Molecular Classification and Clinical Genomics of Medulloblastoma

by

J. H. David Shih

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto

Abstract

Molecular Classification and Clinical Genomics of Medulloblastoma

J. H. David Shih
Doctor of Philosophy
Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto
2015

Background: Children with medulloblastoma have a 72% chance of surviving more than 5 years under current treatment, and those who survive suffer long-term developmental and neurocognitive deficits due to treatment-associated toxicity. We believe that refining the classification of medulloblastoma will facilitate the optimization of treatment intensity and the discovery of novel therapeutics. Recent studies have identified four molecular subgroups of medulloblastoma with distinct expression patterns: WNT, SHH, Group3, and Group4. These subgroups represent different molecular entities that arise through and rely on different oncogenic processes. Accordingly, we aim to improve prediction of patient survival by identifying prognostic markers for each subgroup, and we hope to abrogate non-specific cytotoxic treatments by discovering candidates for targeted intervention against each subgroup.

Methods: We proposed and validated a new classification method for medulloblastoma based on molecular patterns. Using this method, tumours were classified into molecular subgroups. Their DNA copy-number profiles ($n = 1087$) were analyzed to identify somatic copy-number aberrations (SCNAs) and recurrently disrupted genes and pathways. Prognostic SCNAs were identified by Kaplan-Meier survival analyses on a discovery set ($n = 673$), and the candidates were validated by FISH on a tissue microarray of validation samples ($n = 453$).

Results: Tumours of each subgroup harbour recurrent SCNAs disrupting different path-

ways. WNT medulloblastoma is characterized by *CTNNB1* mutation, SHH medulloblastoma by activated Gli signaling, Group3 medulloblastoma by *MYC* activation, and Group4 medulloblastoma by *SNCAIP* duplication. Further, patients of different subgroups exhibit differential response to standard treatments. Incorporating subgroup data into survival models significantly improved predictive performance. Using six FISH biomarkers on FFPE tissues, we reproducibly stratified patients into risk groups with distinct survivorships.

Conclusion: The stark differences in genetic alterations among molecular subgroups of medulloblastoma suggest that each subgroup arises through different biological mechanisms. The molecular classification of medulloblastoma not only improved survival prediction but also revealed pathways for therapeutic intervention. We have identified a panel of prognostic markers that can be used to select patients for therapy de-escalation in future trials, and we have also discovered candidates for targeted therapy.

Dedication

To Professor Hon Kwan.

Acknowledgements

I would like to thank my advisors, Michael and Gary, for all their support and guidance throughout my studies. I would like to acknowledge the contributions of my thesis committee – Dr. Meredith Irwin, Dr. Quaid Morris – for their insightful inputs. I would also like to acknowledge all my labmates for their help and feedback. Most of all, I wish to thank Paul Northcott for his contribution to our collaborative pursuits.

I must thank all members of the Medulloblastoma Advanced Genomics Consortium for their contributions, without which the studies described herein would not be possible. I am especially thankful for work done by Dr. Andrey Korshunov, Dr. Stefan Pfister, and their teams at the German Cancer Research Center.

I would also like to thank all my teachers who imparted to me invaluable knowledge and expertise. I especially thank Dr. Derek Corneil, Dr. Geoffrey Hinton, Dr. Boris Steipe, and Dr. Reinhold Vieth at the University of Toronto. I would also like to highlight Massive Open Online Courses offered by Stanford University, Coursera, and edX, as well as course notes graciously made openly available by instructors at other institutions. I owe a debt of gratitude to inspirational teachers such as Dr. Andrew Ng (Stanford University), Dr. Daphne Koller (Stanford University), Dr. Robert Piché (Tampere University of Technology), Dr. Brain Caffo (Johns Hopkins Bloomberg School of Public Health), Dr. Rafael Irizarry (Harvard School of Public Health), and Dr. Tom Mitchell (Carnegie Mellon University).

I am grateful to open source software communities for openly sharing their work, including authors, contributors, and maintainers on CRAN, Bioconductor, PyPI, GitHub, and numerous other repositories. Their spirit of openness is infectious.

This work is supported by the Frederick Banting and Charles Best Canada Graduate Scholarship, the Michael Smith Foreign Study Supplement, the Ontario Graduate Scholarship, and the University of Toronto Fellowship.

Contents

List of Tables

List of Figures

Abbreviations

aCGH array comparative genomic hybridization.

AIC Akaike information criterion.

Akt Protein kinase B.

ATRT atypical teratoid/rhabdoid tumour.

AUC area under the curve.

CBS circular binary segmentation.

CLIA Clinical Laboratory Improvement Amendments.

CNA copy-number aberration.

CNS central nervous system.

CNS-PNET central nervous system primitive neuroectodermal tumour.

DNA dexoyribonucleic acid.

Egfr Epidermal growth factor receptor.

EMD empirical mode decomposition.

ETMR embryonal tumours with multilayered rosettes.

FFPE formalin-fixed, paraffin-embedded.

FISH fluorescence *in situ* hybridization.

GISTIC Genomic Identification of Significant Targets in Cancer.

GLAD Gain and Loss Analysis of DNA.

HR hazard ratio.

IQ intelligence quotient.

KNN k-nearest neighbour.

LCA large cell/anaplastic medullobalstoma.

LDA linear discriminant analysis.

LOH loss of heterozygosity.

MAGIC Medulloblastoma Advanced Genomics International Consortium.

MB medulloblastoma.

MBEN medulloblastoma with extensive nodularity.

MRI magnetic resonance imaging.

NICD *NOTCH1* intracellular domain.

NMF nonnegative matrix factorization.

NOS not otherwise specified.

PAM prediction analysis of microarrays.

PCA principal component analysis.

PCR polymerase chain reaction.

PI3K phosphoinositide 3-kinase.

PNET primitive neuroectodermal tumour.

RNA ribonucleic acid.

ROC receiver-operating characteristics.

SCNA somatic copy-number aberration.

SCNAs somatic copy-number aberrations.

Shh Sonic hedgehog.

SNP single nucleotide polymorphism.

SVM support-vector machine.

TCAG The Centre for Applied Genomics.

Tgf-β Transforming growth factor β.

WHO World Health Organization.

Wnt Wingless.

Chapter 1

Introduction

One in 285 children are diagnosed with cancer before the age of 20, and cancer is the second leading cause of death in children[25;26]. Owing to advances in treatment, the 5-year survival of children with cancer has steadily increased from 63% in 1975 to 85% in 2006[27]. Improvements in survival, however, vary considerably across cancer types. In 1975, the 5-year survival rates for childhood leukemia and central nervous system (CNS) tumour were 50% and 57%, respectively. In 2006, the survival rate for the former reached 87% while the latter lagged behind at 74%[27]. Indeed, leukemia and CNS tumour are biologically very different types of cancer: they arise from different cells within different organs, hijack different cellular signaling programs to effect uncontrolled proliferation, and reside in different locations with different accessibilities to treatment. It should come as no surprise then that leukemia and CNS tumour respond differently to similar modern anti-cancer treatments (consisting of chemotherapy and radiotherapy). Furthermore, leukemia and CNS tumour can each be classified into additional cancer subtypes, which also respond varyingly to treatment. Within leukemias, the 5-year patient survival of acute lymphocytic leukemia is 92% and that of acute myelogenous leukemia is 66%[27]. Within CNS tumours, the 5-year patient survival of pilocytic astrocytoma is 94% and that of medulloblastoma is 72%[28]. In general, the responses of cancers to therapy depend on not only the affected tissue (blood vs. CNS) but also the cellular origin (lymphoid vs. myeloid and astrocytic vs. embryonal). Using anatomical locations and cellular appearance, clinicians classify cancer into different types in order to predict the responses to treatment. Current post-surgical treatment modalities eradicate cancer largely by one predominant mechanism (inhibiting cellular division and inducing apoptosis), but emerging anti-cancer therapies are increasing in specificity against aberrant cells and diverging in mechanisms of action. Accordingly, the classification of cancer with progressively finer

granularity will become critical for selecting the right treatment for each patient.

While the current classification of cancer based on organ system and cell morphology has been useful for predicting patient response, the advent of molecular profiling and sequencing technologies can refine this classification further and facilitate the development of therapies targeted against specific aberrations within cancer. The World Health Organization (WHO), in its current classification of CNS tumours published in 2007, classifies CNS tumours into 86 distinct entities primarily using histological appearance[29]. In comparison, WHO recognized *only* 36 subtypes of CNS tumours in 1993[30]. Despite a more detailed classification system, the 5-year survival of children with CNS tumours has remained stagnant at about 75% since 1996[27]. Although histological features can help identify aggressive cancers, the discrepancy between the discovery of new CNS tumour histotypes and the lack of therapeutic improvement over the past decade suggests that the histological classification of CNS tumours may be insufficient to identify biologically similar tumours and to facilitate the development of effective novel therapy. In particular, it was noted as early as 1971 that medulloblastoma with desmoplastic histology exhibits longer patient survival in response to radiotherapy[31]; however, it remains unclear why the desmoplastic phenotype is associated with longer survival or which novel therapeutic agent may be effective against this histological variant. In comparison, a medulloblastoma tumour with a loss-of-function mutation in *PTCH1* (endogenous suppressor of Shh signaling) would depend on hyperactive Shh signaling for growth; therefore, the patient would be expected to — and indeed does — respond to inhibition of Shh signaling[32]. Characterizing the genetic mutations and understanding the biology of a tumour can therefore help guide the discovery of novel therapies and the selection of suitable treatment modality or intensity.

In order to identify cancer mutations against which therapeutic intervention may be beneficial, we would need to distinguish between mutations that contribute to tumourigenesis (known as **driver** mutations) from those that do not (known as **passenger** mutations). Cancer cells accumulate somatic mutations by disrupting DNA damage response or DNA repair pathways. Moreover, somatic mutations, in general, occur stochastically (with some exceptions including antibody diversification by activation-induced cytidine deaminase). Thus, most mutations in a cancer cell would be deleterious or neutral to the cell, and mutations that confer selective advantage would increase in **cellular frequency** (proportion of cells that harbour the mutation) within a tumour. There are two caveats, however. First, not all mutations observed at high cellular frequency contribute

to tumourigenesis. Since multiple mutations may occur each time a cancer cell replicates its DNA and multiple cell divisions may occur before a new driver mutation arises, a population of cancer cells arising from the same parental cell will share one driver mutation and many additional mutations that do not contribute to tumourigenesis. Second, not all driver mutations become clonal (i.e. reach 100% cellular frequency) due to complex cell-cell interactions. In glioblastoma multiforme, for example, cancer cells expressing mutant EGFR secrete cytokines to promote the growth of cells expressing wild-type EGFR[33]; hence, tumour heterogeneity in glioblastoma can be sustained by this paracrine mechanism, and *EGFR* mutation does not become clonal. Taking these two scenarios into consideration, high cellular frequency is neither a necessary nor a sufficient condition for a mutation to be a driver. Instead, an alternative, more robust method for distinguish driver from passenger mutations would be to assess its **recurrence frequency**: the frequency that a mutation is observed across samples. If a mutation drives tumour formation, we would expect to find it across tumours from different patients, provided that the collection of tumours is biologically similar and arises through a common molecular mechanism. Above all, in order to use recurrence frequency as a criteria for identifying driver mutations and recognizing genomic patterns, we would need to first classify the tumours into molecularly homogeneous groups.

This thesis will focus on refining the classification of **medulloblastoma**, a malignant brain tumour occurring in the cerebellum and the posterior fossa. WHO classifies medulloblastoma as a grade IV (highly aggressive) embryonal tumour[29]. Its diagnosis is made by the anatomical location of the tumour and the histological morphology of the cells. Medulloblastoma was once a universally fatal disease; today, 58% of patients are expected to live longer than 15 years[25]. To predict whether a patient will respond favourably to treatment, clinicians currently categorize medulloblastoma by such features as metastatic presentation and histological variants. Medulloblastoma may be divided into classic, desmoplastic/nodular, and large cell/anaplastic histotypes. While this histological classification of medulloblastoma can help predict patient outcome, it provides scant insight into potential biological mechanisms. Therefore, several studies generated RNA expression profiles of primary medulloblastoma tumours using microarrays and sought to characterize medulloblastoma by patterns of RNA expression. Since each expression profile captures a snapshot of the overall molecular state of a tumour, biologically similar tumours would exhibit similar expression profiles. Therefore, tumours with similar expression profiles can be **clustered** (grouped) together to discover biologically homogeneous molecular classes. By these clustering analyses, four molecular classes (henceforth

known as **subgroups**) of medulloblastoma were discovered: **WNT**, **SHH**, **Group3**, and **Group4** medulloblastoma[34–38]. As the names suggest, WNT medulloblastoma have hyperactive Wnt signaling and SHH medullobalstoma have increased Shh signaling compared to the other subgroups, while Group3 and Group4 are less well defined[35;36;38]. **We hypothesize that classifying medulloblastoma into these molecular subgroups will shed light on its cancer biology; consequently, this classification will improve prediction of treatment response and point to novel therapeutic targets.**

The remainder of this chapter will discuss genetic and non-genetic factors in medulloblastoma and their implications for the classification and treatment of medullobastoma. At the end of this chapter, the main research objectives of this thesis are outlined. In **Chapter 2**, we will propose and validate a method whereby medulloblastoma tumours may be molecularly classified in the clinical setting. The immediate utility of this molecular classification system for patient risk stratification will be shown in **Chapter 3**, and the potential implication of molecular classification for therapeutic discovery will be presented in **Chapter 4**. Finally, conclusions and future directions arising from this thesis are outlined in **Chapter 5**. To make this thesis accessible to a broader audience, the **Appendix** presents clarifications for topics including gene nomenclature, mouse model notation, signaling pathways, biomarker discovery, and computational algorithms.

I Epidemiology and genetic predisposition

Medulloblastoma occurs at an annual incidence of 4.1 per million children under 20, ranking as the most common type of malignant brain tumour in childhood[28]. The majority of patients with medulloblastoma are diagnosed before the age of 20, and the median age at presentation is 8 years[1]. As medulloblastoma is almost 10 times more likely to afflict children than adults, medulloblastoma is a disease of childhood[39], suggesting that perhaps genetics may play a role in this disease. Consistent with this proposition, medulloblastoma can occur simultaneously in monozygotic twins[40], and having an affected sibling increases a child's risk of developing medulloblastoma by 4 fold[41]. While most medulloblastoma cases are sporadic with unknown genetic contribution, several genetic disorders can predispose children to developing medulloblastoma, as well as many other malignancies (**Table 1.1**). For instance, mutations in genes of DNA damage response and repair pathways, including *TP53*, *ATM*, *BRCA2*, *NBN*, predispose a child to

develop a spectrum of tumours, including medulloblastoma, at a young age[42–55].

Table 1.1: Genetic disorders predisposing to medulloblastoma

Genetic disorder	Mutated genes	Reference
Li-Fraumeni syndrome	*TP53*	42–47
Ataxia telangiectasia (Louis-Bar syndrome)	*ATM*	48; 49
Fanconi anemia	*BRCA2* (*FANCD1*)	50–52
Nijmegen breakage syndrome	*NBN* (*NBS1*)	54; 56; 57
Fragile X syndrome	*FMR1*	46; 55
Neurofibromatosis type 1 (Von Recklinghausen's disease)	*NF1*	46; 58
DICER1 syndrome	*DICER1*	59
Rubinstein-Taybi syndrome	*CREBBP*	60
Basal cell nevus syndrome (Gorlin syndrome)	*PTCH1*, *PTCH2*, *SUFU*	46; 61–66
Turcot syndrome	*APC*	67

Medulloblastoma can also arise as a consequence of germline mutations in developmental signaling pathways. Germline loss-of-function mutations in negative regulators of Shh signaling, including *PTCH1*, *PTCH2*, and *SUFU*, causes basal cell nevus syndrome (Gorlin syndrome) and predisposes patients to medulloblastoma, basal cell carcinoma (most common type of skin cancer), and other cancers. In the context of these mutations, hyperactive Shh signaling during neural development presumably would cause patients to develop SHH medulloblastoma. This notion is supported by mouse models with heterozygous *Ptch1* mutation[68–72] and the $Sufu^{+/-}$ $Trp53^{-/-}$ mouse model[73], which both develop medulloblastoma with active Shh signaling. Further, Gorlin syndrome is highly prevalent in human patients with SHH medulloblastoma[9], though it is yet unclear whether patients with Gorlin syndrome exclusively develop SHH medullobalstoma. Additionally, germline mutations in *APC* activates Wnt signaling, manifests as Turcot syndrome, and predisposes patients to medulloblastoma. The tumours arising in these patients would be expected to be of the WNT subgroup; however, the crosstalk between the Shh and the Wnt signaling pathways in the context of neural development and different genetic backgrounds could modulate the phenotypic outcome of *APC* mutation.

Germline mutations in *SMARCB1* (*SNF5/INI*) had been associated with medulloblastoma[74;75], but brain tumours in these patients are now classified as atypical teratoid/rhabdoid tumour (ATRT), a tumour type that WHO first recognized in its 1993 classification[30]. ATRT and medulloblastoma are similar by histology[76] and by magnetic resonance imaging (MRI)[77]. Currently, *SMARCB1* mutation is widely accepted as a di-

agnostic indicator of ATRT and distinguishes ATRT from medulloblastoma[78;79]. Indeed, *SMARCB1* is an example in which a genetic mutation superseded histological diagnosis and redefined cancer classification.

While mutations are pivotal factors in the formation of medulloblastoma, numerous other factors shape the context under which the mutations exert their effects. The genetic mutations listed in **Table 1.1** all have incomplete penetrance, highlighting the contribution of genetic background (as well as environmental factors). Furthermore, disparate incidence patterns of mouse tumours arising from engineered mutations in negative regulators of the Shh signaling pathway (*SUFU* and *PTCH1*) point to several possible contributing factors, including the limitation of inbred mouse models, involvement of genes in multiple biological pathways, and potential modifying genetic polymorphisms or mutations. On the C57BL/6 genetic background, $SUFU^{+/-}$ mice do not develop medulloblastoma within the same time frame as $PTCH1^{+/-}$ mice, even though both *SUFU* and *PTCH1* negatively regulate Shh signaling[80]. Conversely, in a Manchester cohort of 171 Gorlin syndrome patients, germline *SUFU* mutation had about 20 times higher risk than germline *PTCH1* mutation for developing medulloblastoma[81]. The observation that only 2 (1.7%; $n = 115$) Manchester patients with *PTCH1* mutation developed medulloblastoma[81] starkly contrasts the findings in inbred mouse strains (C57BL/6 and 129X1/Sv) with *PTCH1* mutation, which spontaneously develop medulloblastoma at a frequency of about 18% ($n = 507$; **Table 1.3**)[68;69;80;82]. The difference in the penetrances of mutations in *SUFU* vs. *PTCH1* suggests that these genes may have other functions beyond the Shh pathway or they regulate the pathway in dissimilar manners. However, the inversion in the relative penetrances of *SUFU* and *PTCH1* mutations in a human population compared to mouse models suggest that polymorphisms or mutations in possible modifier genes may influence the manifestation of mutant *SUFU* and *PTCH1*. Alternatively, this observation may indicate that the Shh signaling pathway regulate neural development in critically different ways in mouse and human. Accordingly, these possibilities prompt the need for further studies in additional human populations and mouse strains in order to tease out the intricate interplay between mutant genes and genetic background.

The phenotypic manifestation of genetic mutations is also influenced by the state of the mutation-harbouring cells. *PTEN* (endogenous inhibitor of PI3K signaling) is frequently homozygously deleted in medulloblastoma[83], suggesting that *PTEN* may be a tumour supressor. Heterozygous germline *PTEN* mutation, however, leads to adult tumours (PTEN hamartoma tumour syndrome) but not medulloblastoma[84]. Instead, one

mutant copy of *PTEN* leads to Lhermitte-Duclos disease, which presents as a benign tumour of the cerebellum[85]. The cells in the tumour have complete loss of wildtype *PTEN* expression[85]. Curiously, the granule neurons exhibit dysplasia (disorganization of tissue structure), increased size, aberrant migration, but not abnormal cellular proliferation[86], in contrast to neoplastic neurons in medulloblastoma. This example illustrates that the cell state during nervous system development influences whether the cells progresses to neoplasia (cancer). In the case of *PTEN*, other mutations would need to co-occur with *PTEN* loss-of-function in order for patients to develop medulloblastoma. Indeed, patients with germline *PTEN* and $Pten^{+/-}$ mice do not develop malignant brain tumours, despite predisposition to various other tumours[86]. In contrast to *Pten* loss alone, concomitant overexpression of SHH ligand (human protein) causes medulloblastoma in mouse following irradiation[87]. More generally, the effect of a mutant gene is modulated by the state of the cell harbouring the mutation, and this state can be shaped by genetic background, cooperating mutations, and developmental signaling. As a cell matures and differentiates down various lineages during development, it may reach a state that permits a specific mutation (germline or somatic) to transform it into cancer with supporting signals from the surrounding microenvironment.

II Molecular biology

CNS development involves the coordination of innumerable signaling pathways and the highly controlled proliferation of cells along various lineages. Tumours arise when unfortunate accidents (e.g. germline mutations, genetic background, and DNA damage) coincide with conducive conditions (e.g. cell state and extracellular signals). Simply put, medulloblastoma occurs when normal neural development goes awry. Several developmental signaling pathways are important in medulloblastoma formation, including Wnt, Shh, Notch, and PI3K. (See page 98 in the **Appendix** for general descriptions of these pathways.)

Wnt signaling

Patients with Turcot syndrome (mutation in *APC*) who developed medulloblastoma provided one of the first clues that medulloblastoma involves activation of Wnt signaling[67]. Additionally, CTNNB1 nuclear localization and *CTNNB1* mutation were frequently ob-

served in medulloblastoma[88;89]. CTNNB1 is the transcription factor that serves as the downstream effector of Wnt signaling. As part of the CTNNB1 destruction complex containing AXIN and APC, GSK and CSNK1A1 normally phosphorylates CTNNB1 at multiple serine/threonine residues (S45, T41 S37, and S33) within a region encoded by exon 3, flagging CTNNB1 for ubiquitination and subsequent proteosomal degradation[90]. Mutation in APC therefore aborgates the interaction between the destruction complex and CTNNB1, preventing the degradation of CTNNB1. Similarly, CTNNB1 mutations at or near the phosphorylation sites also stabilizes CTNNB1, permitting its signaling[2;90;91]. Curiously, a mouse model expressing stabilized human (S37F mutant) CTNNB1 under the mouse promoter of *Prnp* did not develop medulloblastoma on a *Trp53*$^{-/-}$ background[92]. Conversely, Gibson *et al.* generated medullobalstoma using a different mouse model expressing stabilized CTNNB1[72]. In this Cre-recombination model, mouse *Ctnnb1* is converted to a stabilized form by Cre-mediated somatic deletion of exon 3 (which was flanked by loxP sequences)[93]. *Ctnnb1* is under endogenous promoter control, but deletion of exon 3 is controlled by expressing Cre under a selected promoter; therefore, expression of stabilized Ctnnb1 can be restricted to cell types that activate the selected promoter. *Ctnnb1*$^{+/\text{lox(ex3)}}$ mice crossed to *Atoh1-Cre* mice did not produce progeny (*Atoh1-Cre*;*Ctnnb1*$^{+/\text{lox(ex3)}}$) that develop hyperplasia or medulloblastoma, suggesting that Atoh1-expressing cells are not the developmental origin of Wnt-activated medulloblastoma[72]. In contrast, *Blbp-Cre*;*Ctnnb1*$^{+/\text{lox(ex3)}}$ mice showed hyperplasia in the dorsal brainstem and eventually developed medulloblastoma in the context of *Trp53* deficiency, which presumably caused additional tumourigenic mutations or inhibited apoptosis[72]. In sum, only *Blbp-Cre*;*Ctnnb1*$^{+/\text{lox(ex3)}}$;*Trp53*$^{\text{flx/flx}}$ mouse developed Wnt-activated medulloblastoma, and these tumours arise from the dorsal brainstem and not the cerebellum (in contrast to the Shh-activated medulloblastoma to be described later). Tumour development not only requires expression of the oncogene (stablized *Ctnnb1*) in a suitable cell type (Blbp-expressing) but may also need additional mutations (secondary to *Trp53* deletion).

Shh signaling

Medulloblastoma can also develop due to activated Shh signaling. Patients with Gorlin syndrome provided the first clue that Shh signaling is activated in a subset of medulloblastoma. Patients with germline loss-of-function mutations in negative regulators of Shh signaling — *PTCH1*, *PTCH2*, or *SUFU* — can develop medullobastoma during

childhood, and basal cell carcinoma and other tumours later in life due to hyperactive Shh signaling[46;61–66]. Moreover, somatic *PTCH1* mutation, activating *SMO* mutation, *PTCH1* homozygous deletions, *GLI2* amplifications are recurrently observed in medulloblastoma tumours[2;83;94;95], which provide further evidence that Shh signaling is activated in some medulloblastoma cases. In mouse, homozygous *Ptch1* loss is embryonic lethal, but heterozygous *Ptch1* mutant mice exhibit hyperplasia in the external granule layer of the cerebellum, and a subset of mutant mice develop cerebellar medulloblastoma[68;69]. Frequently, the external granule layer of the developing cerebellum exhibits hyperplasia (cell mass with increased proliferation) before medulloblastoma[96;97]. The incidence medulloblastoma in *Ptch1*$^{+/-}$ mice ranges from 8% to 45% and is influenced by the genetic background (**Table 1.2**). (Additionally, the incidence of medulloblastoma in *Ptch1*$^{+/-}$ models is affected by the occurrence of rhabdomyosarcoma and other tumours whose early appearance may preclude the development of medulloblastoma[68;98].) This incomplete penetrance suggests that additional events beyond single-copy *Ptch1* loss may be necessary for tumour progression. Indeed, *Ptch1*$^{+/-}$ mice developed tumours with much higher frequency in the context of *Trp53* deficiency (**Table 1.2**) and when subjected to neonatal irradiation, depending on the genetic background (**Table 1.3**). Presumably, *Trp53* deficiency and irradiation generate additional mutations that cooperate with *Ptch1* haploinsuffiency to promote progression of hyperplasia to medulloblastoma. Since *APC* mutation did not enhance tumour incidence in *Ptch1*$^{+/-}$ mouse, the Wnt signalling pathway may not play a major role in the development of Shh-activated medulloblastoma[70]. Similarly, homozygous loss of mouse p19$^{\mathrm{ARF}}$ did not enhance medulloblastoma formation in *Ptch1*$^{+/-}$ mice[70], suggesting that the hyperproliferation of external granule neuron precursors caused by activated Shh signaling does not trigger p19$^{\mathrm{ARF}}$-dependent cell cycle arrest. (Mouse p19$^{\mathrm{ARF}}$ is encoded by *Ckdn2a* and mediates p53-dependent cell cycle arrest, similar to the human homolog p14$^{\mathrm{ARF}}$.) In contrast, concurrent haploinsufficiency of either of two other cell cycle inhibitor, *Cdkn2c* (encodes p18$^{\mathrm{Ink4c}}$) or *Cdkn1b* (encodes p27$^{\mathrm{Kip1}}$), in *Ptch1*+/- mice further increased cell proliferation and enhanced medulloblastoma incidence[99;100]. The tumours in these mice invariably lost the wild-type copy of *Ptch1* [99;100]. Indeed, complete loss of *Ptch1* is highly frequently observed in medulloblastoma tumours and may be a key event for medulloblastoma progression[96;99–104]. Activated Shh signaling in *Ptch1*$^{-/+}$ granule neuron precursors induces cellular proliferation and increases accumulation of DNA breaks, which can lead to the loss of the wildtype *Ptch1* allele, further induction of Shh/Gli signaling independent of Shh ligand or Shh coreceptor Boc, and progression of medulloblastoma[103]. Although homozygous loss of *Ptch1*

is not absolutely required for progression of hyperplasia to medulloblastoma[69;103;105], conditional homozygous loss of *Ptch1* in Atoh1-expressing cells (e.g. external granule neuron precursors) or GFAP-expressing cells (e.g. neural stem cells) is sufficient for developing medulloblastoma from the external granule layer with 100% penetrance[97]. Similarly, activated Smo (inhibited target of Ptch1 and positive regulator of Shh signaling) induced tumour formation in a dose-dependent manner[106]. When one copy of the activated mouse Smo allele habouring the W539L mutation (also known as SmoA1 and corresponds to the human SmoM2 allele) was expressed under the promoter of *Neurod2* (which is expressed in cerebellar granule neuron precursors), the mice developed medulloblastoma with 48% penetrance, and mice carrying two copies of the activated Smo allele (*Neurod2-Smo*$^{W539L/W539L}$) develop medulloblastoma with near complete penetrance[106;107]. Morever, GLI (downstream effector of Shh signaling) promotes cell survival by directly upregulating the prosurvival factor BCL2[108;109] and disabling DNA damage checkpoint[110]. These results support the notion that activated Shh signaling is suffient for medulloblastoma formation and may be necessary for tumour maintainence, highlighting this pathway as a candidate for therapuetic intervention.

Notch signaling

Notch signaling plays important roles in embryonic development; in the central nervous system, Notch signaling prevents precocious neuronal differentiation and maintains the pluripotency and self-renewal of neural stem and progenitor cells[113;114]. While *Notch1*$^{+/-}$ mice are viable, homozygous *Notch1* mutants die before E11.5 and possibly partly due to the depletion of neural stem cell pool[113–115]. The embryonic lethality of *Notch1*$^{-/-}$ also suggests that *Notch1* is not completely functionally redundant with other mammalian Notch homologs (*Notch2*, *Notch3*, and *Notch4*). Conversely, conditional activation of Notch signaling by transgenic expression of a constitutively active Notch, *NOTCH1* intracellular domain (NICD), in *GFAP-CreER*;*ACTB-N1ICD* blocked cell cycle exit of neural progenitors and inhibited neuronal differentiation[114]. (In this model, *N1ICD* is conditionally expressed by Cre excision of an upstream stop cassette.) Conceivably, hyperactive Notch signaling may promote tumour formation by blocking cell cycle exit and differentiation. Consistent with this proposition, Shh-activated medulloblastoma in *Neurod2-Smo*$^{+/W539L}$ mice show increased expression of Notch pathway targets (*HES1* and *HES5*) as well as *Notch1*, suggesting that the Notch pathway is active in medulloblastoma induced by Shh signaling[107]. When *Neurod2-Smo*$^{+/W539L}$ mice were

Table 1.2: Incidence of medulloblastoma in Shh-activated mouse model

Genotype	**Genetic background**	**Incidence** (95% CI)	**Reference**
Ptch1$^{+/-}$	129X1/Sv	**15** (10–22)	68; 69; 96
Ptch1$^{+/-}$	CD-1	**8** (2–19)	82
Ptch1$^{+/-}$	C57BL/6	**45** (36–54)	18; 80; 82; 103
Ptch1$^{+/-}$;*Sufu*$^{+/-}$	C57BL/6	**48** (35–62)	80
Ptch1$^{+/-}$;*Boc*$^{+/-}$	C57BL/6	**54** (37–71)	103
Ptch1$^{+/-}$;*Boc*$^{-/-}$	C57BL/6	**19** (5–42)	103
Ptch1$^{+/-}$	C57BL/6 × DBA/2 (B6D2F1)	**19** (9–33)	68
Ptch1$^{+/-}$	C57BL/6 × 129X1/Sv	**13** (10–16)	70; 99; 111
Ptch1$^{+/-}$;*Ptch2*$^{+/-}$	C57BL/6 × 129X1/Sv	**15** (9–24)	111
Ptch1$^{+/-}$;*Ptch2*$^{-/-}$	C57BL/6 × 129X1/Sv	**17** (9–29)	111
Ptch1$^{+/-}$;*Trp53*$^{+/-}$	C57BL/6 × 129X1/Sv	**14** (7–25)	70
Ptch1$^{+/-}$;*Trp53*$^{-/-}$	C57BL/6 × 129X1/Sv	**95** (85–99)	70; 112
Ptch1$^{+/-}$;*Cdkn2c*$^{+/-}$	C57BL/6 × 129X1/Sv	**37** (19–58)	99
Ptch1$^{+/-}$;*Cdkn2c*$^{-/-}$	C57BL/6 × 129X1/Sv	**46** (33–58)	99
Atoh1-Cre;*Ptch1*$^{\mathrm{flx/flx}}$	C57BL/6 × FVB/N	**100**	97
GFAP-Cre;*Ptch1*$^{\mathrm{flx/flx}}$	C57BL/6 × FVB/N	**100**	97
Sufu$^{+/-}$	C57BL/6	**0** (0–14)	80
Sufu$^{+/-}$;*Trp53*$^{+/-}$	C57BL/6 × 129P2/OlaHsd	**0** (0–6)	112
Sufu$^{+/-}$;*Trp53*$^{-/-}$	C57BL/6 × 129P2/OlaHsd	**58** (44-71)	112
Neurod2-Smo$^{+/\mathrm{W539L}}$	C57BL/6	**48** (38–58)	107
Neurod2-Smo$^{\mathrm{W539L/W539L}}$	C57BL/6	**94** (85–98)	106

crossed to *Atoh1-Cre*;*Notch1*$^{\mathrm{flx/flx}}$ or *Atoh1-Cre*;*Notch2*$^{\mathrm{flx/flx}}$ mice, the progeny (*Neurod2-Smo*$^{+/\mathrm{W539L}}$;*Atoh1-Cre*;*Notch1*$^{\mathrm{flx/flx}}$ and *Neurod2-Smo*$^{+/\mathrm{W539L}}$;*Atoh1-Cre*;*Notch2*$^{\mathrm{flx/flx}}$) did not develop medulloblastoma at a higher incidence[116]. While the conditional knockout of one Notch homolog could perhaps be rescued by another homolog, treatment with multiple gamma-secretase inhibitors (which prevent Notch cleavage and activation) did not affect tumour incidence, tumour size, apoptosis, or cell proliferation in *Neurod2-Smo*$^{+/\mathrm{W539L}}$ mice, nor did it affect xenograft engraftment in immunocompromised mice[116]. Similarly, constitutive knockout of an important downstream target *Hes5* in *Neurod2-Smo*$^{+/\mathrm{W539L}}$;*Hes5*$^{-/-}$ mutants did not reduce medulloblastoma incidence either[116]. Therefore, Notch signaling is not necessary for formation, progression, or maintenance of Shh-activated medulloblastoma[116;117]. However, expression of NICD in *GFAP-Cre*;*ACTB-N1ICD*;*Trp53*$^{-/-}$ mice resulted in medulloblastoma with 69 (39–91) percent

Table 1.3: Incidence of medulloblastoma in irradiated *Ptch1*$^{+/-}$ mouse model

Genetic background	**Treatment**	**Incidence** (95% CI)	**Reference**
CD-1	non-irradiated	**8** (2–19)	82
CD-1	3 Gy at P1	**81** (58–94)	102
CD-1	3 Gy at P4	**51** (37–65)	71
CD-1	3 Gy at P10	**3** (0–16)	102
C57BL/6	non-irradiated	**45** (36–54)	18; 80; 82; 103
C57BL/6	3 Gy at P1	**53** (27–79)	82

incidence, suggesting that Notch signaling is sufficient to induce medulloblastoma formation with concurrent circumvention of Trp53-dependent apoptosis[4;118]. Further, these Notch-activated mouse medulloblastoma tumours exhibited transcriptional profiles resembling Shh-activated medulloblastoma[4], suggest that these tumours may share activation of similar pathways or arise from common origins.

PI3K signaling

Homozygous deletions of *PTEN* in medulloblastoma first highlighted that PI3K signaling may play a role in medulloblastoma[83;104;119]. The PI3K family phosphorylates the plasma membrane lipid phosphatidylinositol-4,5-bisphosphate [PI(4,5)P_2] to produce phosphatidylinositol-3,4,5-trisphosphate [PI(3,4,5)P_3 or PIP_3][120]. In turn, PIP_3 serves as an intermediate signaling molecule for multiple pathways. AKT1 binds to PIP_3 via its pleckstrin-homology domain and becomes anchored to the plasma membrane by PIP_3; consequently, PIP_3-anchored PDPK1 phosphorylates and activates AKT1, which regulates numerous downstream pathways such as inhibition of apoptosis (via BAX[121] and BAD[122]) and cell cycle arrest (via CDKN1A[123;124] and CDKN1B[125]). (PDKP1 is the official symbol for 3-phosphoinositide dependent protein kinase 1, and it is commonly known as PDK1, which also ambiguously refers to pyruvate dehydrogenase kinase, isozyme 1.) Conversely, PTEN (a phosphatase) inhibits the PI3K pathway by dephosphorylating PIP_3 to PI(4,5)P_2. Therefore, loss of *PTEN* may potentiate PI3K signaling and AKT1 activation, inducing cell survival and proliferation. Curiously, germline PTEN loss-of-function mutations underlie a collection of disorders known as PTEN hamartoma tumour syndromes and increase the risk of a diverse array of tumours that usually occur in adulthood[84]. In fact, patients can develop a benign overgrowth of neurons in the cerebellum but not medulloblastoma[86]. Mouse models of conditional homozygous *PTEN* knock-out

reveal that the neuronal cells exhibit migration defect and increased cell size but not proliferation, suggesting that complete PTEN loss is insufficient for childhood malignancy in the cerebellum[86]. Nonethless, *PTEN* loss in Shh-activated mouse models of medulloblastoma accelerates tumour progression and contributes to resistance against inhibition of Shh signaling[126], as well as resistance against irradiation[87].

Crosstalk of signaling pathways

Although the aformentioned pathways are described as distinct linear series of signaling events, the molecular players are in fact highly interconnected by several shared components (e.g. GSK3 and BTRC). Notably, many of the presented signaling pathways converge on Myc family proteins, and this convergence underscores the central role played by the Myc family in medulloblastoma. Similarly, the upstream regulators of Wnt and Shh signaling, SMO and FZD, are both G-protein coupled receptors. Although their G-protein activities were underappreciated in the past, both receptors can regulate intracellular cAMP signaling, and these pathways may cooperate in promoting tumourigenesis. How various pathways are connected would likely depend on cell type or cell lineage, in addition to dynamic extracellular signals during development; hence, the cellular origins of medulloblastoma may dictate the molecular mechanisms of tumour initiation, progression, and maintenance.

III Histological classification

The WHO classification of CNS tumours standardizes the nomenclature for different cancers arising in the CNS and rests on the premise that each tumour type arises from the transformation of a specific cell type[29;30]. The cellular origin, in turn, may predict prognosis and guide treatment decisions. The classification evolves with new discoveries and emerging insights, though it relies primarily on morphological features and protein marker expressions analyzed by immunohistochemistry.

WHO now recognizes medulloblastoma as a distinct entity[29]. Medulloblastoma and central nervous system primitive neuroectodermal tumour (CNS-PNET) used to be grouped collectively as primitive neuroectodermal tumours (PNET), but WHO now classifies them as separate diseases[29]. The WHO classification currently includes four histological variants of medulloblastoma: desmoplastic/nodular medulloblastoma, medul-

loblastoma with extensive nodularity (MBEN), anaplastic medulloblastoma, and large cell medulloblastoma. The last two variants are often grouped together as large cell/anaplastic medullobalstoma (LCA), but some evidence suggest that they may have different prognoses[127]. Medullomyoblastoma and melanocytic medulloblastoma used to be considered histological variants of medulloblastoma in the 1993 classification[30]; however, in the latest WHO classification (2007), they are no longer considered distinct histotypes of medulloblastoma[29]. About 70% of medulloblastoma are of the classic histological variant[1]. Within this histotype, patient response to treatment varies greatly, indicating that the classic histotype encompasses a biologically heterogenous group of tumours[1]. The prognostic significance of the histotypes are discussed further in **Chapter 3**.

IV Molecular classification

More recently, molecular subgroups of medulloblastoma have been identified based on unsupervised clustering of RNA expression profiles of tumours[34]. These classes consist of **WNT**, **SHH**, **Group3**, and **Group4** medulloblastoma. The subgroups have been independently identified in multiple studies[35;37;38;88;128], and this classification system can be applied objectively using a computer algorithm (see **Chapter 2**) so that the process is streamlined and the results are reproducible. Although each of the molecular subgroups may be further subclassified into additional molecular subtypes, the division into four classes provide groups of tumours that are sufficiently homogeneous for predicting treatment outcome and revealing biological insights[1;2;34].

Molecular classification by expression profiles has been described in numerous other cancer types, though the molecular classes are often not reproducibly discovered across studies and the classes often provide little scientific insight or clinical utility. For example, the classification of CNS-PNET is not very robust[129], partly because CNS-PNET is very rare and highly heterogeneous. Another problem occurs when a set of tumours is transcriptionally homogeneous. Although most clustering algorithms will nevertheless produce clusters from the expression profiles of these tumours, the resulting clusters may not represent biologically meaningful classes and will not be reproducible across cohorts[5]. To emphasize, molecular classes discovered by unsupervised clustering should be reproduced across multiple cohorts.

In other cancer types, many other molecular classification approaches have been used. For example, the molecular classes of breast cancer were initially discovered by

expression profiling[130], but it is now predominantly classified by three protein markers — estrogen receptor, progesterone receptor, and ERBB2 (HER2) — as well as a marker for cellular proliferation (Ki67) into four main molecular subtypes: luminal A, luminal B, HER2 type, and basal-like[131–134]. Prostate cancers can be molecularly classified into two classes: positive or negative for a gene fusion involving two ETS transcription factors, *ERG* or *ETV1*, which is found in approximately half of all prostate cancers[135]. In contrast, the classification of leukemia is far more complex and involves both histological features and cytogenetic abnormalities (e.g. chromosomal rearrangements)[136]. Regardless of the methology used, the ultimate objective of molecular classification is the same: separate cancers reproducibly into biologically similar classes so that they may share similar responses to treatments. Given this objective, one may consider classifying cancer types based on patient survival; however, this approach will likely not produce biologically similar classes, because patients can die for various reasons and patient survival can be heavily influenced by the treatments rendered.

Importantly, the molecular subgroups of medulloblastoma are biologically distinct molecular entities whose clinical and genetic differences may require separate therapeutic strategies[35;37;38;88;128]. WNT medulloblastoma responds favourably to standard therapy[2;35]. While targeted therapies based on the genetics of the disease are not currently in use, inhibitors of SMO have shown some early evidence of efficacy in SHH medulloblastoma[32]. With a deeper understanding of the genomics and biology of medulloblastoma subgroups, we hope to herald a new era of medulloblastoma treatment based on selective, specific therapy.

V Risk stratification of patients

Current treatment protocols for medulloblastoma stratify patients based on clinical features: patient age, metastatic stage, and extent of resection. Patients are stratified into standard-risk and high-risk groups based on evidence of metastasis and size of the residual tumour after surgery. Additionally, infants are not irradiated to prevent impairment of neurological function by craniospinal irradiation. Typically, patients whose tumour was incompletely (subtotally) resected by surgery or has metastasized are classified as high-risk[137]. To date, histological features have not been widely used to stratify patients into risk groups in prospective clinical trials.

Stratification schemes can differ across continents, however. For example, whether

a patient should receive radiotherapy is usually determined by an age cutoff, but this threshold can be 2 years[138;139], 3 years[137;140–143], 4 years[36;144;145], or 5 years[146;147]. Conversely, hospitals in Japan (and other places) do not use strict age cutoffs to determine eligibility for radiotherapy[148]. Sometimes, infants of less than 1 year of age can receive reduced doses of chemotherapy as well.

VI Research objectives

Our study will focus on three obstacles that hinder the development of targeted therapy against molecular subgroups of medulloblastoma. First, current methods for classifying medulloblastoma into molecular subgroups are difficult to apply in the clinical setting. Second, current clinical prognostication of medulloblastoma poorly predicts patient survival and does not consider molecular subgroups. Third, few actionable therapeutic targets for WNT, Group3, and Group4 medulloblastomas have been discovered to date. In addressing these problems, we demonstrate the clinical significance of the molecular classification of medulloblastoma.

Aim 1: Molecular classification of medulloblastoma for clinicians

Although the retrospective classification of medulloblastoma has been scientifically informative, molecular classification has not been applied in the context of a prospective clinical trial. One major obstacle is the lack of fresh-frozen samples for most clinical cases. Expression profiling, on which the molecular classification of medulloblastoma was based, depends on the availability of high-quality RNA. In contrast, clinical samples are routinely subjected to formalin-fixation and paraffin-embedding, which preserves tissue integrity but causes nucleic acid degradation. To facilitate the development of therapy specifically targeted against molecular subgroups, we sought to establish a molecular classification assay that can be clinically applied on formalin-fixed, paraffin-embedded (FFPE) samples. We have established an analytic pipeline for molecular classification using expression data generated by nanoString assays, and demonstrated its high classification accuracy on FFPE samples[3]. To further make the assay clinically applicable, we have implemented several quality-control measures that identify cases which cannot be reliably assigned molecular subgroups due to poor specimen quality or assay reaction failure.

