and deletions were identified with SAMtools and bcftools. The InDel discovery pipeline is similar to the SNV pipeline (as described above), but using default bcftools parameters, to reduce the known high false positive rate associated with current InDel detection.
methods for deep sequencing data. To call an indel a germline event, we only required one InDel supporting read in the matching normal sample, again to reduce the high fraction of false positive somatic InDel calls. Calls overlapping simple repeat or microsatellite regions were excluded as such regions are commonly observed to yield false positive calls. Annotation of InDels was identical to SNV annotation.

4.7.6 Computation of recurrently mutated genes

To search for genes mutated at significant frequency, we applied the MutSig algorithm, a method that corrects for background mutation rate and gene length. Details can be found at Broad CGA tools website (http://www.broadinstitute.org/cancer/cga/mutsig) including previously published studies\(^ {48,49} \).

4.7.7 Identification of rearrangements and generating of Circos Plots

Structural rearrangements, namely deletions, tandem duplications, inversions, and translocations, were detected using DELLY\(^ {50} \), which is based on paired-end mapping\(^ {51} \). The structural rearrangement calls were filtered using the corresponding ependymoma germline samples, germline data of additional medulloblastoma samples\(^ {15} \), and phase I 1000 Genomes Project (http://1000genomes.org) genome data to exclude germline structural variants as well as rearrangement calls caused by mapping artefacts. We only considered those rearrangements for further analysis, which were present in at most 0.5% of the 1000 Genomes Project samples assessed and not in the additional germline samples. Two rearrangement calls were considered to be the identical, hence constituting a likely germline variant if they displayed an overlap in
terms of genomic coordinates with their end coordinates differing by less than 5 kb. Furthermore, rearrangement calls with less than 10 supporting pairs as well as supporting pairs with average mapping quality less than 20 were excluded for further analysis. The circular whole-genome plots were generated using Circos.

4.7.8 Identification of pathways affected by SNVs

A Pathway Association Test was used identify groups of functionally related genes that contained a greater than expected number of SNVs in one or the other ependymoma subtype. For each gene set, SNVs were stratified by subtype (PFA vs. PFB). The number of SNVs in all genes and the number of SNVs observed in a given gene set were totaled for each subtype. A Fisher's exact test was performed with the null hypothesis that the frequency of SNVs in a given gene set was equal in the two subtypes. To correct for multiple testing, gene names were randomly shuffled and the analysis repeated to obtain a null p-value for a given gene set. Randomization was done 10,000 times and, for each gene set, the percentage of null p-values that were the same or lower than that obtained from the actual data was used as an estimate of the false discovery rate (FDR).

4.7.9 Generation of copy number profiles from Illumina 450K methylation data

Low-resolution (450K probes) copy number variations were detected from the 450k Infinium methylation array in a custom approach using the sum of both methylated and unmethylated signals. Probes found to be highly variant in the six normal cerebellum samples were excluded from the analysis according to the following criteria: Removal of probes not within the 0.05 and
0.85 Quantile of median summed values or over the 0.8 Quantile of the median absolute deviation. Log-ratios of samples to the median value of control samples were calculated, and sample noisiness was determined as the median absolute deviation of adjacent probes. Probes were then combined by joining 20 adjacent probes, and resulting genomic windows less than 100kb in size were iteratively merged with adjacent windows of smaller size. Windows of more than 5Mb were excluded from analysis, resulting in a total of 8,654 windows throughout the genome. For each window, the median probe value was calculated and shifted to minimize the median absolute deviation from all windows to zero for every sample. Segmentation was performed by applying the circular binary algorithm\textsuperscript{53}, as published previously\textsuperscript{21}.

4.7.10 Methyl-binding domain 2 assisted recovery and sample preparation

Genomic DNA was isolated according to previous methods\textsuperscript{12}. DNA (6µg) was immunoprecipitated using the methyl-binding domain 2 protein and quantified using the Qubit fluorometer (Invitrogen). Enrichment was assessed by quantitative PCR for positive controls: \textit{RASSF1A}, \textit{DLK1}, \textit{H19}, and negative controls: \textit{ACTB} and \textit{GAPDH}. Bound and Unbound fractions from the MBD2 pull-down were whole genome amplified (SIGMA-WGA2) in triplicate, pooled, quantified using Qubit, and subjected to another round of quantitative PCR for the above control targets. DNA was sent to Nimblegen to be hybridized to Nimblegen 385K CpG island promoter plus arrays, in which ‘IP’ fraction was labeled with CY5 and ‘unbound’ fraction with CY3.
4.7.11 Methylation analysis of MBD2-chip data