Aim 2: Clinical prognostication within molecular subgroups of medulloblastoma

Prior clinical prognostication studies in medulloblastoma have identified biomarkers without discriminating between the molecular subgroups of medulloblastoma. Given that medulloblastoma subgroups are biologically and molecularly distinct disease entities, we hypothesized that incorporating molecular subgroup into prognostication can enhance the accuracy of survival prediction and improve the reliability of risk stratification. Practical and reliable identification of risk could allow for therapy intensification in high-risk children to improve survival and therapy de-escalation in low-risk children to avoid complications of therapy. By identifying clinical and molecular biomarkers within medulloblastoma subgroups, we have designed risk stratification schemes for SHH, Group3, and Group4 medulloblastoma that can achieve unprecedented levels of prediction accuracy.

Aim 3: Discovering therapeutic targets by genomic profiling of medulloblastoma

Following the adoption of the molecular classification of medulloblastoma, we then sought to identify molecular targets in medulloblastoma. Unlike SHH medulloblastomas, actionable therapeutic targets for WNT, Group3, and Group4 tumours have yet been identified. Since prior attempts have been underpowered to discriminate the genomic differences among the four molecular subgroups, the Medulloblastoma Advanced Genomics International Consortium (MAGIC), consisting of scientists and physicians from 43 cities across the globe, has gathered > 1200 medulloblastomas. We analyzed the genomic copy-number profiles of the tumours by single nucleotide polymorphism (SNP) arrays and identified genes and pathways that characterize each medulloblastoma subgroup[2].

Chapter 2

Molecular classification of medulloblastoma for clinicians

Objective. We aim to develop a clinically applicable method for classifying medulloblastoma into molecular subgroups.

Medulloblastoma can be classified by RNA expression profiles into four molecular subgroups: WNT, SHH, Group3, and Group4. These four subgroups show some modest association with the histological subtypes of medulloblastoma. The desmosplastic histotype is enriched in SHH medulloblastoma (i.e. it is observed more often than expected by chance), while the large cell/anaplastic histotype is enriched in Group3 medulloblastoma. Nonetheless, the molecular classification is quite different from the histological classification. Despite its relative infancy, the molecular classification of medulloblastoma has gained widespread acceptance in the research community since Taylor *et al.*[34] published a consensus report on classifying medulloblastoma by molecular profiles. Prior to this consensus report, several groups have independently discovered various molecular classes of medulloblastoma using different clustering analyses (hierarchical clustering and non-negative matrix factorization (NMF) consensus clustering) for class discovery[35;37;88;128]. On the surface, these studies discovered different molecular classes; moreover, each study discovered a different number of classes. Upon closer inspection, all the studies reported the delineation of a Shh-activated and a Wnt-activated molecular class. Furthermore, most studies (those with sufficient sample size) discovered at least two additional molecular classes; the exact number of classes discovered depended on the granularity of the clustering or partitioning analysis. Indeed, the number of molecular classes is not as biologically important as the existence of each molecular entity. (Refer to the **Classification** section on page 100 in the **Appendix** for a discussion on the difficulty in ascertaining

the number of classes.) In fact, most studies identified molecular classes bearing semblance to the GroupC and GroupD subtypes discovered by Northcott *et al.*[35], which have been renamed Group3 and Group4, respectively. To emphasize, each of the currently accepted WNT, SHH, and Group3, Group4 molecular subgroups have been identified in multiple studies, and these subgroups represent molecular entities that differ genetically, epidemiologically, and clinically[34].

We sought to develop a method for classifying medulloblastoma samples into the four molecular subgroups. This method consists of two components: an experimental assay for measuring marker expression and a computational classifier for assigning molecular subgroup to an unknown medulloblastoma sample. We used the nanoString nCounter technology[149] to directly measure the expression level of 22 subgroup-specific marker genes, each of which is overexpressed in one of the four molecular subgroups. We then selected the optimal classification algorithm from a panel of algorithms by comparing cross-validation accuracies. Subsequently, we validated the trained classifier on external datasets of medulloblastoma samples with independently assigned molecular subgroups.

We did not use expression microarrays for measuring RNA expression because they have long been considered tools ill-suited for clinical diagnostics[150–160]. While the aforementioned class-discovery studies used expression arrays (Affymetrix U133 or Exon ST 1.0) to measure the expression of all protein-coding genes (more than 20 000), we only needed, for our purpose, to measure the expression of genes that help discriminate between the molecular subgroups. Reducing the number of genes to be measured simultaneously reduces the number of hybridizing probes, which in turn mitigates the potential for cross-hybridization (binding of probes to sequences other than the target sequence)[161;162]. Critically, expression microarrays typically require fresh-frozen samples and perform poorly on FFPE samples. Even on microarray platforms specifically designed for FFPE samples, the signal-to-noise ratios of FFPE samples are generally poor. Formalin fixation and paraffin embedding preserve cellular and tissue architecture but cause extensive degradation of nucleic acids, especially RNA. Furthermore, microarrays are plagued by complex and diverse preprocessing procedures (probe signal normalization and subsequent processing), in addition to experiment-specific effects (unwanted variations). All these limitations preclude the widespread adoption of microarrays in diagnostic laboratories. Conversely, the nanoString technology is less sensitive to RNA degradation and provides high reproducibility between experiments.

The nanoString platform provide high-quality measurements of RNA expression

and differs from expression microarrays in important ways[149]. In nanoString, the probes are custom designed to a relatively small set of target gene transcripts, and the cross-hybridization limitation of microarrays can be circumvented in part by reducing the number of hybridization targets and designing probes against non-homologous sequences[149;161;162]. nanoString measures transcript abundance directly while expression microarrays require *in vitro* transcription (or reverse transcription) and polymerase chain reaction (PCR) amplification. These enzymatic reactions can introduce bias (e.g. GC content bias) and variation (e.g. stochastic amplification by PCR). Additionally, in an nanoString assay, transcripts are detected by the simultaneous binding of a pair of probes (a fluorescent-labeled reporter probe and a biotinylated capture probe) in solution (i.e. 3-dimensional space); for expression arrays in comparison, each target sequence is designed to bind to one immobilized probe at a time, with no cooperativity, on a 2-dimensional surface of a chip or bead. (Newer expression platforms typically contain probes targeting multiple regions of a gene in order to achieve higher redundancy. Notably, mismatch probes present in older microarray designs such as Affymetrix U133 are often ignored during data normalization, and they have been eliminated in newer designs such as Affymetrix Exon ST 1.0.) These properties of nanoString may potentially explain why nanoString can achieve reliable quantitation on FFPE tissue[163].

While clustering analysis has been instrumental in *discovering* the molecular classes of medulloblastoma, it is ill-suited for predicting the class of a new, unknown sample. (See page 100 in the **Appendix** for the distinction between **class discovery** and **class prediction**.) Clustering analysis can be sensitive to changes in sample size: removal or addition of samples can drastically influence the clustering results. The inclusion of samples with poor quality measurements may also completely reorganize the clustering structure. Moreover, clustering analysis is prone to batch effects, where the discovered classes represent different technical batches of samples and do not reflect underlying biology. In comparison, class prediction can mitigate these effects by selecting for features that discriminate between the different classes and by assessing training and testing accuracies separately. Admittedly, clustering algorithms can be adapted for class prediction, but existing classification algorithms are applied more widely, tested more extensively, and understood more deeply. Furthermore, model-based classifiers can be designed under the classification framework in order to exploit specific statistical properties of the input data and thus improve prediction accuracy. Above all, we prefer to use or refine a ready suitable tool rather than to re-purpose a tool designed for another task.

We show that our method can accurately predict the molecular subgroup of medulloblastoma samples. The method is reproducible across different nanoString service centres and across different datasets. The assays only cost about $50 a sample. Thus, we have developed a molecular classification method for medulloblastoma that is rapid, reliable, and reproducible, and this method can be readily adopted for use in a diagnostic laboratory.

I Materials and methods

Patient samples

All samples were obtained in accordance with the Research Ethics Board at the Hospital for Sick Children (Toronto, Canada). Primary medulloblastomas comprising the training series for nanoString ($n = 101$) have been previously described by Northcott *et al.*[35] Samples contributing to the validation series ($n = 131$) were obtained from external collaborators as total RNA extracted from fresh-frozen tissue from the DKFZ (Heidelberg, Germany; Remke series, $n = 56$)[37], the Dana-Farber Cancer Institute (Boston, USA; Cho series, $n = 39$)[38], and Marcel Kool (DKFZ, Heidelberg, Germany; Kool series, $n = 36$)[128]. FFPE cases ($n = 84$) were obtained as paraffin sections from the Hospital for Sick Children (Toronto, Canada; $n = 34$), Johns Hopkins University (Baltimore, USA; $n = 25$), and the DKFZ (Heidelberg, Germany; $n = 25$).

Tissue sample processing

Total RNA was extracted from fresh-frozen tissue using the Trizol method (Invitrogen) according to the manufacturer's instructions. For FFPE samples, 3 to 5 paraffin sections per sample were first deparaffinized with xylene prior to RNA extraction using the RNeasy FFPE kit (Qiagen) as directed by the manufacturer. RNA concentration was measured using a Nanodrop 1000 instrument (Nanodrop). Paul Northcott processed the samples.

RNA integrity assessment

RNA derived from FFPE material was analyzed with the Agilent Bioanalyzer to determine RNA integrity at The Centre for Applied Genomics (TCAG). Smear analysis was performed using the Agilent 2100 expert software to determine the proportion of RNA C300 nucleotides within a given sample.

nanoString CodeSet design and expression quantification

Signature genes for each medulloblastoma subgroup were included in the CodeSet on the basis of their observed subgroup-specific expression as previously determined by Affymetrix exon array analysis. The CodeSet was designed to consist of a total of 25 genes with 5 to 6 signature genes included for each subgroup: WNT (*WIF1*, *TNC*, *GAD1*, *DKK2*, *EMX2*), SHH (*PDLIM3*, *EYA1*, *HHIP*, *ATOH1*, *SFRP1*), Group3 (*IMPG2*, *GABRA5*, *EGFL11*, *NRL*, *MAB21L2*, *NPR3*), Group4 (*KCNA1*, *EOMES*, *KHDRBS2*, *RBM24*, *UNC5D*, *OAS1*). Three housekeeping genes (*ACTB*, *GAPDH*, and *LDHA*) were also included in the CodeSet for biological normalization purposes. Probe sets for each gene in the CodeSet were designed and synthesized at nanoString Technologies. See Northcott *et al.*[3] for details on the subgroup-specific expression of the markers (note that Group C has been renamed Group3 and Group D has been renamed Group4 since the publication of this study).

Total RNA (100 ng) from fresh-frozen tissue and FFPE material was analyzed using the nanoString nCounter Analysis System at the University Health Network Microarray Centre (Toronto, Canada), the Oncogenomics Core Facility at the University of Miami (Miami, USA), and the Frontiers in Genetics Facility at the University of Geneva (Geneva, Switzerland). All procedures related to mRNA quantification including sample preparation, hybridization, detection, and scanning were carried out as recommended by nanoString Technologies.

nanoString data processing and class prediction

Raw nanoString values were log-transformed (zero counts were mapped to zero). For each raw $\log_2$ value $x^{(i)}$ in sample i of n total samples, the normalized value $\tilde{x}^{(i)}$ is calculated

as

$$\tilde{x}^{(i)} = x^{(i)} - p^{(i)} + p - b^{(i)} + b$$

where $p^{(i)}$ is the mean signal from the positive control probes in sample i, and $b^{(i)}$ is the mean signal from the three endogenous biological control probes targeting housekeeping genes (LHDA, GAPDH, and ACTB) in sample i. Additionally, p and b are the mean positive control and biological control signals across all n samples in the dataset.

$$p = \frac{1}{n}\sum_{i}^{n} p^{(i)} \quad b = \frac{1}{n}\sum_{i}^{n} b^{(i)}$$

Following normalization of nanoString counts using all samples, the normalized log2 expression values were used for downstream class prediction analysis.

A series of medulloblastomas with known subgroup affiliation ($n = 101$) were used to establish a training dataset for subsequent class prediction analysis of independent cohorts used in the study. Various class prediction algorithms were assessed by a 10-fold cross-validation scheme, using a set of scoring indices to establish a pipeline for predicting medulloblastoma subgroups with nanoString data derived from the training series. Based on superior performance in cross-validation analysis, the PAM method was selected for all downstream class prediction analyses.

All class prediction analyses were performed in the R statistical programming environment (v2.13). Implementations of the class prediction algorithms were imported from the following R packages: MASS v7.3 (linear discriminant analysis; LDA), class v7.3 (k-nearest neighbor; KNN), e1071 v1.5 (support vector machine; SVM), nnet v7.3 (multinomial log-linear model; MULT), and pamr v1.51 (prediction analysis for microarrays; PAM)[164]. During cross-validation, the training set of 101 samples was randomly split into 10 partitions. Each class predictor was trained on nine of the partitions, and the performance of the predictor was subsequently tested on the one remaining partition. Each of the 10 partitions was used as the testing set once in a round of cross-validation. This cross-validation process was repeated for a total of 10,000 partitionings. The entire experiment was repeated at least 3 times with reproducible results.

The scoring indices used during testing were simple accuracy, Jaccard similarity index, Rand index, adjusted Rand index, and FowlkesMallows index. The latter four indices are different indices for determining the similarity between two groupings, which are the known and predicted classifications of samples in the current analysis. These

indices serve as more stringent measures of accuracy in multiclass prediction. Aside from the aforementioned measures of accuracy, the reliability of a classifier was also determined using Shannon entropy as a measure of uncertainty:

$$H(Y) = -\sum_{y \in Y} P(y)\ log_2 P(y)$$

where Y is taken to be the predicted class label of the classifier for a given sample. Accordingly, $P(y)$ is estimated by the empirical frequency that the classifier predicts the sample to be class y across all cross-validation rounds. The mean entropy value for a classifier is averaged across all training samples. Hence, classifiers with varying predicted classes for the same sample across cross-validation rounds have higher entropy values and are hence less reliable.

Since the model parameters for SVM can affect the prediction performance, these parameters were optimized by a grid search in a separate round of cross-validation for SVMs with linear, radial basis, polynomial, and sigmoid kernels for observations $\mathbf{x}_i$ and $\mathbf{x}_j$ as shown below.

Kernel	$K(\mathbf{x}_i, \mathbf{x}_j)$
linear	$\mathbf{x}_i^\top \mathbf{x}_j$
polynomial	$(\gamma \mathbf{x}_i^\top \mathbf{x}_j + r)^d$
radial basis function	$exp(-\gamma \lVert \mathbf{x}_i - \mathbf{x}_j \rVert^2)$
sigmoid	$tanh(\gamma \mathbf{x}_i^\top \mathbf{x}_j + r)$

The optimal values for the kernel parameters ($\gamma > 0, r, d$) were searched in:

$$\gamma \in \{2^{-15}, 2^{-13}, 2^{-11}, \ldots, 2^3\}$$

$$r \in \{-1, -0.9, -0.8, \ldots, 1\}$$

$$d \in \{2, 3, 4, \ldots, 8\}$$

Furthermore, the optimal value for penalty parameter $C > 0$ was searched among the grid points $\{2^{-5}, 2^{-3}, 2^{-1}, \ldots, 2^{15}\}$. Similarly for KNN, the best model was selected from models with $k \in \{1, 2, \ldots, 10\}$.

Regression analysis of prediction accuracy

Define c_t as the number of correctly classified samples and n_t as the number of samples with age $x_i \leq t$

$$n_t = \sum_i I(x_i \leq t)$$

$$c_t = \sum_i I(x_i \leq t \ \cap \ i \in \mathbb{C})$$

where i iterates over samples, $I(\cdot)$ is the indicator function, $\mathbb{C}$ is the set of correctly classified samples. Cumulative accuracies were calculated for each sample age year bin t as $y_t = c_t/n_t$, and y_t was fitted as a 5-parameter logistic function of t, using the implementation in the drc v2.1 R package:

$$f(t) = \gamma + \frac{\delta - \gamma}{(1 + exp(\beta(log(t) - log(\epsilon))))^{\varsigma}}$$

The maximum asymptote parameter δ was constrained at 1 in order to reflect the high accuracy that the predictor achieved with recent FFPE samples.

Outlier detection

Gaussian mixture models were fitted to the mean endogenous control signals and the mean positive control signals of all collected nanoString data to establish the reference ranges for the endogenous and positive controls. Samples with mean endogenous control or positive control signal that deviate significantly from the respective reference range at a significance level of 0.001 were identified as outliers using the one-sample z test.

II Results

In order to select a classification algorithm that predicts most accurately and reliably, we evaluated the class prediction performance of a panel of well-known classifiers: support-vector machine (SVM), linear discriminant analysis (LDA), multinomial logistic regression, k-nearest neighbour (KNN), prediction analysis of microarrays (PAM). (Note that PAM is to be distinguished from the clustering algorithm, partitioning around medoids; LDA is also to be distinguished from latent Dirichlet allocation, another unsupervised

learning algorithm.) These classifiers were trained on a training set of 101 fresh-frozen medulloblastoma samples with known molecular subgroups. Classifiers that have tuning parameters (SVM, PAM, and KNN) were tuned using a separate round of 10-fold cross-validation. The performance of all classifiers were assessed using repeated, stratified, 10-fold cross-validation using various measures of accuracy. PAM consistently showed superior performance to all other classifiers (**Figure 2.1**). Its predictions are most consistent across multiple rounds of cross-validation, indicating that its predictions are reliable (**Figure 2.1d**). In comparison, multinomial logistic regression predicted different subgroups for the same sample when it was trained on different subsets of the training data, illustrating that it does not generalize well and its predictions are unreliable. Further, the test accuracy of the PAM algorithm as assessed by cross-validation was the highest (**Figure 2.1a-c**). Admittedly, the test accuracies of all classification algorithms were high and roughly in the same range, though PAM consistently outperformed all the other classifiers in repeated rounds of cross-validation.

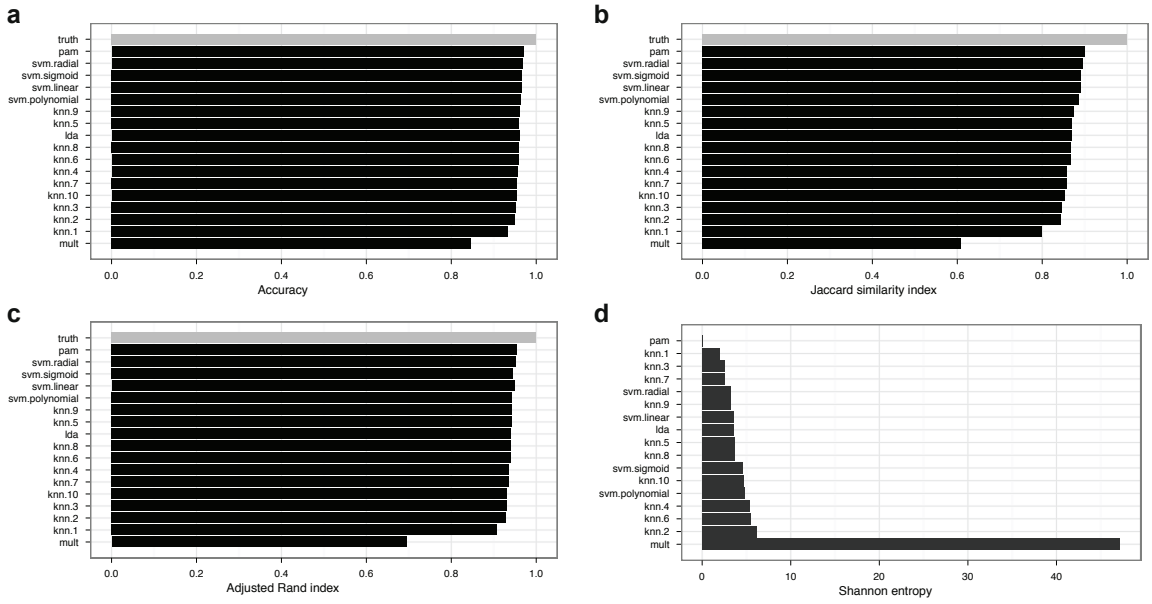

Figure 2.1: Cross-validation comparison of candidate classification algorithms. **a-c**, Accuracy assessment of trained classifiers by repeated, stratified, 10-fold cross-validation, using measures of accuracy: (a) proportion of correctly classified samples, (b) Jaccard similarity index, and (c) adjusted Rand index. **d**, Consistency of trained classifiers during cross-validation, as measured by Shannon entropy. Bars represent mean Shannon entropy values averaged across all training samples.

Therefore, we proceeded to evaluate the performance of the trained PAM classifier on external datasets of medulloblastoma samples with independently assigned molecular

subgroups. The original molecular subtypes from the previous studies were mapped to the consensus molecular subgroups using the mapping detailed in the consensus report[34]. The performance of the trained classifier was tested on an external set of 130 fresh-frozen medulloblastoma samples. By testing on an external validation set that is disjoint from the training set, we show that our classifier generalizes well and is insensitive to irrelevant variability across datasets. Indeed, our method achieved an overall classification accuracy of 98% (**Figure 2.2**).

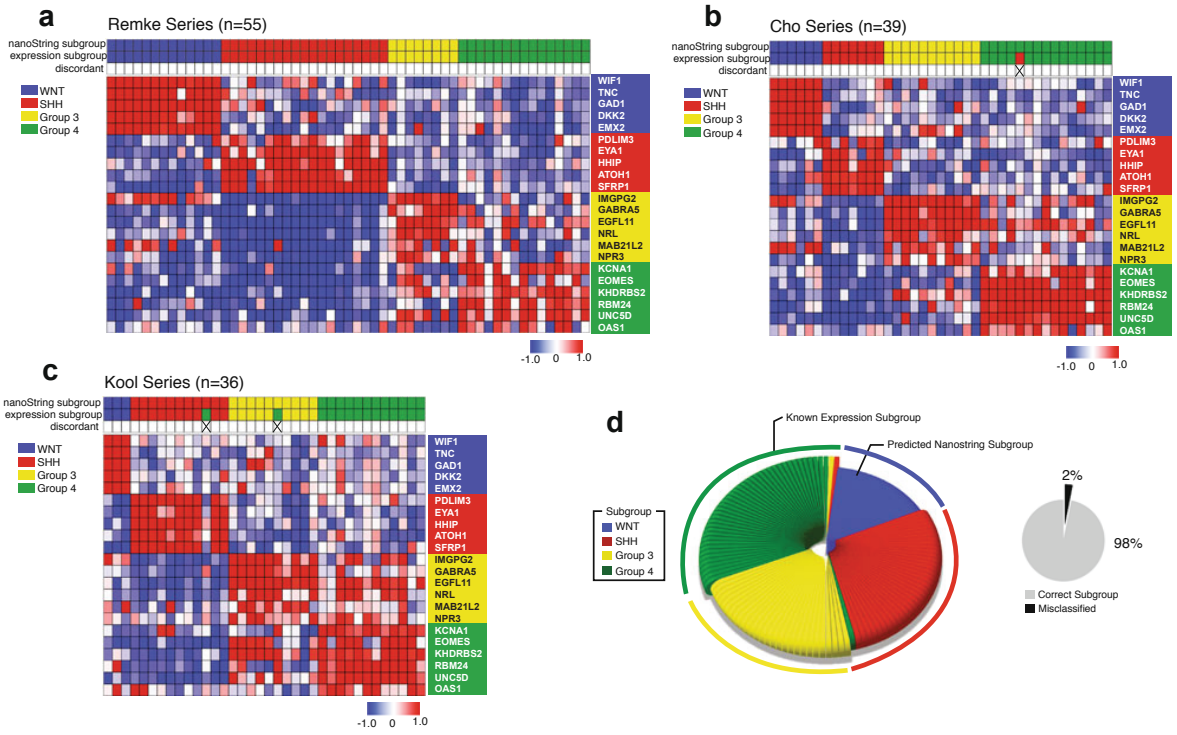

Figure 2.2: Validation of classification assay on independent medulloblastoma cohorts. **a-c**, Expression heatmaps of medulloblastomas of known molecular subgroup as published by Remke et al.[37] (a), Cho et al.[38] (b), and Kool et al.[128] (c). Samples are arranged according to subgroup predictions. Known molecular subgroups and erroneously classified cases are marked above the heatmap. **d**, *Left*, Pie chart depicting the known subgroup distribution of medulloblastomas from the three independent cohorts analyzed in **a-c** ($n = 130$) and the subgroups predicted by nanoString profiling. Misclassified cases are marked within each slice according to the predicted subgroups. *Right*, Pie chart of class prediction accuracy (127/130) on the validation set. Adapted from Northcott et al.[3]

Aside from being highly accurate, our method for determining molecular subgroups is also reproducible across multiple centres. Our method yielded reproducible predictions of molecular subgroups when the same samples were processed in three independent laboratories[3].

Furthermore, our classifier, which was trained on fresh-frozen training samples, continued to predict molecular subgroups accurately on FFPE samples. Since fresh-frozen samples are generally rare in the clinical setting and most samples are only available in FFPE archival form, our method would need to achieve acceptable performance on FFPE samples if it is to be used in diagnostic laboratories. Indeed, the clinical applicability of our method was demonstrated by its high predictive accuracy on FFPE samples of archival ages ≤ 8 years (**Figure 2.3**). The accuracy decreased on older FFPE samples, presumably due to poorer RNA integrity, though standard measurements of RNA quality were not correlated with accuracy[3]. Taken together, these results suggest that any fresh frozen or recent FFPE samples may be reliably assigned molecular subgroups using our classification method.

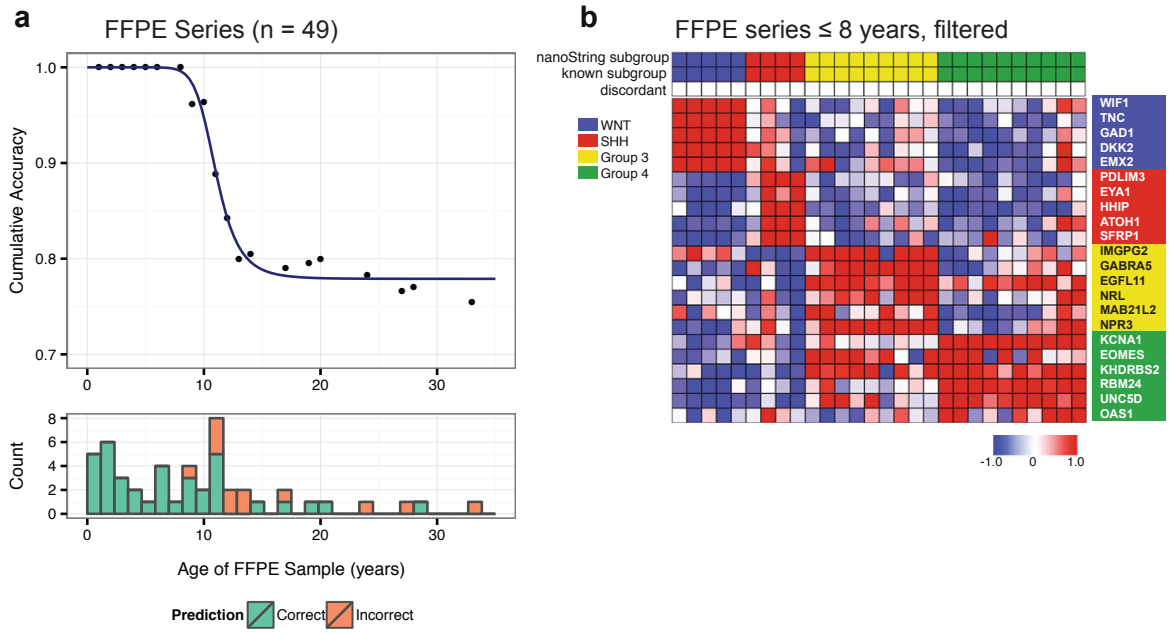

Figure 2.3: Classification performance on FFPE samples. **a**, *Top*, Cumulative class prediction accuracy in relation to sample age of archival medulloblastomas stored as FFPE material ($n = 49$). Samples obtained within the past 8 years exhibit accuracies of $\geq 95\%$. *Bottom*, Frequency distribution of sample age, grouped according to prediction correctness. **b**, Heatmap of nanoString data showing class predictions for FFPE cases of ≤ 8 years confidently predicted by the assay ($n = 28$). Samples are sorted according to subgroup prediction. All cases satisfying prediction probability threshold were assigned to the correct subgroup ($28/28$). Adapted from Northcott et al.[3]

Since the initial publication of this work[3], we have used the method to classify over 1000 medulloblastoma samples, of which 538 samples now have known subgroup affiliations determined from unsupervised clustering analysis on expression profiles. Using the classifier trained on the original training set ($n = 103$) to predict the subgroup of fresh-frozen samples in a non-overlapping new validation set ($n = 453$) yielded a classification accuracy of 92%, which is lower than the accuracy previously achieved (98 %) on the original external validation set of fresh-frozen samples ($n = 131$; **Figure 2.2**). Additionally, a few samples with replicate nanoString assays produced different class predictions. Further examination revealed that poor sample quality and suboptimal assay conditions likely contributed to errors in class prediction.

Therefore, additional quality control measures were implemented to identify unreliable results due to poor RNA quality and nanoString assay failure. Given that standard measurements of RNA quality were insufficient for predicting assay accuracy[3], the mean signals of the endogenous-control probes included in the nanoString assay were used to assess whether sufficient quantities of intact RNA were present in the samples, using an outlier detection method. Samples with low endogenous-control signal, presumably due to extensive RNA degradation, cannot be reliably assigned a molecular subgroup, and they may require repeat RNA extraction or alternative classification methods using DNA copy-number or methylation profiling. Similarly, nanoString assays may fail due to suboptimal reaction conditions (caused by pipetting error or sample contaminants). The mean signals of positive-control probes were therefore used to identify assay failures using outlier detection. The endogenous-control and positive-control quality measures helped improved accuracy on the new validation set to 94%. Although the improvement in prediction accuracy on fresh-frozen samples is very modest, we expect a much greater improvement for FFPE samples, which typically have much poorer RNA quality[3].

III Discussion

We have developed and validated a reliable method for classifying medulloblastoma into molecular subgroups. Importantly, our classification method complements rather than substitutes histological diagnosis. The diagnosis of medulloblastoma from a cerebellar or posterior fossa tumour requires histological examination and exclusion of genetically defined brain tumour entities including ATRT, which is characterized by *SMARCB1* mutation, and embryonal tumours with multilayered rosettes (ETMR), which is characterized by amplification of the chromosome 19 microRNA cluster. Given a medulloblastoma sample, our method produces probabilities of the sample belonging to each subgroup. A sample with a low prediction probability for the most probable subgroup would warrant a second examination of its histological diagnosis. Given an unknown brain tumour sample, we would also need to entertain the possibility that a non-medulloblastoma sample may exclusively express markers that define one of the medulloblastoma subgroups. Notwithstanding these limitations, we show that our method produces reproducible results across different centres and on multiple validation datasets. Furthermore, the method performs well on FFPE material, allowing it to be readily adopted in diagnostic laboratories. Since the publication of our classification method in Northcott *et al.*[3], we have made additional improvements to the method. Currently, we are refining the classification method for Clinical Laboratory Improvement Amendments (CLIA) certification.

In order to reliably guide clinical decision making, we are designing classifiers that emit calibrated prediction probabilities. Presently, the trained classifier can output class probabilities that deviate from true probabilities. For example, samples that are predicted to be SHH medulloblastoma at a class probability of 99% contain less than 99% true SHH medulloblastoma samples; that is, more than 1% of the predicted SHH samples at class probability $\geq$ 99% are in fact not SHH medulloblastoma. Calibrated probabilities are desirable because they can be incorporated into the decision theory framework to achieve the optimal balance between risks and benefits of future treatments targeted against each specific molecular subgroup of medulloblastoma.

As later chapters will show, the classification of medulloblastoma into WNT, SHH, Group3, and Group4 is practically useful and catalyzes research in elucidating the molecular mechanism underlying medulloblastoma. Indeed, the molecular classification of medulloblastoma led to numerous discoveries in the community[1;2;6–21].

Chapter 3

Clinical prognostication within molecular subgroups of medulloblastoma

Objective. We aim to stratify patients into risk groups based on clinical and molecular biomarkers within medulloblastoma subgroups for the purpose of effecting risk-adaptive treatment.

Medulloblastoma was a uniformly fatal disease with a survival duration of mere months until the introduction of systematic irradiation of the entire central nervous system in the 1940s[165]. Prior to the adoption of craniospinal (whole brain and spine) irradiation, medulloblastoma cases treated with surgical resection and localized radiotherapy recur with metastases in the subarachnoid space[166]. Although the propensity of medulloblastoma to metastasize necessitated whole CNS radiotherapy, exposing the developing brain to irradiation led to long-term neuropsychological sequelae that were beginning to be documented in the 1960s[166;167]. Integration of chemotherapy in the 1970s into the standard treatment of medulloblastoma led to a concomitant rise in patient survival[167]. Chemotherapeutic drugs, however, can also have immediate and long-term adverse effects on neurocognitive function[168–170]. Today, patients with medulloblastoma are treated by surgical resection, followed by craniospinal irradiation and combination chemotherapy. While advances in imaging and surgical technologies have largely eliminated operative mortality and minimized damage to the brain during resection, craniospinal irradiation and combination chemotherapy continue to impair neural development and cause debilitating neurocognitive decline of long-term survivors[167;171–173]. With modern treatment, patients with medulloblastoma can be cured, but at great cost to their qualities of life.

Aside from impairing brain development, chemotherapy and radiotherapy can

cause various other side-effects in long-term survivors of childhood cancer. They can cause endocrinological complications, resulting in delayed puberty, hypothyroidism, growth hormone deficiency, and stunted growth; and neurological complications, leading to symptoms including limb weakness, prolonged pain, reduced sense of touch, balance problems, permanent hearing loss, blindness, seizures, tremors, and paralysis[170;174–179]. Ironically, these anti-cancer treatments can also predispose patients to second cancers[170;174;177–180].

The Childhood Cancer Survivor Study reported sobering statistics for adult survivors of childhood cancers and highlighted adverse, long-term socioeconomic consequences of chemotherapeutic and irradiation treatment[181;182]. The survivors, compared to unaffected siblings, are 5 times more likely to suffer from functional impairments that prohibits independent living, 2 times more likely to earn less than $20 000 in annual household income[181]. (Most participants of this study were based in the United States, in which the median household income is more than $50 000 during the same period[183].) Specifically for childhood CNS cancers, the survivors are 18 times more likely to suffer from functional impairments[181]. Moreover, 70% of survivors diagnosed with CNS cancer before the age of 6 years require special education services to cope with learning or emotional difficulties[182]. The survivors' use of special education is directly related to the treatment received: cranial irradiation treatment alone increases the odds of needing special education by 7 times, while methotrexate treatment alone increases this odds by 1.3 times, compared to unaffected siblings[182]. While the long-term neurocognitive effects of chemotherapy, radiotherapy, and the brain tumour itself are intertwined, these findings suggest that cranial irradiation may be the most damaging treatment, and chemotherapy, albeit less harmful, is not entirely innocuous many years after treatment either.

Radiotherapy causes apoptosis (programmed cell death) of dividing cancer cells, but it can also cause normal dividing cells to die, leading to physical, endocrinologic, and neurologic sequelae. In a developing brain, dividing neural progenitors are sensitive to irradiation. Additionally, quiescent neural progenitors or stem cells can also incur radiation-induced DNA damage whose effect may manifest later in life. In patients with acute lymphoblastic leukemia, cranial radiation causes decline in intelligence, and this decline is progressive, showing more impairment of cognitive function with increasing time since radiation therapy[184]. Nowadays, cranial irradiation is reserved for the fewer than 20% of children with acute lymphoblastic leukemia who are considered to be at high risk for CNS relapse, in order to spare the maturing brain of the neurotoxic side-effect of radiotherapy[185]. Conversely, brain tumours usually require irradiation for complete tumour

eradication and long-term patient survival, though clinicians are increasingly aware of neurologic sequelae following radiotherapy. A recently completed prospective trial assessing 54 Gy conformal (targeted against the tumour bed) radiotherapy in low-grade glioma patients revealed a striking correlation between age at treatment and subsequent decline in intelligence quotient (IQ) score: the younger the survivor was during conformal radiotherapy, the more severe was the decline in intelligence[186]. Similarly, younger children with medulloblastoma treated with high-dose irradiation had worse progressive decline in intellectual outcome and academic performance compared to children of older age at diagnosis[171–173;187;188]. Even with a reduced dose of craniospinal radiotherapy, survivors continue to show progressive decline in intellectual and academic outcomes[188–191].

Several attempts have been made over the past three decades to minimize exposure of the developing brain to irradiation. One of the first prospective trials in reducing radiotherapy for patients with medulloblastoma reported an increased rate of tumour recurrence and consequently closed early[192], highlighting the need for a planned strategy for salvaging non-responding disease in order to maintain patient survival. While earlier attempts at reducing craniospinal irradiation led to poorer survival[192–195], other attempts at reducing craniospinal irradiation achieved relative success by incorporating chemotherapy into the treatment regiment[196–202]. Extending this approach, numerous oncology groups sought to postpone or eliminate radiotherapy in young children by using chemotherapy to control or eradicate the tumour[139–141;144–146;148;201;203–214]. Unfortunately, postsurgical chemotherapy alone often cannot achieve complete response of the residual tumour, leading to eventual use of radiotherapy. For example, combination chemotherapy with vinblastine, cisplatin, and etoposide was insufficient by itself to induce complete remission of residual medulloblastoma, and patients often progress during chemotherapy.[204;208;215]. Patients with chemoresistant tumours could be salvaged with subsequent radiotherapy; however, survivors treated with delayed radiotherapy can still suffer from neurodevelopmental deficits[146;215].

In yet other trials, clinicians have successfully used high-dose combination therapy to eliminate irradiation treatment in young children with non-metastatic medulloblastoma. Geyer *et al.* showed that two combination therapy regimens (vincristine, cisplatin, cyclophosphamide, and etoposide; or vincristine, carboplatin, ifosfamide, and etoposide) could both obviate the need for radiotherapy in patients with no metastatic or residual tumour after surgery[140]. Similarly, Grill *et al.* used combination chemotherapy (carboplatin, procarbazine, etoposide, cisplatin, vincristine, cyclophosphamide) and sal-

vaged progressive medulloblastoma with radiation and additional high-dose chemotherapy (busulfan, thiotepa, and melphalan)[146]. Patients without metastasis or residual tumour exhibited favourable outcome, and the survivors in this study also had improved intellectual outcome compared to radiotherapy-treated patients[146]. Rutkowski *et al.* used combination chemotherapy alone (including vincristine, carboplatin, etoposide, cyclophosphamide, and methotrexate) and achieved favourable survival outcomes for children without metastasis or residual tumour[141]. Decline in IQ was still evident among the non-irradiated survivors, albeit less severe than those who had received radiotherapy[141]. Chi *et al.* using combination chemotherapy (vincristine, cisplatin, etoposide, cyclophosphamide, methotrexate, and thiotepa) and autologous stem cell transplant following the consolidation chemotherapy, yielded complete response in macroscopically metastatic medulloblastomas without the need for radiotherapy[216]. Hence, efforts in using chemotherapy to ward off radiotherapy led to the adoption of the practice of withholding or postponing radiotherapy for treating young children with medulloblastoma in most of North America and Europe.

Based on the results of more recent trials, however, clinicians remain divided on the use of radiotherapy in young children. The HIT 2000 trial (2001–2005) confirmed that combination chemotherapy (cyclophosphamide, vincristine, carboplatin, etoposide, and methotrexate) can maintain favourable survivorship without radiotherapy in non-metastatic medulloblastoma[145]. Unless complete remission was achieved following induction, the authors recommended contingent treatment with local radiotherapy, secondary surgery, and consolidation chemotherapy (cisplatin, lomustine, and vincristine)[145]. Conversely, the COG-P9934 trial (2000–2006) brought back unconditional, planned irradiation and showed that conformal radiotherapy (localized to posterior fossa and tumour bed) in addition to chemotherapy achieved superior progression-free survival than chemotherapy alone by comparing against the POG-9233 trial[142]. Despite known risk for long-term neurotoxicity in children, the authors contended that radiotherapy is still required for optimal survival of patients with nonmetastatic medulloblastoma[142].

Given the limited capability for chemotherapy to replace radiotherapy in all patients, it is important to select the patients who are at low risk of progressive or recurrent medulloblastoma and evaluate their candidacy for therapy de-escalation. Currently, the main risk factors for medulloblastoma relapse are residual tumour after subtotal surgical resection and presentation with metastasis either macroscopically or in the cerebral spinal fluid. As earlier trials have shown, patients with residual disease or metastasis are poor

candidates for reduced or withheld radiotherapy[141;146], though one trial had some success with metastatic medulloblastoma[216]. Additionally, the 5-year overall survival rate for nonmetastatic, completely resected medulloblastoma in young children (age < 3) of the HITSKK92 trial (1992–1997) was an impressive 93% ($n = 17$), while the POG-9233 trial (1992–ongoing) reported a 5-year overall survival rate of 43% ($n = 37$) for nonmetastatic, completely resected medulloblastoma. Indeed, the results from the two trials are difficult to compare, given differences in the chemotherapy regimen, surgical resection, imaging technologies, and supportive care. Nonetheless, the striking difference in survival for what should be similar cases of medulloblastoma suggest that the two cohorts are, in fact, biologically dissimilar. Accordingly, the prognostic factors currently used in patient risk stratification, metastasis at diagnosis and extent of resection, fail to accurately identify favourable responders to chemotherapeutic treatment.

What has been missing in past clinical trials is the classification of medulloblastoma into molecular subgroups. As medulloblastoma subgroups exhibit different survivorships (**Figure 3.1**), we believe that subgroups may be useful for risk stratification of patients. Further, given the distinct origins of subgroups[72], we hypothesize that prognostic markers would be influenced by subgroups. That is, some markers may be prognostic only in specific subgroups while others may be surrogate markers of subgroup status and have no prognostic value themselves. Therefore, by incorporating molecular subgroups into risk stratification, we believe that we would be able to more accurately predict favourable responders to treatment and obviate the need for indiscriminate administration of intensive treatment to children, who will suffer long-term treatment-induced toxicities. By improving risk stratification and tuning treatment intensity, we hope to minimize collateral damage to the patient's developing brain and preserve the survivor's quality of life.

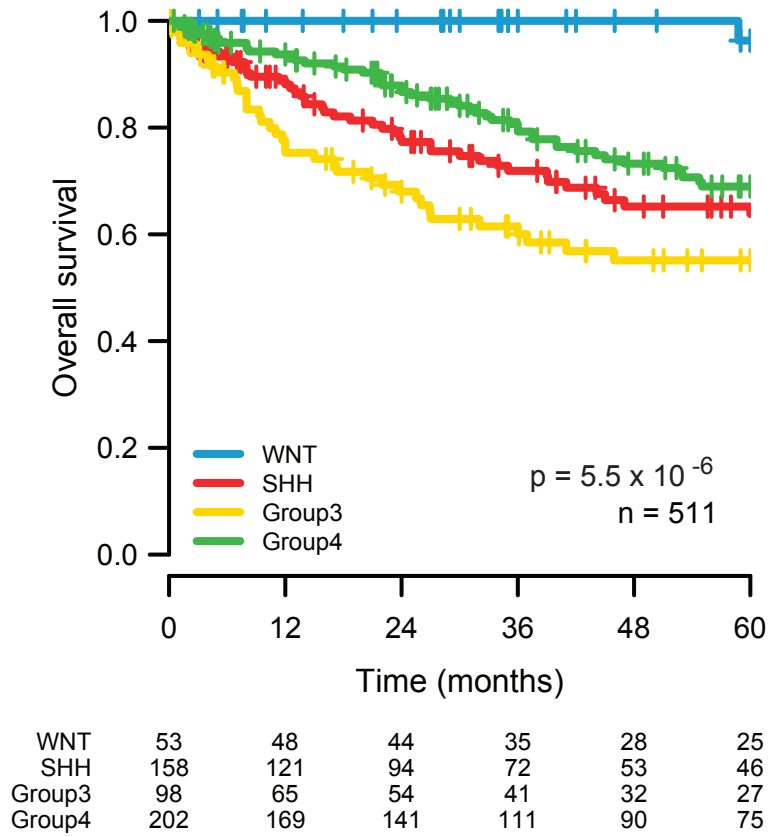

Figure 3.1: Overall survival curves for molecular subgroups of medulloblastoma. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; hazard ratio (HR) estimates are derived from Cox proportional-hazards analyses.

The current paucity of markers used in risk stratification is not due to a lack of biomarker studies. Indeed, the medulloblastoma literature is rife with reports of prognostic markers. Most of the purported markers, however, do not reproducibly predict survival in different cohorts due to small sample sizes and distributional differences in underlying covariates[1]. We purport that disagreements in prior biomarker identification attempts may be explained by differences in the composition of medulloblastoma subgroups in the different cohorts. For example, patients with desmoplastic medulloblastoma often exhibit better survival[31;36;141;142;145;209;211;217], and the discordant survival outcomes between the POG-9233 and the COG-P9934 trials could be due the latter having a higher proportion of desmoplastic medulloblastoma, leading to better patient survival[142]. The aforementioned difference in survival outcomes between the POG-9233 and the HIT-SKK92 trial could also be due to a similar reason. Indeed, the small sample sizes in these trials (POG-9233, $n = 112$; HITSKK92, $n = 62$; COG-P9934, $n = 82$) makes uneven distribution of covariates likely, and these covariates may be responsible for the outcome differences. Further, while desmoplastic histology has been proposed to be one such covariate, it is likely that unobserved covariates such as molecular subgroups and genetic mutations may better explain the differences in response. In addition to its status as a favourable prognostic factor, desmoplasia has also been reported to be a prognostic factor for poor survival in medulloblastoma by some[147;218;219] and an insignificant factor by others[217;220]. High interobserver and intraobserver variability in histological examination may be responsible for these discrepancies. Therefore, difficulty and variability in assessment can limit the utility of a biomarker, and competing covariates should be assessed carefully.