Microarray data was Quantile normalized using the LIMMA Bioconductor package. Log2 ratios were then imported into Agilent Genomics Workbench (Agilent Technologies), following which the BATMAN algorithm was used to infer the methylation statuses associated with each probe. Mean methylation states were calculated for probes within a 1000 base pair window and termed a region of interest (ROI)\(^{42}\). ROIs were then filtered to those with greater than 4 probes and mapped to autosomal chromosomes. ROIs exhibiting a standard deviation greater than 0.65 were used for subgroup assignment as described below. Comparisons between subgroups were performed using a Wilcoxon-rank sum test, and p-values were corrected for multiple testing using the Benjamini-Hochberg method. For comparisons of DNA methylation and other factors in this manuscript a Wilcoxon test was used and corrected for multiple testing, such that no assumptions were made regarding the normality of the data distributions.

4.7.12 Illumina Infinium 450K methylation sample preparation and data analysis

Genomic DNA was isolated according to previous methods\(^{12}\). DNA (1µg) was used for bisulfite treatment (Qiagen – EpiTect plus) with the use of DNA protect buffer, particularly in the case of DNA from formalin fixed paraffin embedded tissue. Bisulfite treated DNA was then quantified using spectrophotometry (Nanodrop). >500 ng was sent to The Centre for Applied Genomics (TCAG – Toronto) for hybridization to Illumina 450K Methylation Arrays. Array preprocessing was performed using GenomeStudio (Illumina) with background subtraction adjustment applied. Arrays were also normalized using the BMIQ method, which produced the same finding of a Group A specific CIMP. Methylation values were then exported as Beta-
values (estimate of actual CpG methylation levels). Probes that overlapped with known single nucleotide polymorphisms, which mapped to Chromosome X and Y, and were Illumina control probes, were removed from the analysis. Methylation probes were then filtered to CpG sites, which mapped to promoters containing CpG islands. A Wilcoxon-Rank Sum Test (Mann-Whitney) was used to identify the differentially methylated CpG sites between Group A (CIMP+) and Group B (CIMP-). P-values for differentially methylated CpG sites identified were then corrected for multiple testing using the Benjamini-Hochberg method. Significant differences between numbers of CpG sites, genes, or methylated and silenced genes, was calculated using a Binominal Distribution Test. Methylated and Silenced Genes were identified in two ways 1) By identifying genes which were methylated and down-regulated following comparison between Group A and B using a Wilcoxon-Rank Sum test or 2) By Performing a Pearson correlation between the methylation status of a CpG site with the corresponding down-stream gene. Methylated and Silenced genes (within the same tumor) were identified by genes demonstrating significant and preferential methylation in a particular subgroup and evidence of down-regulation as compared to a collection of normal brain samples. Gene expression data for these samples can be found in our previous publication\textsuperscript{12}.

4.7.13 Subgroup analysis of gene expression and methylation data

To detect robust sample clusters from the gene expression data (Affymetrix Exon 1.0ST) we performed hierarchical clustering using the top 1000 varying probes as described previously in Witt and Mack., \textit{et al}\textsuperscript{12}. For clustering of MBD2-chip data we performed consensus hierarchical clustering with agglomerative average linkage as our method for consensus clustering. (R package: ConsensusClusterPlus)\textsuperscript{54}. The change in area under the cumulative distribution
function curve was used to identify the principal number of subgroups for a given clustering method. Silhouette analysis was performed to evaluate cluster representation of samples (R package: cluster)\(^{55}\). To evaluate the concordance between gene expression and DNA methylation subgroup stratification we calculated the Rand Index, with the significance assessed by permutation of sample labels and computing the Rand index over 10000 iterations in order to generate a null distribution. Illumina 450K methylation data was clustered using the probes exhibiting a standard deviation of >0.2 as described previously\(^ {56} \). A variety of consensus clustering methods was performed including K-means, non-negative matrix factorization, hierarchical clustering, and self-organizing maps were used\(^ {57} \). The distance metric used in the case of K-means was Euclidean, whereas a Pearson correlation was used for all other methods. Principal component analysis was performed within Partek Genomics Suite (Partek Inc.) to compare Group A (CIMP+) and Group B (CIMP-) posterior fossa subtypes with the same genes or CpG sites used for consensus HCL and consensus K-means respectively.

### 4.7.14 H3K27me3 and EZH2 ChIP-Seq profiling and analysis in PF ependymoma

10-20mg of fresh frozen primary tumour samples were homogenized in 1% Formaldehyde and allowed to incubate at room temperature for 6 – 10 min. Cross-linking was stopped with the addition of 125mM of Glycine, and samples were washed twice with ice-cold PBS containing (1% BSA and 10% FBS). Samples were then sonicated to ~200 bp fragments using a Biorupter (Diagenode). The chromatin immunoprecipitation was then performed using 5 mg of EZH2 antibody (#39875-Active Motif) or H3K27me3 antibody (C15410069-Diagenode) overnight at 4C as described previously\(^ {45} \). DNA was quantified using PicoGreen (Invitrogen) and libraries were prepared using NEBNext ChIP-seq Illumina Sequencing library preparation kit (NEB).
Samples were barcoded (NEB Next Barcodes) and pooled in equimolar amounts such that up to 6 samples could be sequenced by paired-end Illumina HiSeq 2000 sequencing (Illumina).

ChIP-seq reads were aligned using the BWA algorithm with removal of redundant reads (Picard Algorithm) likely to represent ChIP-seq PCR library artifacts, yielding uniquely mapped ChIP-seq reads. Peaks were identified using MACS (Version 2) with a p-value cutoff of 0.01. Differential peaks were identified using the R: Bioconductor DiffBind package (p < 0.05) and annotated to the nearest gene +/- 5kb using Cistrome (http://cistrome.org/ap/root). Overlap analysis between H3K27me3 genes or EZH2 target genes was assessed statistically using a binomial distribution test. Unsupervised consensus clustering of H3K27me3 predicted target genes was performed using the top 1000 genes exhibiting the greatest standard deviation. Supervised analysis of predicted H3K27me3 target genes was also performed using Significant Analysis of Microarrays with FDR cutoff of 0.01.

4.7.15 Sequenom analysis of ependymoma samples

Validation of gene methylation was performed using Sequenom Mass Spectrometry. Primers were designed using Sequenom: EpiDesigner and tested on bisulfite-treated universally methylated DNA (Invitrogen) by standard PCR (Qiagen) followed by Sanger Sequencing. For Bisulfite treated tumour samples, following PCR amplification, amplicons were sent to Genome Quebec for quantification using Sequenom Mass Spectrometry.

4.7.16 Subgroup stratification of ependymoma samples in a validation cohort
Sequenom primers were designed to three highly methylated genes in Group A (CIMP), *PKP1*, *CRIP1*, *CYP26C1*, as selected by CpG coverage and PCR efficiency. PCR amplification was performed in a training data set consisting of the samples, which were analyzed by Illumina 450K methylation arrays. These three methylated genes were used to train a classification model using the Prediction Analysis for Microarrays algorithm. Class prediction was performed on a non-overlapping cohort of 82 samples collected from The Hospital for Sick Children, Children’s Hospital Boston, University of Michigan, and the MD Anderson Cancer Centre. Posterior probabilities corresponding to Group A (CIMP+) or Group B (CIMP-) were calculated for each sample, and an odds ratio > 2-fold (Probability Group A / Probability Group B) for either subgroup was used to classify tumours. Survival was graphed throughout the manuscript using Kaplan-Meier curves and assessed statistically using a Log-Rank test.

### 4.7.17 Pathway analysis of DNA methylation data

A Pathway Association Test was used to identify groups of functionally related genes that contained a greater than expected number of methylation events in one or the other ependymoma subtype. A gene was considered to be methylated if all profiled sites within 1000bp upstream of the TSS showed a variance of no more than 0.1 and a mean score of >0.5. For each gene set, methylation events were classified by subtype. The number of methylation events in all genes and the number of methylation events observed in a given gene set were totaled for each subtype. A Fisher’s exact test was performed with the null hypothesis that the frequency of methylation events in a given gene set was equal in the two subtypes. To correct for multiple testing, gene names were randomly shuffled and the analysis repeated to obtain a null *p*-value for a given gene set. Randomization was done 10,000 times and, for each
gene set, the percentage of null $p$-values that were the same or lower than that obtained from the actual data was used as an estimate of the false discovery rate (FDR).