In order to mitigate the problems of small sample size and competing unobserved covariates, we assembled an large, international cohort of 673 medulloblastoma samples with clinical annotation. The cytogenetic and focal copy-number events were determined using copy-number profiling on this discovery cohort. We identified subgroup-specific cytogenetic events and integrated them with clinical variables to develop subgroup-specific, multivariate risk-stratification models based on the discovery cohort. In order to validate the models and ensure that the technique was generalizable to routine pathology laboratories, we then studied a panel of six cytogenetic biomarkers (*GLI2*, *MYC*, 11, 14, 17p, and 17q) using interphase fluorescence *in situ* hybridization (FISH) on an FFPE medulloblastoma tissue microarray that includes a set of 453 medulloblastomas that were treated at a single center and does not overlap with the discovery cohort.

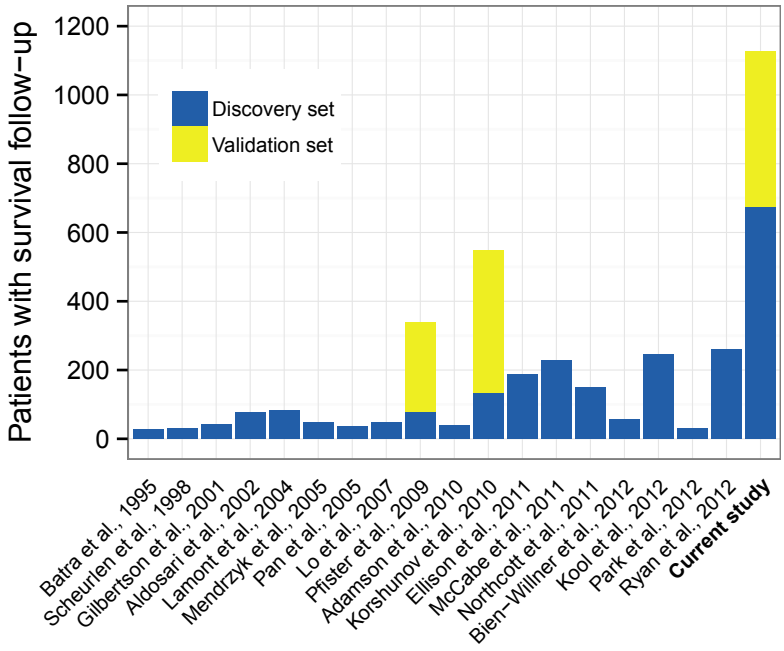

Figure 3.2: Sample sizes of recent prognostic marker studies. This meta-analysis was performed by Marc Remke.

The size of our discovery and validation cohorts provides unprecedented power for clinical prognostication and enables comprehensive, multivariate modeling of patient survival to identify robust prognostic markers **Figure 3.2**. In this retrospective study, we wish to comprehensively assess cytogenetic markers in the context of the molecular subgroups of medulloblastoma and determine whether subgroup affiliation could complement clinical variables for more accurate risk stratification of patients and predict favourable responders for de-escalation of radiotherapy, in order to improve the quality of life for survivors. Being retrospective, our cohort is subject to recall bias (for cases with frozen samples), and it encompasses heterogeneously treated patients from multiple centres and continents. Our histology records were not centrally reviewed; nevertheless, our study reflects the typical clinical experience more closely and implicitly reveals the weakness of histological diagnosis in decentralized clinical practice. Further, the lack of surgical details and treatment protocols precludes analyses on how specific treatments and extent of surgical resection affect survival outcome. Notwithstanding the limitations of our discovery cohort, the findings are highly reproducible in an independent cohort of patients.

I Materials and methods

Patient information

All tissues and clinicopathological information were collected in accordance with institutional review boards from various contributing centers. In the discovery set, although precise treatment dates were often unavailable, at least 95% of the patients were treated within the past 15 years using modern treatment protocols, including surgical resection, craniospinal (whole brain and spine) irradiation, or chemotherapy. Discovery set samples were collected between 2005 and 2013, with a focus on samples with available fresh-frozen material. Among the samples with treatment details, the earliest diagnosis is July 1997 and the latest is August 2012. Samples in the validation set were all obtained from the Burdenko institute with no selection criterion applied. All patients in the validation set were treated between 1995 and 2010 according to standardized therapy protocols of the German HIT study group.

Tumor material and patient characteristics

A discovery set of 673 medulloblastoma samples with clinical follow-up was acquired retrospectively from 43 cities around the globe. These samples were copy-number profiled on the Affymetrix SNP6 array platform in order to identify potential molecular biomarkers[2]. An independent validation set of 453 samples with clinical follow-up on a medulloblastoma tissue microarray was analyzed by FISH as previously described[35]. The validation set consisted only of patients treated in Burdenko, Moscow. Tumors were classified based on signature marker expression into molecular subgroups as previously described[3]; additional tumours were classified based on cytogenetic aberrations using standard conditional probability models. Subgroup affiliation was not available for 162 discovery samples. The validation set includes an additional set of 50 WNT tumours that were not on the tissue microarray. Nucleic acid isolation, tissue microarray construction, and β-catenin mutation analysis were performed as previously described[2]. Tissue microarray analysis was performed by collaborators Andrey Korshunov and Stefan Pfister.

Prognostic biomarker identification

Cytogenetic events and copy-number aberrations were identified as previously described in the discovery set[2]. All chromosomal events (or chromosome arm events) were compared against reference samples with balanced copy-number for the chromosome (or chromosome arm); samples with copy-number changes in the opposite direction were specifically excluded from each comparison. Subsequent to biomarker discovery, cross-validation with correction for multiple hypothesis testing was performed to estimate the reproducibility and generalizability of the potential biomarkers in an independent cohort. During cross-validation, the discovery set was split randomly into two subsets. First, the biomarkers are tested by the log-rank test on the first subset. Then, statistically significant biomarkers ($p < 0.05$) are tested again by the log-rank test on the second subset, with correction for multiple hypotheses testing. This process was repeated 10 000 times to estimate the expected validation rate of each biomarker. The expected validation rate of each biomarker is n_v/n_d, where n_d is the number of times a biomarker is significant in the first subset and n_v is the number of times a discovered biomarker is also significant in the second subset. The final set of biomarkers was further validated in the external validation set. Additionally, sample size estimates for prospective trials of each biomarker were calculated under univariate Cox models based on the observed hazard ratios. (See page 105 in the **Appendix** for more details on prognostic biomarker discovery.)

Multiple hypothesis testing correction

Within each biomarker identification analysis, correction for multiple hypothesis testing was performed by the Benjamini-Hochberg method[221] during cross-validation. Independent analyses were corrected for multiple hypotheses testing independently: clinical biomarker identification across medulloblastoma, within WNT medulloblastoma, within SHH medulloblastoma, within Group3, and within Group4; molecular biomarker identification across medulloblastoma, within WNT medulloblastoma, within SHH medulloblastoma, within Group3 medulloblastoma, and within Group4 medulloblastoma.

Time-dependent ROC analysis

Time-dependent receiver-operating characteristics (ROC) analyses were performed using the `CoxWeights` function provided in the risksetROC (v1.0.4) R package. This func-

tion calculates areas under time-dependent ROC curves as described by Heagerty and Zheng[222]. Estimates of area under the curve (AUC) for the fitted multivariate Cox model being assessed were calculated every month, from 1 month to 60 months, in order to determine the collective predictive performance of the biomarkers in the Cox models. Differences in AUC estimates among Cox models across time points were tested by Friedman rank sum tests.

Risk-stratification model selection

Candidate prognostic markers, including all cytogenetic events, focal copy-number events, and all clinical features, were first tested by univariate survival analyses (log-rank tests) individually. Significant univariate markers were tested under multivariate Cox proportional-hazards models including age and gender as covariates. Significant multivariate markers (including age) were included in the model selection step.

The candidate survival models for each medulloblastoma subgroup were determined by unbiased model selection procedures: stepwise regression using forward selection, backward elimination, and bidirectional elimination. To make the final model practical for use with FISH in diagnostic laboratories, up to a maximum of three molecular (cytogenetic or copy-number) markers were accepted in a candidate survival model. Therefore, if a candidate model contains more than three moelcular markers, only the three most significant multivariate markers were kept. In most cases, all model selection procedures produced the same survival model. When more than one candidate models were generated, the candidate models were compared by analysis of deviance tests[223] (as implemented in the `anova.glm` function of the R stats package, v2.15) to determine the model that best fits the data.

The survival data was re-analyzed by the final model (for each medulloblastoma subgroup) in order to assess the survival of patients with each combination of variable levels; the risk-stratification trees were manually designed in order to group patients into distinct risk groups. For example, suppose a final model consists of prognostic binary variables A, B, and C. The survival data is re-analyzed by fitting a multivariate Cox model with variables A, B, and C. The most significant variable (say A) of the three is chosen for the initial risk-stratification tree split, and the outcomes of patients with $A = 0$ and $A = 1$ are compared and the two branches are sorted in ascending order of survival. Then, the next significant marker (say B) is added to the tree to both branches;

however, a split must result in two branches each with combinations of variable levels that are observed. For instance, the splitting of the $A = 1$ branch with the B variable is only considered if the combinations $(A = 1, B = 1)$ and $(A = 1, B = 0)$ are both observed. The outcomes of patients with different combinations of variable levels are compared, and the combinations are again sorted in ascending order of survival. The last significant marker (C) is added in the same manner. Finally, combinations of variable levels at the leaves of the tree are collapsed into three risk groups by comparing survivals of adjacent combinations using the log-rank test. The three risk groups may be further collapsed to two risk groups if adjacent groups do not have different survival.

Statistical analysis

Patient survival characteristics were right-censored at 5 years (or 10 years) and analyzed by the Kaplan-Meier method. Univariate comparison of two or more survival curves were performed using log-rank tests and the Cox proportional-hazards regression models. The predictive values of biomarkers were assessed by analyses of deviance tests under multivariate Cox models and by time-dependent ROC analyses. Associations between covariates and risk groups were tested by the Fisher's exact test. All statistical analyses were performed in the R software environment (v2.15), using R packages survival (v2.36), risksetROC (v1.0.4), powerSurvEpi (v0.0.6), and ggplot2 (v0.9.3).

II Results

Prognostic significance of clinical variables within medulloblastoma subgroups

Many prior medulloblastoma biomarker publications were limited by sample size, a problem that will only be exacerbated once cohorts are divided into their molecular subgroups. The current study includes 1126 medulloblastoma patients (673 discovery plus 453 validation patients), which is more than double the sample size of any prior medulloblastoma biomarker publication, and one of only a very few that includes a validation cohort (**Figure 3.2**). Although the discovery cohort accumulated by MAGIC consists of medulloblastomas gathered from 43 different treating centers from around the world, the subgroup-specific outcome mirrors what has been previously published with very good

outcomes for WNT patients, poor outcomes for Group3 patients, and intermediate outcomes for SHH and Group4 patients (**Figure 3.1**) suggesting that the discovery cohort is a representative sample[1].

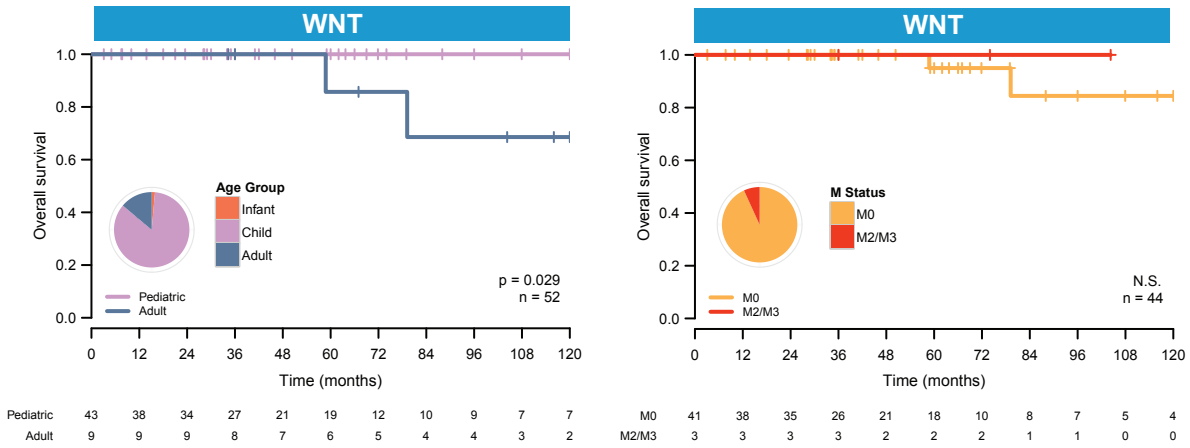

Figure 3.3: Ten-year overall survival curves for WNT medulloblastoma, split by age group or metastatic status. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

In order to assess long-term survivors, WNT patients were followed for up to 10 years, and only two deaths were observed, both late in the follow-up period and due to recurrence of medulloblastoma (**Figure 3.3**). Among the SHH tumours, there is a significantly better outcome in the adult patients as compared to children or infants (**Figure 3.4**). There is a trend towards a worse outcome for infants with Group3 tumours that is not statistically significant (**Figure 3.4**). Infants with Group4 tumours have a significantly worse outcome than children or adults (**Figure 3.4**), suggesting that radiation therapy is critical in the treatment of Group4 medulloblastoma. There is no reproducible association between gender and prognosis in any of the four subgroups ([1]). Desmoplastic histology portends a more favorable prognosis than classic histology, which is more favorable than anaplastic histology among SHH tumours[1]. Large cell/anaplastic histology has prognostic significance for Group3 medulloblastomas in the discovery cohort, but does not validate as significant in the validation cohort.

While metastatic status is not prognostic for patients with WNT medulloblastoma, macroscopic metastasis (M2/M3) is consistently associated with poor survival in all non-WNT subgroups, though the clinical effect is very slight among patients with Group4 disease (**Figure 3.5**). While the prognostic significance of M0 disease as compared to

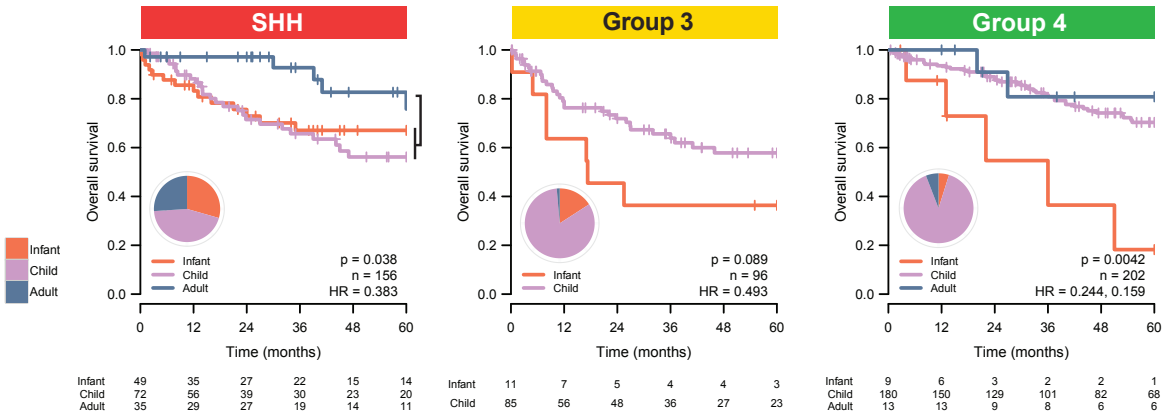

Figure 3.4: Overall survival curves for age groups within SHH, Group3, and Group4 subgroups. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

M2/3 disease is very clear across SHH, Group3, and Group4, the prognostic significance of isolated M1 disease is less clear (**Figure 3.5**). Isolated M1 disease is associated with increased risk in Group3 in the discovery cohort, but not the validation cohort, with the opposite pattern seen in the SHH patients. However, for both discovery and validation cohorts, there are no survival differences survival between M0 and M1 patients with Group4 disease. There are no CNAs in any of the subgroups that are associated with an increased risk of leptomeningeal dissemination (not shown). Overall, many clinical biomarkers continue to exhibit prognostic significance when medulloblastoma is analyzed in a subgroup-specific fashion.

Subgroup and metastatic status are the most predictive markers

Multivariate survival analyses were conducted in order to dissect the relative predictive value of clinical variables (age, gender, metastatic status, and histotype) and molecular subgroup affiliation. Stepwise Cox proportional-hazards regressions revealed that molecular subgroup significantly contributes to multivariate survival prediction, on top of a regression model already parameterized by clinical variables gender, age, metastatic status, and histology (**Figure 3.6a**). Further, Cox models parameterized with both clinical biomarkers and molecular subgroup achieve higher prediction accuracy in time-dependent ROC analyses (**Figure 3.6b**). In isolation, each biomarker has modest prediction accuracy (**Figure 3.6c**) compared to the complete multivariate model (**Figure 3.6b**). In the com-

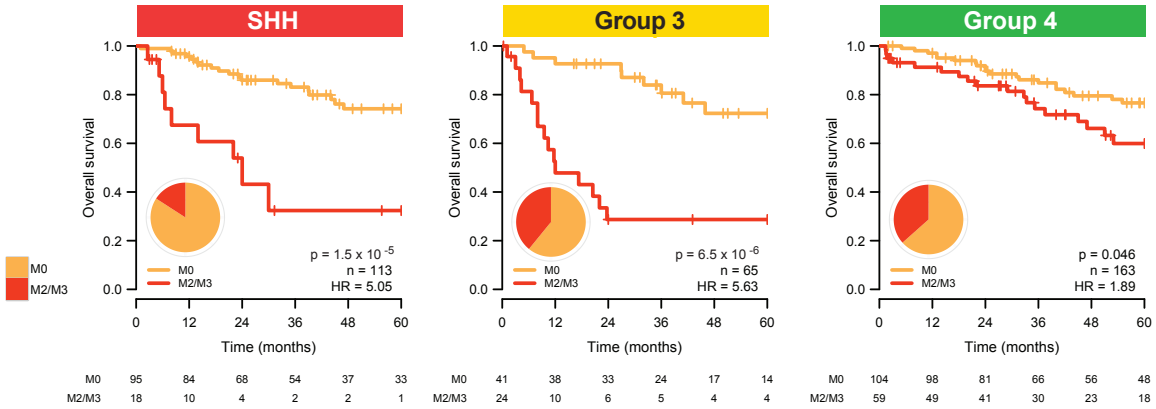

Figure 3.5: Overall survival curves for metastatic status within SHH, Group3, and Group4 subgroups. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

plete model, the removals of metastatic status and subgroup lead to the greatest decreases in predictive accuracy (**Figure 3.6d**). Taken together, these results suggest that subgroup affiliation and metastatic status are the most important predictive biomarkers, and that they make non-redundant contributions to the prediction of survival. We conclude that combining both clinical biomarkers (metastatic status) and molecular biomarkers (subgroup affiliation) will make the optimal tool for predicting survival of medulloblastoma patients.

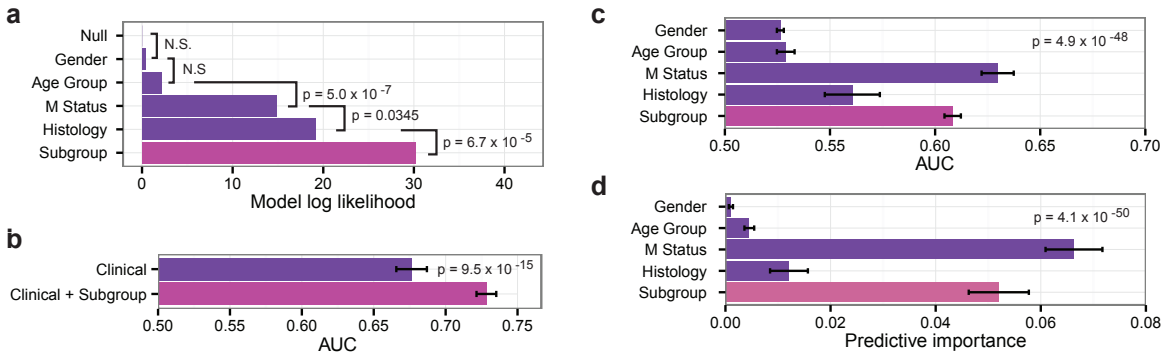

Figure 3.6: Molecular subgroup and metastatic status are the most important prognostic biomarkers. **a**, Multivariate Cox proportional-hazards survival analysis of predictor variables. Starting with the null model, each variable is added stepwise (from top to bottom) to the survival model. Model likelihood values assess the degree to which each Cox model fits the survival data. Increments in model likelihoods are tested by analysis of deviance. **b**, Average areas under time-dependent receiver operating characteristic curves (AUC) for multivariate Cox models parameterized by only clinical variables, or both clinical and subgroup variables. **c**, Average time-dependent AUCs for univariate Cox models parameterized by each variable. **d**, Predictive importance of each variable in the fully-parameterized multivariate Cox models, as determined by the average decrease in time-dependent AUC when the variable is omitted from the model. Differences in time-dependent AUC and predictive importance are evaluated by the Friedman rank sum test.

Subgroup specificity of published molecular biomarkers

Several cytogenetic biomarkers have been previously reported to be associated with patient survival across medulloblastoma, but their prognostic values have seldom been assessed in the context of medulloblastoma subgroups[1]. Loss of chromosome 6 is significantly associated with improved survival across all medulloblastoma (**Figure 3.7a**). However, the prognostic value of chr6 loss can be completely attributed to its enrichment in WNT medulloblastomas, as chr6 loss has no prognostic value among WNT patients, or among non-WNT patients, when compared to their respective controls with balanced chr6 (**Figure 3.7b**). We would suggest that chr6 loss is a subgroup-driven biomarker in that its prognostic significance is driven by its enrichment in a particular subgroup, and it thus has no further significance in subgroup-specific analysis. Further, based on these results, we would caution against using chr6 loss as the lone diagnostic criteria for WNT medulloblastoma, since it is also observed in non-WNT medulloblastoma (7/49 chr6 loss medulloblastomas were not WNT (14%)), and chr6 loss is only present in 42/53 WNT tumours (79%). The prognostic role of isochromosome 17q (iso17q) has been very controversial; in our cohort, iso17q is a statistically significant predictor of poor outcome overall (**Figure 3.7c**). However, subgroup-specific analysis demonstrates that iso17q is highly prognostic for Group3 medulloblastoma, but not for Group4 medulloblastoma (**Figure 3.7d**), indicating that it is a subgroup-specific molecular biomarker. Similarly, while chr10q loss is a modestly significant predictor of poor outcome across medulloblastoma subgroups (**Figure 3.7e**), its prognostic power is limited to the SHH subgroup of tumours in a subgroup-specific analysis (**Figure 3.7f**). We conclude that determination of molecular subgroup is crucial in the evaluation and implementation of molecular biomarkers for patients with medulloblastoma, as some putative biomarkers are merely enriching for a specific subgroup (**subgroup driven**) while most others are only significant within a specific subgroup (**subgroup specific**).

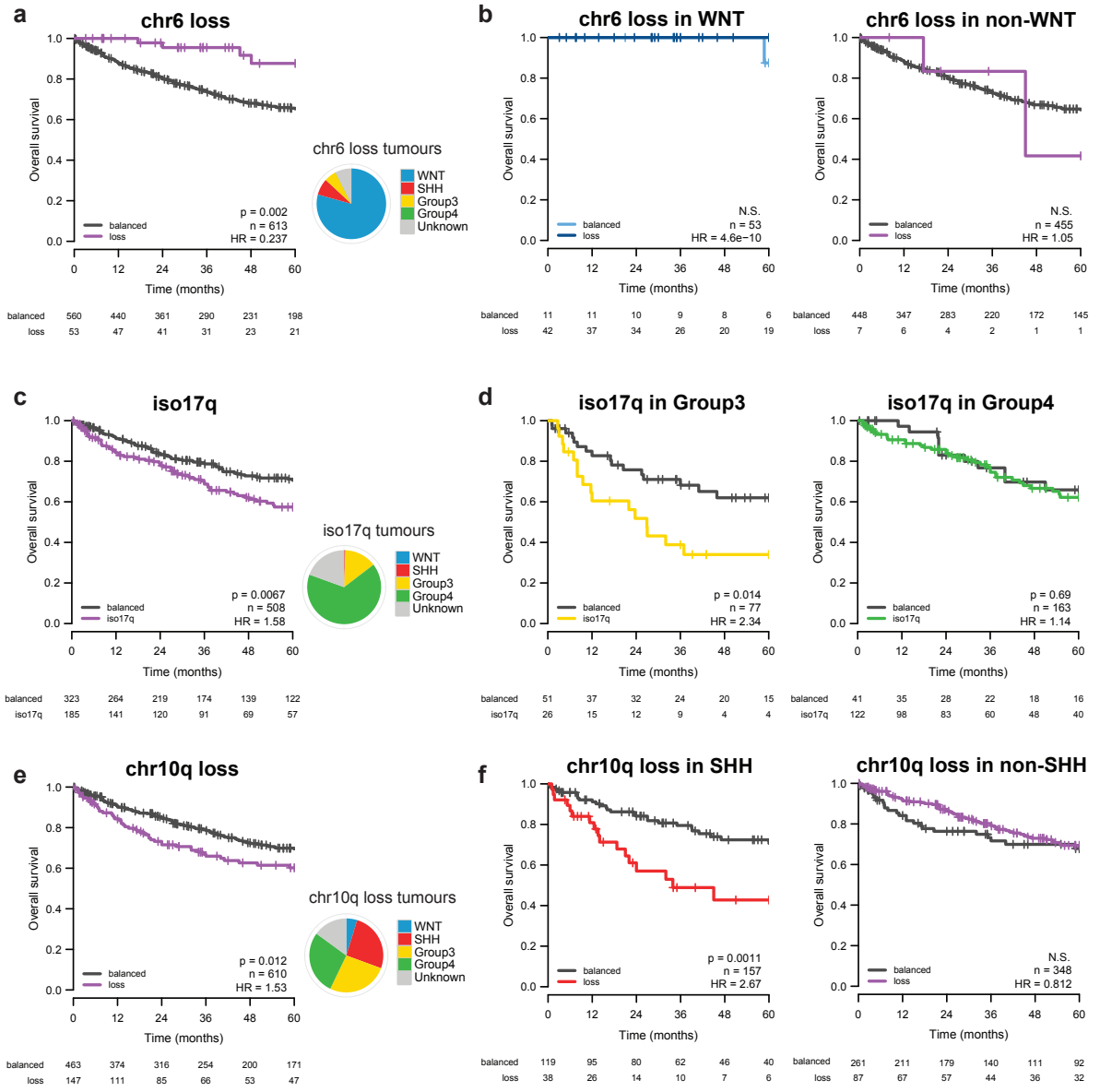

Figure 3.7: Subgroup-driven and subgroup-specific molecular biomarkers. **a**, Overall survival curves and frequency distribution of chr6 status across the entire cohort. **b**, Overall survival curves for chr6 status in WNT and non-WNT medulloblastomas. **c**, Overall survival curves and frequency distribution of isolated chr17q gain across the entire cohort. **d**, Overall survival curves for chr17q status in Group3 and Group4 subgroups. **e**, Overall survival curves for chr10q status across the entire cohort. **f**, Overall survival curves for chr10q status in SHH and non-SHH medulloblastomas. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

SHH patients can be stratified into three distinct risk groups

We identified 11 CNAs that are prognostically significant in our SHH medulloblastoma discovery set in univariate survival analyses (**Figure 3.8**, **Figure 3.9**). Given the considerable number of candidates, the reproducibility of the identified biomarkers was assessed by cross-validation. Furthremore, the expected sample sizes required for validation in future prospective trials were estimated using power analyses under Cox proportional-hazards models, in order to guide candidate prioritization in future prospective trials. Specific amplifications but not broad gains encompassing *GLI2* or *MYCN* are associated with bleak prognosis (**Figure 3.8a–b**). Loss of chr14q confers significantly inferior survival (**Figure 3.8c**). There is no minimal region of deletion on chr14 in SHH patients, and recent medulloblastoma re-sequencing efforts have not identified any recurrent SNVs on chr14 in SHH medulloblastoma[1]. The presence of chromothripsis (chromosome shattering) is associated with worse survival in SHH patients (**Figure 3.8d**).

To integrate the individual biomarkers into a risk stratification model, multivariate Cox proportional-hazards analyses were performed on all significant prognostic markers. Through multiple stepwise regression procedures, a consensus set of biomarkers was selected for inclusion in the model in an unbiased manner. The proposed risk stratification scheme represents the model that was most consistent with available data in the discovery cohort, from among many possible alternatives (**Figure 3.10a**)[1]. *GLI2* amplification, chr14q loss, and leptomeningeal dissemination (M+ disease) identify high and standard risk patients. Specifically, *GLI2* amplification alone can identify patients with bleak prognosis (**Figure 3.10a**)[1]. Absence of these markers demarcates a low-risk group of patients who exhibit survivorship reminiscent of WNT patients. Importantly, none of the covariates, particularly age and anaplastic histology, can explain the survival differences observed among the risk groups (**Figure 3.10a**)[1]. Direct application of the proposed risk stratification scheme on the independent validation cohort yields distinct survivorships for the three risk groups, thereby validating the model (**Figure 3.10c**).

Two additional stratification schemes were constructed using only clinical biomarkers or only cytogenetic markers; however, the proposed model, which combines both types of biomarkers, yields the highest prediction accuracy (**Figure 3.10b**)[1]. Furthermore, the accuracy of the combined risk model is reduced when applied across non-SHH patients, further underscoring the importance of taking subgroup into consideration during risk stratification. By using two molecular biomarkers (*GLI2* and 14q FISH) and metastatic status, we can reliably predict prognosis for patients with SHH medulloblastoma.

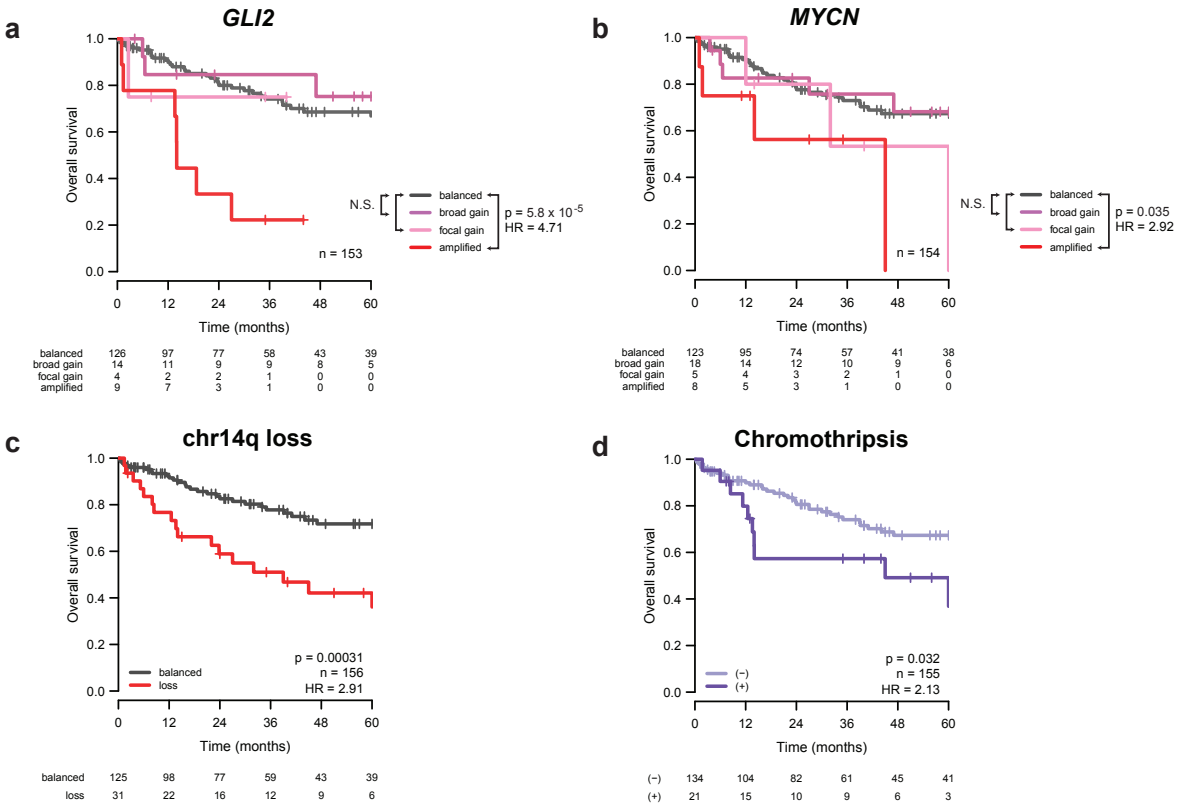

Figure 3.8: Overall survival curves for molecular biomarkers in SHH medulloblastoma: **a**, *GLI2* copy number status; **b**, *MYCN* copy number status; **c**, chr14q status; and **d**, chromothripsis status. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

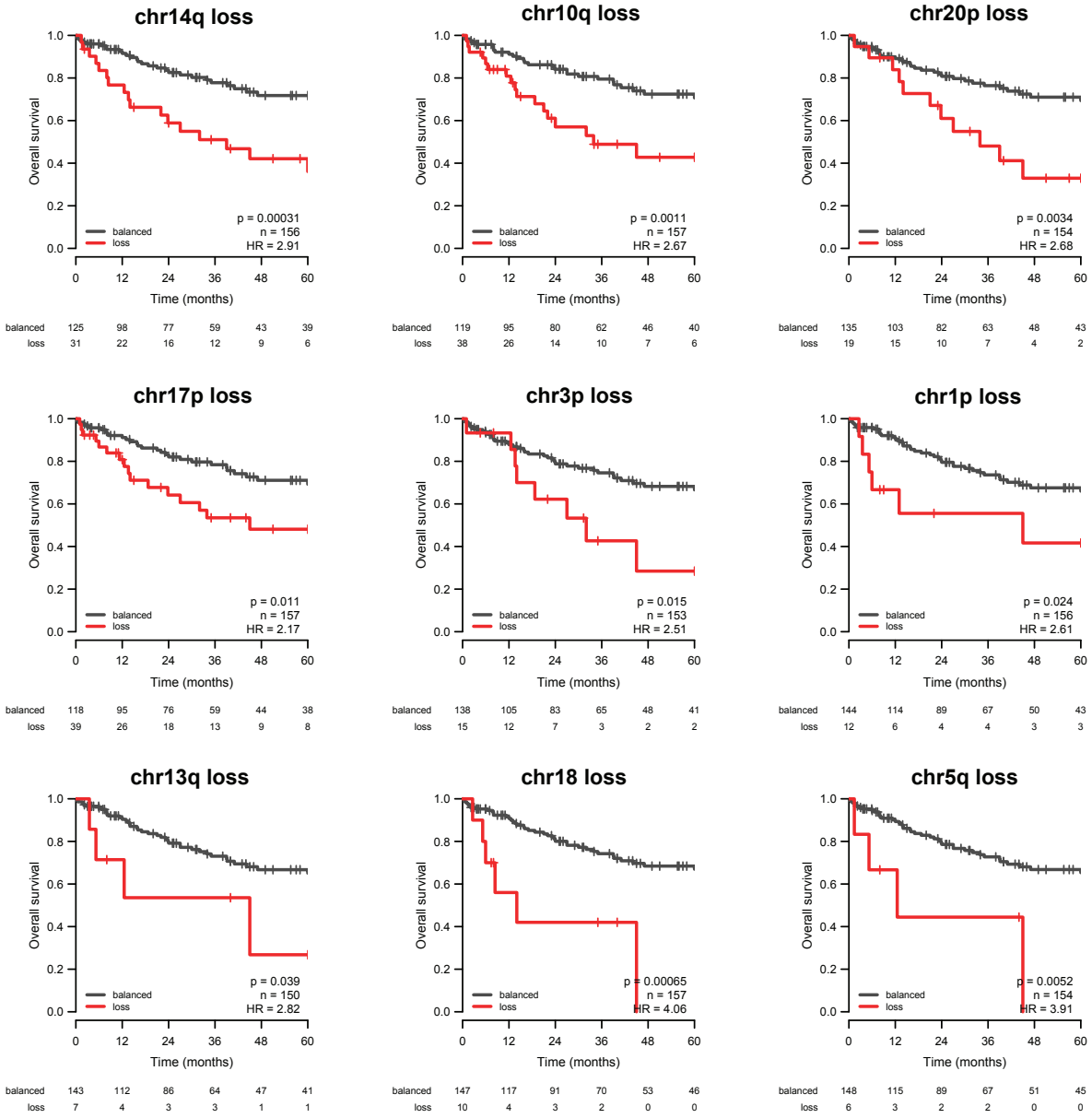

Figure 3.9: Overall survival curves for significant cytogenetic biomarkers in SHH medulloblastoma. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

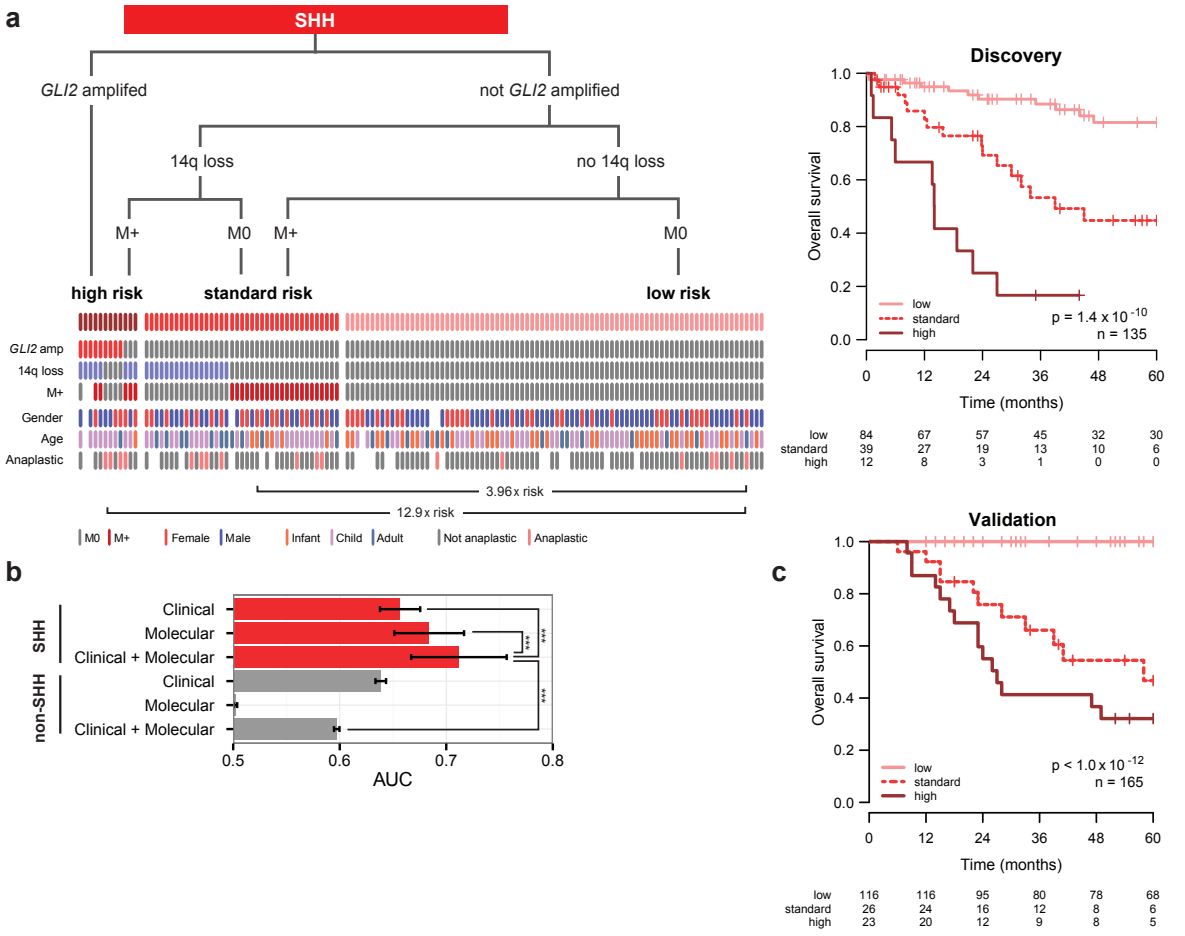

Figure 3.10: Combined clinical and molecular biomarkers improve risk-stratification of SHH patients. **a**, Risk stratification of SHH medulloblastomas by molecular and clinical prognostic markers. *Top-left*, decision tree; *bottom-left*, events plot depicting status of molecular and clinical markers across the risk groups; *right*, overall survival curves for SHH risk groups. **b**, Average time-dependent AUCs for risk groups stratified using only clinical or molecular markers, or both. Risk stratification regimens are applied to SHH and non-SHH medulloblastomas. ***, $p < 0.001$, Friedman rank sum tests. **c**, Survival curves for SHH risk groups in the validation cohort. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

Three biomarkers demarcate high-risk Group3 patients

In Group3 patients, iso17q and *MYC* amplification remain the only cytogenetic markers associated with poor survival, whereas chr8q loss and chr1q gain are the only good prognosis markers (**Figure 3.11**)[1]. In multivariate survival analyses, patients with metastasis, iso17q, or MYC amplification represent the high-risk group (**Figure 3.12a**). Critically, absence of these markers can identify a population of Group3 patients who have a survivorship much longer than Group3 taken as a whole. The risk groups are not associated with any clinical covariates, including age (**Figure 3.12a**)[1]. Consistent with the findings in SHH patients, optimal risk stratification in Group3 patients requires the use of both clinical and molecular prognostic markers, which have reduced or no prognostic value outside of Group3 (**Figure 3.12b**)[1]. Our proposed risk stratification scheme was validated on the non-overlapping validation cohort using three molecular biomarkers (*MYC*, 17p, and 17q FISH) and metastatic status (**Figure 3.12c**).