### 4.7.18 Whole-genome bisulfite sequencing, DNA preparation, and differentially methylated region (DMR) analysis

Strand-specific MethylC-seq libraries were prepared according to Lister et al. 2011 with minor modifications. Adaptor-ligated DNA fragments with insert lengths of 200–250 bp were converted using the EZ DNA Methylation kit (Zymo Research). After PCR amplification in six parallel reactions using the FastStart High Fidelity PCR kit (Roche), library aliquots were pooled per sample and sequenced using the Illumina HiSeq 2000 platform. This yielded an average of 513 million (±102 million (s.d.)) 101 bp paired-end reads per sample.

For analysis of DMR enrichment in specific genomic sites, we first extracted genomic features from UCSC genome browser. Then the percentage of total genomic CpGs for each genomic feature was calculated as a background value. Thereafter, the percentage of total hyper/hypomethylated CpGs in each genomic feature was calculated based on the DMR list. The enrichment fold change was then set as the ratio between the two percentages above. In order to test the significance of the enrichment /depletion, we randomly permuted the CpGs from all DMRs in the whole genome for 10,000 times and used Fisher's exact test to determine the significance of the difference between the observed and simulated results.
4.7.19 Whole-genome bisulfite sequencing DMR and PMD calling

WGBS data was mapped to hg19 using BSMAP (version 2.74)\textsuperscript{46}. The potential duplications were removed afterwards using Picard tools\textsuperscript{104}. BisSNP (Version 0.82.2) was then used to detect and remove SNPs and CpGs with potential technical biases before DMR calling. BSmooth was used to smooth bisulphite sequencing data and call candidate DMRs as described previously\textsuperscript{47}. PMDs were detected using MethylSeekR\textsuperscript{48}.

4.7.20 Ependymoma short-term primary cell culture and in vitro drug treatment

Primary ependymoma cells were isolated from patients and cultured on Laminin (Sigma) coated plates in Neurobasal media (Invitrogen) consisting of N2 (Invitrogen), B27 (Invitrogen), Glutamine (Invitrogen), BSA (Sigma), heparin (Sigma), human EGF (Invitrogen), and human basic FGF (Invitrogen). Media was replenished every other day while leaving ~50% conditioned media to encourage continued cell proliferation. Cell viability assays were performed in 96 wells using an Alamar Blue stain (Invitrogen) or MTS Aqueous One (Promega) according to manufacturer’s instructions. 5-aza-2’-deoxycytidine (decitabine-Sigma) was dissolved to a stock concentration of 2mM in PBS and stored in aliquots at -20 C. DAC was prepared fresh and added to treatment media on a daily basis at the appropriate final concentration, for a total of 7 days. 3-dezaneplanocin A (DZNep) (Cayman Chemical) was dissolved to a stock concentration of 25 mM in DMSO and stored in aliquots at -20 C. DZNep treatments were performed every other day along with replenishment of cell culture medium for a total of 7 days. GSK343 (active compound) and GSK669 (inactive compound) were dissolved in DMSO and used to treat cells at varying concentrations with media replenishment every other day for a total of 11 days.
4.7.21 *Gene expression profiling of DAC and GSK343 treated cultures*

Primary cell cultures were treated for 5 days in Decitabine (500nM) or GSK343 (500nM), following which RNA was isolated using the Trizol (Invitrogen) method. RNA libraries were prepared according to manufacturers recommendations and hybridized to Affymetrix Gene 1.0ST arrays. The RMA method with Quantile normalization was used for gene expression array normalization. Differentially expressed genes were detected using Significance of Microarray Analysis (FDR < 0.01).

4.7.22 *Western blot analysis*

Ependymoma cell cultures were lysed in PLC lysis buffer containing deoxycholate, with sonication to facilitate the release of nuclear histones. SDS-PAGE analysis was performed in a 12% gel, loading 20 ug of protein, as quantified by BCA (Pierce). Membranes were blocked with 5% bovine serum albumin (Roche) diluted in TBST. Western blot antibodies were used at the following concentrations in overnight incubations (2% BSA): EZH2 (Abcam: ab110646, 1:5000), H3K27me3 (Cell Signaling: #9733, 1:5000), H3K4me3 (Cell Signaling: #9751, 1:5000), cleaved PARP (Cell Signaling: #5625, 1:1000) and Alpha Tubulin (Cell Signaling: #2148, 1:20000). Secondary antibodies were used at a concentration of 1:5000 for all primary antibodies, and 1:20000 for alpha-tubulin.
4.7.23 Flank injections and in vivo treatments of immunodeficient mice

For all animal studies, following engraftment of tumor cells, mice were then randomly assigned a treatment of vehicle versus treatment to control for assignment biases and other confounding factors. 50,000 E520 – PF1 ependymoma cells were injected subcutaneously into flanks of immunodeficient NOD-scid gamma mice. Tumours were allowed to develop for 7 days until either visible or palpable. DZNep or vehicle (Sigma: Cremaphor) was administered 3 consecutive days a week at a dosage of 3 mg/kg/day via intraperitoneal injections. Tumours were monitored and measured continuously using a caliper. Experimental endpoint was determined when tumours reached 15 mm in size. Final tumour volumes were determined using caliper measurements. Investigators were blinded during measurement of tumour volumes. A comparison between tumour volumes of DZNep versus vehicle treated mice was calculated using a Wilcoxon-rank sum test. For all animal studies, adequate sample sizes were chosen such that any result could be appropriately evaluated statistically using a two-sided non-parametric test. For flank xenograft experiments this entailed a Wilcoxon-Rank sum test, and for intracranial experiments this involved a Log-Rank test.

4.7.24 Cerebellar xenografts and in vivo treatments of immunodeficient mice

10,000 cells were xenographed by stereotactic injection into posterior fossas of immunodeficient NODscid gamma mice of 5-8 weeks old. Tumours were allowed to develop for 7 days, following which DZNep (3mg/kg/d) or vehicle (Sigma: cremaphor) administered by intra-peritoneal injection. Mice were treated according to the same protocol for flank tumor bearing
mice (above). Survival of mice was visualized using a Kaplan-Meier curve and quantified using a log rank test.

4.7.25 Limiting dilution assays (LDAs) of primary ependymoma patient samples or ependymoma xenografts

Cells from a lung metastasis resection or tumour xenograft were dissociated according to previously published protocols\textsuperscript{49}. LDAs were performed in a 96-well plate format. LDAs from xenografts were not treated with inhibitors but monitored for neurosphere colony formation. For primary patient samples serial dilutions of cells were performed to reach cell doses of 2000 cells per well at the highest dose and 4 cells per well at the lowest dose. A total of 10 cell doses were tested with 6 technical replicates per dose. Cells were treated with selected compounds about 5 h post-surgery. GSK343 was used at a concentration of 3 uM, and Decitabine was used at 0.5 uM. Fresh media and drugs were added to the cells after 7 days. Wells were scored for sphere formation on day 14. Statistical analysis was performed with the Extreme Limiting Dilution Analysis web-based software\textsuperscript{50}.
5 Chapter 5: Future Directions

5.1 Further examination of tumour heterogeneity in ependymoma

We and other groups have demonstrated that supratentorial, posterior fossa, and spinal ependymomas are distinct molecular and clinical entities, which can be further subdivided into ST-RELA–FUSION \(^{+/-}\), and PF-CIMP \(^{+/-}\). One possible argument for the lack of detectable genetic alterations in genomically stable PF-CIMP \(^{+}\) ependymoma could be a significant degree of inter-tumoral heterogeneity within this subgroup. This was perhaps unrecognized in the past because tumours were grouped with histologically similar ST and SP ependymomas. One might argue that genomic characterization of a larger cohort of PF-CIMP \(^{+}\) ependymomas could reveal novel and recurrent somatic mutations, however these events are predicted to be exceedingly rare. As a simple thought process, combination of PF ependymomas from both Mack et al., 2014 and Parker et al., 2014 (Combined cohort n=73) suggests that the frequency of a recurrent somatic mutation, if identified, would be lower than 1.3%. In other words, 98.7% of PF ependymomas would most likely harbour no recurrent mutations in coding space.