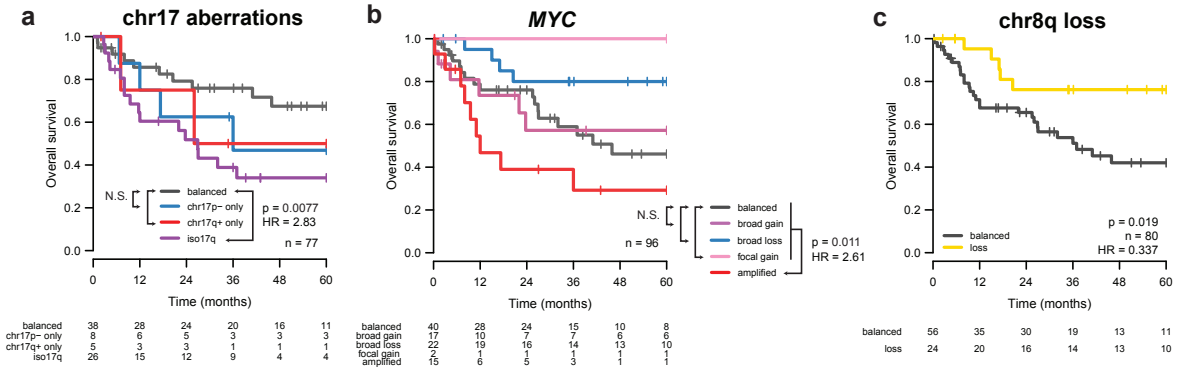

Figure 3.11: Overall survival curves for molecular biomarkers in Group3 medulloblastoma: **a**, chr17 copy number aberrations; **b**, *MYC* copy number status; and **c**, chr8q status. **d**, Risk stratification of Group3 medulloblastomas by molecular and clinical prognostic markers. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

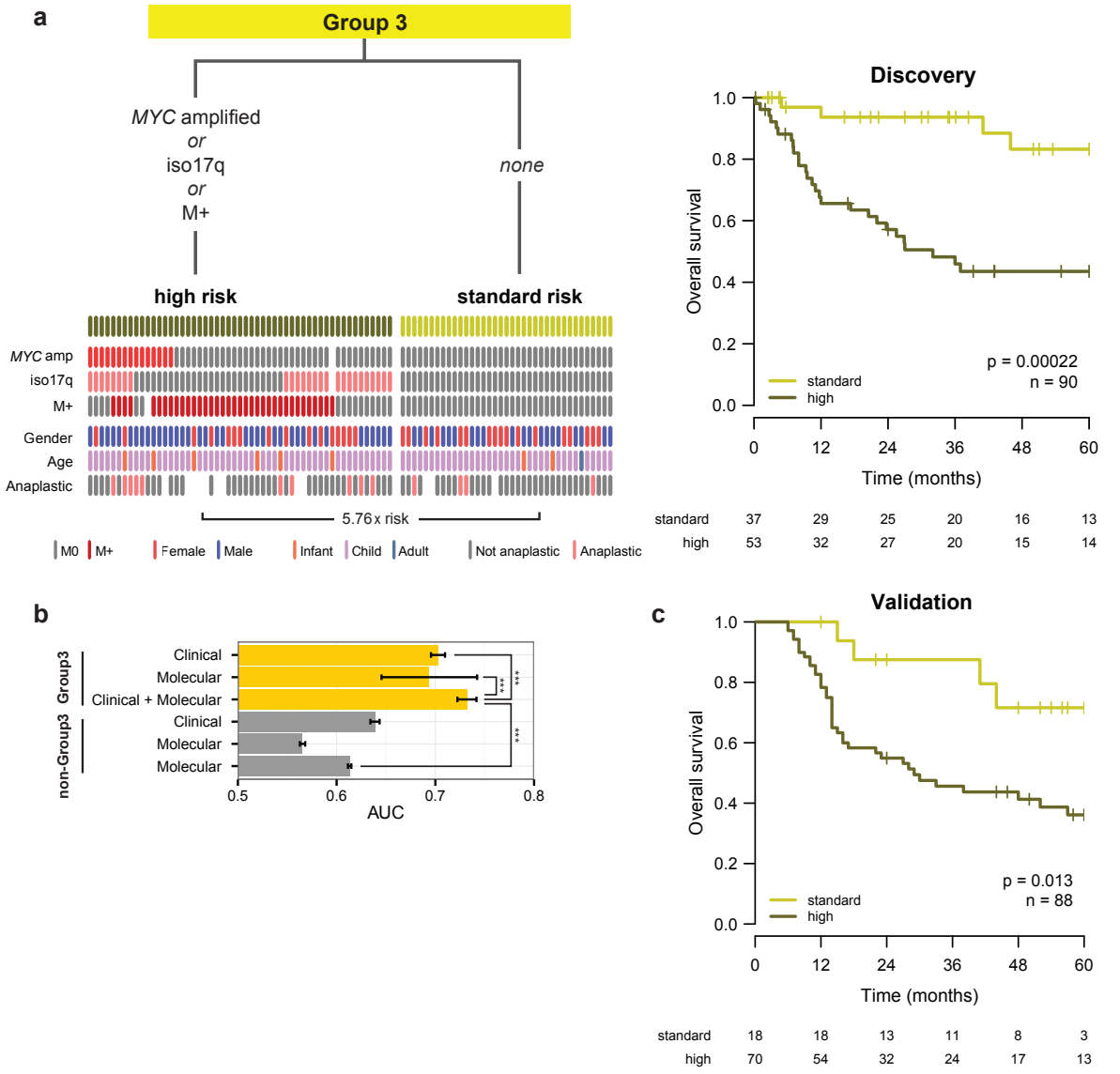

Figure 3.12: Combined clinical and molecular biomarkers improve risk-stratification of Group3 patients. **a**, Risk stratification of Group3 medulloblastomas by molecular and clinical prognostic markers. *Top-left*, decision tree; *bottom-left*, events plot depicting status of molecular and clinical markers across the risk groups; *right*, overall survival curves for Group3 risk groups. **b**, Average time-dependent AUCs for risk groups stratified using only clinical or molecular markers, or both. Risk stratification regimens are applied to Group3 and non-Group3 medulloblastomas. ***, $p < 0.001$, Friedman rank sum tests. **c**, Survival curves for Group3 risk groups in the validation cohort. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

Identification of a low-risk group of metastatic Group4 patients

Group4 patients with whole chromosome loss of chr11 or gain of chr17 exhibit better survival under univariate Cox models (**Figure 3.13a**), in addition to chr10p loss[1]. There is no cytogenetic marker associated with poor prognosis[1]. Specifically, neither *MYCN* gain nor amplification is associated with poorer survival in Group4, in stark contrast to SHH patients, reinforcing the distinction in their underlying biology (**Figure 3.13b**)[1]. Similarly, none of the cytogenetic biomarkers identified for Group3 patients (e.g. iso17q) have any prognostic value in Group4[1]. Following unbiased model selection, the consensus set of biomarkers results in a risk stratification scheme in which leptomeningeal dissemination identifies high-risk Group4 patients, except in the context of chr11 loss or chr17 gain (**Figure 3.14a**). The biology underlying chr11 loss is not apparent as there is no obvious minimal common region of deletion, nor are there any recurrent SNVs on chr11 reported in the recent medulloblastoma re-sequencing publications[1]. Group4 patients with either chr17 gain or chr11 loss, irrespective of their metastatic statuses, exhibit survivorship that is characteristic of WNT patients in both the discovery and validation cohorts (**Figure 3.14a**,**c**), and the survival differences are not explainable by covariates[1]. Significantly, the low-risk Group4 cohort also included some patients with anaplastic histology, suggesting that anaplasia may not be universally prognostic for poor outcome. Consistent with other subgroups, the risk stratification model using both clinical and molecular biomarkers achieve the highest accuracy (**Figure 3.14b**). Critically, the cytogenetic biomarkers identify low-risk Group4 patients whom would be otherwise designated as high-risk by evidence of metastasis and/or anaplastic histology, and this finding cannot be extrapolated to SHH and Group3 patients (**Figure 3.14**). We conclude that through the use of three molecular biomarkers (chr11, 17p, and 17q FISH) and metastatic status, we can accurately and reliably predict the survival of patients with Group4 medulloblastoma.

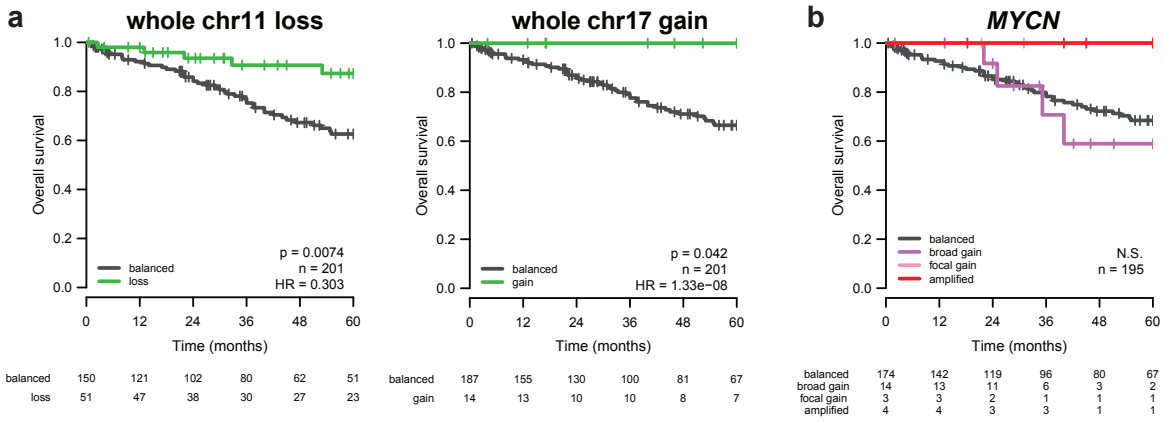

Figure 3.13: Overall survival curves for molecular biomarkers in Group4 medulloblastoma: **a**, whole chr11 status and whole chr17 status; and **b**, *MYCN* copy number status. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

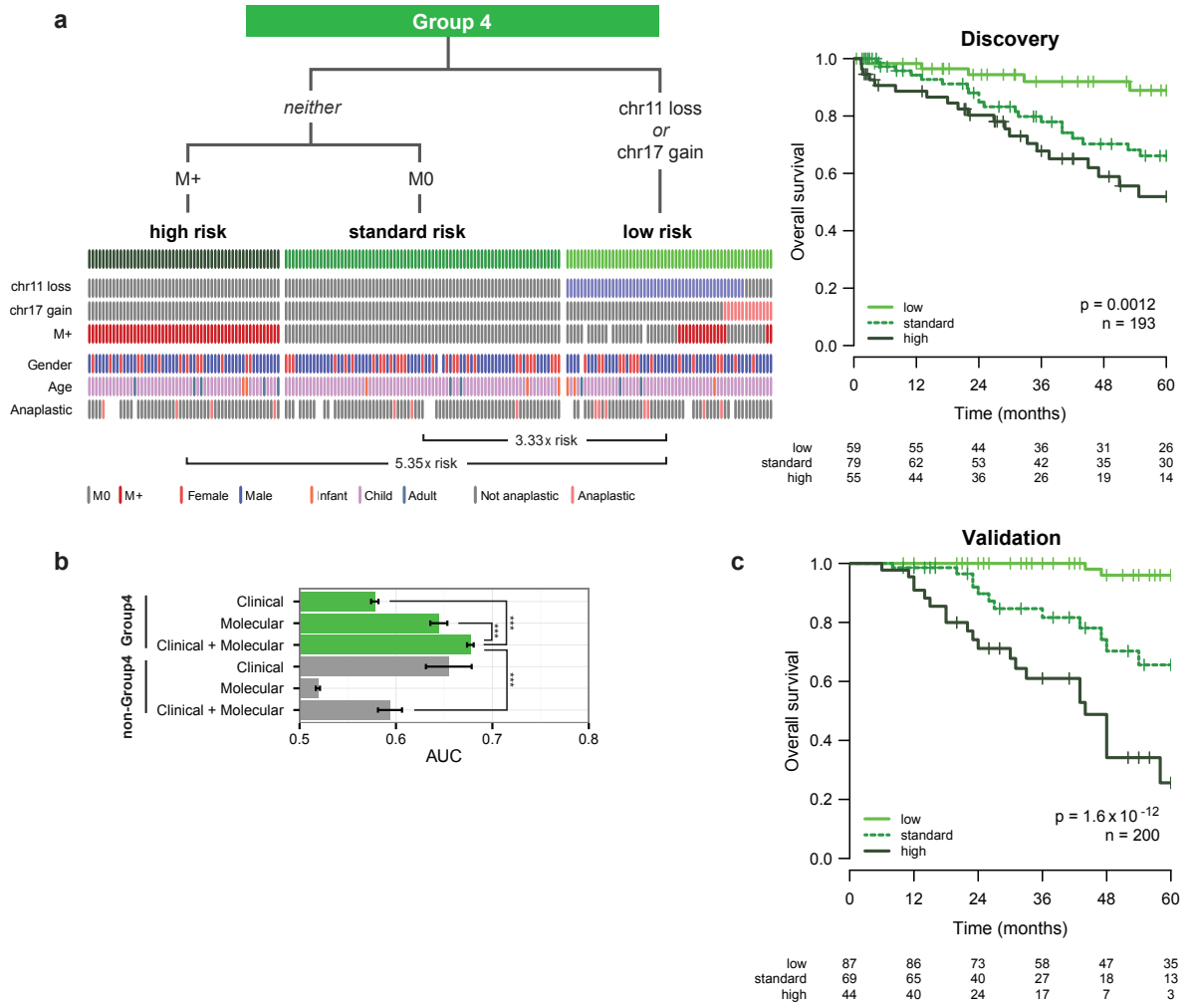

Figure 3.14: Combined clinical and molecular biomarkers improve risk-stratification of Group4 patients. **a**, Risk stratification of Group4 medulloblastomas by molecular and clinical prognostic markers. *Top-left*, decision tree; *bottom-left*, events plot depicting status of molecular and clinical markers across the risk groups; *right*, overall survival curves for Group4 risk groups. **b**, Average time-dependent AUCs for risk groups stratified using only clinical or molecular markers, or both. Risk stratification regimens are applied to Group4 and non-Group4 medulloblastomas. ***, $p < 0.001$, Friedman rank sum tests. **c**, Survival curves for Group4 risk groups in the validation cohort. Numbers below x-axis represent patients at risk of event; statistical significances are evaluated by log-rank tests; HR estimates are derived from Cox proportional-hazards analyses.

III Discussion

The analysis of > 1000 medulloblastoma patients clearly demonstrates that subgroup affiliation enhances prognostication with clinical biomarkers and that the majority of published molecular biomarkers are only relevant in the setting of a single subgroup. The combination of clinical variables, molecular subgroup, and six cytogenetic markers analyzed on FFPE tissues can achieve an unprecedented level of prognostic prediction for medulloblastoma patients that is practical, reliable, and reproducible. The proposed risk stratification models represent those that best fit the available data in the discovery cohort. Despite the large size of our discovery cohort, missing data and the complexity of multivariate analyses may necessitate the use of even larger cohorts to assess the inclusion of additional prognostic markers. Moreover, while we strive to include the most important markers in multivariate models, we cannot exclude the possibility that alternative markers may perform equally well. Our results nonetheless elucidate the prognostic potential of known and novel markers and highlight clinically useful risk-stratification schemes.

The prognostic significance of M1 status (presence of cells in the cerebrospinal fluid) has long been controversial. Most reports agree that presence of metastasis portends poor prognosis and warrants intensified treatment[36;141;146;147;217;224–227]; however, it is unclear whether M1 disease has the same prognosis as M2/M3 (macroscopic metastasis). Kortmann *et al.* contended in a prospective trial that M2/M3 status were indicators of poor outcome in medulloblastoma, but residual disease or M1 status were not[228]. In another prospective trial, Zeltzer *et al.* maintained that both M1 and M2/M3 statuses were prognostically unfavourable[227]. In subsequent studies, some investigators group M0 and M1 together in one category[141;217], while others group M1, M2 and M3 together as M+[36;143;147;217]. In a retrospective review, Sanders *et al.* reported that M1 patients do not have better survival than M2/M3 patients under the same treatment[229]. In our cohorts, the prognostic significance of M1 disease may be subgroup specific, though the small sample size of M1 patients hinders a definitive conclusion[1]. Based on our data, it is unlikely that M1 status is a universal indicator of poor outcome. Nevertheless, irrespective of whether M1 was categorized with M0 or M2/M3, our risk-stratification models can reproducibly and robustly predict patient survival.

Controversy also surrounds the prognostic value of anaplastic histology. Large cell and anaplastic histologies are often grouped together because these histological features often co-occur and can be difficult to distinguish. Several studies reveal that large

cell/anaplastic histology is prognostically unfavourable[36;217;230;231]. von Hoff *et al.* distinguished between large cell and anaplastic histologies and reported that large cell histology was a negative prognostic factor while anaplasia was not[127]. The authors further suggested that *MYC* amplification, which co-occurred with large cell histology, may be the underlying cause of poor prognosis; however, the precise definition of *MYC* amplification remains contentious[127;232]. We clarify this issue by demonstrating that only high-level *MYC* amplification but not single copy gains of *MYC* (focal or broad) is prognostically significant. Additionally, large cell/anaplastic histology has no prognostic value in a multivariate model accounting for *MYC* amplification. Possibly, *MYC* amplification may be a marker for apoptotic resistance, leading to resistance against radiotherapy and chemotherapy. Although MYC promotes cell proliferation, it also normally induces apoptosis; therefore, *MYC* amplification is incompatible with tumour formation except in the context of apoptotic pathway disruption[233].

Another marker notably absent from our risk-stratification schemes is *TP53* mutation, which is a well-known indicator of poor prognosis[14]. Loss of TP53 function abrogates the apoptotic pathway and contributes to resistance against chemotherapy and radiotherapy. While we had some *TP53* mutation data in our cohorts, a substantial proportion of samples were not interrogated for *TP53* status. Additionally, *TP53* mutation appears to be predominately prognostic for long-term survivors[14], and the follow-up lengths in our cohorts were insufficient to evaluate the long-term prognostic impact of *TP53* mutation. Therefore, the utility of *TP53* mutation in multivariate patient risk-stratification should be further assessed in a cohort with more complete data and longer follow-up.

The absence of age in our risk stratification schemes bodes well for the modern practice of restricting the role of radiotherapy in the treatment of young children. Indeed, age groups are not associated with the risk groups defined by our models, and the survival differences among the risk groups cannot be explained by differences in age distributions. Furthermore, age (discretized or otherwise) has no prognostic value in multivariate survival models once such prognostic factors such as metastatic status is included[1]. The univariate significance of age for poor outcome may be explained by the notation that early age at diagnosis is simply a proxy for tumour aggressiveness. While some aggressive tumours arise early, not all early arising tumours are aggressive. Accordingly, age has little independent prognostic value in a multivariate survival model, despite that infants often receive less intensive treatment in contemporary protocols. Our results suggest that

elimination of radiotherapy and dose reduction in chemotherapy has not contributed to poorer survival of infants; rather, aggressive tumours that respond poorly to treatment sometimes present in infants.

Above all, our risk stratification models identify patient groups who are promising candidates for de-escalation or elimination of irradiation during treatment. In particular, WNT patients exhibit excellent long term survival; with careful monitoring, these patients may respond well to reduced radiotherapy with postsurgical chemotherapy. For SHH medulloblastoma, the finding that infants in the low-risk group under our model respond favourably to multimodal treatment (with presumably reduced or eliminated radiotherapy) points to the tantalizing possibility that the remaining patients in this low risk group (defined by absence of all unfavourable markers) may similarly respond well to chemotherapy alone. Given that SHH medulloblastoma tends not to recur with metastasis[11], localized radiotherapy may be sufficient to prevent recurrence. Among patients with Group4 medulloblastoma, some patients with metastatic disease show excellent survival. Since patients presenting with metastasis are traditionally considered high risk, their apparent favourable outcome in our cohorts begs the question: Did these patients need the intensified radiotherapy for tumour eradication? If their favourable survival is not attributable to intensified treatment, these patients may benefit from radiotherapy de-escalation and survive with improved qualities of life. Encouragingly, recent findings suggest that the dose of craniospinal irradiation might be reduced in high-risk medulloblastoma (metastatic or residual disease) without compromising survival by supplementing the treatment with tandem high dose chemotherapy (and autologous stem cell transplantation)[202].

To conclude, we demonstrate that medulloblastoma molecular subgroups are highly informative for predicting patient outcome, and we can dramatically improve the accuracy of survival prediction by incorporating molecular subgroup with conventional clinical parameters for patient risk stratification. Moreover, we proposed, tested, and validated novel subgroup-specific risk stratification models that consider both clinical and molecular variables. These models perform robustly and reproducibly both in the discovery cohort consisting of a heterogeneously treated group of patients and in a large, non-overlapping validation cohort of patients treated at a single institution according to a single treatment protocol. Given that we do not have detailed treatment information for patients in these cohorts, it is highly possible that treatment effects (type, duration, or intensity) could impact our results. We would suggest that this can only be accounted

through examination of our stratification model in a sufficiently large prospectively followed cohort of patients with medulloblastoma. While the current study uses either SNP arrays, or interphase FISH on FFPE sections, it is possible that other approaches such as array comparative genomic hybridization (aCGH) could also be used to determine the copy-number status of the six cytogenetic markers. Our results demonstrate the utility of incorporating tumour biology into clinical decision-making and offer a novel perspective on risk stratification using FISH (applicable on paraffin sections), and thus should be validated in prospective multi-centre trials and be translated into routine clinical practice.

Chapter 4

Discovering therapeutic targets by genomic profiling of medulloblastoma

Objective. We hypothesize that each medulloblastoma molecular subgroup is characterized by specific genomic alterations, and we aim to identify potential therapeutic targets specific to each molecular subgroup.

Increased awareness of long-term neurotoxcities of craniospinal irradiation has motivated the de-escalation of radiotherapy by incorporating dose-intensive combination chemotherpay into the treatment regime. In order to eradicate tumour cells, chemotherapy is often intensified to the point where autologous stem cell transplants are required to circumvent patient mortality. Albeit less harmful, chemotherapy is not without neurocognitive and functional consequences for survivors. In one trial, addition of chemotherapy to craniospinal irradiation caused long-term survivors of medulloblastoma to suffer poorer health status[234]. In another trial, patients with medulloblastoma who were treated with chemotherapy alone (vincristine, carboplatin, etoposide, cyclophosphamide, and methotrexate) suffered from declines in neurocognitive function[141]. Perhaps part of the decline could be due to the disease itself; however, even for patients with acute lymphoplastic leukemia, chemotherapy alone (vincristine, methotrexate, prednisone, and antracyclines) can cause reduction in volumes of several neuroanatomic structures of the brain and consequent decline in processing speed, executive function, learning and memory[169]. Furthermore, chemotherapy causes a myarid of other side-effects; for example, it increases the risk of infection for patients due to suppression of immune system. Chemotherapeutic drugs such as cisplatin can cause permanent hearing loss and peripheral neurotoxicity[170].

Aside from mounting evidence for long-term adverse effects, combination chemotherapy encountered yet another setback when a prospective trial reported that prolonged dose-intensive chemotherapy (with cisplatin, cyclophosphamide, vincristine, etoposide) for patients with medulloblastoma yielded no improvement in survival (and also resulted in several treatment-associated deaths)[143]. In this trial, the authors treated the patients with 72 weeks of dose-intensity combination therapy, and patients with medulloblastoma showed no improvement in survival (though patients with ependymoma did)[143]. Similar to prior attempts, dose-intensive chemotherapy also caused patient death (toxic mortality) about 10 times more frequently than standard dose[143;209]. These negative findings indicate that patient survival cannot be improved by simply increasing the dose of chemotherapy, whose arsenal has essentially remained unchanged for decades.

To faciliate the discovery of novel therapeutic targets, we sought to identify recurrently dysregulated genes and pathways by analyzing the DNA copy-number profiles of > 1000 primary medulloblastoma samples (before radiotherapy and chemotherapy). Genes amplified recurrently across tumours are candidate proto-oncogenes, against which therapeutic agents may be developed. Conversely, genes deleted recurrently are candidate tumour suppressors, and therapeutic interventions may be designed against downstream pathways. Earlier studies aimed at identifying recurrent copy-number aberration (CNA) or mutations did not identify any highly recurrent focal genetic lesions across medulloblastoma tumours (without classification into subtypes): most focal genetic lesions occur at frequency of less than 10%, with *MYC* amplification and *PTCH1* mutation representing the most commonly observed events in medulloblastoma[83;235]. These observations suggest that medulloblastoma is genetically heterogeneous, and classifying medulloblastoma into molecular subgroups may aid the identification of genes and pathways that are recurrently disrupted above the background mutation rate in medulloblastoma.

The four molecular subgroups of medulloblastoma exhibit activation of different transcriptional progrms, which may be shaped by different genetic abnormalities. We therefore stratify our medulloblastoma samples by subgroup status into more homogeneous groups, and we sought to identify recurrent CNS enriched in or specific to each molecular subgroup. We hypothesize that each medulloblastoma subgroup incurs different genetic events. Accordingly, we hope to reveal dysregulated genes and pathways underlying the tumourigenesis of each subgroup in order to facilitate the development of precise, targeted therapeutic interventions.

I Materials and methods

Patient samples and nucleic acid extraction

All patient samples were procured in accordance with the Research Ethics Board at The Hospital for Sick Children (Toronto, Canada). Samples were obtained as frozen tissue biopsies at the time of diagnosis and stored at -80 °C until processed for purification of nucleic acids. Frozen tissue was available for 75-80% of cases included in the study; the remaining cases were shipped as pre-isolated DNA and/or RNA. Whenever possible, tumour isolates were partitioned for both standard DNA and RNA extraction. Tissues were either manually homogenized using a mortar and pestle in the presence of liquid nitrogen or in an automated manner using a Precellys 24 tissue homogenizer (Bertin Technologies, France), according to the manufacturers instructions. High molecular weight DNA was extracted by SDS/Proteinase K digestion followed by 2-3 phenol extractions and ethanol precipitation. Total RNA was isolated using the Trizol method (Invitrogen, USA) using standard protocols. DNA and RNA were quantified using a NanoDrop 1000 instrument (Thermo Scientific, USA) and integrity assessed either by agarose gel electrophoresis (DNA) or Agilent 2100 Bioanalyzer (RNA; Agilent, USA) at The Centre for Applied Genomics (TCAG, Toronto, Canada). RNA with an RNA Integrity Number (RIN) ≥ 7.0 was required for analysis by either Affymetrix Gene array or RNASeq. Paul Northcott performed the nucleic acid extractions.

DNA copy number analysis

SNP array processing and quality control

Genotyping and copy-number profiling of DNA samples was performed on the Affymetrix Genome-Wide Human SNP Array 6.0 platform, which includes more than 906 600 probes for the genotyping of SNP loci and more than 946 000 probes for the detection of copy number variations. With a total of 1.8 million probe markers, the median distance between markers is less than 7000 bases. DNA was prepared, labeled, and hybridized to the Affymetrix SNP 6.0 arrays as previously described[2]. Sample quality control was assessed using Affymetrix Genotyping Console as previously described[2]. TCAG performed the SNP array processing and quality control.

Generation and normalization of copy number profiles

Affymetrix SNP6 CEL files were processed in dChip[236] to obtain raw copy number estimates. Arrays were normalized by quantile normalization and signal intensities computed using the MBEI method (PM-only). To generate a diploid reference baseline for copy number analysis of medulloblastoma samples, we used Affymetrix SNP6 data from 132 individuals from the Ontario Population Genomics Platform epidemiological project and the HapMap project. Germline DNA from these samples was genotyped in the same microarray facility as the tumour samples, using identical experimental protocols as described above.

The normalized copy number estimates from dChip were subsequently imported into the R environment, and the copy number profiles were segmented using the circular binary segmentation (CBS) algorithm from the DNAcopy (v1.24) package[237], with the undo split option enabled (`method = sdundo, undo.SD = 1`). The resulting segmentation profiles were further processed to reduce artificial segments. Segments with fewer than 10 markers were removed. Adjacent segments whose copy number states differed by less than 0.25 were merged together using their size-weighted mean. This merging step is repeated iteratively until no more segments can be merged. The segmentation profile of each sample was then median-centered. Further, normal CNVs reported in the Database of Genomic Variants (DGV, v10)[238] were filtered from the segmentation profiles. Segments that reciprocally overlapped (Dice coefficient > 0.5) with normal CNVs were removed. CNVs reported on BAC End Sequencing, BAC Array CGH, ROMA, and FISH were excluded from this filter. Upon removing a segment, the upstream segment was merged to the downstream segment using a size-weighted mean. The aforementioned merging and CNV-filtering steps helped reduce the occurrence of broad segments being broken into non-contiguous pieces by artificial segments or normal CNVs (which influences the downstream GISTIC2 analysis and segment classification). The copy number segments were classified as balanced or one of 6 copy number aberrations based on the criteria in **Table 4.1**.

Table 4.1: Criteria for DNA copy number aberrations

Class	**Log_2 R ratio (r)**	**Size in Mbp (s)**
Balanced	$\|r\| \leq 0.2$	—
Gain	$r > 0.2$	—
Loss	$r < -0.2$	—
Focal gain	$r > 0.2$	$s < 12$
Focal loss	$r < -0.2$	$s < 12$
High level amplification	$r > log_2(5/2)$	$s < 12$
Homozygous deletion	$r < -log_2(0.7/2)$	$s < 12$

Identification of recurrent copy number aberrations

The post-processed segmentation files profiles were analyzed using two algorithms, GISTIC2 and modified CMDS, to identify recurrent copy number events. GISTIC2 requires prior single-sample segmentation, and hence may be affected by the presence of segmentation artifacts. In contrast, CMDS requires no single-sample segmentation and works with the raw copy number profiles directly. Many post-processing steps were lacking from the distributed CMDS program (e.g. multiple hypothesis correction, peak calling, CNV filtering); therefore, a number of post-processing steps were added.

GISTIC2 (v2.0.12)[239] was run with default parameters unless specified otherwise (`brlen = 0.5, conf = 0.9`), on the segmentation profiles of the entire cohort and of each subgroup separately. The significant peaks were filtered based on the gene content ($ge1$ gene spanned), size of the wide peak (< 12 Mbp), and additionally based on containment within a normal CNV region (one-way overlap $> 90\%$) reported in the filtered DGV database. This second CNV-filtering step ensures that amalgamated regions reported in DGV are considered, as these regions would often have poor reciprocal overlap with individual query segments and would be missed by the previous CNV-filtering procedure. It is not possible to apply this secondary CNV-filtering procedure directly on the segmentation file, as the CNV regions are often very large and this filtering can cause the loss of many informative CNA segments.

CMDS (v1.0)[240] was run using default parameters unless specified otherwise (`w = 40, s = 1`), on the unstratified and subgroup-stratified segmentation profiles, for each chromosome separately. Since the raw outputs had a high false positive rate, the outputs

were processed further. The z-scores were recalculated (separately for each chromosome) from the Fishers z-transformed of the Pearson correlations, using the means of the dominant components from Gaussian mixture models ($k = 3$). This recalculation ensured that the z-scores were not skewed due to the presence of multiple modes in the raw output. The z-scores were further detrended using empirical mode decomposition (EMD)[241], which corrects for trends due to recurring broad events. The p-values derived from the z-scores were then corrected for multiple hypothesis correction using the qvalue R package (v1.1)[242]. Finally, the peak regions were identified using a simple q-value threshold ($FDR = 0.05$).

Estimation of the copy number states of genes

The copy number states of all RefSeq genes in each sample were inferred from the respective post-processed segmentation profiles. The state for a given gene was determined by the copy number segment (classified as described earlier) that spans the greatest proportion of the gene (if multiple segments span the gene). Further, a gene is considered lost or deleted if any portion of its coding region is spanned by a loss or deletion segment; a gene is considered gained or amplified if at least 50% of the coding region is spanned by a gained or amplified segment.

Identification of recurrent broad events

Identification of recurrent CNAs above explicitly excluded broad events (based on a broad length cutoff in GISTIC and by a detrending procedure after CMDS). Therefore, broad events in the segmentation profiles were analyzed separately, using an approach similar to GISTIC's broad event analysis[239]. The log2 R ratio (LRR) of each chromosome was calculated using a size-weighted mean of all segments mapping to the chromosome. A chromosome was declared gained if its LRR was greater than 0.2, lost if the LRR was less than -0.2, and balanced otherwise. Unlike GISTIC, gained and lost broad events were analyzed together. The significance of the frequency of each broad event was tested using the exact binomial test. Each broad event frequency was compared to the background frequency, which was determined from a robust regression of the observed frequencies with respect to gene content (i.e. number of RefSeq genes) across all chromosomes.

Identification of chromothripsis

The occurrence of chromothripsis was identified using tumour copy number profiles as previously described[20]. Using the post-processed segmentation profiles, chromothripsis was identified on a chromosome based on the presence of greater than 10 copy number state changes. The enrichment of chromothripsis on a chromosome for a subgroup was determined using the hypergeometric test, comparing the observed incidence to the background incidence (tallied across all samples). Select samples inferred to have chromothripsis were confirmed by whole-genome sequencing to identify rearrangements.

Subgroup enrichment analysis of recurrent copy-number events

Recurrent SCNAs identified by GISTIC2 in the unstratified and subgroup-stratified analysis were tested for subgroup enrichment. In the case of stratified analysis, regions identified in each subgroup were combined together. Since the reported region coordinates of common SCNAs events can differ among the strata, regions that reciprocally overlap (Dice coefficient > 0.2) were merged together using their union. In the enrichment analysis, the frequency of recurrent SCNAs in each subgroup was compared against the remaining subgroups, and odds ratios were estimated using Fisher's conditional maximum likelihood estimate. SCNAs with odds ratios greater than 2.0 were considered subgroup-enriched. A similar enrichment analysis was repeated for comparing combinations of two subgroups against the remaining, in order to identify SCNAs that are enriched in the ensemble of two subgroups (and are not otherwise enriched in a single subgroup).

Integration of gene expression and copy-number events

To directly assess the correlation between significant somatic copy-number aberration (SCNA) and gene expression, expression profiles were generated on 285 medulloblastomas from our study, including samples from SHH ($n = 51$), Group3 ($n = 46$), and Group4 ($n = 188$), using the Affymetrix Gene 1.1 ST platform. Global integration of expression data was performed by comparing expression levels of amplified or deleted genes relative to genes in balanced regions (Mann-Whitney tests). Further, specific integration of expression data at each significant SCNA locus was done by comparing expression of genes in samples with the aberration against those that do not, within each medulloblas-

toma subgroup. Multiple hypothesis correction by the false discovery rate method was applied to each locus independently, and the false discovery rate threshold was adaptively tuned at each locus so that no false positives are expected. The resulting lists of genes with copy-number driven expression were used in candidate identification.

Identification of candidate driver genes in each significant region

Many of the significant regions identified using GISTIC and CMDS span multiple genes. Therefore, multiple lines of evidence were used to prioritize putative target genes within each region. Evidences collected from integrated expression data, the literature, and multiple datasets were classified into the tiers in **Table 4.2**.

Table 4.2: Tiered evidence-based framework for identifying candidate driver genes

Type	Description	Tier
Correlated expression	Gene expression is driven by SCNA in the integrated analysis	I
Medulloblastoma literature	Implicated in medulloblastoma based on the literature (PubMed)	I
Cancer Gene Census	Documented in the Cancer Gene Census	I
Parsons	Reported to be somatically mutated in the Parsons *et al.*[235] study on SNVs identified in medulloblastoma	III
ICGC	Reported to be somatically mutated in the ICGC medulloblastoma study	III
Northcott signature gene	Reported in the Northcott *et al.*[35] medulloblastoma expression study	IV
Cho signature gene	Reported in the Cho *et al.*[38] medulloblastoma expression subgroup study	IV
RNASeq SNV	Identified in the pilot RNASeq screen for SNV (Group 3 and 4 only)	IV
Gli1 target	Identified as a Gli1 target in Lee *et al.*[243] (SHH subgroup only)	IV
Shh-inducible target	Identified as a Shh-inducible gene by a screen in cerebellar granule neuron precursors (SHH subgroup only)	IV
COSMIC	Documented in the Catalogue of Somatic Mutations in Cancer	V

The priority of each gene in a region was ranked based on the total score of the above lines of evidence, weighted by their respective tiers. A higher tier (lower number) is assigned a higher weight. The weights were assigned so that supporting evidence from the next tier is only considered for breaking ties at previous tiers. At most two genes from each region were selected for network analysis using this evidence-based ranking.

Mutual exclusivity analysis

The significant gene lists were analyzed to detect mutual exclusive relationships by iterating through all possible subsets of genes in the list (for set sizes from 2 to 6). Within each subgroup, sets of genes with the highest total exclusivity scores were identified. We define the total exclusivity score as the fraction of all samples that harbour exactly one aberration among the genes in the set; in other words, the total exclusivity score is the product of proportion coverage and proportion exclusivity as defined by Miller *et al*[244]. Proportion coverage is the proportion of samples that contain at least one aberration within a given set of genes, and proportion exclusivity is the proportion of covered samples that contain exactly one aberration within the set of genes. For a set of genes $\mathbb{G}$ and samples $\mathbb{S}$, we designate total exclusivity as $TExcl(\mathbb{G})$, proportion exclusivity as $Excl(\mathbb{G})$, and proportion coverage as $Cover(\mathbb{G})$; then,

$$TExcl(\mathbb{G}) = \frac{\sum_{s \in \mathbb{S}} I\left(\sum_{g \in \mathbb{G}} \mathbf{M}[g,s] \;=\; 1\right)}{|\mathbb{S}|}$$

$$Cover(\mathbb{G}) = \frac{\sum_{s \in \mathbb{S}} I\left(\sum_{g \in \mathbb{G}} \mathbf{M}[g,s] \;>\; 0\right)}{|\mathbb{S}|}$$

$$Excl(\mathbb{G}) = \frac{\sum_{s \in \mathbb{S}} I\left(\sum_{g \in \mathbb{G}} \mathbf{M}[g,s] \;=\; 1\right)}{\sum_{s \in \mathbb{S}} I\left(\sum_{s \in \mathbb{G}} \mathbf{M}[g,s] \;>\; 0\right)}$$

where $I(\cdot)$ is the indicator function, and $\mathbf{M}[g,s] = 1$ if gene g is aberrant in sample s and $\mathbf{M}[g,s] = 0$ otherwise.

Network analysis

Pathway enrichment analysis of copy number aberrations was carried out using the g:Profiler web server[245] and visualized in Cyotscape as an Enrichment Map network[246]. Candidate driver genes from selected GISTIC2 regions were compiled for SHH, Group3, and Group4, ranked by frequency, and subsequently queried for significantly enriched functional categories with g:Profiler using the ordered list algorithm (FDR-corrected cutoff $p = 0.05$, hypergeometric test). Detected categories were filtered with a custom R script to include only Gene Ontology, Reactome, and KEGG pathway terms, using an upper limit of 500 genes per gene set and the requirement that at least two putative driver genes intersect with the gene set. A small number of non-informative KEGG

pathways were removed from the final list. The overlap coefficient value 0.6 was used in Enrichment map visualization. Enrichment maps were manually adjusted to highlight the most significant themes for visualization purposes. Jüri Reimand performed the network analysis.

Unsupervised clustering analysis of copy number events

All significant broad events (spanning chromosome arms) and focal events identified in the pan-cohort analysis were used in the agglomerative hierarchical clustering of medulloblastoma samples, by the Ward linkage method and the Euclidean distance metric, as implemented in the R environment. The copy number states were converted to absolute values and samples with unknown subgroup affiliation were removed prior to clustering. The agreement between the observed clusters and the medulloblastoma subgroups were assessed by the adjusted Rand index and tested by the χ^2 test.

Expression array processing and data analysis

For gene expression array profiling, 400 ng total RNA was processed and hybridized to the Affymetrix Gene 1.1 ST array at TCAG according to the manufacturers instructions. The CEL files were quantile normalized using Expression Console (v1.1.2; Affymetrix, USA) and signal estimates determined using the RMA algorithm. Prior to clustering analysis of Group 4 medulloblastomas, 500–1000 genes with the highest standard deviations were selected and the expression signals anti-log transformed. Unsupervised clustering was carried out using the NMFConsensus module available on the GenePattern public server (Broad Institute, USA) with default parameters[247].

NMF consensus clustering[247] generates clusterings for a range of predefined number of clusters, k. Clustering results with $k = 2$ up to $k = 5$ were generated, and the resulting fit of the NMF consensus clusters to the data for each value of k was assessed using the cophenetic correlation coefficient[248], and the value of k (the number of clusters) that led to the highest cophenetic correlation coefficient was chosen.

The cophenetic correlation coefficient is a measure of the degree to which cophenetic distances preserve the original distances among the observations, and it is most commonly used in hierarchical clustering. Given distances $d(x, y)$, as measured by some distance metric such as Euclidean distance, among all pairs of data points, a cophenetic

distance $c(x, y)$ between two observations x and y is the linkage distance at which the x and y are first merged into a single cluster. The linkage distance is defined by the linkage method used (e.g. average linkage, single linkage, complete linkage). For example, average linkage between two disjoint clusters A and B is

$$l(A, B) = \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b)$$

Accordingly, if $x \in A$ and $y \in B$, then $c(x, y) = l(A, B)$. Due to cluster merging, the cophenetic distance $c(x, y)$ and the original distance $d(x, y)$ may differ. The cophenetic correlation coefficient $\tilde{\rho}$ is the Pearson correlation $cor(\cdot)$ between $c(x, y)$ and $d(x, y)$ for all pairs of observations:

$$\tilde{\rho} = cor(\mathbf{c}, \mathbf{d})$$

where $\mathbf{c}$ is a vector of the cophenetic correlations between all pairs of observations x_i and x_j for $i < j$ and $\mathbf{d}$ is the corresponding vector of the original distances. Ordinarily, cophenetic correlation does not measure clustering stability; however, Brunet *et al.* consider cophenetic correlation to be a measure of clustering stability when hierarchical clustering is applied to a consensus clustering matrix. (Brunet *et al.*, in their NMF consensus clustering algorithm, generate this consensus clustering matrix by applying multiple rounds of non-negative matrix factorization on the data).

nanoString CodeSets and data analysis

To determine subgroup affiliation of the MAGIC cohort, a custom nanoString CodeSet was designed to assess the expression status of 22 medulloblastoma signature genes and samples processed as described previously[3]. Samples were processed as recommended by nanoString at the University Health Network Microarray Facility using an input of 100 ng total RNA. Raw nanoString counts were normalized as described in **Chapter 2**, and the normalized log2 expression data was used as input for class prediction analysis using the PAM algorithm. A series of 101 medulloblastomas with known subgroup affiliation were used as a training dataset for class prediction[35].

Statistical and bioinformatic analyses

Statistical and bioinformatic analyses were performed in the R statistical environment (v2.13) or using custom programs/scripts written in Python, C++, or Go. Enrichment analyses were done using the hypergeometric test. The significance of chromosome arm frequencies were done using the exact binomial test, comparing the observed frequency to the expected frequency derived from a robust regression of event frequency and gene content, in a manner similar to the broad analysis in GISTIC2. Comparisons of event frequencies across medulloblastoma subgroups were performed using Fisher's exact test. Expressions of genes across samples were compared using the Mann-Whitney test (and confirmed with the Student's independent t-test). Where applicable, multiple hypothesis corrections were performed using the false discovery rate method[242]. Univariate survival analyses were done using the log-rank test, as implemented in the survival R package (v2.36).

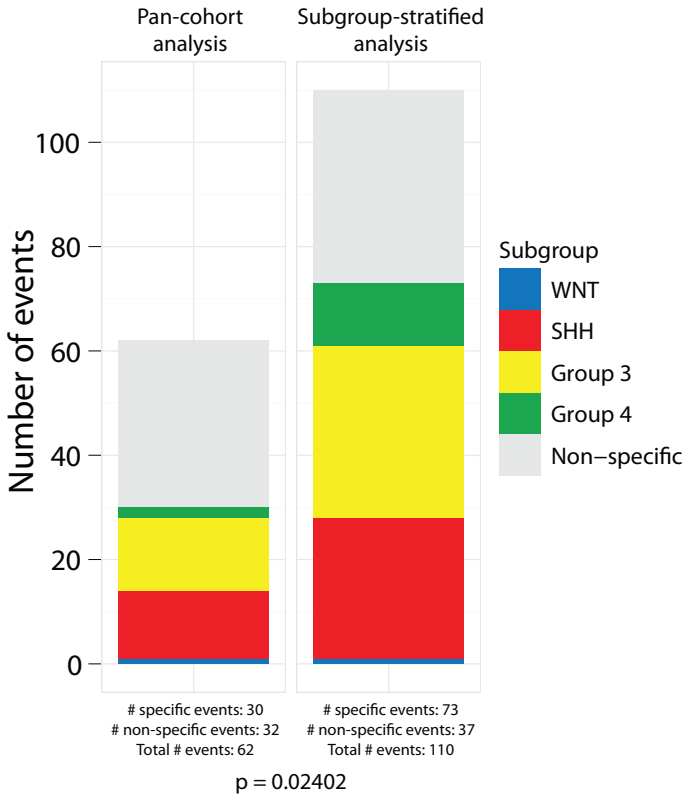

Figure 4.1: Significant regions of focal SCNA identified by GISTIC2 in pan-cohort or subgroup-stratified analyses. A total of 62 significant regions were identified when the cohort was analyzed as a single group, whereas 110 significant regions were captured when the cohort was analyzed according to subgroup. The number of significant subgroup-enriched regions identified more than doubled (73 vs. 30) when the subgroups were analyzed independently.

II Results

Molecular subgroups have disparate patterns of genomic events

Copy-number profiles were generated on $>$ 1200 medulloblastomas using the Affymetrix Genome-wide SNP6 platform. After quality control and clinical criteria filtering, copy-number profiles of 1087 primary medulloblastomas were available for further analysis in identifying SCNA events: regions of aberrant gains and losses in the tumour genome. The tumours were stratified based on molecular subgroups, as determined by the method described in **Chapter 2**. The copy-number and cytogenetic profiles of medulloblastoma subgroups were highly divergent, demonstrating that medulloblastoma subgroups are genomically heterogeneous[2]. Indeed, when the cohort was analyzed by each subgroup independently, an increased number of SCNAs were identified, many of which were subgroup-enriched (**Figure 4.1**).

Recurrent events target known cancer-associated genes

Among the recurrent high-level amplifications (copy-number ≥ 5) identified (**Figure 4.2**), the most prevalent events targeted members of the Myc family (*MYCN*, *MYC*, *MYCL*), with *MYCN* predominantly amplified in SHH and Group4, *MYC* in Group3, and *MYCL* in SHH medulloblastomas. The most common homozygous deletions targeted known tumour suppressors *PTEN*, *PTCH1*, and *CDKN2A/B*, all of which were enriched in SHH tumours (**Figure 4.3**). A selected set of genes were assessed using custom DNA copy-number nanoString assays, and 90.9% of events were verified (**Figure 4.4**). Additional genes were validated on external cohorts by FISH[2].

Chromothripsis is rare in WNT medulloblastoma

Chromothripsis (chromosome shattering) leads to co-occurrence of high-level amplifications and disruption of several genes localized to specific regions in one or multiple chromosomes. Several samples exhibited genomic aberrations reminiscent of chromothripsis, which has recently been implicated in cancer formation[249–254], as well as in medulloblastoma[20]. In our cohort of chromothripsis, the incidence of chromothripsis is not uniform across subgroups ($p = 0.015$, Fisher's exact test). While the incidence of chromothripsis is depleted in WNT tumours compared to non-WNT tumours ($p = 0.0028$), chromothripsis is neither enriched nor depleted in the SHH, Group3, or Group4 ($p > 0.05$). The incidence of chromothripsis is about 12% among SHH, Group3, and Group4 tumours. These findings are consistent with the observation that WNT medulloblastoma shows no recurrent SCNA other than chr6 loss and appear more genomically stable (**Figure 4.5**)[2].