Spatial intra-tumoural heterogeneity is another issue that will need to be addressed in future studies, particularly for design of clinical trials. As shown in other solid cancers such as renal cell carcinoma or glioblastoma multiforme mutational and/or subgroup diversity can be seen in distinct regions of the same tumour sample\(^{105,106}\). This information will be important for accurate subgroup classification of ependymoma samples in future clinical trials, based upon a single, or if needed, multiple biopsies from different locations of the same tumour.
5.2 Further investigation of genomically balanced posterior fossa ependymoma

Despite a balanced genome observed in PF ependymoma, it is possible that gains or losses of entire genomes, or specific chromosomes, might have occurred and were undetectable by copy number analysis using array CGH as presented in Chapter 2. The answer to this hypothesis could be addressed with current published data, by extracting allele frequency calls in 500K SNP array data by Johnson et al., 2010, and mutant allele frequency calls in WGS data by Parker et al., 2014. Furthermore, the analysis presented for PF ependymoma in this thesis has focused primarily on coding or exonic regions. New evidence in solid cancers such as melanoma and medulloblastoma has demonstrated that inter-genic mutations can have a direct influence upon gene expression through mutation of transcriptional regulatory elements\(^99,100\). This has been demonstrated by TERT promoter somatic mutations, which are associated with increased TERT expression and therefore avoidance of replication induced senescence. Although our study, in addition to Parker et al., 2014, present whole-genome sequencing from 37 PF ependymomas, the sequencing coverage (~30X) was significantly less than achieved through exome sequencing (~150X). Future sequencing coverage of PF ependymoma genomes will need to be increased in order to potentially identify non-coding somatic SNVs, or the absence of these events across a broad tumour cohort. If non-coding somatic SNVs are identified understanding the functional relevance of these events will require additional layers of genomic evidence. As one possible approach, WGS data could be overlaid with techniques such as FAIRE-seq (Formaldehyde assisted isolation of regulatory elements), MNase-seq
(Micrococcal nuclease), or DNase1-seq, which chart genomic regions of open versus closed chromatin, to determine if sequence alterations affect the higher-order structure of chromatin. Based upon these findings, additional layers of histone modifications could be potentially examined such as enhancer marks: H3K27ac, H3K4me1, or repressive marks: H3K27me3, H3K9me2/3 among many others. In addition, this also raises a question as to whether non-coding mutations need to be recurrent within the same genomic region, or could non-coding mutations be situated at distinct locations but regulate the same gene or sets of genes through DNA looping. Delineating these scenarios could be potentially addressed with complementary techniques such as Hi-C, which reveal DNA-DNA interactions across the entire genome.

Another hypothesis is that the analysis presented in this study, and several other cancer sequencing studies, have been focused upon non-synonymous (non-silent) somatic mutations, which cause alterations in the amino acid sequence\(^{40-45}\). Analysis presented by Supek et al., 2014 demonstrate that synonymous mutations (silent) in genes can create alternative splice variants, which may be bonafide drivers of tumourigenesis\(^{107}\). Potential events such as these could be mined in public data sets by leveraging the WGS sequencing data from PF ependymoma and combining this with RNA-seq data from the same samples.

5.3 Foundations for the PFA ependymoma CIMP

In Chapter 2 of this thesis, I presented evidence suggesting that Group A PF ependymoma may be characterized by a CpG island methylator phenotype (CIMP). In contrast to other cancers,
PFA ependymomas lack a clear genetic basis for the CIMP, with absence of IDH1 mutations (GBM-CIMP), DNMT3A/TET1/2 mutations (AML-CIMP), or SDH mutations (Paraganglioma-CIMP)\textsuperscript{40,108,109}. Furthermore, PFA ependymomas do not significantly over-express DNA methyltransferases as observed in other solid tumours\textsuperscript{110}. One possible explanation for a PFA CIMP could be developmental, owing to a lack of necessary chemical and/or electrical stimuli in proliferating cells of origin of ependymoma that could be required for differentiation and cell cycle exit. Properly addressing this hypothesis would require a better understanding of the hierarchy of neuroepithelial cells giving rise to radial glial cells (RGCs), the progeny of RGCs, the transcriptional networks that govern RGC differentiation programs, and the chemical/electrical stimuli needed for induction and regulation. These programs could then be exploited in developing RGC populations by transgenic mouse models and/or chemical modulators that target key regulators of RGC differentiation. As one hypothesis, in a manner similar to GBM-CIMP and AML-CIMP, metabolite imbalance related to aberrant differentiation signals could potentially lead to a CIMP phenotype through pathogenic accumulation of 2-hydroxyglutarate, an inhibitor of DNA de-methylation. An alternate mechanism for establishing the PFA ependymoma CIMP might involve re-distribution of the DNA methyltransferases DNMT1/3A/3B during tumour initiation and/or progression. Despite no significant changes in the gene expression levels of these DNMT enzymes in PFA ependymoma, the locations of the DNMTs has yet to be investigated, and could be readily assessed using ChIP-seq. Furthermore, because the DNMT proteins dimerize, investigating the coordinated locations of these enzymes could reveal possible mechanisms leading to aberrant DNA methylation.
5.4 Further elucidating the epigenomic landscape of posterior fossa ependymoma

Our study has demonstrated that DNA methylation patterns in PFA ependymoma are highly connected with higher-order chromatin structure, specifically the repressive mark H3K27me3. Furthermore, the genes DNA hypermethylated in PFA ependymoma, and occupied by H3K27me3, overlap significantly with genes known to be marked by the polycomb repressive complex 2 (PRC2) in embryonic stem cells (ESC). One important aspect that could be addressed in the future is examining whether the H3K27me3 mark is established first, followed by long-term silencing mediated by DNA hypermethylation. Alternatively, could DNA hypermethylation occur first followed by histone-mediated gene silencing, as directed by EZH2's physical association with DNMT enzymes? Another important and related aspect is that the PRC2 marked ESC genes are bi-valent, and harbor the ‘active’ H3K4me3 mark. During fertilization CpG islands are kept un-methylated by recognition of H3K4 by an enzymatically inactive DNA methyltransferase, DNMT3L. A possible explanation for DNA methylation spreading into CpG islands could be because of inhibition of DNMT3L binding through H3K4 tri-methylation, thus leading to a possible CIMP phenotype.

Although this thesis focused primarily upon the histone modification H3K27me3, global examination of other marks such as H3K4me1, H3K27ac, H3K36me3, H3K9me2/3, among many others remains to be explored. Understanding the functional consequences of these marks may reveal mechanisms of disease pathogenesis and targets for therapeutic intervention, by leveraging the recent development of numerous small-molecule inhibitors of epigenetic
regulators. The importance of elucidating epigenetic landscapes and the functional plasticity of tumour cells in response to treatment is highlighted in a publication by Knoechel et al., 2014. Here they demonstrate that tumour resistance to targeted therapies, specifically gamma-secretase inhibitors, can be mediated through epigenetic adaptation, specifically modulation of binding of the histone acetylation ‘reader’ BRD4 protein. This resistance mechanism could be circumvented with up-front combinatory treatment of gamma-secretase inhibitors with the BRD4 inhibitor, JQ1. Along the same lines, one hypothesis for the chemo-resistant nature of PF ependymoma could be cellular adaptation mediated by epigenetic modifications. In a similar approach, ependymoma primary cultures could be exposed to a panel of chemotherapeutic agents and synthetic lethal epigenetic targets could be identified using chemical and/or genetic library screens.

5.5 Mouse models and cellular origins of posterior fossa ependymoma

In Chapter 3, I emphasized that PFA ependymoma are characterized by distinct DNA and histone methylation patterns, which may be involved in tumour development. Potential animal modeling of these genome-wide epigenetic alterations could be achieved by perturbation of DNA and histone methyltransferases in murine RGCs. As one possibility, the EZH2-Y641 mutant, which promotes the conversion of histone H3 lysine 27 di-methylation to a trimethylated state, could be over-expressed in RGCs through transgenic models or ex-vivo approaches\textsuperscript{111}. To potentially model the CIMP phenotype a hyperactive DNMT1 isoform could be expressed in the RGCs, or genetic ablation of DNA de-methylation machinery such as TET proteins or AID/APOBEC proteins. These approaches however rely on the concept that PF ependymomas arise from the RGC compartment, which encompass a heterogeneous group of
neural progenitor cells both anatomically and temporally. Accurate modeling of PF ependymoma might first require elucidation of specific RGCs that give rise to PF ependymoma and the markers needed for cellular isolation. One strategy for pinpointing the exact cell of origin of posterior fossa ependymoma, might involve generating whole-genome DNA methylation profiles in mice of early neuroepithelial cells, radial glial cells, differentiated neurons, astrocytes, ependymal cells, and oligodendrocytes in different CNS regions of the mouse developing embryo. Since DNA methylation profiles are highly stable, it is possible that genomic regions of methylation in PF ependymoma tumours may be capturing signatures of putative cells of origin of ependymoma. These signatures could then be compared against mouse methylation profiles by mapping conserved CpG sites and predicting neural compartments most similar to PF ependymomas.