Subgroup-specific events converge on oncogenic pathways

The disparate genomic landscapes of medulloblastoma subgroups lead to the identification of a multitude of focal SCNAs that characterize each molecular subgroup[2]. Novel genes identified in this study include: *PPM1D*, *PIK3C2B*, and *MDM4* in SHH (**Figure 4.6**); *ACVR2A*, *ACVR2B*, and *TGFBR1* in Group3 (**Figure 4.8**); and *NFKBIA* and *USP4* in Group4 (**Figure 4.10**). Conversely, WNT medulloblastoma have few recurrent SCNAs (**Figure 4.5**).

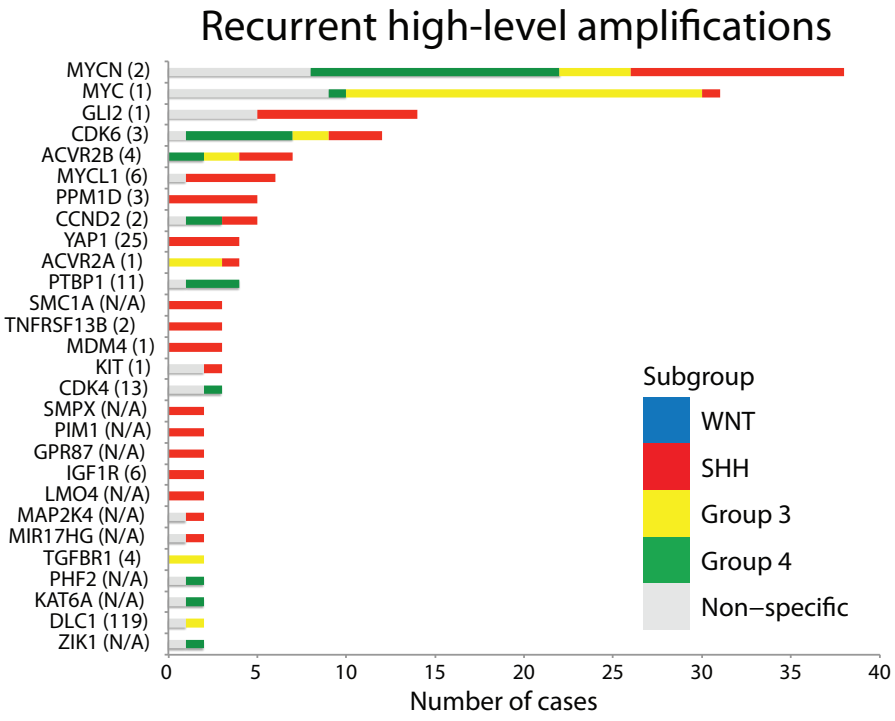

Figure 4.2: Recurrent high-level amplifications in medulloblastoma. Frequency of genes amplified (segmented copy-number ≥ 5) in at least two samples are shown with the distribution of the event across subgroups. The number of genes mapping to the peak region as defined by GISTIC2 (where applicable) are listed in parentheses after the candidate driver gene.

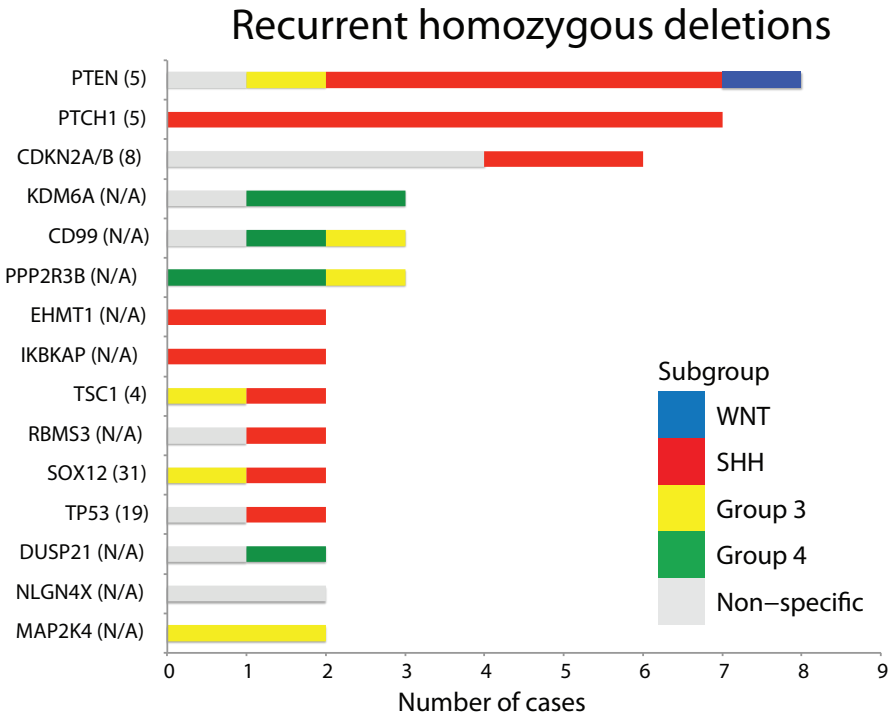

Figure 4.3: Recurrent homozygous deletions in medulloblastoma. Frequency of genes targeted by homozygous deletion (segmented copy-number ≤ 0.7) in at least two samples are shown.

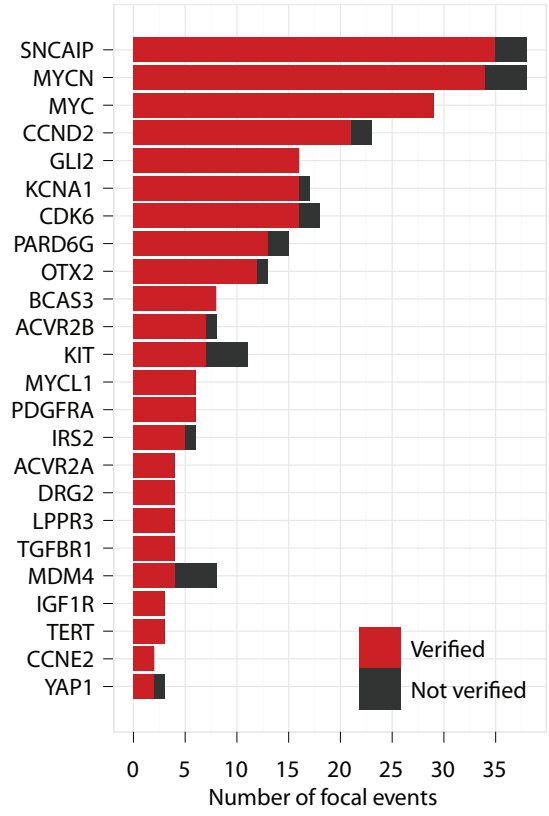

Figure 4.4: Verification of focal SCNAs by nanoString. Genes inferred to be focally amplified by SNP6 were interrogated using a custom nanoString CodeSet across a set of 192 medulloblastomas selected from our cohort. Bar-plot shows the number of samples for which each gene is verified (red) or not (black). An overall verification rate of 90.9% was achieved.

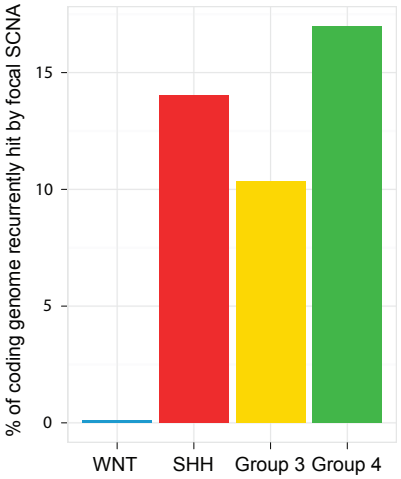

Figure 4.5: WNT medulloblastomas sustain a paucity of recurrent focal SCNAs. Bar-plots of the proportion of genome recurrently disrupted by focal SCNAs are depicted for each medulloblastoma subgroup.

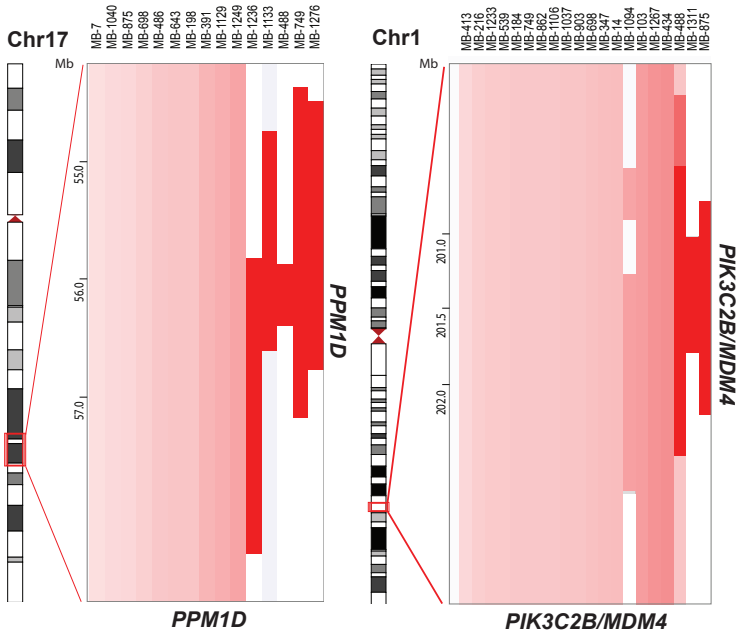

Figure 4.6: Recurrent high-level amplifications of *PPMID* and co-amplification of *MDM4* and *PIK3C2B* in SHH medulloblastoma. Segmented copy-number tracks are shown for the amplified loci (17q23 and 1q23).

SHH medulloblastoma, which is characterized by activation of Shh signaling[34;35;37;38;128], exhibits frequent SCNAs in the Shh pathway. Genes involved in focal SCNAs amplifications are significantly associated with SHH medulloblastoma signatures genes[2], suggesting that copy-number changes contribute in part to the altered expression signatures previously observed in SHH tumours. Accordingly, positive regulators of Shh signaling (*MYCN* and *GLI2*) were recurrently amplified, while a negative regulator of Shh signaling (*PTCH1*) was recurrently lost. Consistent with their functions in the same pathway, these events were mutually exclusive; however, they lead to different clinical outcomes[2]. In additional to Shh signaling, other core pathways recurrently disrupted in SHH medulloblastoma are TP53 signaling and PI3K signaling (**Figure 4.7**).

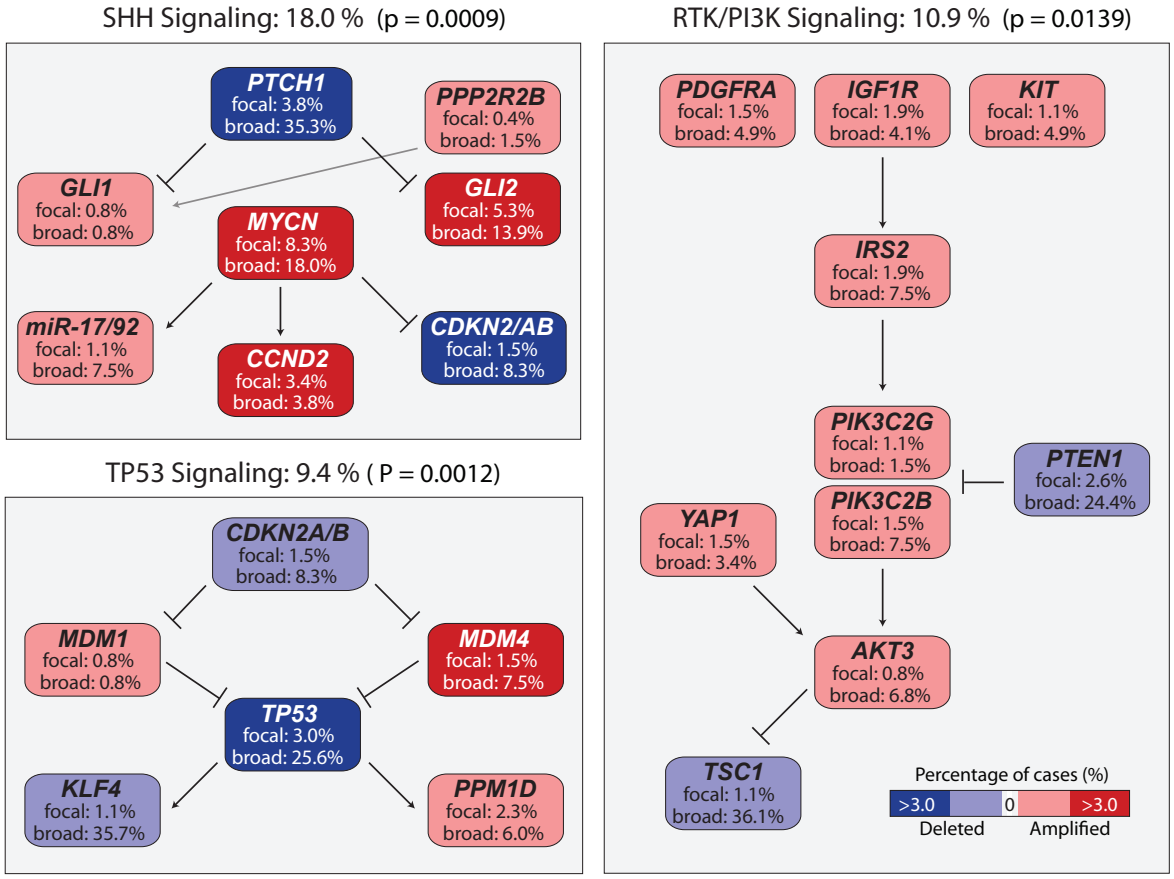

Figure 4.7: Core pathways genetically disrupted in SHH medulloblastoma. Summary of SCNAs affecting components of Shh signaling, TP53 signaling, and PI3K signaling are depicted. Colours reflect the frequency by which the respective genes are targeted by focal or broad events in SHH medulloblastomas (red for amplification, blue for deletion). Significance values indicate the prevalence with which each pathway is targeted in SHH vs. non-SHH cases (Fisher's exact test).

The signaling pathways involved in Group3 and Group4 medulloblastomas are less well understood, as suggested by their names. Nonetheless, at the copy-number level, distinct pathways are dysregulated in Group3 and Group4[2]. Group3 tumours are characterized by amplification of *MYC* and *OTX2*, which occur in a mutually exclusive pattern[2]. This observation is consistent with the tendency of the two oncogenic transcription factors to bind the same promoter regions[255]. Further, the TGF-β signaling pathway is frequently disrupted by SCNAs in Group3 (**Figure 4.9**). Conversely, the NF-κB pathway appear to be genetically targeted in Group4 medulloblastomas (**Figure 4.10**).

While *MYC* amplification is a known pivotal player for Group3, our data indicate that other genes in close proximity to the *MYC* locus may also play cooperative roles. This locus was frequently disrupted by a multitude of high-level amplicons[2] as well as massively genomic rearrangements termed chromothripsis (**Figure 4.12**). As a consequence of these events, the adjacent *PVT1* gene and miR-1204 are frequently co-amplified with *MYC*. Moreover, amplifications of the *MYC*/*PVT1* locus frequently result in the formation of fusion transcripts. Concurrent with MYC-PTV1 fusion expression, miR-1204 (hosted within PTV1) is upregulated[2]. *MYC*, *PVT1*, and miR-1204 all have been previously shown to play independent functional roles in other tumours[256–259], and may synergistically promote tumourigenesis.

The most prevalent focal gain in Group4 is the somatic tandem duplication of the *SNCAIP* gene[2]. SNCAIP expression is highly elevated in Group4 medulloblastomas (**Figure 4.13**). In fact, *SNCAIP* duplication is restricted to the Group4α subtype and may play a functional role in this medulloblastoma subtype (**Figure 4.14**).

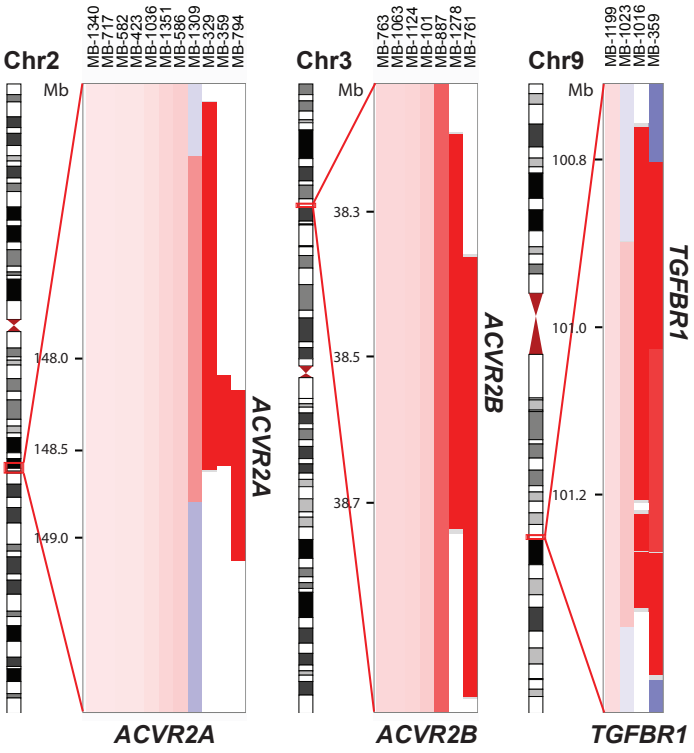

Figure 4.8: Recurrent amplifications target receptors of the TGFβ superfamily in Group3. Segmented copy-number tracks of Group3 medulloblastomas show recurrent high-level amplifications affecting *ACVR2A* (2q22), *ACVR2B* (3p22), and *TGFBR1* (9q22).

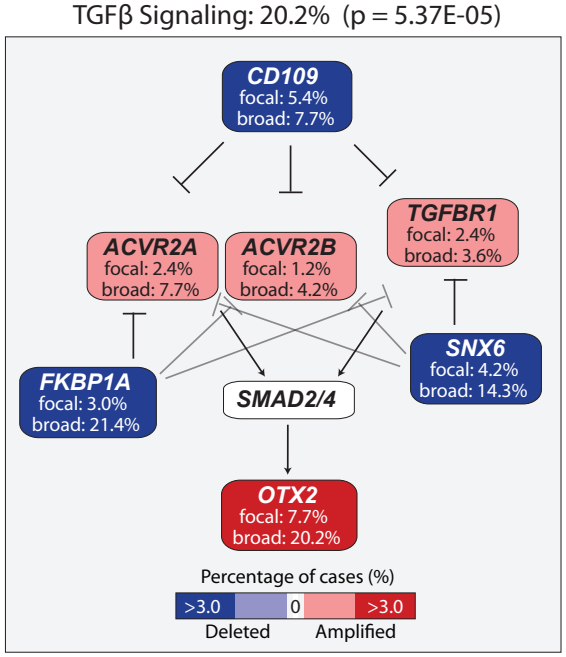

Figure 4.9: TGF-β signaling is recurrently disrupted by SCNAs in Group3. SCNAs affecting the TGF-β pathway comprise 20.2% of Group3 cases and are significantly enriched in Group3 compared to non-Group3 cases (Fisher's exact test).

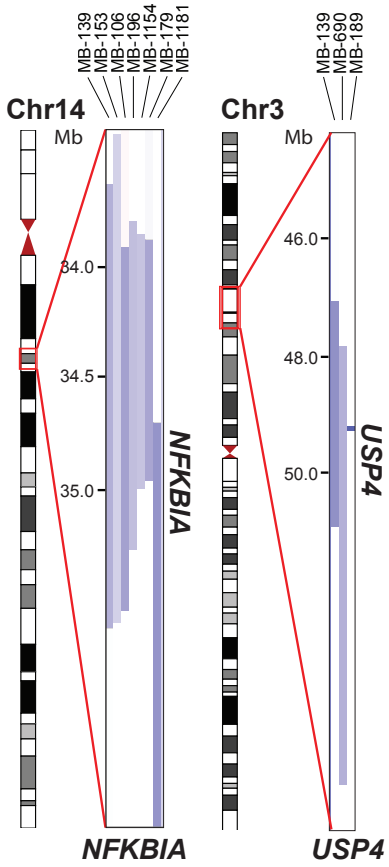

Figure 4.10: NF-κB pathway is recurrently disrupted in Group4. Recurrent focal deletions disrupt *NFKBIA* and *USP4*, negative regulators of the NF-κB pathway, in Group4 medulloblastoma.

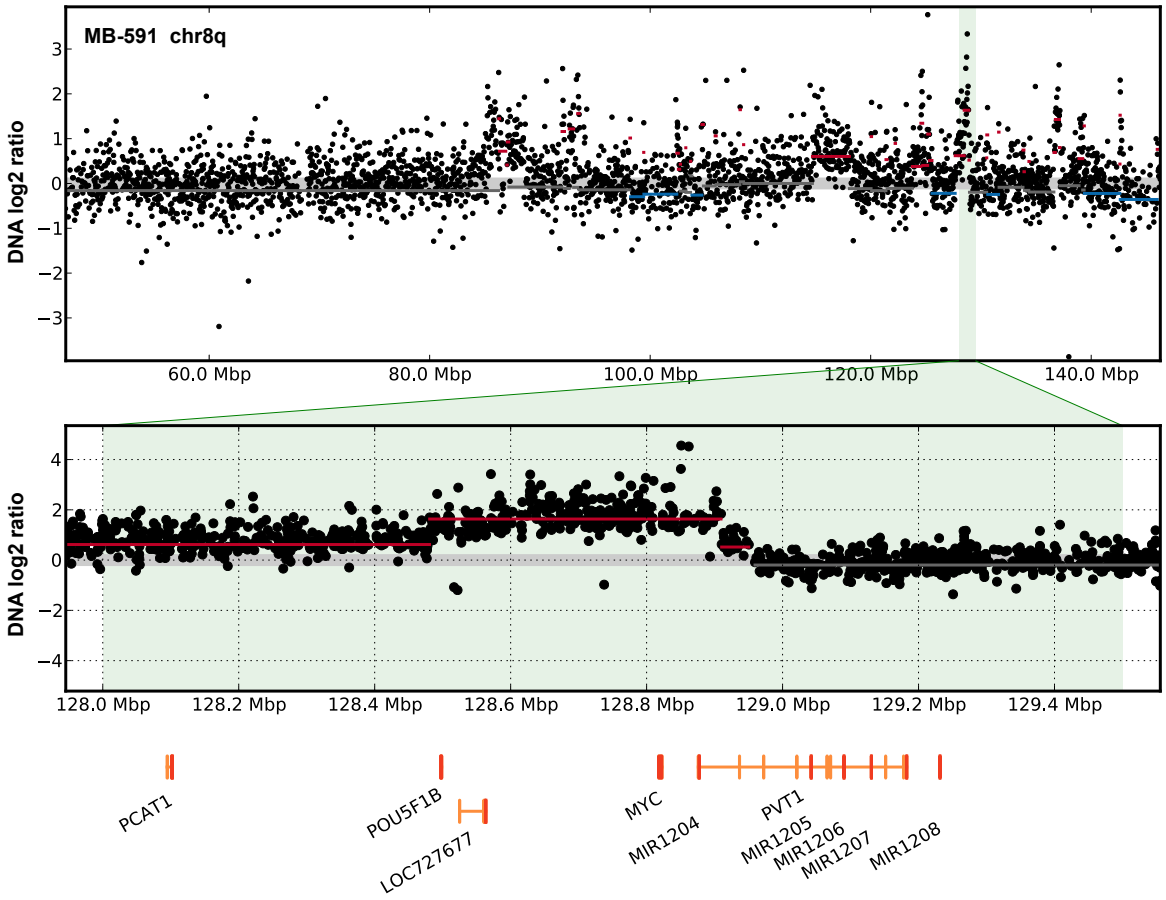

Figure 4.11: A multitude of amplicons disrupt the *MYC*/*PVT1* locus. Copy-number plot of 8q24.21 is shown for a representative sample. Dots represent raw copy-number estimates and lines denote copy-number segments and state (red: gain, blue: loss). 71.4% of *MYC*-amplified (20/28) cases exhibit (partial) co-amplification of adjacent non-coding *PVT1* gene and miR-1204. PVT1-MYC fusion transcripts were detected by RNA-seq, qRT-PCR, and Sanger sequencing in 8/20 samples[2].

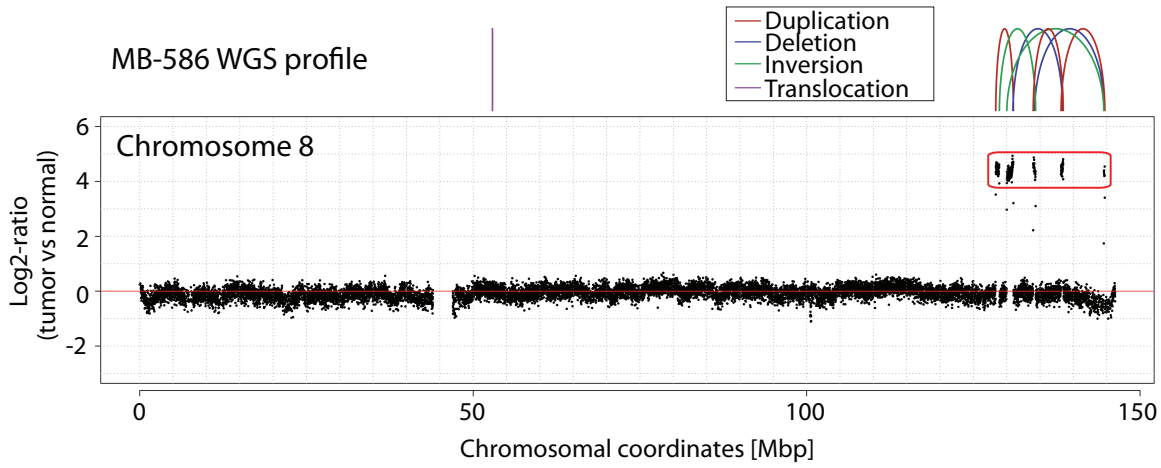

Figure 4.12: Chromothripsis disrupts the *MYC*/*PVT1* locus. Whole-genome sequencing confirms a complex pattern of rearrangements on 8q24 in a representative sample, reminiscent of chromothripsis (chromosome shattering). Plot shows chromosome 8 copy-number estimates (derived from read depth ratio of tumour vs. matched germline). Complex rearrangements are observed in concordance with a multitude of amplicons in the 8q24 region.

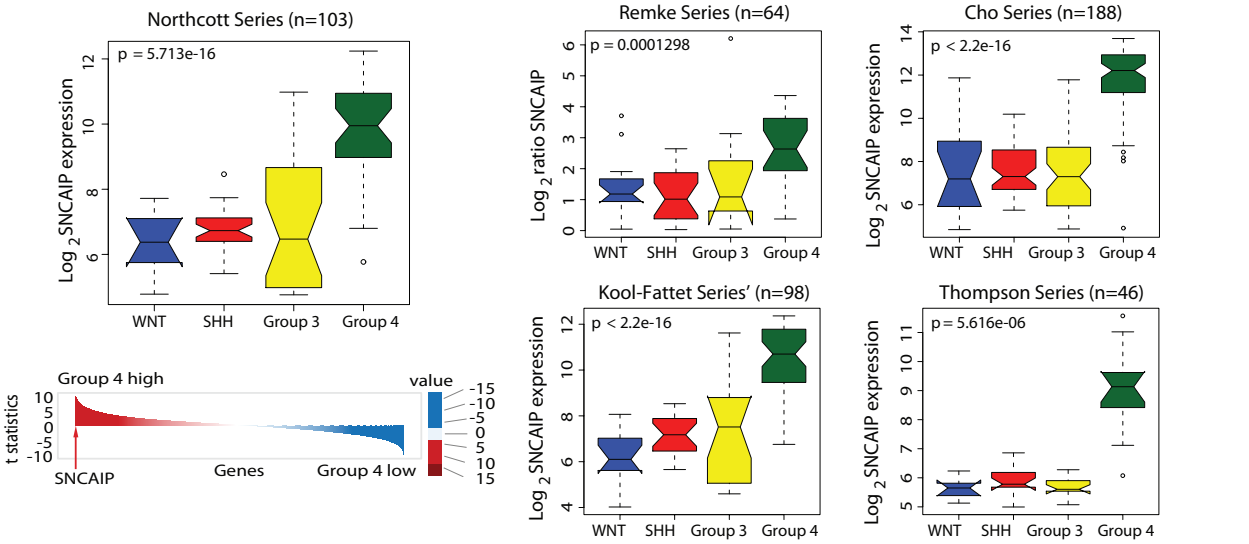

Figure 4.13: *SNCAIP* is a Group4 signature gene. *Left*, Box-plot depicting SNCAIP significant upregulation in Group4 (Mann-Whitney test), as determined by expression analysis of a previously published cohort of 103 primary medulloblastoma[35]. SNCAIP ranks among the top 1% of most highly expressed genes in Group4 medulloblastoma (rank 39 out of 16758). *Right*, Validation of SNCAIP as a Group4 signature gene across five published medulloblastoma expression datasets: Thompson[88], Kool[128], Fattet, Cho[38], and Remke[37]. Expression datasets total 396 cases on four different array platforms. In all datasets, SNCAIP exhibits the highest expression in Group4.

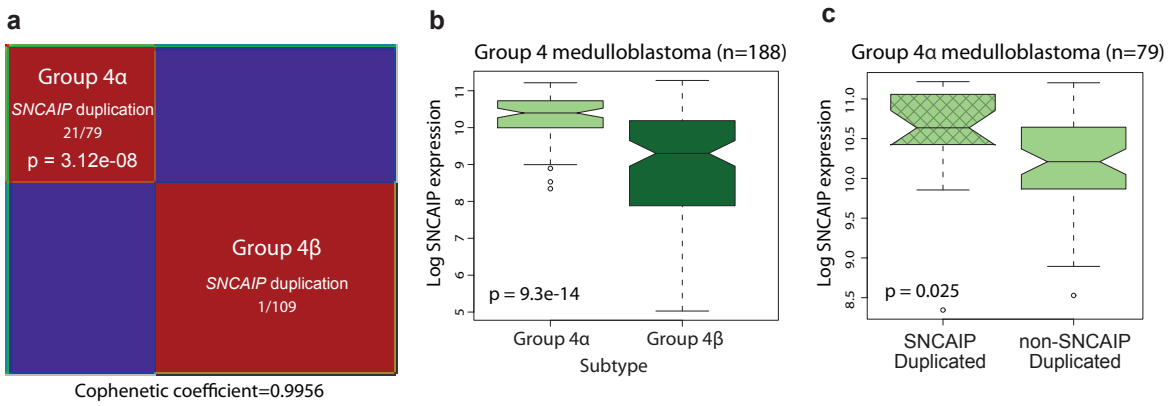

Figure 4.14: *SNCAIP* duplication is restricted to one subtype of Group4. a, Non-negative matrix factorization (NMF) consensus clustering performed on expression profiles of Group4 cases ($n = 188$) reveal two transcriptionally distinct subtypes of Group4, designated 4α and 4β. SNCAIP duplicated is significantly enriched in the Group 4α subtype (Fisher's exact test). b, SNCAIP expression is significantly elevated in Group 4α compared to 4β (Mann-Whitney test). c, SNCAIP expression is copy number-driven in Group 4α. Group 4α cases were stratified by *SNCAIP* copy-number status. Samples harbouring SNCAIP duplication exhibit a significant $\tilde{1}$.5-fold increase in SNCAIP expression (Mann-Whitney test).

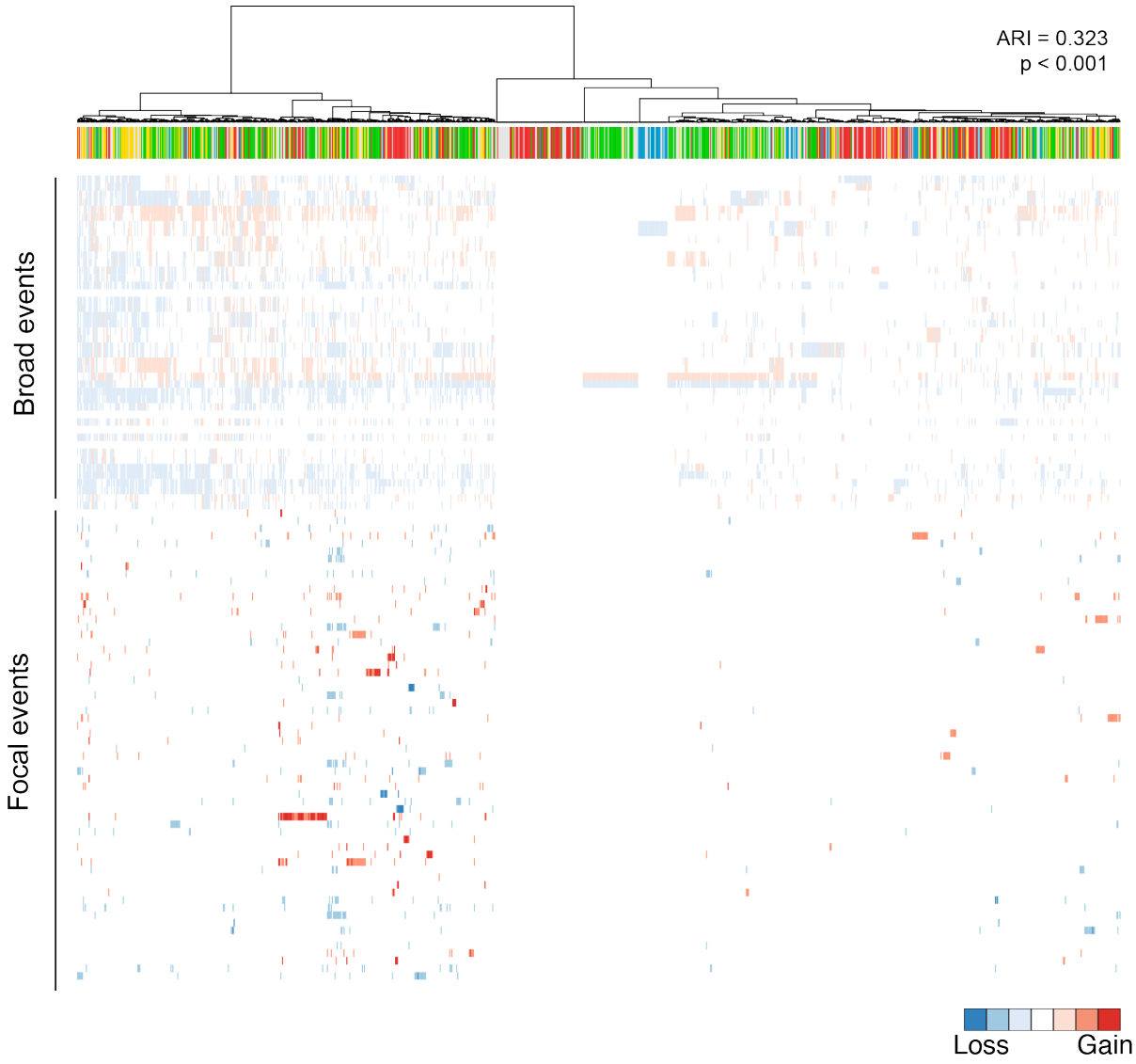

Figure 4.15: Hierarchical clustering of broad and focal SCNAs in medulloblastoma. Agglomerative hierarchical clustering was performed on 827 primary medulloblastoma samples with known molecular subgroups, in order to assess the association between the clusters driven by DNA copy-number profiles and the classes defined by expression signatures. Coloured top-side bar indicates known expression subgroups (WNT: blue, SHH: red, Group 3: yellow, Group 4: green). Only focal events identified in the pan-cohort GISTIC2 analysis were included in the clustering. Resulting clusters show significant agreement with known expression subgroups ($p < 0.001$, χ^2 test, $ARI = 0.323$). Adjusted Rand index (ARI) is a measure in $[0, 1]$ of the effect size of the agreement.

III Discussion

This study raised several bioinformatic challenges particularly due to the lack of germline samples. Owing to current and historical practices, the medulloblastoma samples amassed in this study were often not paired with germline samples. Fortunately, somatic and germline copy-number events are distinguishable on account of the rarity of large germline copy-number variants (spanning > 1 Mbp or > 10 probes) in the control population[2]. Additionally, we also dismissed copy-number variants observed in both cases and controls at high frequency ($> 90\%$). Identified germline copy-number variants were removed from copy-number profiles along with the breakpoints they introduced. However, despite our best efforts, some likely germline events and hybridization artifacts (unrelated to submission batch but possibly related to reagent lot) remain in the copy-number profiles of the samples. Nonetheless, with additional post-processing, careful curation, focus on high frequency events, and integration with multiple sources of evidence, we successfully identified recurrently disrupted genomic loci, though we caution against the isolated use of individual tumour copy-number profiles for future studies. The poor interpretability of individual copy-number events in a single tumour sample stems from non-systematic hybridization biases, poor sample DNA quality, the general genomic instability of some samples, and the notion that most genetic events during cancer evolution do not contribute to cancer cell survival and tumour formation. In light of these limitations, we prioritized genomic events observed frequently across multiple samples. Furthermore, we split samples into more homogeneous groups in order to identify such events above a background mutation rate. Indeed, we discovered significantly more recurrent CNA when the samples were split by subgroup than when they were analyzed together, reinforcing the utility of classifying medulloblastoma as four distinct diseases. To emphasize, we discovered patterns hidden in complex genomic profiles of tumours by comparing and contrasting biologically similar groups of samples.

Another common issue that arises in cancer genomics studies is the role that genomic instability plays in tumour formation, progression, or recurrence. We contend that genomic instability in itself does not impact outcome, though the underlying cause or the consequent disruption of specific genes may affect tumour aggressiveness. For example, *TP53* mutation may predispose to chromothripsis, but only the former portends poor outcome in multivariate survival analyses[1]. Chromothripsis does not exclusively occur in *TP53* mutated cases; instead, it may be associated with mutations in other DNA damage response or repair genes, not all of which may contribute to tumour resistance

against chemotherapy or radiotherapy. Further, chromothripsis in SHH medulloblastoma often results in amplification of *GLI2*, but only the latter is independently prognostic for poor survival[1]. These findings in our medulloblastoma cohort contrast the association of chromothripsis with poorer survival (by univariate instead of multivariate analysis) in neuroblastoma[254]. Further, our pattern of chromothripsis incidence also contrasts that of a smaller cohort of medulloblastoma[20], in which the authors observed significantly higher frequency of chromothripsis in SHH medulloblastoma. In contrast, the incidence of chromothripsis is fairly uniform across SHH, Group3, and Group4 medulloblastomas and significantly depleted in WNT medulloblastoma. These results could be due to characteristic differences between the two cohorts of medulloblastoma examined. Both studies used the same method and algorithm to identify chromothripsis (the same researcher processed the raw SNP array data and made the chromothripsis calls in both studies). Interactions among competing covariates may obscure meaningful interpretation and may explain the observed discrepancies across studies; furthermore, recall biases in retrospective studies may skew distributions of specific variables, especially in studies with small sample sizes.

Notwithstanding the technical, biological, and statistical challenges, we have successfully identified recurrent amplifications and deletions of genes that converge on specific signaling pathways contributing to tumourigenesis within each molecular subgroup of medulloblastoma. WNT medulloblastoma is surprisingly devoid of recurrent genomic aberrations aside from chr6 loss and *CTNNB1* activating mutation. While the functional consequence of chr6 is unclear, the high frequency of *CTNNB1* mutation suggest that activated Wnt signaling may be the predominant tumourigenic mechanism in WNT medulloblastoma. Similarly, recurrent CNA events converge on and conspire to activate Shh signaling in SHH medulloblastoma. In a subset of SHH cases, TP53 signaling and PI3K signaling are also recurrently disrupted by CNA, and the latter may be a candidate pathway for therapeutic intervention against tumours with aberrations upstream of the drug target. The dominant theme in Group3 medulloblastoma is recurrent disruption of signaling of the super Transforming growth factor β (Tgf-β) family; in particular, CNA converge on activation of activin signaling, and this pathway may be the first rational candidate pathway for therapeutic intervention in this aggressive subgroup. While no signaling pathway is highly recurrently dysregulated in Group4 medulloblastoma, tandem duplications of *SNCAIP* is frequently observed and may play an important role in the pathogenesis of Group4 medulloblastoma. Taken together, copy-number profiling of medulloblastoma samples have led to the discovery of several frequently disrupted genes

and pathways that may serve as candidates for targeted therapeutic intervention.

The disparate patterns of recurrent CNA among medulloblastoma subgroups not only support the distinct etiologies of the subgroups but also raise questions regarding the origin of the molecular subgroups of medulloblastoma. Most recurrent CNA were enriched (observed at higher frequency) in specific subgroups but were not exclusively found in one subgroup. Further, unsupervised clustering by copy-number profiles produced groups that showed only modest agreement with the molecular subgroups of medulloblastoma based on transcriptional profiles (**Figure 4.15**). These results taken together suggest that the CNA events do not determine the molecular subgroups, though they may modulate the expression patterns of the subgroups. In contrast, DNA methylation patterns can identify, without supervision, four medulloblastoma classes that show high concordance with the subgroups identified by expression patterns[260;261]. This finding indicates that the epigenetic landscape in the cell of origin controls the expression patterns of the developing tumour and may also provide contexts that favours specific genomic aberrations, thus forming disparate genetic landscapes for each medulloblastoma subgroup.

The Myc family of proto-oncogenes is amplified in a curious pattern across the medulloblastoma subgroups, and this pattern poses a question regarding whether and how epigenetic patterns influence amplification patterns of the Myc family, comprising *MYC*, *MYCN*, and *MYCL*. These homologous genes were found to be amplified in lymphoma, neuroblastoma, and lung carcinoma, respectively, which possibly points to functional differences in Myc family proteins in diverse cancer types. Similarly, Myc family genes are amplified in a non-random pattern in medulloblastoma: *MYC* is amplified predominately (albeit not exclusively) in Group3, *MYCN* is amplified preferentially in SHH and Group4 medulloblastomas, and rare *MYCL* amplifications are observed only in SHH medulloblastoma. At the expression level, most medulloblastomas of either WNT or Group3 subgroup upregulate RNA expression of MYC independent of DNA amplification. Similarly, SHH medulloblastomas express MYCN at a higher level than other subgroups and normal cerebellums even without *MYCN* amplification. In neuroblastoma, *MYCN* amplification is found in about 20% of cases[254;262], but amplification of *MYC* or *MYCL* is very rarely observed[254]. These observations raise the question of why *MYC*, *MYCN*, and *MYCL* is preferentially dysregulated in one cancer type or subtype as compared to another. Conceivably, Myc family genes could interact with different sets of co-transcription factors, drive dissimilar transcriptional programs, and cooperate with dysregulation of disparate signaling pathways in promoting tumour formation. Ac-

cordingly, the amplification of different Myc family genes could modulate the expression patterns of medulloblastoma subgroups. Additionally, genes nearby the *MYC* or *MYCN* locus can modulate the effect of amplifications spanning these loci.

An alternative hypothesis is that differential accessibilities at genomic loci in the cell of origin govern which of the three human Myc family gene is amplified. Under this model, *MYCN* amplification preferentially occurs in SHH medulloblastoma because *MYCN* is expressed in the cell of origin for SHH medulloblastoma (e.g. external granule neural precursor) during normal development, rendering the genomic locus accessible to such enzymes as deaminases (e.g. *AICDA*, *APOBEC*), which initiate double-strand breaks that can ultimately lead to DNA rearrangement. Similarly, *MYC* amplification may preferentially occur in Group3 medulloblastoma because the *MYC* locus is permissible to structural rearrangement in the cell of origin of Group3. In other words, the observed patterns of Myc family gene amplification can arise due to variations in accessibility of these genomic loci in different cellular origins instead of functional differences among Myc family members. Consistent with this model, MYCN is functionally redundant with MYC, and knock-in of *MYCN* into the *MYC* locus rescues the loss of *MYC* in mouse development, cell growth, and cell differentiation[263]. *MYC* and *MYCN* double knock-out mice exhibit much more severe phenotype in neurogenesis than either one alone, suggesting that loss of one Myc family member can be partially rescued by another member with overlapping expression pattern[264]. Further, *MYC* and *MYCN* are interchangeable in promoting tumour formation: *MYCN* coding region can substitute for that of *MYC* in the transformation of pro-B cell[265]. Similarly, MYC can cooperate with SHH ligand to enhance formation of SHH medulloblastoma in a virus-induced mouse tumour model, despite that *MYC* amplification is rarely observed in human SHH medulloblastoma[2;266]. Therefore, Myc family genes may indeed encode functionally similar proteins (whose transcriptional regulation may differ) that promote medulloblastoma formation by a common mechanism.

Despite being an important player in transforming normal cells into cancer or in inducing pluripotency in differentiated cells, a Myc family gene may exert different effects in different tumour contexts. In Group3 medulloblastoma, *MYC* amplification is associated with poorer survival. However, *MYCN* amplification is associated with poorer survival only in SHH medulloblastoma but not in Group4 medulloblastoma[1]. The amplification of a Myc family gene is thus not universally associated with poorer outcome, and the function of Myc in medulloblastoma may be modulated by cooperating

disruption of other pathways in the tumour-initiating cell and by extracellular signals in the context of neural development.

The story of the Myc family illustrates a broader theme of how genomic alterations lead to different subgroups of medulloblastoma. *MYC* amplifications can drive tumour formation, but only in the suitable cellular context and with other cooperating mutations[233;266]. Indeed, *MYC* and *SHH* overexpression are sufficient to induce tumour formation in the external granule layer of the cerebellum but not in the cerebrum[266]. Accordingly, MYC overexpression alone likely does not create Group3 medulloblastoma. Rather, (possibly multiple) cellular origins and cooperating alterations provide the necessary context for the amplification of *MYC* and consequent induction of medulloblastoma, and they dictate the molecular subgroup identity of the resulting tumour. Similarly, *MYCN* amplification is a common mechanism of tumourigenesis in subsets of SHH and Group4 medulloblastoma cases, and this genetic event does not induce an overwhelming transcriptional change that by itself justifies the definition of an distinct *MYCN*-amplified molecular subgroup. Indeed, gene amplification of the Myc family does not create a transcriptionally homogeneous entity that is distinct from other molecular subgroups. The functional outcome of Myc amplification very much depends on the molecular subgroup of the tumour in which Myc amplification occurs[1]. In experimental models, Myc family overexpression, along with cooperating mutations, can likely transform multiple cells of origin, and the subgroup of the resulting tumour would likely be determined by the cellular origin rather than the Myc family member. More generally, tumours of specific subtypes arise due to interaction between genetics and context. The cellular origin provides the permissive context for genomic alterations to exert their effects, and the nuclear organization, epigenetic landscape, chromatin architecture within the cell of origin jointly influence the accessibility of genomic loci and shape the mutational landscape of the cancer cell[267].

In summary, medulloblastoma subgroups are characterized by distinct genomic alterations that disrupt disparate signaling pathways. The mutational profiles do not define each subgroup; instead, they are shaped by the molecular and cellular context within each subgroup, and the functional consequences of the mutations may also depend on the molecular subgroup. Hence, genes and pathways that are recurrently disrupted by genomic alterations within each subgroup likely play pivotal roles in the formation and perhaps maintenance of the tumours, and such genes and pathways serve as prime candidates for design of rational, targeted therapeutic intervention.

Chapter 5

Conclusions and future directions

The discovery of the four molecular subgroups of medulloblastoma has paved the road for the informed development of specific targeted therapy and promises hope for achieving personalized medicine in the treatment of medulloblastoma, where each patient will receive an individualized regimen that maximizes efficacy and minimizes side-effects. Medulloblastomas, not unlike many other malignancies, arise due to a multitude of genetic aberrations, leading to disruptions of biological processes that differ from patient to patient. These aberrations begin to form clear patterns when viewed against the backdrop of molecular subgroup. To streamline scientific research and clinical practice, we hope to introduce molecular classification in the clinic. Currently, two major obstacles in the adoption of molecular subgroup classification in the clinic are the lack of applicable specimens for genomic analysis and the discrepancy in quality control standards between research and clinical labs. To overcome these challenges, we have tested and validated an assay suitable for subgroup classifcation of FFPE samples, and we have implemented rigorous quality controls to ensure that the assay results are reproducible, credible, and suitable for guiding clinical decision-making. If molecular subgroup information does not influence clinical treatment, however, the assay would have little clinical utility. We have therefore identified, through genomic profiling, actionable signalling pathways within medulloblastoma subgroups that may serve as rational targets for future therapeutic development. To further fuel the motivation for classifying medulloblastoma subgroups in the clinic, we have identified molecular biomarkers that, together with clinical biomarkers, successfully stratify patients into risk groups using schemes specific to each molecular subgroup and attain unprecedented prognostic accuracy. Accordingly, we have addressed some of the challenges that face the treatment of patients with medulloblastoma and provided evidence that supports the classification of medulloblastoma by

molecular subgroups in the clinic. We hope that the adoption of molecular classification will inform the next generation of clinical trials and facilitate the development of personalized targeted therapy.

While in pursuit of this long-term goal, we need to address some remaining questions regarding the classification of medulloblastoma. While anatomical location and histology will likely continue to be an integral core of CNS tumour classification, numerous other ways of categorizing cancer present the problem of choosing or appropriately integrating classification schemes. One alternative classification of medulloblastoma could be based on genetic alterations, as in many other cancer types. While some genes identify specific cancer types – *RB1* (retinoblastoma) and *SMARCB1* (ATRT) – other genes may be less useful for defining cancer types. For example, *TP53* mutation or loss leads to a spectrum of tumours, and restoring TP53 function will not restore the genomic damages already incurred. Additionally, given that most observed mutations in cancer likely do not contribute to tumourigenesis, identifying and validating tumourigenic mutations may be difficult without first grouping cancers into sufficiently homogeneous subtypes. Moreover, disruption of multiple genes in the same pathway may lead to the same molecular phenotype, and it can be difficult to identify recurrently disrupted pathways prior to molecular classification[2]. The most problematic issue for defining medulloblastoma based on genetic alterations is the relatively low frequency of most mutations. Conversely, epigenetic profiles may reflect the cellular origin that shape the genomic landscape of the tumour, and it may be useful for the classification of medulloblastoma. Encouragingly, DNA methylation profiles define similar subtypes as RNA expression profiles, suggesting that both may be integrated to develop a robust classification of medulloblastoma.

Since medulloblastoma is now classified into four molecular subgroups, mouse models of medulloblastoma would also need to be classified into the same subgroups. Numerous mouse models purport to recapitulate a specific molecular subgroup of human medulloblastoma,[4;22–24] but the molecular subgroup of some mouse models are contested[268]. Better comparative transcriptomics may therefore be needed to resolve controversies surrounding mouse models of medulloblastoma (and similarly for other cancers or diseases). We would need to move beyond general descriptions of the conservation or divergence of transcriptomes and attempt to draw parallels between individual human and mouse transcriptional programs in order to more precisely identify conserved molecular mechanisms and enable specific hypotheses regarding human diseases to be tested in mouse models.

Concurrent with ongoing scientific inquiries, the search for more effective therapy for medulloblastoma continues. With the recognition that medulloblastoma comprises four different diseases, many prospective trials are now testing emerging therapies for specific medulloblastoma subgroups, consistent with the spirit of precision medicine. For Group3 medulloblastoma, an *in vitro* drug screened identified pemetrexed and gemcitabine as a potential combination therapy, and this combination showed efficacy in mouse models of Group3 medulloblastoma (but not SHH medulloblastoma)[269]. Similarly, BET bromodomain inhibition of *MYC*-amplified medulloblastoma is currently under investigation[270]. For SHH medulloblastoma, SMO inhibitors (e.g. vismodegib and sonidegib) showed some efficacy and present a promising avenue for further development[271–274]. In addition, mouse models of SHH medulloblastoma appear sensitive to inhibition of Auroa and Polo-like kinases[13] or inhibition of BIRC5 (survivin)[275]. In order to expand the arsenal of anti-cancer drugs, it may be prudent to consider administering candidate drugs before standard combination chemotherapy in future clinical trials, as precedent exists for drugs to be effective against untreated tumours but not recurrent tumours. For example, topotecan (topoisomerase inhibitor) is ineffective in recurrent medulloblastoma[276;277] but is effective upfront in untreated, high-risk medulloblastoma[278]. In trials of novel therapy on untreated medulloblastoma, salvage treatment with chemotherapy should of course be planned so that patient survival is not compromised; furthermore, prior trials on replacing radiotherapy with chemotherapy would provide invaluable insight for planning salvage treatments. Past clinical experiences and novel scientific knowledge will together help usher in a new era of medulloblastoma treatment based on individualized targeted therapy that enhances the quality of care and preserves the quality of life for the patients.

Chapter 6

Appendix

I Nomenclature

Gene nomenclature

Although a discussion of nomenclature may be pedantic, unambiguous nomenclature is important to avoid confusion and misinterpretation of scientific results, as exemplified by a retracted Nature publication[279]. This thesis uses standard notations commonly seen in the literature to distinguish between genes and proteins from different species, but not all publications (including those cited herein) use the notations presented here. For the most part, gene symbols serve merely as identifiers for genes in this thesis; accordingly, full gene names are usually not provided. (In particular, some genes such as *AKT1* have no full names due to the systematic nature of their discoveries.) Genes are often known by several names owing to multiple discoveries in different research laboratories. Although one unique official symbol is used to identify a gene, genes often have pleiotropic functions, and some researchers prefer to refer to a gene by a particular alias that reflects one facet of the gene's function of interest. On other occasions, genes were initially assigned generic names such as p21, p27, p53 (referring to proteins with masses 21 kDa, 27 kDa, and 53 kDa, respectively), and these names have become standard in the literature. This thesis instead uses official gene symbols to ensure that genes can be uniquely identified and that no unfortunate misinterpretation will arise, as in the study of Kawasaki *et al*[279]. To learn more about a particular gene, readers may wish to query the Entrez Gene database (http://www.ncbi.nlm.nih.gov/gene).

As much as possible, official gene symbols as approved by the Human Genome Or-

ganisation Gene Nomenclature Committee or the Mouse Genome Informatics are used, and where appropriate, followed by common gene name aliases in parentheses. Human gene symbols appear in uppercase, and mouse gene symbols, in title case. Italicized names refer to genes, whereas non-italicized names refer to gene products. For example, the human gene on chromosome 9 (at location 98205264–98279247 of the GRCh37 human genome assembly) encoding the first family member of the patched 12-pass transmembrance receptor is denoted *PTCH1*, and its protein counterpart, PTCH1. The mouse gene encoding the patched homologue gene is denoted *Ptch1* and its protein, Ptch1. Specifically in this thesis, SHH and WNT usually refer to the molecular subgroups of medulloblastoma and not the protein or protein family.

Animal model nomenclature

The nomenclature for genetically engineered mouse models is extensive[280;281]. This thesis will use the following simplified notation: *Ptch1*$^{-/-}$ is homozygous mutant, *Ptch1*$^{+/-}$ is heterozygous mutant, and *Ptch*$^{+/+}$ is homozygous wildtype at the *Ptch* locus. The (+) symbol denotes a wildtype or non-mutated allele, and the (-) symbol denotes a null or loss-of-function allele. Multiple genetic modifications are separated by a semicolon: *Ptch1*$^{+/-}$;*Trp53*$^{+/-}$ is heterozygous mutant for both *Ptch* and *Trp53*. While these notations omit the mouse strain, the genetic background of the mouse model can indeed modulate the mutant phenotype[82;282;283].

Gene function may be disrupted constitutively (**knocked out**). Homozygous knock-outs, which abrogate gene function completely, may cause embryonic lethality and preclude further study (e.g. *Ptch1*$^{-/-}$ mice die before birth)[68]. In addition to studying heterozygous knock-outs (e.g. *Ptch1*$^{+/-}$), researchers may also conditionally disrupt of gene function using Cre recombination. A sequence flanked by loxP sites is knocked into the native gene locus. Cre recombinase recognizes loxP sites and recombines the DNA, causing deletion of the region **f**lanked by **lox**P sites (the region is **floxed** and subsequently deleted). Whenever and wherever Cre is expressed, the floxed region of the gene would be deleted. Such a model requires two genetic contructs: the floxed gene (e.g. *Ptch1*$^{\text{flx/flx}}$ where flx denotes 'floxed') and Cre with a promoter (e.g. *Atoh1-Cre*). The floxed region may encompass the entire gene or (usually) critical exons of the gene so that the no functional products would result from the recombined gene. The promoter is often named after the gene from which it was derived. That is, *Atoh1-Cre*

indicates that Cre is placed downstream of the promoter for *Atoh1* (and not the coding sequence of the *Atoh1* gene itself). Therefore, in the *Atoh1-Cre*;*Ptch1*$^{\text{flx/flx}}$ mouse[97], Cre would be expressed in Atoh1-expressing cells and delete both of the floxed *Ptch1* alleles; consequently, the Atoh1-expressing cells would incur a homozygous deletion of *Ptch1* and lose Ptch1 function completely.

A transgenic construct may also be integrated (randomly) into the genome to express a gene. For example, the *Neurod2-Smo*$^{+/\text{W539L}}$ mouse has the *Neurod2-Smo*$^{\text{W539L}}$ construct integrated into the genome once (later mapped to chr14)[106], the activated Smo (with a substitution mutation from tryptophan to leucine at residue 539) is expressed under the promoter of *Neurod2*. The Cre recombination system may also be used to conditionally express a gene using a more complex design, such as the *CAGGS-CreER*;*Rosa26-SMO*$^{\text{W535L}}$ mouse[284]. In this mouse, the human *SMO* gene with an activating W535L mutation is knocked into the ubiquitously expressed *Rosa26* locus, whose official symbol is Gt(ROSA)26Sor. (This human *SMO* gene is known as SmoM2[285] and corresponds to mouse *Smo*$^{\text{W539L}}$, which is known as SmoA1[107].) However, since a floxed polyadenylation stop sequence cassette is placed upstream of the *SMO*$^{\text{W535L}}$ gene, the latter is only expressed subsequent to Cre-mediated deletion of the cassette. Cre is expressed here under the synthetic CAGGS promoter (which drives high-level, generalized expression), and its activity is (somewhat) tamoxifen-dependent because it is fused to the estrogen receptor. Therefore, the expression of *SMO*$^{\text{W535L}}$ is ultimately controlled by the administration of tamoxifen (at least in design).

II Signaling Pathways

Wnt signaling (CTNNB1-dependent)

In the absence of WNT ligand, CTNNB1 (β-catenin) is continuously marked for degradation by the CTNNB1 destruction complex, consisting of AXIN, APC, GSK, and CSNK1A1 (CK1). GSK and CSNK1A1 primes CTNNB1 by phosphorylation, which leads to ubiquinylation by the CUL1-containing E3 ligase complex and subsequent complete degradation by the proteome. Upon binding of WNT to FZD, DVL1 is activated via an unknown mechanism, and DVL1 phosphorylates LRP5/6, which then sequesters AXIN and frees CTNNB1 from the destruction complex. CTNNB1 accumulates in the cytoplasm and translocates to the nucleus to activate target genes such as *MYC*, *CCND1*,

and *AXIN2*. Two *AXIN* genes exist in humans: *AXIN1* and *AXIN2*. GSK consists of two subunits: GSK3A and GSK3B (catalytic). CTNNB1 regulates expression in concert with the TCF/LEF family of co-transcription factors, such as TCF7, TCF7L1, TCF7L2, and LEF1. WNT signaling also activates many other pathways independently of CTNNB1; these pathways include: planar cell polarity, WNT-Ca^{2+}, and others. Both WNT and FZD encompass a large family of proteins.

Shh signaling

GLI transcription factors, including GLI1, GLI2, and GLI3, are downstream effectors of Shh signaling. The proteome can completely degrade GLI proteins or proteolytically process them into activators or repressors forms, depending on post-translational modifications on the full-length GLI proteins. GLI1 and GLI2 predominantly function in the activator form and activates transcription, while GLI3 mainly functions in the repressor form and represses transcription. Shh signaling modulates the marks on full-length GLI, thereby influencing downstream transcription. In the absence of SHH ligand, SMO activity is repressed by PTCH. Upon SHH binding to PTCH, this repression is relieved, leading to active SMO signaling, which favours the processing of GLI into the active form; consequently, transcription of target genes including cell cycle genes (*CCND1* and *CCNE1*), *MYC*, and negative regulators of Shh signaling (*PTCH1* and *HHIP*) is induced. The precise mechanism whereby SMO (indirectly) activates GLI is not conserved between fly and mammals and remains unknown in mammals. Four proteins bind to and regulate GLI processing: KIF7, SUFU, SPOP, and BTRC. Both KIF7 and SUFU are negative regulators of Shh signaling via unclear mechanisms. The role of SUFU remains contentious; possible roles include: nuclear export of GLI, protection of GLI from degradation, recruitment of GLI3 for processing into the repressor form. Conversely, SPOP and BTRC (βTRCP) serve to recognize GLI and they function as subunits of E3 ligase complexes, which mark (by ubiquitination) GLI for proteolytic processing by the proteome. The CUL1 containing complex is also known as the Skp, Cullin1, F-box (SCF) complex, Via BTRC-mediated recognition, this complex ubiquitinates many other proteins, including CTNNB1 from the Wnt pathway. Mammals have three hedgehog ligands: DHH, IHH, and SHH. PTCH encompasses PTCH1 and PTCH2. PKA is a tetramer composed of two catalytic subunits (protein family includes PRKACA, PRKACB, and PRKACG) and two regulatory subunits (protein family includes PRKAR1A, PRKAR1B, PRKAR2A, and PRKAR2B).

Notch signaling

When NOTCH binds DLL that is expressed on the surface of an adjacent cell, NOTCH is extracellularly cleaved by ADAM10 or ADAM17 (TACE), and it is subsequently cleaved intracellularly by the gamma-secretase/presenilin complex. This proteolytic event releases the NOTCH intracellular domain (NICD) from the cell surface, allowing it to translocate to the nucleus in order to activate transcription of target genes such as *MYC*, *HES*, *CDKN1A*, *CCND3*. Mammals have four Notch family members (NOTCH1, NOTCH2, NOTCH3, and NOTCH4), and these receptors can bind ligands DLL1, DLL3, DLL4, JAG1, and JAG2.

PI3K signaling

Kinases from the PI3K family phosphorylate PIP_2 to produce PIP_3, which serves as the intermediate signaling molecule for many pathways. One well-studied binding target for PIP_3 is AKT1, which in turn inhibits apoptosis and promotes cell cycle progression. PI3K signaling is activated downstream of many cell-surface receptors, including receptor tyrosine kinases, cytokine receptors, integrins, and G-protein coupled receptors. PTEN is a negative regulator of PI3K signaling and functions by dephosphorylating PIP_3 to form PIP_2.

III Classification

Classification is the central preoccupation of one of the oldest disciplines, **taxonomy**, which seeks to place organisms into groups based on shared characteristics. Just as classifying organisms facilitates their study, classifying diseases also helps clinicians understand common mechanisms underlying disease and develop rational treatment against each type of disease. Classification encompasses two parts: first, the classes are established by exploring features and grouping the samples based on similarity of features; second, a method is created to classify new sample using features that discriminate between classes. In bioinformatics, the first part is referred to as **class discovery**, and the second, **class prediction**. In taxonomy, the term **classification** only refers to the first part while the term **identification** refers to the second part. In machine learning, the first part is equivalent to **unsupervised learning**, and the second, **supervised learning**.

In statistics, the term **classification** only refers to the second part. Notwithstanding the different terminologies in various fields and the conflation of different meanings of the term 'classification', the two components have distinct objectives. While some algorithms can be used for both class discovery and class prediction, algorithms specifically designed for prediction usually outperform other algorithms in the class prediction problem. Furthermore, class prediction algorithms are well developed for the case in which the classes are known for a set of samples. Conversely, the class discovery process may be contentious: when using different features results in discovering different sets of classes, which set of classes is correct? Indeed, a classification system is practically useful insofar as it is widely accepted and stable. Once such a classification system is discovered and becomes established, one can focus on predicting the class of a new sample (as least until a new classification system arises).

Class discovery

The objective of class discovery is to group (or split) samples into disjoint classes. One systematic and unbiased way of achieving this goal is by **clustering** (or **partitioning**) a set of samples. Given measurements for various features (characteristics), clustering groups samples with similar features together, whereas partitioning split the samples into dissimilar groups. The distinction between clustering and partitioning is moot because the same overall objectives are achieved; however, different algorithms often produce different results. A variety of clustering (or partitioning) algorithms are available, including hierarchical clustering, k-means, self-organizing map, affinity propagation, spectral clustering, graph clustering, mixture models, and consensus clustering. The two most popular clustering algorithms are hierarchical clustering and k-means (these are also among the oldest). These clustering algorithms have many additional variants. Hierarchical clustering may be divisive or agglomerative; it may calculate distance between groups using various **linkage methods** (single, complete, average or others). k-means have variants such as k-medians or k-medoids, where different methods are used to represent the center of a group. k-means uses the mean of all the group members (i.e. centroid), k-medians uses the median, and k-medoids uses the group member whose average distance to all other group members is minimal. A variant of k-medoids is **partitioning around medoids**, which uses a particular method for initializing the groups and updating the clusters. Furthermore, the aforementioned clustering methods may use different measures of distance between individual samples. Self-organizing map, affinity propa-

gation, and spectral clustering are less popular in computational biology (at least for now). Graph clustering encompasses a myriad of algorithms that cluster a **graph** (a collection of nodes with interconnecting edges, known more informally as a 'network'). These algorithms rely on known independence (lack of edges) between samples (nodes), whereas most of the other algorithms consider connections between all pairs of samples as presented in a distance matrix or similarity matrix. Mixture models attempt to model the data as a mixture of statistical distributions, such as multivariate Gaussian distributions, and the class assignments are probabilistic. Consensus clustering runs multiple clustering algorithms or variants (with different parameters) to generate multiple clustering results; then, it define classes based on the census of the clustering results. A type of consensus clustering algorithm that has gained popularity recently is **NMF consensus clustering**. In this method, the data matrix is randomly factorized into two matrices using NMF and samples are assigned preliminary groups based on one of the factor matrices (using an arbitrary rule); subsequent to multiple NMF runs, the clustering results are combined using consensus clustering, by apply hierarchical clustering on the census matrix, which counts the number of times that each pair of samples were assigned the same group.

The foremost decision in class discovery is deciding which type of features to use. For example, one could run clustering analysis with expression data, DNA methylation data, gene mutation data, patient, and other data. One could also combine different types of features and perform clustering in an integrative manner (while taking care to weight the different types appropriately). Due to the exploratory nature of class discovery, there are no formal guidelines for the choice of features (though interval-scale features are more mathematically amendable to clustering analysis than nominal-scale features). Often, the type of features can be chosen to discover classes for specific objectives, such as using expression classes to identify tumour types with activation of different biological pathways. Indeed, a rational researcher guided by the same objective would be unlikely to group patients into classes based on their birth days, months, and years.

Another challenge of clustering analysis is to determine the number of classes represented in the data, k. Some algorithms begin by putting all samples into one class and proceed to partition the samples into smaller classes recursively (e.g. divisive hierarchical clustering), some algorithms begin by putting each sample into its own class and proceed to cluster the samples into bigger classes recursively (e.g. agglomerative hierarchical clustering). Either way, a decision must be made as to when to stop parti-

tioning or clustering. In the case of hierarchical clustering, this decision is usually made informally. Other algorithms require the number of classes k to be specified *a priori* (e.g. k-means, finite mixture models). One recent algorithm that circumvents this requirement is **Dirichlet Process infinite mixture model**, which allows for all possible values of k and discovers k from the data. However, Dirichlet Process can be very computationally intensive. An alternative approach is evaluate the clustering results for different values of k. (For algorithms in which k is not specified *a prior*, one runs the algorithm until k clusters are created.)

There is no concensus on the best method for evaluating the quality of the clusters. Often, practitioners evaluate clustering results by considering external information. This external evaluation of the data should be done with care in order to avoid overfitting and optimistic bias. Additionally, since the objective of class discovery is to discover new classes, comparison with known classes would not likely help guide the evaluation of clustering results. As an example, suppose we evaluate clustering results by assessing the association of the discovered classes with survival (i.e. how different are the survival times among the classes). Using this criterion, we compare the association with survival among different clustering algorithms or parameter settings, and we choose the algorithm with the parameter setting that achieves the most significant association with survival. Unless we have additional samples not used during clustering to validate the association of the classes with survival, we may be *overfitting* the algorithm (or parameters) to the existing data, and the optimistically biased association may disappear in a new dataset. In other words, we may have discovered classes that are spuriously associated with survival in our available data, and we may not observe this association again in a new dataset. We may mitigate this problem by using a subset of the data for optimizing the clustering algorithm and evaluating the results on the remaining data. Alternatively, we may also evaluate the clustering results without using external data. This internal evaluation of the data involves measuring the distance between sample pairs within each cluster as well as the distance between sample pairs from different clusters. The best cluster assignments should minimize the within-cluster distance and maximize the between-cluster distance. Different measures for evaluating and combining these two objectives have been proposed, including the silhouette width and the Dunn index. However, these internal measures depend on the definition of distance between samples: the results would change if a different distance measure were used. Indeed, evaluating the results of a clustering algorithm can be difficult and far from straightforward. Due to this difficulty in evaluating clustering results, there is also no concensus on the best clustering

method. There is simply no substitute for applying class discovery on multiple datasets and confirming that the same classes are discovered in each dataset.

Class prediction

Given a set of samples with known classes (**labeled data**), the objective of class prediction is to learn the association between the labels and the features of the samples and subsequently predict the class of a new sample. The class labels could be defined by expert opinion, discovered by systematic clustering, or determined by other means. Class prediction involves three steps: **training**, **testing**, and **application**. That is, a classifier must be trained on a labeled dataset and tested on another labeled dataset before it can be applied on samples with unknown classes (**unlabeled data**). During training, the classifier learns the association between labels and features in a first labeled dataset (**training data**). During testing, we evaluate how well the classifier generalizes its learned association to a separate labeled dataset (**testing data**). Finally, only when we are certain that the classifier can predict classes accurately, we apply it to unknown samples and make downstream decisions.

One simple measure of classifier performance is **accuracy**: the percentage of samples classified correctly. A high accuracy on the training data indicates the classifier has sufficient capacity to learn from the data, and a high accuracy on the testing data implies that the classifier can generalize well. When a classifier has high training accuracy but low testing accuracy, the classifier is likely overfitted to the training data. In such an event, possible remedies include simplifying the complexity of the classification algorithm (by permitting fewer free parameters) or collect more training data.

For a classification problem with two classes (positive and negative), common evaluation methods include receiver-operating characteristics curve or precision-recall curve. These methods simultaneously evaluate how often a classifier falsely assigns a positive label to a true negative sample and how often a classifier falsely assign a negative label to a true positive sample. For multiclass problems, common evaluation methods include Rand index or Jaccard index. Alternatively, a multiclass problem can be split into multiple two-class sub-problems, in which each sub-problem addresses whether a sample belongs to a given class.

IV Cancer treatment

Chemotherapy

Chemotherapy is the use of drugs that target general cellular processes such as cell division in order to eradicate cancer cells. The drugs may damage DNA directly (by alkylation or forming DNA adducts), interfere with DNA synthesis (by inhibiting folic acid, substituting for a nucleotide, intercalating into DNA, or inhibiting topoisomerase), prevent mitosis (by inhibiting microtubules), or various other mechanisms. Often, chemotherapy relies on triggering the (hopefully intact) apoptotic program of the cell in response to recognition of DNA damage. For example, loss-of-function mutation in *TP53* (responsible for triggering apoptosis) can lead to resistance against chemotherapy. Overexpression of efflux pumps is another major mechanism of chemoresistance.

Chemotherapeutic treatment typically lasts from weeks to months, and it is given at different stages. The purpose of induction chemotherapy is to induce a remission, while consolidation chemotherapy is given, typically at high dose, at the end of induction to complete remission. Conversely, maintenance chemotherapy is given at low dose over a long period to prevent cancer recurrence.

Sometimes, chemotherapy are described as adjuvant or neoadjuvant. Adjuvant therapy is given to assist the main treatment in eradicating the cancer and is given during or after the latter; neoadjuvant therapy is given before the main treatment. In solid tumours, some literature consider surgical resection as the main treatment while others consider radiotherapy to be of prime importance.

V Prognostic biomarker discovery

Log-rank tests vs. Cox proportional-hazards test

The survival analyses presented were based on log-rank tests and Cox proportional-hazards tests, which may yield considerably different p-values. As log-rank tests do not assume proportional hazards, their results were presented instead of those of Cox proportional-hazards tests. Univariate Cox proportional-hazards analyses were performed to estimate hazard ratios and sample sizes required for prospective studies.

Construction and validation of risk stratification models

In order to identify novel and robust prognostic biomarkers, the present study examined a discovery set and a validation set of medulloblastoma cases. The discovery set consisted of cases with patient survival follow-up, whole-genome copy-number profiles, and varying degree of clinical details, including age, gender, metastatic status, and histological subtype. This set of cases was acquired from several hospitals and tumour banks around the globe. Therefore, patients in the discovery set represent a heterogeneously treated group with diverse ethnic backgrounds. In contrast, the validation set consisted of medulloblastoma patients who were uniformly treated at a single institution in Moscow (Burdenko Hospital).

All available clinical variables and molecular markers were tested for prognostic association in the discovery set. Several clinical variables, such as metastatic status and age group, were categorized in multiple different ways, due to disagreements in the literature and clinical practice across continents. Due to the large number of candidate markers tested, a rigorous selection procedure was applied in order to select a small number of candidates to be validated in the external validation set using fluorescence in situ hybridization (FISH), which is routinely performed in modern pathology laboratories within hospitals.

Accordingly, the clinical and molecular candidate biomarkers were assessed by three approaches. First, the candidates were assessed by a cross-validation method, in order to estimate the expected validation rate of the biomarker. That is, whether the biomarker will likely validate in an independent cohort. Second, the sample size required for further validation in a prospective study was estimated for each candidate. Prognostic markers with small effect size (i.e. hazard ratio) or with low frequency may need impractically large sizes and are thus clinically irrelevant. Third, the candidates were combined in multivariate Cox proportional-hazards models in order to assess whether the biomarkers have prognostic values independent of one another. Biomarkers were prioritized by high validation rates, reasonably small sample sizes, and/or prognostically significance in multivariate models. The selected biomarkers were then used to construct the risk stratification models for each medulloblastoma subgroup.

The proposed risk stratification models represent promising candidates for future prospective trials. The constituent biomarkers were selected based on analyses within a heterogeneous discovery set, and are likely generalizable to different patient popula-

tions. For a specific treatment protocol within a specific patient population, there may be prognostic markers that have better prognostic value, particularly those that were not assessed in the present study due to scope. Notwithstanding these limitations, the proposed risk stratification models have been validated in an independent cohort, and can serve as the basis for the informed design of a future prospective trial.

Rare cytogenetic events

Some molecular biomarker candidates (e.g. *MYC* amplification, chr17 gain) have only been observed in a relatively small number of patients (≈ 10). Notwithstanding their infrequency in specific subgroups of medulloblastoma (a rare disease), their prognostic significances are supported by log-rank tests, likely due to their large 'effect size' (i.e. hazard ratio). Such biomarker candidates, however, have low expected validation rates from cross-validation and large estimated sample sizes from power analysis. On accounts of their potential therapeutic impact, these candidates were nonetheless included in the risk stratification models based on their independent prognostic significance under multivariate Cox models. Indeed, the candidates were ultimately validated to be *bona fide* prognostic biomarkers in the external validation set.

Isolated vs. non-isolated events

Isolated arm events occur in the absence of whole-chromosome event; non-isolated events may occur in the context of a whole-chromosome event.

Publications Arising from Thesis

[1] **Shih***, D. J., Northcott*, P. A., Remke*, M., Korshunov, A., Ramaswamy, V., Kool, M., Luu, B., Yao, Y., Wang, X., Dubuc, A., Garzia, L., Peacock, J., Mack, S., Wu, X., Rolider, A., orrissy, S. M., Cavalli, F., Jones, D. T., Zitterbart, K., Faria, C. C., Sch'uller, U., Kren, L., Kumabe, T., Tominaga, T., Ra, Y. S., Garami, M., Hauser, P., Chan, J. A., Robinson, S., Bognár, L., Klekner, A., Saad, A. G., Liau, L. M., Albrecht, S., Fontebasso, A., Cinalli, G., Antonellis, P. D., Zollo, M., Cooper, M. K., Thompson, R. C., Bailey, S., Lindsey, J. C., Rocco, C. D., Massimi, L., Michiels, E. M., Scherer, S. W., Phillips, J. J., Gupta, N., Fan, X., Muraszko, K. M., Vibhakar, R., Eberhart, C. G., Fouladi, M., Lach, B., Jung, S., Wechsler-Reya, R. J., Fèvre-Montange, M., Jouvet, A., Jabado, N., Pollack, I. F., Weiss, W. A., Lee, J.-Y., Cho, B.-K., Kim, S.-K., Wang, K.-C., Leonard, J. R., Rubin, J. B., de Torres, C., Lavarino, C., Mora, J., Cho, Y.-J., Tabori, U., Olson, J. M., Gajjar, A., Packer, R. J., Rutkowski, S., Pomeroy, S. L., French, P. J., Kloosterhof, N. K., Kros, J. M., Meir, E. G. V., Clifford, S. C., Bourdeaut, F., Delattre, O., Doz, F., Hawkins, C. E., Malkin, D., Grajkowska, W. A., Perek-Polnik, M., Bouffet, E., Rutka, J. T., Pfister, S. M., and Taylor, M. D. Cytogenetic Prognostication within Medulloblastoma Subgroups. *Journal of Clinical Oncology* **32**(9), 886–96 (2014).

[2] **Shih***, D. J. H., Northcott*, P. A., Peacock, J., Garzia, L., Morrissy, A. S., Zichner, T., Stutz, A. M., Korshunov, A., Reimand, J., Schumacher, S. E., Beroukhim, R., Ellison, D. W., Marshall, C. R., Lionel, A. C., Mack, S., Dubuc, A., Yao, Y., Ramaswamy, V., Luu, B., Rolider, A., Cavalli, F. M. G., Wang, X., Remke, M., Wu, X., Chiu, R. Y. B., Chu, A., Chuah, E., Corbett, R. D., Hoad, G. R., Jackman, S. D., Li, Y., Lo, A., Mungall, K. L., Nip, K. M., Qian, J. Q., Raymond, A. G. J., Thiessen, N. T., Varhol, R. J., Birol, I., Moore, R. A., Mungall, A. J., Holt, R., Kawauchi, D., Roussel, M. F., Kool, M., Jones, D. T. W., Witt, H., Fernandez-L, A., Kenney, A. M., Wechsler-Reya, R. J., Dirks, P., Aviv, T., Grajkowska, W. A., Perek-Polnik, M., Haberler, C. C., Delattre, O., Reynaud, S. S., Doz, F. F., Pernet-Fattet, S. S., Cho, B.-K., Kim, S.-K., Wang, K.-C., Scheurlen, W., Eberhart, C. G., Fèvre-Montange, M., Jouvet, A., Pollack, I. F., Fan, X., Muraszko, K. M., Gillespie, G. Y., Di Rocco, C., Massimi, L., Michiels, E. M. C., Kloosterhof, N. K., French, P. J., Kros, J. M., Olson, J. M., Ellenbogen, R. G., Zitterbart, K., Kren, L., Thompson, R. C., Cooper, M. K., Lach, B., McLendon, R. E., Bigner, D. D., Fontebasso, A., Albrecht, S., Jabado, N., Lindsey, J. C., Bailey, S., Gupta, N., Weiss, W. A., Bognár, L., Klekner, A., Van Meter, T. E., Kumabe, T., Tominaga, T., Elbabaa, S. K., Leonard, J. R., Rubin, J. B., Liau, L. M., Van Meir, E. G., Fouladi, M., Nakamura, H., Cinalli, G., Garami, M., Hauser, P., Saad, A. G., Iolascon, A., Jung, S., Carlotti, C. G., Vibhakar, R., Ra, Y. S., Robinson, S., Zollo, M., Faria, C. C., Chan, J. A., Levy, M. L., Sorensen, P. H. B., Meyerson, M., Pomeroy, S. L., Cho, Y.-J., Bader, G. D., Tabori, U., Hawkins, C. E., Bouffet, E., Scherer, S. W., Rutka, J. T., Malkin, D., Clifford, S. C., Jones, S. J. M., Korbel, J. O., Pfister, S. M., Marra, M. A., and Taylor, M. D. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**(7409), 49–56 (2012).

[3] Northcott, P. A., **Shih**, D. J. H., Remke, M., Cho, Y.-J., Kool, M., Hawkins, C., Eberhart, C. G., Dubuc, A., Guettouche, T., Cardentey, Y., Bouffet, E., Pomeroy, S. L., Marra, M., Malkin, D., Rutka, J. T., Korshunov, A., Pfister, S., and Taylor, M. D. Rapid, reliable, and reproducible

molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathol* **123**(4), 615–26 (2012).

[4] Natarajan, S., Li, Y., Miller, E. E., **Shih**, D. J., Taylor, M. D., Stearns, T. M., Bronson, R. T., Ackerman, S. L., Yoon, J. K., and Yun, K. Notch1-induced brain tumor models the sonic hedgehog subgroup of human medulloblastoma. *Cancer Res* **73**(17), 5381–90 (2013).

[5] Agnihotri, S., Gugel, I., Remke, M., Bornemann, A., Pantazis, G., Mack, S. C., **Shih**, D., Singh, S. K., Sabha, N., Taylor, M. D., Tatagiba, M., Zadeh, G., and Krischek, B. Gene-expression profiling elucidates molecular signaling networks that can be therapeutically targeted in vestibular schwannoma. *J Neurosurg* **121**(6), 1434–45 (2014).

[6] Diaz, R. J., Golbourn, B., Faria, C., Picard, D., **Shih**, D., Raynaud, D., Leadly, M., MacKenzie, D., Bryant, M., Bebenek, M., Smith, C. A., Taylor, M. D., Huang, A., and Rutka, J. T. Mechanism of action and therapeutic efficacy of Aurora kinase B inhibition in MYC overexpressing medulloblastoma. *Oncotarget* **6**(5), 3359–74 (2015).

[7] Zhukova, N., Ramaswamy, V., Remke, M., Martin, D. C., Castelo-Branco, P., Zhang, C. H., Fraser, M., Tse, K., Poon, R., **Shih**, D. J. H., Baskin, B., Ray, P. N., Bouffet, E., Dirks, P., von Bueren, A. O., Pfaff, E., Korshunov, A., Jones, D. T. W., Northcott, P. A., Kool, M., Pugh, T. J., Pomeroy, S. L., Cho, Y.-J., Pietsch, T., Gessi, M., Rutkowski, S., Bognar, L., Cho, B.-K., Eberhart, C. G., Conter, C. F., Fouladi, M., French, P. J., Grajkowska, W. A., Gupta, N., Hauser, P., Jabado, N., Vasiljevic, A., Jung, S., Kim, S.-K., Klekner, A., Kumabe, T., Lach, B., Leonard, J. R., Liau, L. M., Massimi, L., Pollack, I. F., Ra, Y. S., Rubin, J. B., Van Meir, E. G., Wang, K.-C., Weiss, W. A., Zitterbart, K., Bristow, R. G., Alman, B., Hawkins, C. E., Malkin, D., Clifford, S. C., Pfister, S. M., Taylor, M. D., and Tabori, U. WNT activation by lithium abrogates TP53 mutation associated radiation resistance in medulloblastoma. *Acta Neuropathol Commun* **2**, 174 (2014).

[8] Perreault, S., Ramaswamy, V., Achrol, A. S., Chao, K., Liu, T. T., **Shih**, D., Remke, M., Schubert, S., Bouffet, E., Fisher, P. G., Partap, S., Vogel, H., Taylor, M. D., Cho, Y. J., and Yeom, K. W. MRI surrogates for molecular subgroups of medulloblastoma. *AJNR Am J Neuroradiol* **35**(7), 1263–9 (2014).

[9] Kool, M., Jones, D. T. W., Jager, N., Northcott, P. A., Pugh, T. J., Hovestadt, V., Piro, R. M., Esparza, L. A., Markant, S. L., Remke, M., Milde, T., Bourdeaut, F., Ryzhova, M., Sturm, D., Pfaff, E., Stark, S., Hutter, S., Seker-Cin, H., Johann, P., Bender, S., Schmidt, C., Rausch, T., **Shih**, D., Reimand, J., Sieber, L., Wittmann, A., Linke, L., Witt, H., Weber, U. D., Zapatka, M., Konig, R., Beroukhim, R., Bergthold, G., van Sluis, P., Volckmann, R., Koster, J., Versteeg, R., Schmidt, S., Wolf, S., Lawerenz, C., Bartholomae, C. C., von Kalle, C., Unterberg, A., Herold-Mende, C., Hofer, S., Kulozik, A. E., von Deimling, A., Scheurlen, W., Felsberg, J., Reifenberger, G., Hasselblatt, M., Crawford, J. R., Grant, G. A., Jabado, N., Perry, A., Cowdrey, C., Croul, S., Zadeh, G., Korbel, J. O., Doz, F., Delattre, O., Bader, G. D., McCabe, M. G., Collins, V. P., Kieran, M. W., Cho, Y.-J., Pomeroy, S. L., Witt, O., Brors, B., Taylor, M. D., Sch´uller, U., Korshunov, A., Eils, R., Wechsler-Reya, R. J., Lichter, P., and Pfister, S. M. Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25**(3), 393–405 (2014).

[10] Ramaswamy, V., Remke, M., **Shih**, D., Wang, X., Northcott, P. A., Faria, C. C., Raybaud, C., Tabori, U., Hawkins, C., Rutka, J., Taylor, M. D., and Bouffet, E. Duration of the pre-diagnostic interval in medulloblastoma is subgroup dependent. *Pediatr Blood Cancer* **61**(7), 1190–4 (2014).

[11] Ramaswamy, V., Remke, M., Bouffet, E., Faria, C. C., Perreault, S., Cho, Y.-J., **Shih**, D. J., Luu, B., Dubuc, A. M., Northcott, P. A., Sch´uller, U., Gururangan, S., McLendon, R., Bigner, D., Fouladi, M., Ligon, K. L., Pomeroy, S. L., Dunn, S., Triscott, J., Jabado, N., Fontebasso, A., Jones, D. T. W., Kool, M., Karajannis, M. A., Gardner, S. L., Zagzag, D., Nunes, S., Pimentel,

J., Mora, J., Lipp, E., Walter, A. W., Ryzhova, M., Zheludkova, O., Kumirova, E., Alshami, J., Croul, S. E., Rutka, J. T., Hawkins, C., Tabori, U., Codispoti, K.-E. T., Packer, R. J., Pfister, S. M., Korshunov, A., and Taylor, M. D. Recurrence patterns across medulloblastoma subgroups: an integrated clinical and molecular analysis. *Lancet Oncol* **14**(12), 1200–7 (2013).

[12] Dey, J., Dubuc, A. M., Pedro, K. D., Thirstrup, D., Mecham, B., Northcott, P. A., Wu, X., **Shih**, D., Tapscott, S. J., LeBlanc, M., Taylor, M. D., and Olson, J. M. MyoD is a tumor suppressor gene in medulloblastoma. *Cancer Res* **73**(22), 6828–37 (2013).

[13] Markant, S. L., Esparza, L. A., Sun, J., Barton, K. L., McCoig, L. M., Grant, G. A., Crawford, J. R., Levy, M. L., Northcott, P. A., **Shih**, D., Remke, M., Taylor, M. D., and Wechsler-Reya, R. J. Targeting sonic hedgehog-associated medulloblastoma through inhibition of Aurora and Polo-like kinases. *Cancer Res* **73**(20), 6310–22 (2013).

[14] Zhukova, N., Ramaswamy, V., Remke, M., Pfaff, E., **Shih**, D. J. H., Martin, D. C., Castelo-Branco, P., Baskin, B., Ray, P. N., Bouffet, E., von Bueren, A. O., Jones, D. T. W., Northcott, P. A., Kool, M., Sturm, D., Pugh, T. J., Pomeroy, S. L., Cho, Y.-J., Pietsch, T., Gessi, M., Rutkowski, S., Bognár, L., Klekner, A., Cho, B.-K., Kim, S.-K., Wang, K.-C., Eberhart, C. G., Fèvre-Montange, M., Fouladi, M., French, P. J., Kros, M., Grajkowska, W. A., Gupta, N., Weiss, W. A., Hauser, P., Jabado, N., Jouvet, A., Jung, S., Kumabe, T., Lach, B., Leonard, J. R., Rubin, J. B., Liau, L. M., Massimi, L., Pollack, I. F., Shin Ra, Y., Van Meir, E. G., Zitterbart, K., Sch´uller, U., Hill, R. M., Lindsey, J. C., Schwalbe, E. C., Bailey, S., Ellison, D. W., Hawkins, C., Malkin, D., Clifford, S. C., Korshunov, A., Pfister, S., Taylor, M. D., and Tabori, U. Subgroup-specific prognostic implications of TP53 mutation in medulloblastoma. *J Clin Oncol* **31**(23), 2927–35 (2013).