5.6 Clinical implications for treatment of posterior fossa ependymoma

In this thesis, I have demonstrated that Group A and B are molecularly distinct entities and can be robustly stratified by routine immunohistochemistry (IHC), gene expression profiling, and DNA methylation profiling. Given the clinical disparity between these subgroups, it will be important for future ependymoma clinical trials to stratify PF ependymomas based on subgroup, or at least collect tissue so that outcomes can then be analyzed in the setting of molecular subgroups. This is particularly important in the case of ependymoma for which histopathological grading has limited utility, and molecular markers are unavailable. There are several tools for which rapid and robust PFA versus PFB stratification could be achieved such
as IHC, gene expression arrays, DNA methylation arrays, or examination of specific DNA methylation markers using Sequenom technology. The advantage of molecular approaches that measure DNA methylation is that it is a stable mark, DNA from FFPE tissue can be used, and very small quantities of DNA are required. Indeed, the 3-marker PFA-CIMP+ assay described in Chapter 3 of this manuscript requires less than 3ng of DNA from fresh frozen tissue for accurate sample classification.

Given the distinct clinical outcomes and behavior of PFA versus PFB ependymoma, numerous prospective clinical trials could be designed to improve PF ependymoma patient outcomes. In the case of PFB, which have a very good prognosis, a clinical trial could be designed to address whether de-escalation of radiotherapy could be achieved in order to reduce treatment-associated morbidity. In the case of PFA, which have a very poor prognosis, a clinical trial could be designed to address whether more aggressive surgery and/or radiotherapy might be beneficial, or whether novel biological agents such as Decitabine and/or Vorinostat should be investigated. Given the fact that ependymomas are highly chemo-resistant, an existing FDA approved compound such as Decitabine or Vorinostat could be rapidly translated in a Phase I/II clinical trial. These epigenetic agents could also be tested in combination with chemotherapeutic agents, at least in pre-clinical models, to identify synergistic combinations that may be effective in PFA ependymoma treatment.
The research described in the majority of this thesis stems from two publications for which I was co-first author, and played a central role in the conception, analysis, and interpretation of both studies under the direct supervision of Dr. Michael D. Taylor. Specifically, I organized sample collection and processing, performed gene expression analysis, copy number analysis, DNA methylation analysis, and ChIP-seq analysis. I also performed and/or managed in vitro and in vivo drug treatments, ChIP experiments, and immunoblot experiments. These functional studies were performed in collaboration with Dr. Peter B. Dirks and Dr. Marco Gallo. Our partners in both studies, Dr. Hendrik Witt, Dr. Stefan M. Pfister, and Dr. Andrey Korshunov, were vital for generating and analyzing whole-genome, exome, and whole-genome bisulfite sequencing, in addition to TMA staining, and additional array processing. Because ependymoma is such a rare tumour, this study would not have been possible without many collaborators across the world whom provided both tissue samples, study design advice, and manuscript feedback.
7 Copyright Permissions

Copyright release for the following articles was obtained using RightsLink:

1) Licensed content title: *The genetic and epigenetic basis of ependymoma*
   License number: 3322680594487
   License date: Feb 05, 2014
   Licensed content publisher: Springer
   Licensed content publication: Child's Nervous System
   Licensed content author: Stephen C. Mack
   Licensed content date: Jan 1, 2009

2) Licensed content title: *Emerging Insights into the Ependymoma Epigenome*
   License number: 3322680193331
   License date: Feb 05, 2014
   Licensed content publisher: John Wiley and Sons
   Licensed content publication: Brain Pathology
   Licensed content author: Stephen C. Mack
   Licensed content date: Feb 25, 2013

3) Licensed content title: *Delineation of Two Clinically and Molecularly Distinct Subgroups of Posterior Fossa Ependymoma*
   License number: 3322670685501
   License date: Feb 05, 2014
   Licensed content publisher: Elsevier
   Licensed content publication: Cancer Cell
   Licensed content author: Stephen C. Mack
   Licensed content date: 16 August 2011
8 Reference List