[15] Remke, M., Ramaswamy, V., Peacock, J., **Shih**, D. J. H., Koelsche, C., Northcott, P. A., Hill, N., Cavalli, F. M. G., Kool, M., Wang, X., Mack, S. C., Barszczyk, M., Morrissy, A. S., Wu, X., Agnihotri, S., Luu, B., Jones, D. T. W., Garzia, L., Dubuc, A. M., Zhukova, N., Vanner, R., Kros, J. M., French, P. J., Van Meir, E. G., Vibhakar, R., Zitterbart, K., Chan, J. A., Bognár, L., Klekner, A., Lach, B., Jung, S., Saad, A. G., Liau, L. M., Albrecht, S., Zollo, M., Cooper, M. K., Thompson, R. C., Delattre, O. O., Bourdeaut, F., Doz, F. F., Garami, M., Hauser, P., Carlotti, C. G., Van Meter, T. E., Massimi, L., Fults, D., Pomeroy, S. L., Kumabe, T., Ra, Y. S., Leonard, J. R., Elbabaa, S. K., Mora, J., Rubin, J. B., Cho, Y.-J., McLendon, R. E., Bigner, D. D., Eberhart, C. G., Fouladi, M., Wechsler-Reya, R. J., Faria, C. C., Croul, S. E., Huang, A., Bouffet, E., Hawkins, C. E., Dirks, P. B., Weiss, W. A., Sch´uller, U., Pollack, I. F., Rutkowski, S., Meyronet, D., Jouvet, A., Fèvre-Montange, M., Jabado, N., Perek-Polnik, M., Grajkowska, W. A., Kim, S.-K., Rutka, J. T., Malkin, D., Tabori, U., Pfister, S. M., Korshunov, A., von Deimling, A., and Taylor, M. D. TERT promoter mutations are highly recurrent in SHH subgroup medulloblastoma. *Acta Neuropathol* **126**(6), 917–29 (2013).

[16] Dubuc, A. M., Remke, M., Korshunov, A., Northcott, P. A., Zhan, S. H., Mendez-Lago, M., Kool, M., Jones, D. T. W., Unterberger, A., Morrissy, A. S., **Shih**, D., Peacock, J., Ramaswamy, V., Rolider, A., Wang, X., Witt, H., Hielscher, T., Hawkins, C., Vibhakar, R., Croul, S., Rutka, J. T., Weiss, W. A., Jones, S. J. M., Eberhart, C. G., Marra, M. A., Pfister, S. M., and Taylor, M. D. Aberrant patterns of H3K4 and H3K27 histone lysine methylation occur across subgroups in medulloblastoma. *Acta Neuropathol* **125**(3), 373–84 (2013).

[17] Dubuc, A. M., Morrissy, A. S., Kloosterhof, N. K., Northcott, P. A., Yu, E. P. Y., **Shih**, D., Peacock, J., Grajkowska, W., van Meter, T., Eberhart, C. G., Pfister, S., Marra, M. A., Weiss, W. A., Scherer, S. W., Rutka, J. T., French, P. J., and Taylor, M. D. Subgroup-specific alternative splicing in medulloblastoma. *Acta Neuropathol* **123**(4), 485–99 (2012).

[18] Wu, X., Northcott, P. A., Dubuc, A., Dupuy, A. J., **Shih**, D. J. H., Witt, H., Croul, S., Bouffet, E., Fults, D. W., Eberhart, C. G., Garzia, L., Van Meter, T., Zagzag, D., Jabado, N., Schwartzentruber,

J., Majewski, J., Scheetz, T. E., Pfister, S. M., Korshunov, A., Li, X.-N., Scherer, S. W., Cho, Y.-J., Akagi, K., MacDonald, T. J., Koster, J., McCabe, M. G., Sarver, A. L., Collins, V. P., Weiss, W. A., Largaespada, D. A., Collier, L. S., and Taylor, M. D. Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature* **482**(7386), 529–33 (2012).

[19] Jones, D. T. W., Jager, N., Kool, M., Zichner, T., Hutter, B., Sultan, M., Cho, Y.-J., Pugh, T. J., Hovestadt, V., Stutz, A. M., Rausch, T., Warnatz, H.-J., Ryzhova, M., Bender, S., Sturm, D., Pleier, S., Cin, H., Pfaff, E., Sieber, L., Wittmann, A., Remke, M., Witt, H., Hutter, S., Tzaridis, T., Weischenfeldt, J., Raeder, B., Avci, M., Amstislavskiy, V., Zapatka, M., Weber, U. D., Wang, Q., Lasitschka, B., Bartholomae, C. C., Schmidt, M., von Kalle, C., Ast, V., Lawerenz, C., Eils, J., Kabbe, R., Benes, V., van Sluis, P., Koster, J., Volckmann, R., **Shih**, D., Betts, M. J., Russell, R. B., Coco, S., Tonini, G. P., Schüller, U., Hans, V., Graf, N., Kim, Y.-J., Monoranu, C., Roggendorf, W., Unterberg, A., Herold-Mende, C., Milde, T., Kulozik, A. E., von Deimling, A., Witt, O., Maass, E., Rossler, J., Ebinger, M., Schuhmann, M. U., Fruhwald, M. C., Hasselblatt, M., Jabado, N., Rutkowski, S., von Bueren, A. O., Williamson, D., Clifford, S. C., McCabe, M. G., Collins, V. P., Wolf, S., Wiemann, S., Lehrach, H., Brors, B., Scheurlen, W., Felsberg, J., Reifenberger, G., Northcott, P. A., Taylor, M. D., Meyerson, M., Pomeroy, S. L., Yaspo, M.-L., Korbel, J. O., Korshunov, A., Eils, R., Pfister, S. M., and Lichter, P. Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**(7409), 100–5 (2012).

[20] Rausch, T., Jones, D. T. W., Zapatka, M., Stutz, A. M., Zichner, T., Weischenfeldt, J., Jager, N., Remke, M., **Shih**, D., Northcott, P. A., Pfaff, E., Tica, J., Wang, Q., Massimi, L., Witt, H., Bender, S., Pleier, S., Cin, H., Hawkins, C., Beck, C., von Deimling, A., Hans, V., Brors, B., Eils, R., Scheurlen, W., Blake, J., Benes, V., Kulozik, A. E., Witt, O., Martin, D., Zhang, C., Porat, R., Merino, D. M., Wasserman, J., Jabado, N., Fontebasso, A., Bullinger, L., Rucker, F. G., Dohner, K., Dohner, H., Koster, J., Molenaar, J. J., Versteeg, R., Kool, M., Tabori, U., Malkin, D., Korshunov, A., Taylor, M. D., Lichter, P., Pfister, S. M., and Korbel, J. O. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**(1-2), 59–71 (2012).

[21] Northcott, P. A., Hielscher, T., Dubuc, A., Mack, S., **Shih**, D., Remke, M., Al-Halabi, H., Albrecht, S., Jabado, N., Eberhart, C. G., Grajkowska, W., Weiss, W. A., Clifford, S. C., Bouffet, E., Rutka, J. T., Korshunov, A., Pfister, S., and Taylor, M. D. Pediatric and adult sonic hedgehog medulloblastomas are clinically and molecularly distinct. *Acta Neuropathol* **122**(2), 231–40 (2011).

[22] He, X., Zhang, L., Chen, Y., Remke, M., **Shih**, D., Lu, F., Wang, H., Deng, Y., Yu, Y., Xia, Y., Wu, X., Ramaswamy, V., Hu, T., Wang, F., Zhou, W., Burns, D. K., Kim, S. H., Kool, M., Pfister, S. M., Weinstein, L. S., Pomeroy, S. L., Gilbertson, R. J., Rubin, J. B., Hou, Y., Wechsler-Reya, R., Taylor, M. D., and Lu, Q. R. The G protein alpha subunit Galphas is a tumor suppressor in Sonic hedgehog-driven medulloblastoma. *Nat Med* **20**(9), 1035–42 (2014).

[23] Chow, K.-H., Shin, D.-M., Jenkins, M. H., Miller, E. E., **Shih**, D. J., Choi, S., Low, B. E., Philip, V., Rybinski, B., Bronson, R. T., Taylor, M. D., and Yun, K. Epigenetic states of cells of origin and tumor evolution drive tumor-initiating cell phenotype and tumor heterogeneity. *Cancer Res* **74**(17), 4864–74 (2014).

[24] Northcott, P. A., Lee, C., Zichner, T., Stutz, A. M., Erkek, S., Kawauchi, D., **Shih**, D. J. H., Hovestadt, V., Zapatka, M., Sturm, D., Jones, D. T. W., Kool, M., Remke, M., Cavalli, F. M. G., Zuyderduyn, S., Bader, G. D., VandenBerg, S., Esparza, L. A., Ryzhova, M., Wang, W., Wittmann, A., Stark, S., Sieber, L., Seker-Cin, H., Linke, L., Kratochwil, F., Jager, N., Buchhalter, I., Imbusch, C. D., Zipprich, G., Raeder, B., Schmidt, S., Diessl, N., Wolf, S., Wiemann, S., Brors, B., Lawerenz, C., Eils, J., Warnatz, H.-J., Risch, T., Yaspo, M.-L., Weber, U. D., Bartholomae, C. C., von Kalle, C., Turanyi, E., Hauser, P., Sanden, E., Darabi, A., Siesjo, P., Sterba, J., Zitterbart, K.,

Sumerauer, D., van Sluis, P., Versteeg, R., Volckmann, R., Koster, J., Schuhmann, M. U., Ebinger, M., Grimes, H. L., Robinson, G. W., Gajjar, A., Mynarek, M., von Hoff, K., Rutkowski, S., Pietsch, T., Scheurlen, W., Felsberg, J., Reifenberger, G., Kulozik, A. E., von Deimling, A., Witt, O., Eils, R., Gilbertson, R. J., Korshunov, A., Taylor, M. D., Lichter, P., Korbel, J. O., Wechsler-Reya, R. J., and Pfister, S. M. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**(7510), 428–34 (2014).

[*] These authors contributed equally.

References

[25] Ward, E., DeSantis, C., Robbins, A., Kohler, B., and Jemal, A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin* **64**(2), 83–103 (2014).

[26] Murphy, S. L., Xu, J., and Kochanek, K. D. Deaths: Final data for 2010. national vital statistics reports. Technical Report 4, National Center for Health Statistics, Hyattsville, MD, (2013).

[27] Howlader, N., Noone, A. M., Krapcho, M., Garshell, J., Miller, D., Altekruse, S. F., Kosary, C. L., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D. R., Chen, H. S., Feuer, E. J., and A, C. K. Seer cancer statistics review, 1975-2011. Technical report, National Cancer Institute, Bethesda, MD, , April (2014).

[28] Ostrom, Q. T., Gittleman, H., Liao, P., Rouse, C., Chen, Y., Dowling, J., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007-2011. *Neuro Oncol* **16 Suppl 4**, iv1–63 (2014).

[29] Lous, D. N., Ohgaki, H., Wiestler, O. D., and Cavenee, W. K., editors. *World Health Organization Classification of Tumours of the Central Nervous System*. World Health Organization, (2007).

[30] Kleihues, P., Burger, P. C., and Scheithauer, B. W. The new WHO classification of brain tumours. *Brain Pathol* **3**(3), 255–68 (1993).

[31] Chatty, E. M. and Earle, K. M. Medulloblastoma. A report of 201 cases with emphasis on the relationship of histologic variants to survival. *Cancer* **28**(4), 977–83 (1971).

[32] Rudin, C. M., Hann, C. L., Laterra, J., Yauch, R. L., Callahan, C. A., Fu, L., Holcomb, T., Stinson, J., Gould, S. E., Coleman, B., LoRusso, P. M., Von Hoff, D. D., de Sauvage, F. J., and Low, J. A. Treatment of medulloblastoma with hedgehog pathway inhibitor GDC-0449. *N Engl J Med* **361**(12), 1173–8 (2009).

[33] Inda, M.-d.-M., Bonavia, R., Mukasa, A., Narita, Y., Sah, D. W. Y., Vandenberg, S., Brennan, C., Johns, T. G., Bachoo, R., Hadwiger, P., Tan, P., Depinho, R. A., Cavenee, W., and Furnari, F. Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. *Genes Dev* **24**(16), 1731–45 (2010).

[34] Taylor, M. D., Northcott, P. A., Korshunov, A., Remke, M., Cho, Y.-J., Clifford, S. C., Eberhart, C. G., Parsons, D. W., Rutkowski, S., Gajjar, A., Ellison, D. W., Lichter, P., Gilbertson, R. J., Pomeroy, S. L., Kool, M., and Pfister, S. M. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol* **123**(4), 465–72 (2012).

[35] Northcott, P. A., Korshunov, A., Witt, H., Hielscher, T., Eberhart, C. G., Mack, S., Bouffet, E., Clifford, S. C., Hawkins, C. E., French, P., Rutka, J. T., Pfister, S., and Taylor, M. D. Medulloblastoma comprises four distinct molecular variants. *J Clin Oncol* **29**(11), 1408–14 (2011).

[36] Kool, M., Korshunov, A., Remke, M., Jones, D. T. W., Schlanstein, M., Northcott, P. A., Cho, Y.-J., Koster, J., Schouten-van Meeteren, A., van Vuurden, D., Clifford, S. C., Pietsch, T., von Bueren, A. O., Rutkowski, S., McCabe, M., Collins, V. P., Backlund, M. L., Haberler, C., Bourdeaut, F., Delattre, O., Doz, F., Ellison, D. W., Gilbertson, R. J., Pomeroy, S. L., Taylor, M. D., Lichter, P., and Pfister, S. M. Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathol* **123**(4), 473–84 (2012).

[37] Remke, M., Hielscher, T., Northcott, P. A., Witt, H., Ryzhova, M., Wittmann, A., Benner, A., von Deimling, A., Scheurlen, W., Perry, A., Croul, S., Kulozik, A. E., Lichter, P., Taylor, M. D., Pfister, S. M., and Korshunov, A. Adult medulloblastoma comprises three major molecular variants. *J Clin Oncol* **29**(19), 2717–23 (2011).

[38] Cho, Y.-J., Tsherniak, A., Tamayo, P., Santagata, S., Ligon, A., Greulich, H., Berhoukim, R., Amani, V., Goumnerova, L., Eberhart, C. G., Lau, C. C., Olson, J. M., Gilbertson, R. J., Gajjar, A., Delattre, O., Kool, M., Ligon, K., Meyerson, M., Mesirov, J. P., and Pomeroy, S. L. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J Clin Oncol* **29**(11), 1424–30 (2011).

[39] Smoll, N. R. and Drummond, K. J. The incidence of medulloblastomas and primitive neurectodermal tumours in adults and children. *J Clin Neurosci* **19**(11), 1541–4 (2012).

[40] Scheurlen, W., Sorensen, N., Roggendorf, W., and Kuhl, J. Molecular analysis of medulloblastomas occurring simultaneously in monozygotic twins. *Eur J Pediatr* **155**(10), 880–4 (1996).

[41] Hemminki, K., Tretli, S., Sundquist, J., Johannesen, T. B., and Granstrom, C. Familial risks in nervous-system tumours: a histology-specific analysis from Sweden and Norway. *Lancet Oncol* **10**(5), 481–8 (2009).

[42] Pearson, A. D., Craft, A. W., Ratcliffe, J. M., Birch, J. M., Morris-Jones, P., and Roberts, D. F. Two families with the Li-Fraumeni cancer family syndrome. *J Med Genet* **19**(5), 362–5 (1982).

[43] Barel, D., Avigad, S., Mor, C., Fogel, M., Cohen, I. J., and Zaizov, R. A novel germ-line mutation in the noncoding region of the p53 gene in a Li-Fraumeni family. *Cancer Genet Cytogenet* **103**(1), 1–6 (1998).

[44] Guran, S., Tunca, Y., and Imirzalioglu, N. Hereditary TP53 codon 292 and somatic P16INK4A codon 94 mutations in a Li-Fraumeni syndrome family. *Cancer Genet Cytogenet* **113**(2), 145–51 (1999).

[45] Yamazaki, Y., Sugita, K., Kurosawa, H., Ozawa, T., Eguchi, M., Wakai, S., and Hata, J. Malignant fibrous histiocytoma preceding medulloblastoma. *J Pediatr Hematol Oncol* **22**(5), 480–1 (2000).

[46] Garre, M. L., Cama, A., Bagnasco, F., Morana, G., Giangaspero, F., Brisigotti, M., Gambini, C., Forni, M., Rossi, A., Haupt, R., Nozza, P., Barra, S., Piatelli, G., Viglizzo, G., Capra, V., Bruno, W., Pastorino, L., Massimino, M., Tumolo, M., Fidani, P., Dallorso, S., Schumacher, R. F., Milanaccio, C., and Pietsch, T. Medulloblastoma variants: age-dependent occurrence and relation to Gorlin syndrome–a new clinical perspective. *Clin Cancer Res* **15**(7), 2463–71 (2009).

[47] Villani, A., Tabori, U., Schiffman, J., Shlien, A., Beyene, J., Druker, H., Novokmet, A., Finlay, J., and Malkin, D. Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: a prospective observational study. *Lancet Oncol* **12**(6), 559–67 (2011).

[48] Hart, R. M., Kimler, B. F., Evans, R. G., and Park, C. H. Radiotherapeutic management of medulloblastoma in a pediatric patient with ataxia telangiectasia. *Int J Radiat Oncol Biol Phys* **13**(8), 1237–40 (1987).

[49] Reiman, A., Srinivasan, V., Barone, G., Last, J. I., Wootton, L. L., Davies, E. G., Verhagen, M. M., Willemsen, M. A., Weemaes, C. M., Byrd, P. J., Izatt, L., Easton, D. F., Thompson, D. J., and Taylor, A. M. Lymphoid tumours and breast cancer in ataxia telangiectasia; substantial protective effect of residual ATM kinase activity against childhood tumours. *Br J Cancer* **105**(4), 586–91 (2011).

[50] Offit, K., Levran, O., Mullaney, B., Mah, K., Nafa, K., Batish, S. D., Diotti, R., Schneider, H., Deffenbaugh, A., Scholl, T., Proud, V. K., Robson, M., Norton, L., Ellis, N., Hanenberg, H., and Auerbach, A. D. Shared genetic susceptibility to breast cancer, brain tumors, and Fanconi anemia. *J Natl Cancer Inst* **95**(20), 1548–51 (2003).

[51] Hirsch, B., Shimamura, A., Moreau, L., Baldinger, S., Hag-alshiekh, M., Bostrom, B., Sencer, S., and D'Andrea, A. D. Association of biallelic BRCA2/FANCD1 mutations with spontaneous chromosomal instability and solid tumors of childhood. *Blood* **103**(7), 2554–9 (2004).

[52] Reid, S., Renwick, A., Seal, S., Baskcomb, L., Barfoot, R., Jayatilake, H., Pritchard-Jones, K., Stratton, M. R., Ridolfi-Luthy, A., and Rahman, N. Biallelic BRCA2 mutations are associated with multiple malignancies in childhood including familial Wilms tumour. *J Med Genet* **42**(2), 147–51 (2005).

[53] Dewire, M. D., Ellison, D. W., Patay, Z., McKinnon, P. J., Sanders, R. P., and Gajjar, A. Fanconi anemia and biallelic BRCA2 mutation diagnosed in a young child with an embryonal CNS tumor. *Pediatr Blood Cancer* **53**(6), 1140–2 (2009).

[54] Ciara, E., Piekutowska-Abramczuk, D., Popowska, E., Grajkowska, W., Barszcz, S., Perek, D., Dembowska-Baginska, B., Perek-Polnik, M., Kowalewska, E., Czajnska, A., Syczewska, M., Czornak, K., Krajewska-Walasek, M., Roszkowski, M., and Chrzanowska, K. H. Heterozygous germ-line mutations in the NBN gene predispose to medulloblastoma in pediatric patients. *Acta Neuropathol* **119**(3), 325–34 (2010).

[55] Alexiou, G. A., Siozos, G., Stefanaki, K., Vlachakis, E., Sfakianos, G., Prodromou, N., and Moschovi, M. Medulloblastoma in a child with fragile X syndrome. *Neuropediatrics* **43**(3), 155–8 (2012).

[56] Bakhshi, S., Cerosaletti, K. M., Concannon, P., Bawle, E. V., Fontanesi, J., Gatti, R. A., and Bhambhani, K. Medulloblastoma with adverse reaction to radiation therapy in nijmegen breakage syndrome. *J Pediatr Hematol Oncol* **25**(3), 248–51 (2003).

[57] Distel, L., Neubauer, S., Varon, R., Holter, W., and Grabenbauer, G. Fatal toxicity following radio- and chemotherapy of medulloblastoma in a child with unrecognized Nijmegen breakage syndrome. *Med Pediatr Oncol* **41**(1), 44–8 (2003).

[58] Martinez-Lage, J. F., Salcedo, C., Corral, M., and Poza, M. Medulloblastomas in neurofibromatosis type 1. Case report and literature review. *Neurocirugia (Astur)* **13**(2), 128–31 (2002).

[59] Slade, I., Bacchelli, C., Davies, H., Murray, A., Abbaszadeh, F., Hanks, S., Barfoot, R., Burke, A., Chisholm, J., Hewitt, M., Jenkinson, H., King, D., Morland, B., Pizer, B., Prescott, K., Saggar, A., Side, L., Traunecker, H., Vaidya, S., Ward, P., Futreal, P. A., Vujanic, G., Nicholson, A. G., Sebire, N., Turnbull, C., Priest, J. R., Pritchard-Jones, K., Houlston, R., Stiller, C., Stratton, M. R., Douglas, J., and Rahman, N. DICER1 syndrome: clarifying the diagnosis, clinical features and management implications of a pleiotropic tumour predisposition syndrome. *J Med Genet* **48**(4), 273–8 (2011).

[60] Bourdeaut, F., Miquel, C., Richer, W., Grill, J., Zerah, M., Grison, C., Pierron, G., Amiel, J., Krucker, C., Radvanyi, F., Brugieres, L., and Delattre, O. Rubinstein-Taybi syndrome predisposing to non-WNT, non-SHH, group 3 medulloblastoma. *Pediatr Blood Cancer* **61**(2), 383–6 (2014).

[61] Wolter, M., Reifenberger, J., Sommer, C., Ruzicka, T., and Reifenberger, G. Mutations in the human homologue of the Drosophila segment polarity gene patched (PTCH) in sporadic basal cell carcinomas of the skin and primitive neuroectodermal tumors of the central nervous system. *Cancer Res* **57**(13), 2581–5 (1997).

[62] Taylor, M. D., Liu, L., Raffel, C., Hui, C.-c., Mainprize, T. G., Zhang, X., Agatep, R., Chiappa, S., Gao, L., Lowrance, A., Hao, A., Goldstein, A. M., Stavrou, T., Scherer, S. W., Dura, W. T., Wainwright, B., Squire, J. A., Rutka, J. T., and Hogg, D. Mutations in SUFU predispose to medulloblastoma. *Nat Genet* **31**(3), 306–10 (2002).

[63] Crawford, J. R., Rood, B. R., Rossi, C. T., and Vezina, G. Medulloblastoma associated with novel PTCH mutation as primary manifestation of Gorlin syndrome. *Neurology* **72**(18), 1618 (2009).

[64] Brugieres, L., Pierron, G., Chompret, A., Paillerets, B. B.-d., Di Rocco, F., Varlet, P., Pierre-Kahn, A., Caron, O., Grill, J., and Delattre, O. Incomplete penetrance of the predisposition to medulloblastoma associated with germ-line SUFU mutations. *J Med Genet* **47**(2), 142–4 (2010).

[65] Jones, E. A., Sajid, M. I., Shenton, A., and Evans, D. G. Basal cell carcinomas in gorlin syndrome: a review of 202 patients. *J Skin Cancer* **2011**, 217378 (2011).

[66] Brugieres, L., Remenieras, A., Pierron, G., Varlet, P., Forget, S., Byrde, V., Bombled, J., Puget, S., Caron, O., Dufour, C., Delattre, O., Bressac-de Paillerets, B., and Grill, J. High frequency of germline SUFU mutations in children with desmoplastic/nodular medulloblastoma younger than 3 years of age. *J Clin Oncol* **30**(17), 2087–93 (2012).

[67] Hamilton, S. R., Liu, B., Parsons, R. E., Papadopoulos, N., Jen, J., Powell, S. M., Krush, A. J., Berk, T., Cohen, Z., and Tetu, B. The molecular basis of Turcot's syndrome. *N Engl J Med* **332**(13), 839–47 (1995).

[68] Goodrich, L. V., Milenkovic, L., Higgins, K. M., and Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**(5329), 1109–13 (1997).

[69] Wetmore, C., Eberhart, D. E., and Curran, T. The normal patched allele is expressed in medulloblastomas from mice with heterozygous germ-line mutation of patched. *Cancer Res* **60**(8), 2239–46 (2000).

[70] Wetmore, C., Eberhart, D. E., and Curran, T. Loss of p53 but not ARF accelerates medulloblastoma in mice heterozygous for patched. *Cancer Res* **61**(2), 513–6 (2001).

[71] Pazzaglia, S., Mancuso, M., Atkinson, M. J., Tanori, M., Rebessi, S., Majo, V. D., Covelli, V., Hahn, H., and Saran, A. High incidence of medulloblastoma following X-ray-irradiation of newborn Ptc1 heterozygous mice. *Oncogene* **21**(49), 7580–4 (2002).

[72] Gibson, P., Tong, Y., Robinson, G., Thompson, M. C., Currle, D. S., Eden, C., Kranenburg, T. A., Hogg, T., Poppleton, H., Martin, J., Finkelstein, D., Pounds, S., Weiss, A., Patay, Z., Scoggins, M., Ogg, R., Pei, Y., Yang, Z.-J., Brun, S., Lee, Y., Zindy, F., Lindsey, J. C., Taketo, M. M., Boop, F. A., Sanford, R. A., Gajjar, A., Clifford, S. C., Roussel, M. F., McKinnon, P. J., Gutmann, D. H., Ellison, D. W., Wechsler-Reya, R., and Gilbertson, R. J. Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**(7327), 1095–9 (2010).

[73] Heby-Henricson, K., Bergstrom, A., Rozell, B., Toftgard, R., and Teglund, S. Loss of Trp53 promotes medulloblastoma development but not skin tumorigenesis in Sufu heterozygous mutant mice. *Mol Carcinog* **51**(9), 754–60 (2012).

[74] Sevenet, N., Sheridan, E., Amram, D., Schneider, P., Handgretinger, R., and Delattre, O. Constitutional mutations of the hSNF5/INI1 gene predispose to a variety of cancers. *Am J Hum Genet* **65**(5), 1342–8 (1999).

[75] Lee, H.-Y., Yoon, C.-S., Sevenet, N., Rajalingam, V., Delattre, O., and Walford, N. Q. Rhabdoid tumor of the kidney is a component of the rhabdoid predisposition syndrome. *Pediatr Dev Pathol* **5**(4), 395–9 (2002).

[76] Utsuki, S., Oka, H., Tanaka, S., Kondo, K., Tanizaki, Y., and Fujii, K. Importance of re-examination for medulloblastoma and atypical teratoid/rhabdoid tumor. *Acta Neurochir (Wien)* **145**(8), 663–6; discussion 666 (2003).

[77] Koral, K., Gargan, L., Bowers, D. C., Gimi, B., Timmons, C. F., Weprin, B., and Rollins, N. K. Imaging characteristics of atypical teratoid-rhabdoid tumor in children compared with medulloblastoma. *AJR Am J Roentgenol* **190**(3), 809–14 (2008).

[78] Biegel, J. A., Fogelgren, B., Zhou, J. Y., James, C. D., Janss, A. J., Allen, J. C., Zagzag, D., Raffel, C., and Rorke, L. B. Mutations of the INI1 rhabdoid tumor suppressor gene in medulloblastomas and primitive neuroectodermal tumors of the central nervous system. *Clin Cancer Res* **6**(7), 2759–63 (2000).

[79] Kraus, J. A., Oster, C., Sorensen, N., Berthold, F., Schlegel, U., Tonn, J. C., Wiestler, O. D., and Pietsch, T. Human medulloblastomas lack point mutations and homozygous deletions of the hSNF5/INI1 tumour suppressor gene. *Neuropathol Appl Neurobiol* **28**(2), 136–41 (2002).

[80] Svard, J., Rozell, B., Toftgard, R., and Teglund, S. Tumor suppressor gene co-operativity in compound Patched1 and suppressor of fused heterozygous mutant mice. *Mol Carcinog* **48**(5), 408–19 (2009).

[81] Smith, M. J., Beetz, C., Williams, S. G., Bhaskar, S. S., O'Sullivan, J., Anderson, B., Daly, S. B., Urquhart, J. E., Bholah, Z., Oudit, D., Cheesman, E., Kelsey, A., McCabe, M. G., Newman, W. G., and Evans, D. G. R. Germline mutations in SUFU cause Gorlin syndrome-associated childhood medulloblastoma and redefine the risk associated with PTCH1 mutations. *J Clin Oncol* **32**(36), 4155–61 (2014).

[82] Pazzaglia, S., Pasquali, E., Tanori, M., Mancuso, M., Leonardi, S., di Majo, V., Rebessi, S., and Saran, A. Physical, heritable and age-related factors as modifiers of radiation cancer risk in patched heterozygous mice. *Int J Radiat Oncol Biol Phys* **73**(4), 1203–10 (2009).

[83] Northcott, P. A., Nakahara, Y., Wu, X., Feuk, L., Ellison, D. W., Croul, S., Mack, S., Kongkham, P. N., Peacock, J., Dubuc, A., Ra, Y.-S., Zilberberg, K., McLeod, J., Scherer, S. W., Sunil Rao, J., Eberhart, C. G., Grajkowska, W., Gillespie, Y., Lach, B., Grundy, R., Pollack, I. F., Hamilton, R. L., Van Meter, T., Carlotti, C. G., Boop, F., Bigner, D., Gilbertson, R. J., Rutka, J. T., and Taylor, M. D. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nat Genet* **41**(4), 465–72 (2009).

[84] Tan, M.-H., Mester, J. L., Ngeow, J., Rybicki, L. A., Orloff, M. S., and Eng, C. Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res* **18**(2), 400–7 (2012).

[85] Zhou, X.-P., Marsh, D. J., Morrison, C. D., Chaudhury, A. R., Maxwell, M., Reifenberger, G., and Eng, C. Germline inactivation of PTEN and dysregulation of the phosphoinositol-3-kinase/Akt pathway cause human Lhermitte-Duclos disease in adults. *Am J Hum Genet* **73**(5), 1191–8 (2003).

[86] Endersby, R. and Baker, S. J. PTEN signaling in brain: neuropathology and tumorigenesis. *Oncogene* **27**(41), 5416–30 (2008).

[87] Hambardzumyan, D., Becher, O. J., Rosenblum, M. K., Pandolfi, P. P., Manova-Todorova, K., and Holland, E. C. PI3K pathway regulates survival of cancer stem cells residing in the perivascular niche following radiation in medulloblastoma in vivo. *Genes Dev* **22**(4), 436–48 (2008).

[88] Thompson, M. C., Fuller, C., Hogg, T. L., Dalton, J., Finkelstein, D., Lau, C. C., Chintagumpala, M., Adesina, A., Ashley, D. M., Kellie, S. J., Taylor, M. D., Curran, T., Gajjar, A., and Gilbertson, R. J. Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J Clin Oncol* **24**(12), 1924–31 (2006).

[89] Ellison, D. W., Onilude, O. E., Lindsey, J. C., Lusher, M. E., Weston, C. L., Taylor, R. E., Pearson, A. D., Clifford, S. C., and United Kingdom Children's Cancer Study Group Brain Tumour Committee. beta-catenin status predicts a favorable outcome in childhood medulloblastoma: the united kingdom children's cancer study group brain tumour committee. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **23**(31), 79517957, November (2005).

[90] Liu, C., Li, Y., Semenov, M., Han, C., Baeg, G. H., Tan, Y., Zhang, Z., Lin, X., and He, X. Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **108**(6), 837–47 (2002).

[91] Legoix, P., Bluteau, O., Bayer, J., Perret, C., Balabaud, C., Belghiti, J., Franco, D., Thomas, G., Laurent-Puig, P., and Zucman-Rossi, J. Beta-catenin mutations in hepatocellular carcinoma correlate with a low rate of loss of heterozygosity. *Oncogene* **18**(27), 4044–6 (1999).

[92] Kratz, J. E., Stearns, D., Huso, D. L., Slunt, H. H., Price, D. L., Borchelt, D. R., and Eberhart, C. G. Expression of stabilized beta-catenin in differentiated neurons of transgenic mice does not result in tumor formation. *BMC cancer* **2**, 33, December (2002).

[93] Harada, N., Tamai, Y., Ishikawa, T., Sauer, B., Takaku, K., Oshima, M., and Taketo, M. M. Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. *EMBO J* **18**(21), 5931–42 (1999).

[94] Raffel, C., Jenkins, R. B., Frederick, L., Hebrink, D., Alderete, B., Fults, D. W., and James, C. D. Sporadic medulloblastomas contain PTCH mutations. *Cancer Res* **57**(5), 842–5 (1997).

[95] Reifenberger, J., Wolter, M., Weber, R. G., Megahed, M., Ruzicka, T., Lichter, P., and Reifenberger, G. Missense mutations in SMOH in sporadic basal cell carcinomas of the skin and primitive neuroectodermal tumors of the central nervous system. *Cancer Res* **58**(9), 1798–803 (1998).

[96] Oliver, T. G., Read, T. A., Kessler, J. D., Mehmeti, A., Wells, J. F., Huynh, T. T. T., Lin, S. M., and Wechsler-Reya, R. J. Loss of patched and disruption of granule cell development in a pre-neoplastic stage of medulloblastoma. *Development* **132**(10), 2425–39 (2005).

[97] Yang, Z.-J., Ellis, T., Markant, S. L., Read, T.-A., Kessler, J. D., Bourboulas, M., Schuller, U., Machold, R., Fishell, G., Rowitch, D. H., Wainwright, B. J., and Wechsler-Reya, R. J. Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. *Cancer Cell* **14**(2), 135–45 (2008).

[98] Hahn, H., Wojnowski, L., Zimmer, A. M., Hall, J., Miller, G., and Zimmer, A. Rhabdomyosarcomas and radiation hypersensitivity in a mouse model of Gorlin syndrome. *Nat Med* **4**(5), 619–22 (1998).

[99] Uziel, T., Zindy, F., Xie, S., Lee, Y., Forget, A., Magdaleno, S., Rehg, J. E., Calabrese, C., Solecki, D., Eberhart, C. G., Sherr, S. E., Plimmer, S., Clifford, S. C., Hatten, M. E., McKinnon, P. J., Gilbertson, R. J., Curran, T., Sherr, C. J., and Roussel, M. F. The tumor suppressors Ink4c and p53 collaborate independently with Patched to suppress medulloblastoma formation. *Genes Dev* **19**(22), 2656–67 (2005).

[100] Ayrault, O., Zindy, F., Rehg, J., Sherr, C. J., and Roussel, M. F. Two tumor suppressors, p27Kip1 and patched-1, collaborate to prevent medulloblastoma. *Mol Cancer Res* **7**(1), 33–40 (2009).

[101] Berman, D. M., Karhadkar, S. S., Hallahan, A. R., Pritchard, J. I., Eberhart, C. G., Watkins, D. N., Chen, J. K., Cooper, M. K., Taipale, J., Olson, J. M., and Beachy, P. A. Medulloblastoma growth inhibition by hedgehog pathway blockade. *Science* **297**(5586), 1559–61 (2002).

[102] Pazzaglia, S., Tanori, M., Mancuso, M., Rebessi, S., Leonardi, S., Di Majo, V., Covelli, V., Atkinson, M. J., Hahn, H., and Saran, A. Linking DNA damage to medulloblastoma tumorigenesis in patched heterozygous knockout mice. *Oncogene* **25**(8), 1165–73 (2006).

[103] Mille, F., Tamayo-Orrego, L., Levesque, M., Remke, M., Korshunov, A., Cardin, J., Bouchard, N., Izzi, L., Kool, M., Northcott, P. A., Taylor, M. D., Pfister, S. M., and Charron, F. The Shh receptor Boc promotes progression of early medulloblastoma to advanced tumors. *Dev Cell* **31**(1), 34–47 (2014).

[104] Frappart, P.-O., Lee, Y., Russell, H. R., Chalhoub, N., Wang, Y.-D., Orii, K. E., Zhao, J., Kondo, N., Baker, S. J., and McKinnon, P. J. Recurrent genomic alterations characterize medulloblastoma arising from DNA double-strand break repair deficiency. *Proc Natl Acad Sci U S A* **106**(6), 1880–5 (2009).

[105] Zurawel, R. H., Allen, C., Wechsler-Reya, R., Scott, M. P., and Raffel, C. Evidence that haploinsufficiency of Ptch leads to medulloblastoma in mice. *Genes Chromosomes Cancer* **28**(1), 77–81 (2000).

[106] Hatton, B. A., Villavicencio, E. H., Tsuchiya, K. D., Pritchard, J. I., Ditzler, S., Pullar, B., Hansen, S., Knoblaugh, S. E., Lee, D., Eberhart, C. G., Hallahan, A. R., and Olson, J. M. The Smo/Smo model: hedgehog-induced medulloblastoma with 90incidence and leptomeningeal spread. *Cancer Res* **68**(6), 1768–76 (2008).

[107] Hallahan, A. R., Pritchard, J. I., Hansen, S., Benson, M., Stoeck, J., Hatton, B. A., Russell, T. L., Ellenbogen, R. G., Bernstein, I. D., Beachy, P. A., and Olson, J. M. The SmoA1 mouse model reveals that notch signaling is critical for the growth and survival of sonic hedgehog-induced medulloblastomas. *Cancer Res* **64**(21), 7794–800 (2004).

[108] Regl, G., Kasper, M., Schnidar, H., Eichberger, T., Neill, G. W., Philpott, M. P., Esterbauer, H., Hauser-Kronberger, C., Frischauf, A.-M., and Aberger, F. Activation of the BCL2 promoter in response to Hedgehog/GLI signal transduction is predominantly mediated by GLI2. *Cancer Res* **64**(21), 7724–31 (2004).

[109] Bar, E. E., Chaudhry, A., Farah, M. H., and Eberhart, C. G. Hedgehog signaling promotes medulloblastoma survival via Bc/II. *Am J Pathol* **170**(1), 347–55 (2007).

[110] Leonard, J. M., Ye, H., Wetmore, C., and Karnitz, L. M. Sonic Hedgehog signaling impairs ionizing radiation-induced checkpoint activation and induces genomic instability. *J Cell Biol* **183**(3), 385–91 (2008).

[111] Lee, Y., Miller, H. L., Russell, H. R., Boyd, K., Curran, T., and McKinnon, P. J. Patched2 modulates tumorigenesis in patched1 heterozygous mice. *Cancer Res* **66**(14), 6964–71 (2006).

[112] Lee, Y., Kawagoe, R., Sasai, K., Li, Y., Russell, H. R., Curran, T., and McKinnon, P. J. Loss of suppressor-of-fused function promotes tumorigenesis. *Oncogene* **26**(44), 6442–7 (2007).

[113] de la Pompa, J. L., Wakeham, A., Correia, K. M., Samper, E., Brown, S., Aguilera, R. J., Nakano, T., Honjo, T., Mak, T. W., Rossant, J., and Conlon, R. A. Conservation of the Notch signalling pathway in mammalian neurogenesis. *Development* **124**(6), 1139–48 (1997).

[114] Breunig, J. J., Silbereis, J., Vaccarino, F. M., Sestan, N., and Rakic, P. Notch regulates cell fate and dendrite morphology of newborn neurons in the postnatal dentate gyrus. *Proc Natl Acad Sci U S A* **104**(51), 20558–63 (2007).

[115] Swiatek, P. J., Lindsell, C. E., del Amo, F. F., Weinmaster, G., and Gridley, T. Notch1 is essential for postimplantation development in mice. *Genes Dev* **8**(6), 707–19 (1994).

[116] Hatton, B. A., Villavicencio, E. H., Pritchard, J., LeBlanc, M., Hansen, S., Ulrich, M., Ditzler, S., Pullar, B., Stroud, M. R., and Olson, J. M. Notch signaling is not essential in sonic hedgehog-activated medulloblastoma. *Oncogene* **29**(26), 3865–72 (2010).

[117] Julian, E., Hallahan, A. R., and Wainwright, B. J. RBP-J is not required for granule neuron progenitor development and medulloblastoma initiated by Hedgehog pathway activation in the external germinal layer. *Neural Dev* **5**, 27 (2010).

[118] Yang, X., Klein, R., Tian, X., Cheng, H.-T., Kopan, R., and Shen, J. Notch activation induces apoptosis in neural progenitor cells through a p53-dependent pathway. *Dev Biol* **269**(1), 81–94 (2004).

[119] Inda, M. M., Mercapide, J., Munoz, J., Coullin, P., Danglot, G., Tunon, T., Martinez-Penuela, J. M., Rivera, J. M., Burgos, J. J., Bernheim, A., and Castresana, J. S. PTEN and DMBT1 homozygous deletion and expression in medulloblastomas and supratentorial primitive neuroectodermal tumors. *Oncol Rep* **12**(6), 1341–7 (2004).

[120] Cantley, L. C. The phosphoinositide 3-kinase pathway. *Science* **296**(5573), 1655–7 (2002).

[121] Gardai, S. J., Hildeman, D. A., Frankel, S. K., Whitlock, B. B., Frasch, S. C., Borregaard, N., Marrack, P., Bratton, D. L., and Henson, P. M. Phosphorylation of Bax Ser184 by Akt regulates its activity and apoptosis in neutrophils. *J Biol Chem* **279**(20), 21085–95 (2004).

[122] Datta, S. R., Dudek, H., Tao, X., Masters, S., Fu, H., Gotoh, Y., and Greenberg, M. E. Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. *Cell* **91**(2), 231–41 (1997).

[123] Rossig, L., Jadidi, A. S., Urbich, C., Badorff, C., Zeiher, A. M., and Dimmeler, S. Akt-dependent phosphorylation of p21(Cip1) regulates PCNA binding and proliferation of endothelial cells. *Mol Cell Biol* **21**(16), 5644–57 (2001).

[124] Li, Y., Dowbenko, D., and Lasky, L. A. AKT/PKB phosphorylation of p21Cip/WAF1 enhances protein stability of p21Cip/WAF1 and promotes cell survival. *J Biol Chem* **277**(13), 11352–61 (2002).

[125] Liang, J., Zubovitz, J., Petrocelli, T., Kotchetkov, R., Connor, M. K., Han, K., Lee, J.-H., Ciarallo, S., Catzavelos, C., Beniston, R., Franssen, E., and Slingerland, J. M. PKB/Akt phosphorylates p27, impairs nuclear import of p27 and opposes p27-mediated G1 arrest. *Nat Med* **8**(10), 1153–60 (2002).

[126] Metcalfe, C., Alicke, B., Crow, A., Lamoureux, M., Dijkgraaf, G. J. P., Peale, F., Gould, S. E., and de Sauvage, F. J. PTEN loss mitigates the response of medulloblastoma to Hedgehog pathway inhibition. *Cancer Res* **73**(23), 7034–42 (2013).

[127] von Hoff, K., Hartmann, W., von Bueren, A. O., Gerber, N. U., Grotzer, M. A., Pietsch, T., and Rutkowski, S. Large cell/anaplastic medulloblastoma: outcome according to myc status, histopathological, and clinical risk factors. *Pediatr Blood Cancer* **54**(3), 369–76 (2010).

[128] Kool, M., Koster, J., Bunt, J., Hasselt, N. E., Lakeman, A., van Sluis, P., Troost, D., Meeteren, N. S.-v., Caron, H. N., Cloos, J., Mrsic, A., Ylstra, B., Grajkowska, W., Hartmann, W., Pietsch, T., Ellison, D., Clifford, S. C., and Versteeg, R. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS One* **3**(8), e3088 (2008).

[129] Li, M., Lee, K. F., Lu, Y., Clarke, I., **Shih**, D., Eberhart, C., Collins, V. P., Van Meter, T., Picard, D., Zhou, L., Boutros, P. C., Modena, P., Liang, M.-L., Scherer, S. W., Bouffet, E., Rutka, J. T., Pomeroy, S. L., Lau, C. C., Taylor, M. D., Gajjar, A., Dirks, P. B., Hawkins, C. E., and Huang, A. Frequent amplification of a chr19q13.41 microRNA polycistron in aggressive primitive neuroectodermal brain tumors. *Cancer Cell* **16**(6), 533–46 (2009).

[130] Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. Molecular portraits of human breast tumours. *Nature* **406**(6797), 747–52 (2000).

[131] Carey, L. A., Perou, C. M., Livasy, C. A., Dressler, L. G., Cowan, D., Conway, K., Karaca, G., Troester, M. A., Tse, C. K., Edmiston, S., Deming, S. L., Geradts, J., Cheang, M. C. U., Nielsen, T. O., Moorman, P. G., Earp, H. S., and Millikan, R. C. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* **295**(21), 2492–502 (2006).

[132] Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., Livasy, C., Carey, L. A., Reynolds, E., Dressler, L., Nobel, A., Parker, J., Ewend, M. G., Sawyer, L. R., Wu, J., Liu, Y., Nanda, R., Tretiakova, M., Ruiz Orrico, A., Dreher, D., Palazzo, J. P., Perreard, L., Nelson, E., Mone, M., Hansen, H., Mullins, M., Quackenbush, J. F., Ellis, M. J., Olopade, O. I., Bernard, P. S., and Perou, C. M. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**, 96 (2006).

[133] Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S. A., Nobel, A. B., van't Veer, L. J., and Perou, C. M. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* **355**(6), 560–9 (2006).

[134] O'Brien, K. M., Cole, S. R., Tse, C.-K., Perou, C. M., Carey, L. A., Foulkes, W. D., Dressler, L. G., Geradts, J., and Millikan, R. C. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res* **16**(24), 6100–10 (2010).

[135] Tomlins, S. A., Rhodes, D. R., Perner, S., Dhanasekaran, S. M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J. E., Shah, R. B., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**(5748), 644–8 (2005).

[136] Falini, B., Tiacci, E., Martelli, M. P., Ascani, S., and Pileri, S. A. New classification of acute myeloid leukemia and precursor-related neoplasms: changes and unsolved issues. *Discov Med* **10**(53), 281–92 (2010).

[137] Polkinghorn, W. R. and Tarbell, N. J. Medulloblastoma: tumorigenesis, current clinical paradigm, and efforts to improve risk stratification. *Nat Clin Pract Oncol* **4**(5), 295–304 (2007).

[138] Pezzotta, S., Cordero di Montezemolo, L., Knerich, R., Arrigoni, M., Barbara, A., Besenzon, L., Brach del Prever, A., Fidani, P., Locatelli, D., Loiacono, G., Magrassi, L., Perilongo, G., Rigobello, L., Urgesi, A., and Madon, E. CNS-85 trial: a cooperative pediatric CNS tumor study–results of treatment of medulloblastoma patients. *Childs Nerv Syst* **12**(2), 87–96 (1996).

[139] Geyer, J. R., Zeltzer, P. M., Boyett, J. M., Rorke, L. B., Stanley, P., Albright, A. L., Wisoff, J. H., Milstein, J. M., Allen, J. C., and Finlay, J. L. Survival of infants with primitive neuroectodermal tumors or malignant ependymomas of the CNS treated with eight drugs in 1 day: a report from the Childrens Cancer Group. *J Clin Oncol* **12**(8), 1607–15 (1994).

[140] Geyer, J. R., Sposto, R., Jennings, M., Boyett, J. M., Axtell, R. A., Breiger, D., Broxson, E., Donahue, B., Finlay, J. L., Goldwein, J. W., Heier, L. A., Johnson, D., Mazewski, C., Miller, D. C., Packer, R., Puccetti, D., Radcliffe, J., Tao, M. L., and Shiminski-Maher, T. Multiagent chemotherapy and deferred radiotherapy in infants with malignant brain tumors: a report from the Children's Cancer Group. *J Clin Oncol* **23**(30), 7621–31 (2005).

[141] Rutkowski, S., Bode, U., Deinlein, F., Ottensmeier, H., Warmuth-Metz, M., Soerensen, N., Graf, N., Emser, A., Pietsch, T., Wolff, J. E. A., Kortmann, R. D., and Kuehl, J. Treatment of early childhood medulloblastoma by postoperative chemotherapy alone. *N Engl J Med* **352**(10), 978–86 (2005).

[142] Ashley, D. M., Merchant, T. E., Strother, D., Zhou, T., Duffner, P., Burger, P. C., Miller, D. C., Lyon, N., Bonner, M. J., Msall, M., Buxton, A., Geyer, R., Kun, L. E., Coleman, L., and Pollack, I. F. Induction chemotherapy and conformal radiation therapy for very young children with nonmetastatic medulloblastoma: Children's Oncology Group study P9934. *J Clin Oncol* **30**(26), 3181–6 (2012).

[143] Strother, D. R., Lafay-Cousin, L., Boyett, J. M., Burger, P., Aronin, P., Constine, L., Duffner, P., Kocak, M., Kun, L. E., Horowitz, M. E., and Gajjar, A. Benefit from prolonged dose-intensive chemotherapy for infants with malignant brain tumors is restricted to patients with ependymoma: a report of the Pediatric Oncology Group randomized controlled trial 9233/34. *Neuro Oncol* **16**(3), 457–65 (2014).

[144] White, L., Kellie, S., Gray, E., Toogood, I., Waters, K., Lockwood, L., Macfarlane, S., and Johnston, H. Postoperative chemotherapy in children less than 4 years of age with malignant brain tumors: promising initial response to a VETOPEC-based regimen. A Study of the Australian and New Zealand Children's Cancer Study Group (ANZCCSG). *J Pediatr Hematol Oncol* **20**(2), 125–30 (1998).

[145] von Bueren, A. O., von Hoff, K., Pietsch, T., Gerber, N. U., Warmuth-Metz, M., Deinlein, F., Zwiener, I., Faldum, A., Fleischhack, G., Benesch, M., Krauss, J., Kuehl, J., Kortmann, R. D., and Rutkowski, S. Treatment of young children with localized medulloblastoma by chemotherapy alone: results of the prospective, multicenter trial HIT 2000 confirming the prognostic impact of histology. *Neuro Oncol* **13**(6), 669–79 (2011).

[146] Grill, J., Sainte-Rose, C., Jouvet, A., Gentet, J.-C., Lejars, O., Frappaz, D., Doz, F., Rialland, X., Pichon, F., Bertozzi, A.-I., Chastagner, P., Couanet, D., Habrand, J.-L., Raquin, M.-A., Le Deley,

M.-C., and Kalifa, C. Treatment of medulloblastoma with postoperative chemotherapy alone: an SFOP prospective trial in young children. *Lancet Oncol* **6**(8), 573–80 (2005).

[147] Rutkowski, S., von Hoff, K., Emser, A., Zwiener, I., Pietsch, T., Figarella-Branger, D., Giangaspero, F., Ellison, D. W., Garre, M.-L., Biassoni, V., Grundy, R. G., Finlay, J. L., Dhall, G., Raquin, M.-A., and Grill, J. Survival and prognostic factors of early childhood medulloblastoma: an international meta-analysis. *J Clin Oncol* **28**(33), 4961–8 (2010).

[148] Yasuda, K., Taguchi, H., Sawamura, Y., Ikeda, J., Aoyama, H., Fujieda, K., Ishii, N., Kashiwamura, M., Iwasaki, Y., and Shirato, H. Low-dose craniospinal irradiation and ifosfamide, cisplatin and etoposide for non-metastatic embryonal tumors in the central nervous system. *Jpn J Clin Oncol* **38**(7), 486–92 (2008).

[149] Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., Fell, H. P., Ferree, S., George, R. D., Grogan, T., James, J. J., Maysuria, M., Mitton, J. D., Oliveri, P., Osborn, J. L., Peng, T., Ratcliffe, A. L., Webster, P. J., Davidson, E. H., Hood, L., and Dimitrov, K. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **26**(3), 317–25 (2008).

[150] Shi, L., Perkins, R. G., Fang, H., and Tong, W. Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential. *Curr Opin Biotechnol* **19**(1), 10–8 (2008).

[151] de Ronde, J., Wessels, L., and Wesseling, J. Molecular subtyping of breast cancer: ready to use? *Lancet Oncol* **11**(4), 306–7 (2010).

[152] Weigelt, B., Mackay, A., A'hern, R., Natrajan, R., Tan, D. S. P., Dowsett, M., Ashworth, A., and Reis-Filho, J. S. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol* **11**(4), 339–49 (2010).

[153] Ein-Dor, L., Zuk, O., and Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* **103**(15), 5923–8 (2006).

[154] Frantz, S. An array of problems. *Nat Rev Drug Discov* **4**(5), 362–3 (2005).

[155] Michiels, S., Koscielny, S., and Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**(9458), 488–92 (2005).

[156] Ioannidis, J. P. A. Microarrays and molecular research: noise discovery? *Lancet* **365**(9458), 454–5 (2005).

[157] Marshall, E. Getting the noise out of gene arrays. *Science* **306**(5696), 630–1 (2004).

[158] Check, E. Geneticists chip away at unruly data. *Nature* **427**(6970), 91 (2004).

[159] Tan, P. K., Downey, T. J., Spitznagel, E. L. J., Xu, P., Fu, D., Dimitrov, D. S., Lempicki, R. A., Raaka, B. M., and Cam, M. C. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* **31**(19), 5676–84 (2003).

[160] Tilstone, C. DNA microarrays: vital statistics. *Nature* **424**(6949), 610–2 (2003).

[161] Bishop, J., Chagovetz, A. M., and Blair, S. Kinetics of multiplex hybridization: mechanisms and implications. *Biophys J* **94**(5), 1726–34 (2008).

[162] Li, X., He, Z., and Zhou, J. Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. *Nucleic Acids Res* **33**(19), 6114–23 (2005).

[163] Reis, P. P., Waldron, L., Goswami, R. S., Xu, W., Xuan, Y., Perez-Ordonez, B., Gullane, P., Irish, J., Jurisica, I., and Kamel-Reid, S. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol* **11**, 46 (2011).

[164] Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* **99**(10), 6567–72 (2002).

[165] Paterson, E. and Farr, R. F. Cerebellar medulloblastoma: treatment by irradiation of the whole central nervous system. *Acta radiol* **39**(4), 323–36 (1953).

[166] McFarland, D. R., Horwitz, H., Saenger, E. L., and Bahr, G. K. Medulloblastoma–a review of prognosis and survival. *Br J Radiol* **42**(495), 198–214 (1969).

[167] Gudrunardottir, T., Lannering, B., Remke, M., Taylor, M. D., Wells, E. M., Keating, R. F., and Packer, R. J. Treatment developments and the unfolding of the quality of life discussion in childhood medulloblastoma: a review. *Childs Nerv Syst* **30**(6), 979–90 (2014).

[168] Khong, P.-L., Leung, L. H. T., Fung, A. S. M., Fong, D. Y. T., Qiu, D., Kwong, D. L. W., Ooi, G.-C., McAlonan, G., Cao, G., and Chan, G. C. F. White matter anisotropy in post-treatment childhood cancer survivors: preliminary evidence of association with neurocognitive function. *J Clin Oncol* **24**(6), 884–90 (2006).

[169] Zeller, B., Tamnes, C. K., Kanellopoulos, A., Amlien, I. K., Andersson, S., Due-Tonnessen, P., Fjell, A. M., Walhovd, K. B., Westlye, L. T., and Ruud, E. Reduced neuroanatomic volumes in long-term survivors of childhood acute lymphoblastic leukemia. *J Clin Oncol* **31**(17), 2078–85 (2013).

[170] Avan, A., Postma, T. J., Ceresa, C., Avan, A., Cavaletti, G., Giovannetti, E., and Peters, G. J. Platinum-Induced Neurotoxicity and Preventive Strategies: Past, Present, and Future. *Oncologist* **20**(4), 411–432 (2015).

[171] Palmer, S. L., Armstrong, C., Onar-Thomas, A., Wu, S., Wallace, D., Bonner, M. J., Schreiber, J., Swain, M., Chapieski, L., Mabbott, D., Knight, S., Boyle, R., and Gajjar, A. Processing speed, attention, and working memory after treatment for medulloblastoma: an international, prospective, and longitudinal study. *J Clin Oncol* **31**(28), 3494–500 (2013).

[172] Schreiber, J. E., Gurney, J. G., Palmer, S. L., Bass, J. K., Wang, M., Chen, S., Zhang, H., Swain, M., Chapieski, M. L., Bonner, M. J., Mabbott, D. J., Knight, S. J., Armstrong, C. L., Boyle, R., and Gajjar, A. Examination of risk factors for intellectual and academic outcomes following treatment for pediatric medulloblastoma. *Neuro Oncol* **16**(8), 1129–36 (2014).

[173] Knight, S. J., Conklin, H. M., Palmer, S. L., Schreiber, J. E., Armstrong, C. L., Wallace, D., Bonner, M., Swain, M. A., Evankovich, K. D., Mabbott, D. J., Boyle, R., Huang, Q., Zhang, H., Anderson, V. A., and Gajjar, A. Working memory abilities among children treated for medulloblastoma: parent report and child performance. *J Pediatr Psychol* **39**(5), 501–11 (2014).

[174] Armstrong, G. T., Liu, Q., Yasui, Y., Huang, S., Ness, K. K., Leisenring, W., Hudson, M. M., Donaldson, S. S., King, A. A., Stovall, M., Krull, K. R., Robison, L. L., and Packer, R. J. Long-term outcomes among adult survivors of childhood central nervous system malignancies in the Childhood Cancer Survivor Study. *J Natl Cancer Inst* **101**(13), 946–58 (2009).

[175] Xu, W., Janss, A., and Moshang, T. Adult height and adult sitting height in childhood medulloblastoma survivors. *J Clin Endocrinol Metab* **88**(10), 4677–81 (2003).

[176] Edelstein, K., Spiegler, B. J., Fung, S., Panzarella, T., Mabbott, D. J., Jewitt, N., D'Agostino, N. M., Mason, W. P., Bouffet, E., Tabori, U., Laperriere, N., and Hodgson, D. C. Early aging in adult survivors of childhood medulloblastoma: long-term neurocognitive, functional, and physical outcomes. *Neuro Oncol* **13**(5), 536–45 (2011).

[177] Christopherson, K. M., Rotondo, R. L., Bradley, J. A., Pincus, D. W., Wynn, T. T., Fort, J. A., Morris, C. G., Mendenhall, N. P., Marcus, R. B. J., and Indelicato, D. J. Late toxicity following craniospinal radiation for early-stage medulloblastoma. *Acta Oncol* **53**(4), 471–80 (2014).

[178] Boman, K. K., Hoven, E., Anclair, M., Lannering, B., and Gustafsson, G. Health and persistent functional late effects in adult survivors of childhood CNS tumours: a population-based cohort study. *Eur J Cancer* **45**(14), 2552–61 (2009).

[179] Duffner, P. K., Krischer, J. P., Horowitz, M. E., Cohen, M. E., Burger, P. C., Friedman, H. S., and Kun, L. E. Second malignancies in young children with primary brain tumors following treatment with prolonged postoperative chemotherapy and delayed irradiation: a Pediatric Oncology Group study. *Ann Neurol* **44**(3), 313–6 (1998).

[180] Packer, R. J., Zhou, T., Holmes, E., Vezina, G., and Gajjar, A. Survival and secondary tumors in children with medulloblastoma receiving radiotherapy and adjuvant chemotherapy: results of Children's Oncology Group trial A9961. *Neuro Oncol* **15**(1), 97–103 (2013).

[181] Hudson, M. M., Mertens, A. C., Yasui, Y., Hobbie, W., Chen, H., Gurney, J. G., Yeazel, M., Recklitis, C. J., Marina, N., Robison, L. R., and Oeffinger, K. C. Health status of adult long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *JAMA* **290**(12), 1583–92 (2003).

[182] Mitby, P. A., Robison, L. L., Whitton, J. A., Zevon, M. A., Gibbs, I. C., Tersak, J. M., Meadows, A. T., Stovall, M., Zeltzer, L. K., and Mertens, A. C. Utilization of special education services and educational attainment among long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. *Cancer* **97**(4), 1115–26 (2003).

[183] DeNavas-Walt, C. and Proctor, B. D. Income and poverty in the united states: 2013. In *Current Population Reports*, 60–249. U.S. Government Printing Office, Washington, DC (2014).

[184] Krull, K. R., Zhang, N., Santucci, A., Srivastava, D. K., Krasin, M. J., Kun, L. E., Pui, C.-H., Robison, L. L., Hudson, M. M., and Armstrong, G. T. Long-term decline in intelligence among adult survivors of childhood acute lymphoblastic leukemia treated with cranial radiation. *Blood* **122**(4), 550–3 (2013).

[185] Pui, C.-H. and Howard, S. C. Current management and challenges of malignant disease in the CNS in paediatric leukaemia. *Lancet Oncol* **9**(3), 257–68 (2008).

[186] Merchant, T. E., Conklin, H. M., Wu, S., Lustig, R. H., and Xiong, X. Late effects of conformal radiation therapy for pediatric patients with low-grade glioma: prospective evaluation of cognitive, endocrine, and hearing deficits. *J Clin Oncol* **27**(22), 3691–7 (2009).

[187] Radcliffe, J., Bunin, G. R., Sutton, L. N., Goldwein, J. W., and Phillips, P. C. Cognitive deficits in long-term survivors of childhood medulloblastoma and other noncortical tumors: age-dependent effects of whole brain radiation. *Int J Dev Neurosci* **12**(4), 327–34 (1994).

[188] Mulhern, R. K., Kepner, J. L., Thomas, P. R., Armstrong, F. D., Friedman, H. S., and Kun, L. E. Neuropsychologic functioning of survivors of childhood medulloblastoma randomized to receive conventional or reduced-dose craniospinal irradiation: a Pediatric Oncology Group study. *J Clin Oncol* **16**(5), 1723–8 (1998).

[189] Ris, M. D., Walsh, K., Wallace, D., Armstrong, F. D., Holmes, E., Gajjar, A., Zhou, T., and Packer, R. J. Intellectual and academic outcome following two chemotherapy regimens and radiotherapy for average-risk medulloblastoma: COG A9961. *Pediatr Blood Cancer* **60**(8), 1350–7 (2013).

[190] Mulhern, R. K., Palmer, S. L., Merchant, T. E., Wallace, D., Kocak, M., Brouwers, P., Krull, K., Chintagumpala, M., Stargatt, R., Ashley, D. M., Tyc, V. L., Kun, L., Boyett, J., and Gajjar, A. Neurocognitive consequences of risk-adapted therapy for childhood medulloblastoma. *J Clin Oncol* **23**(24), 5511–9 (2005).

[191] Ris, M. D., Packer, R., Goldwein, J., Jones-Wallace, D., and Boyett, J. M. Intellectual outcome after reduced-dose radiation therapy plus adjuvant chemotherapy for medulloblastoma: a Children's Cancer Group study. *J Clin Oncol* **19**(15), 3470–6 (2001).

[192] Deutsch, M., Thomas, P. R., Krischer, J., Boyett, J. M., Albright, L., Aronin, P., Langston, J., Allen, J. C., Packer, R. J., Linggood, R., Mulhern, R., Stanley, P., Stehbens, J. A., Duffner, P., Kun, L., Rorke, L., Cherlow, J., Freidman, H., Finlay, J. L., and Vietti, T. Results of a prospective randomized trial comparing standard dose neuraxis irradiation (3,600 cGy/20) with reduced neuraxis irradiation (2,340 cGy/13) in patients with low-stage medulloblastoma. A Combined Children's Cancer Group-Pediatric Oncology Group Study. *Pediatr Neurosurg* **24**(4), 167–176; discussion 176–7 (1996).

[193] Thomas, P. R., Deutsch, M., Kepner, J. L., Boyett, J. M., Krischer, J., Aronin, P., Albright, L., Allen, J. C., Packer, R. J., Linggood, R., Mulhern, R., Stehbens, J. A., Langston, J., Stanley, P., Duffner, P., Rorke, L., Cherlow, J., Friedman, H. S., Finlay, J. L., Vietti, T. J., and Kun, L. E. Low-stage medulloblastoma: final analysis of trial comparing standard-dose with reduced-dose neuraxis irradiation. *J Clin Oncol* **18**(16), 3004–11 (2000).

[194] Bailey, C. C., Gnekow, A., Wellek, S., Jones, M., Round, C., Brown, J., Phillips, A., and Neidhardt, M. K. Prospective randomised trial of chemotherapy given before radiotherapy in childhood medulloblastoma. International Society of Paediatric Oncology (SIOP) and the (German) Society of Paediatric Oncology (GPO): SIOP II. *Med Pediatr Oncol* **25**(3), 166–78 (1995).

[195] Bouffet, E., Bernard, J. L., Frappaz, D., Gentet, J. C., Roche, H., Tron, P., Carrie, C., Raybaud, C., Joannard, A., and Lapras, C. M4 protocol for cerebellar medulloblastoma: supratentorial radiotherapy may not be avoided. *Int J Radiat Oncol Biol Phys* **24**(1), 79–85 (1992).

[196] Packer, R. J., Gajjar, A., Vezina, G., Rorke-Adams, L., Burger, P. C., Robertson, P. L., Bayer, L., LaFond, D., Donahue, B. R., Marymont, M. H., Muraszko, K., Langston, J., and Sposto, R. Phase III study of craniospinal radiation therapy followed by adjuvant chemotherapy for newly diagnosed average-risk medulloblastoma. *J Clin Oncol* **24**(25), 4202–8 (2006).

[197] Packer, R. J., Sutton, L. N., Elterman, R., Lange, B., Goldwein, J., Nicholson, H. S., Mulne, L., Boyett, J., D'Angio, G., and Wechsler-Jentzsch, K. Outcome for children with medulloblastoma treated with radiation and cisplatin, CCNU, and vincristine chemotherapy. *J Neurosurg* **81**(5), 690–8 (1994).

[198] Packer, R. J., Goldwein, J., Nicholson, H. S., Vezina, L. G., Allen, J. C., Ris, M. D., Muraszko, K., Rorke, L. B., Wara, W. M., Cohen, B. H., and Boyett, J. M. Treatment of children with medulloblastomas with reduced-dose craniospinal radiation therapy and adjuvant chemotherapy: A Children's Cancer Group Study. *J Clin Oncol* **17**(7), 2127–36 (1999).

[199] Kim, S. Y., Sung, K. W., Hah, J. O., Yoo, K. H., Koo, H. H., Kang, H. J., Park, K. D., Shin, H. Y., Ahn, H. S., Im, H. J., Seo, J. J., Lim, Y. J., Lee, Y. H., Shin, H. J., Lim, D. H., Cho, B. K., Ra, Y. S., and Choi, J. U. Reduced-dose craniospinal radiotherapy followed by high-dose chemotherapy and autologous stem cell rescue for children with newly diagnosed high-risk medulloblastoma or supratentorial primitive neuroectodermal tumor. *Korean J Hematol* **45**(2), 120–6 (2010).

[200] Halberg, F. E., Wara, W. M., Fippin, L. F., Edwards, M. S., Levin, V. A., Davis, R. L., Prados, M. B., and Wilson, C. B. Low-dose craniospinal radiation therapy for medulloblastoma. *Int J Radiat Oncol Biol Phys* **20**(4), 651–4 (1991).

[201] Oyharcabal-Bourden, V., Kalifa, C., Gentet, J. C., Frappaz, D., Edan, C., Chastagner, P., Sariban, E., Pagnier, A., Babin, A., Pichon, F., Neuenschwander, S., Vinchon, M., Bours, D., Mosseri, V., Le Gales, C., Ruchoux, M., Carrie, C., and Doz, F. Standard-risk medulloblastoma treated by adjuvant chemotherapy followed by reduced-dose craniospinal radiation therapy: a French Society of Pediatric Oncology Study. *J Clin Oncol* **23**(21), 4726–34 (2005).

[202] Sung, K. W., Lim, D. H., Son, M. H., Lee, S. H., Yoo, K. H., Koo, H. H., Kim, J. H., Suh, Y.-L., Joung, Y. S., and Shin, H. J. Reduced-dose craniospinal radiotherapy followed by tandem high-dose chemotherapy and autologous stem cell transplantation in patients with high-risk medulloblastoma. *Neuro Oncol* **15**(3), 352–9 (2013).

[203] Mulhern, R. K., Horowitz, M. E., Kovnar, E. H., Langston, J., Sanford, R. A., and Kun, L. E. Neurodevelopmental status of infants and young children treated for brain tumors with preirradiation chemotherapy. *J Clin Oncol* **7**(11), 1660–6 (1989).

[204] Jeng, M. J., Chang, T. K., Wong, T. T., Hsien, Y. L., Tang, R. B., and Hwang, B. Preirradiation chemotherapy for very young children with brain tumors. *Childs Nerv Syst* **9**(3), 150–3 (1993).

[205] Duffner, P. K., Horowitz, M. E., Krischer, J. P., Friedman, H. S., Burger, P. C., Cohen, M. E., Sanford, R. A., Mulhern, R. K., James, H. E., and Freeman, C. R. Postoperative chemotherapy and delayed radiation in children less than three years of age with malignant brain tumors. *N Engl J Med* **328**(24), 1725–31 (1993).

[206] Gentet, J. C., Bouffet, E., Doz, F., Tron, P., Roche, H., Thyss, A., Plantaz, D., Stephan, J. L., Mottolese, C., and Ponvert, D. Preirradiation chemotherapy including "eight drugs in 1 day" regimen and high-dose methotrexate in childhood medulloblastoma: results of the M7 French Cooperative Study. *J Neurosurg* **82**(4), 608–14 (1995).

[207] Duffner, P. K., Horowitz, M. E., Krischer, J. P., Burger, P. C., Cohen, M. E., Sanford, R. A., Friedman, H. S., and Kun, L. E. The treatment of malignant brain tumors in infants and very young children: an update of the Pediatric Oncology Group experience. *Neuro Oncol* **1**(2), 152–61 (1999).

[208] Walter, A. W., Mulhern, R. K., Gajjar, A., Heideman, R. L., Reardon, D., Sanford, R. A., Xiong, X., and Kun, L. E. Survival and neurodevelopmental outcome of young children with medulloblastoma at St Jude Children's Research Hospital. *J Clin Oncol* **17**(12), 3720–8 (1999).

[209] Dhall, G., Grodman, H., Ji, L., Sands, S., Gardner, S., Dunkel, I. J., McCowage, G. B., Diez, B., Allen, J. C., Gopalan, A., Cornelius, A. S., Termuhlen, A., Abromowitch, M., Sposto, R., and Finlay, J. L. Outcome of children less than three years old at diagnosis with non-metastatic medulloblastoma treated with chemotherapy on the "Head Start" I and II protocols. *Pediatr Blood Cancer* **50**(6), 1169–75 (2008).

[210] Sands, S. A., Oberg, J. A., Gardner, S. L., Whiteley, J. A., Glade-Bender, J. L., and Finlay, J. L. Neuropsychological functioning of children treated with intensive chemotherapy followed by myeloablative consolidation chemotherapy and autologous hematopoietic cell rescue for newly diagnosed CNS tumors: an analysis of the Head Start II survivors. *Pediatr Blood Cancer* **54**(3), 429–36 (2010).

[211] Grundy, R. G., Wilne, S. H., Robinson, K. J., Ironside, J. W., Cox, T., Chong, W. K., Michalski, A., Campbell, R. H. A., Bailey, C. C., Thorp, N., Pizer, B., Punt, J., Walker, D. A., Ellison, D. W., and Machin, D. Primary postoperative chemotherapy without radiotherapy for treatment of brain tumours other than ependymoma in children under 3 years: results of the first UKCCSG/SIOP CNS 9204 trial. *Eur J Cancer* **46**(1), 120–33 (2010).

[212] Saha, A., Salley, C. G., Saigal, P., Rolnitzky, L., Goldberg, J., Scott, S., Olshefski, R., Hukin, J., Sands, S. A., Finlay, J., and Gardner, S. L. Late effects in survivors of childhood CNS tumors treated on Head Start I and II protocols. *Pediatr Blood Cancer* **61**(9), 1644–52; quiz 1653–72 (2014).

[213] Kellie, S. J., Wong, C. K. F., Pozza, L. D., Waters, K. D., Lockwood, L., Mauger, D. C., and White, L. Activity of postoperative carboplatin, etoposide, and high-dose methotrexate in pediatric CNS embryonal tumors: results of a phase II study in newly diagnosed children. *Med Pediatr Oncol* **39**(3), 168–74 (2002).

[214] Strauss, L. C., Killmond, T. M., Carson, B. S., Maria, B. L., Wharam, M. D., and Leventhal, B. G. Efficacy of postoperative chemotherapy using cisplatin plus etoposide in young children with brain tumors. *Med Pediatr Oncol* **19**(1), 16–21 (1991).

[215] Gajjar, A., Mulhern, R. K., Heideman, R. L., Sanford, R. A., Douglass, E. C., Kovnar, E. H., Langston, J. A., Jenkins, J. J., and Kun, L. E. Medulloblastoma in very young children: outcome of definitive craniospinal irradiation following incomplete response to chemotherapy. *J Clin Oncol* **12**(6), 1212–6 (1994).

[216] Chi, S. N., Gardner, S. L., Levy, A. S., Knopp, E. A., Miller, D. C., Wisoff, J. H., Weiner, H. L., and Finlay, J. L. Feasibility and response to induction chemotherapy intensified with high-dose methotrexate for young children with newly diagnosed high-risk disseminated medulloblastoma. *J Clin Oncol* **22**(24), 4881–7 (2004).

[217] Pietsch, T., Schmidt, R., Remke, M., Korshunov, A., Hovestadt, V., Jones, D. T. W., Felsberg, J., Kaulich, K., Goschzik, T., Kool, M., Northcott, P. A., von Hoff, K., von Bueren, A. O., Friedrich, C., Mynarek, M., Skladny, H., Fleischhack, G., Taylor, M. D., Cremer, F., Lichter, P., Faldum, A., Reifenberger, G., Rutkowski, S., and Pfister, S. M. Prognostic significance of clinical, histopathological, and molecular characteristics of medulloblastomas in the prospective HIT2000 multicenter clinical trial cohort. *Acta Neuropathol* **128**(1), 137–49 (2014).

[218] Park, T. S., Hoffman, H. J., Hendrick, E. B., Humphreys, R. P., and Becker, L. E. Medulloblastoma: clinical presentation and management. Experience at the hospital for sick children, toronto, 1950-1980. *J Neurosurg* **58**(4), 543–52 (1983).

[219] Gajjar, A., Chintagumpala, M., Ashley, D., Kellie, S., Kun, L. E., Merchant, T. E., Woo, S., Wheeler, G., Ahern, V., Krasin, M. J., Fouladi, M., Broniscer, A., Krance, R., Hale, G. A., Stewart, C. F., Dauser, R., Sanford, R. A., Fuller, C., Lau, C., Boyett, J. M., Wallace, D., and Gilbertson, R. J. Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *Lancet Oncol* **7**(10), 813–20 (2006).

[220] Lannering, B., Rutkowski, S., Doz, F., Pizer, B., Gustafsson, G., Navajas, A., Massimino, M., Reddingius, R., Benesch, M., Carrie, C., Taylor, R., Gandola, L., Bjork-Eriksson, T., Giralt, J., Oldenburger, F., Pietsch, T., Figarella-Branger, D., Robson, K., Forni, M., Clifford, S. C., Warmuth-Metz, M., von Hoff, K., Faldum, A., Mosseri, V., and Kortmann, R. Hyperfractionated versus conventional radiotherapy followed by chemotherapy in standard-risk medulloblastoma: results from the randomized multicenter HIT-SIOP PNET 4 trial. *J Clin Oncol* **30**(26), 3187–93 (2012).

[221] Benjamini, Y. and Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* **57**(1), 289–300 (1995).

[222] Heagerty, P. J. and Zheng, Y. Survival model predictive accuracy and ROC curves. *Biometrics* **61**(1), 92–105 (2005).

[223] Hastie, T. J. and Pregibon, D. Generalized linear models. In Statistical Models in S, Chambers, J. M. and Hastie, T. J., editors, chapter 6. Wadsworth & Brooks/Cole (1992).

[224] Salama, M. M., Ghorab, E. M., Al-Abyad, A. G., and Al-Bahy, K. M. Concomitant weekly vincristine and radiation followed by adjuvant vincristine and carboplatin in the treatment of high risk medulloblastoma: Ain Shams University Hospital and Sohag Cancer Center study. *J Egypt Natl Canc Inst* **18**(2), 167–74 (2006).

[225] von Hoff, K., Hinkes, B., Gerber, N. U., Deinlein, F., Mittler, U., Urban, C., Benesch, M., Warmuth-Metz, M., Soerensen, N., Zwiener, I., Goette, H., Schlegel, P. G., Pietsch, T., Kortmann, R. D., Kuehl, J., and Rutkowski, S. Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT'91. *Eur J Cancer* **45**(7), 1209–17 (2009).

[226] Bouffet, E., Gentet, J. C., Doz, F., Tron, P., Roche, H., Plantaz, D., Thyss, A., Stephan, J. L., Lasset, C., and Carrie, C. Metastatic medulloblastoma: the experience of the French Cooperative M7 Group. *Eur J Cancer* **30A**(10), 1478–83 (1994).

[227] Zeltzer, P. M., Boyett, J. M., Finlay, J. L., Albright, A. L., Rorke, L. B., Milstein, J. M., Allen, J. C., Stevens, K. R., Stanley, P., Li, H., Wisoff, J. H., Geyer, J. R., McGuire-Cullen, P., Stehbens, J. A., Shurin, S. B., and Packer, R. J. Metastasis stage, adjuvant treatment, and residual tumor are prognostic factors for medulloblastoma in children: conclusions from the Children's Cancer Group 921 randomized phase III study. *J Clin Oncol* **17**(3), 832–45 (1999).

[228] Kortmann, R. D., Kuhl, J., Timmermann, B., Mittler, U., Urban, C., Budach, V., Richter, E., Willich, N., Flentje, M., Berthold, F., Slavc, I., Wolff, J., Meisner, C., Wiestler, O., Sorensen, N., Warmuth-Metz, M., and Bamberg, M. Postoperative neoadjuvant chemotherapy before radiotherapy as compared to immediate radiotherapy followed by maintenance chemotherapy in the treatment of medulloblastoma in childhood: results of the German prospective randomized trial HIT '91. *Int J Radiat Oncol Biol Phys* **46**(2), 269–79 (2000).

[229] Sanders, R. P., Onar, A., Boyett, J. M., Broniscer, A., Morris, E. B., Qaddoumi, I., Armstrong, G. T., Boop, F. A., Sanford, R. A., Kun, L. E., Merchant, T. E., and Gajjar, A. M1 Medulloblastoma: high risk at any age. *J Neurooncol* **90**(3), 351–5 (2008).

[230] Gajjar, A., Hernan, R., Kocak, M., Fuller, C., Lee, Y., McKinnon, P. J., Wallace, D., Lau, C., Chintagumpala, M., Ashley, D. M., Kellie, S. J., Kun, L., and Gilbertson, R. J. Clinical, histopathologic, and molecular markers of prognosis: toward a new disease risk stratification system for medulloblastoma. *J Clin Oncol* **22**(6), 984–93 (2004).

[231] Jakacki, R. I., Burger, P. C., Zhou, T., Holmes, E. J., Kocak, M., Onar, A., Goldwein, J., Mehta, M., Packer, R. J., Tarbell, N., Fitz, C., Vezina, G., Hilden, J., and Pollack, I. F. Outcome of children with metastatic medulloblastoma treated with carboplatin during craniospinal radiotherapy: a Children's Oncology Group Phase I/II study. *J Clin Oncol* **30**(21), 2648–53 (2012).

[232] Raabe, E. H. and Eberhart, C. G. High-risk medulloblastoma: does c-myc amplification overrule histopathology? *Pediatr Blood Cancer* **54**(3), 344–5 (2010).

[233] Pei, Y., Moore, C. E., Wang, J., Tewari, A. K., Eroshkin, A., Cho, Y.-J., Witt, H., Korshunov, A., Read, T.-A., Sun, J. L., Schmitt, E. M., Miller, C. R., Buckley, A. F., McLendon, R. E., Westbrook, T. F., Northcott, P. A., Taylor, M. D., Pfister, S. M., Febbo, P. G., and Wechsler-Reya, R. J. An animal model of MYC-driven medulloblastoma. *Cancer Cell* **21**(2), 155–67 (2012).

[234] Bull, K. S., Spoudeas, H. A., Yadegarfar, G., and Kennedy, C. R. Reduction of health status 7 years after addition of chemotherapy to craniospinal irradiation for medulloblastoma: a follow-up study in PNET 3 trial survivors on behalf of the CCLG (formerly UKCCSG). *J Clin Oncol* **25**(27), 4239–45 (2007).

[235] Parsons, D. W., Li, M., Zhang, X., Jones, S., Leary, R. J., Lin, J. C.-H., Boca, S. M., Carter, H., Samayoa, J., Bettegowda, C., Gallia, G. L., Jallo, G. I., Binder, Z. A., Nikolsky, Y., Hartigan, J., Smith, D. R., Gerhard, D. S., Fults, D. W., VandenBerg, S., Berger, M. S., Marie, S. K. N., Shinjo, S. M. O., Clara, C., Phillips, P. C., Minturn, J. E., Biegel, J. A., Judkins, A. R., Resnick, A. C., Storm, P. B., Curran, T., He, Y., Rasheed, B. A., Friedman, H. S., Keir, S. T., McLendon, R., Northcott, P. A., Taylor, M. D., Burger, P. C., Riggins, G. J., Karchin, R., Parmigiani, G., Bigner, D. D., Yan, H., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**(6016), 435–9 (2011).

[236] Lin, M., Wei, L.-J., Sellers, W. R., Lieberfarb, M., Wong, W. H., and Li, C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**(8), 1233–40 (2004).

[237] Venkatraman, E. S. and Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**(6), 657–63 (2007).

[238] Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W., and Lee, C. Detection of large-scale variation in the human genome. *Nat Genet* **36**(9), 949–51 (2004).

[239] Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., and Getz, G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**(4), R41 (2011).

[240] Zhang, Q., Ding, L., Larson, D. E., Koboldt, D. C., McLellan, M. D., Chen, K., Shi, X., Kraja, A., Mardis, E. R., Wilson, R. K., Borecki, I. B., and Province, M. A. CMDS: a population-based method for identifying recurrent DNA copy number aberrations in cancer from high-resolution data. *Bioinformatics* **26**(4), 464–9 (2010).

[241] Wu, Z., Huang, N. E., Long, S. R., and Peng, C.-K. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proc Natl Acad Sci U S A* **104**(38), 14889–94 (2007).

[242] Storey, J. D. and Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**(16), 9440–5 (2003).

[243] Lee, E. Y., Ji, H., Ouyang, Z., Zhou, B., Ma, W., Vokes, S. A., McMahon, A. P., Wong, W. H., and Scott, M. P. Hedgehog pathway-regulated gene networks in cerebellum development and tumorigenesis. *Proc Natl Acad Sci U S A* **107**(21), 9736–41 (2010).

[244] Miller, C. A., Settle, S. H., Sulman, E. P., Aldape, K. D., and Milosavljevic, A. Discovering functional modules by identifying recurrent and mutually exclusive mutational patterns in tumors. *BMC Med Genomics* **4**, 34 (2011).

[245] Reimand, J., Arak, T., and Vilo, J. g:Profiler–a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* **39**(Web Server issue), W307–15 (2011).

[246] Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**(11), e13984 (2010).

[247] Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* **101**(12), 4164–9 (2004).

[248] Sneath, P. H. A. and Sokal, R. R. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*, 278. Freeman, San Francisco (1973).

[249] Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A., McLaren, S., Lin, M.-L., McBride, D. J., Varela, I., Nik-Zainal, S., Leroy, C., Jia, M., Menzies, A., Butler, A. P., Teague, J. W., Quail, M. A., Burton, J., Swerdlow, H., Carter, N. P., Morsberger, L. A., Iacobuzio-Donahue, C., Follows, G. A., Green, A. R., Flanagan, A. M., Stratton, M. R., Futreal, P. A., and Campbell, P. J. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**(1), 27–40 (2011).

[250] Liu, P., Erez, A., Nagamani, S. C. S., Dhar, S. U., Kolodziejska, K. E., Dharmadhikari, A. V., Cooper, M. L., Wiszniewska, J., Zhang, F., Withers, M. A., Bacino, C. A., Campos-Acevedo, L. D., Delgado, M. R., Freedenberg, D., Garnica, A., Grebe, T. A., Hernandez-Almaguer, D., Immken, L., Lalani, S. R., McLean, S. D., Northrup, H., Scaglia, F., Strathearn, L., Trapane, P., Kang, S.-H. L., Patel, A., Cheung, S. W., Hastings, P. J., Stankiewicz, P., Lupski, J. R., and Bi, W. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* **146**(6), 889–903 (2011).

[251] Kloosterman, W. P., Hoogstraat, M., Paling, O., Tavakoli-Yaraki, M., Renkens, I., Vermaat, J. S., van Roosmalen, M. J., van Lieshout, S., Nijman, I. J., Roessingh, W., van 't Slot, R., van de Belt, J., Guryev, V., Koudijs, M., Voest, E., and Cuppen, E. Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. *Genome Biol* **12**(10), R103 (2011).

[252] Magrangeas, F., Avet-Loiseau, H., Munshi, N. C., and Minvielle, S. Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. *Blood* **118**(3), 675–8 (2011).

[253] Crasta, K., Ganem, N. J., Dagher, R., Lantermann, A. B., Ivanova, E. V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D., and Pellman, D. DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**(7383), 53–8 (2012).

[254] Molenaar, J. J., Koster, J., Zwijnenburg, D. A., van Sluis, P., Valentijn, L. J., van der Ploeg, I., Hamdi, M., van Nes, J., Westerman, B. A., van Arkel, J., Ebus, M. E., Haneveld, F., Lakeman, A., Schild, L., Molenaar, P., Stroeken, P., van Noesel, M. M., Ora, I., Santo, E. E., Caron, H. N., Westerhout, E. M., and Versteeg, R. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**(7391), 589–93 (2012).

[255] Bunt, J., Hasselt, N. E., Zwijnenburg, D. A., Koster, J., Versteeg, R., and Kool, M. Joint binding of OTX2 and MYC in promotor regions is associated with high gene expression in medulloblastoma. *PLoS One* **6**(10), e26058 (2011).

[256] Guan, Y., Kuo, W.-L., Stilwell, J. L., Takano, H., Lapuk, A. V., Fridlyand, J., Mao, J.-H., Yu, M., Miller, M. A., Santos, J. L., Kalloger, S. E., Carlson, J. W., Ginzinger, D. G., Celniker, S. E., Mills, G. B., Huntsman, D. G., and Gray, J. W. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clin Cancer Res* **13**(19), 5745–55 (2007).

[257] Carramusa, L., Contino, F., Ferro, A., Minafra, L., Perconti, G., Giallongo, A., and Feo, S. The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. *J Cell Physiol* **213**(2), 511–8 (2007).

[258] Huppi, K., Volfovsky, N., Runfola, T., Jones, T. L., Mackiewicz, M., Martin, S. E., Mushinski, J. F., Stephens, R., and Caplen, N. J. The identification of microRNAs in a genomically unstable region of human chromosome 8q24. *Mol Cancer Res* **6**(2), 212–21 (2008).

[259] Barsotti, A. M., Beckerman, R., Laptenko, O., Huppi, K., Caplen, N. J., and Prives, C. p53-Dependent induction of PVT1 and miR-1204. *J Biol Chem* **287**(4), 2509–19 (2012).

[260] Hovestadt, V., Remke, M., Kool, M., Pietsch, T., Northcott, P. A., Fischer, R., Cavalli, F. M. G., Ramaswamy, V., Zapatka, M., Reifenberger, G., Rutkowski, S., Schick, M., Bewerunge-Hudler, M., Korshunov, A., Lichter, P., Taylor, M. D., Pfister, S. M., and Jones, D. T. W. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol* **125**(6), 913–6 (2013).

[261] Schwalbe, E. C., Williamson, D., Lindsey, J. C., Hamilton, D., Ryan, S. L., Megahed, H., Garami, M., Hauser, P., Dembowska-Baginska, B., Perek, D., Northcott, P. A., Taylor, M. D., Taylor, R. E., Ellison, D. W., Bailey, S., and Clifford, S. C. DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. *Acta Neuropathol* **125**(3), 359–71 (2013).

[262] Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E., and Bishop, J. M. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **224**(4653), 1121–4 (1984).

[263] Malynn, B. A., de Alboran, I. M., O'Hagan, R. C., Bronson, R., Davidson, L., DePinho, R. A., and Alt, F. W. N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. *Genes Dev* **14**(11), 1390–9 (2000).

[264] Wey, A. and Knoepfler, P. S. c-myc and N-myc promote active stem cell metabolism and cycling as architects of the developing brain. *Oncotarget* **1**(2), 120–30 (2010).

[265] Gostissa, M., Ranganath, S., Bianco, J. M., and Alt, F. W. Chromosomal location targets different MYC family gene members for oncogenic translocations. *Proc Natl Acad Sci U S A* **106**(7), 2265–70 (2009).

[266] Rao, G., Pedone, C. A., Coffin, C. M., Holland, E. C., and Fults, D. W. c-Myc enhances sonic hedgehog-induced medulloblastoma formation from nestin-expressing neural progenitors in mice. *Neoplasia* **5**(3), 198–204 (2003).

[267] Polak, P., Karlic, R., Koren, A., Thurman, R., Sandstrom, R., Lawrence, M. S., Reynolds, A., Rynes, E., Vlahovicek, K., Stamatoyannopoulos, J. A., and Sunyaev, S. R. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**(7539), 360–4 (2015).

[268] Poschl, J., Stark, S., Neumann, P., Grobner, S., Kawauchi, D., Jones, D. T. W., Northcott, P. A., Lichter, P., Pfister, S. M., Kool, M., and Schuller, U. Genomic and transcriptomic analyses match medulloblastoma mouse models to their human counterparts. *Acta Neuropathol* **128**(1), 123–36 (2014).

[269] Morfouace, M., Shelat, A., Jacus, M., Freeman, B. B. r., Turner, D., Robinson, S., Zindy, F., Wang, Y.-D., Finkelstein, D., Ayrault, O., Bihannic, L., Puget, S., Li, X.-N., Olson, J. M., Robinson, G. W., Guy, R. K., Stewart, C. F., Gajjar, A., and Roussel, M. F. Pemetrexed and gemcitabine as combination therapy for the treatment of Group3 medulloblastoma. *Cancer Cell* **25**(4), 516–29 (2014).

[270] Bandopadhayay, P., Bergthold, G., Nguyen, B., Schubert, S., Gholamin, S., Tang, Y., Bolin, S., Schumacher, S. E., Zeid, R., Masoud, S., Yu, F., Vue, N., Gibson, W. J., Paolella, B. R., Mitra, S. S., Cheshier, S. H., Qi, J., Liu, K.-W., Wechsler-Reya, R., Weiss, W. A., Swartling, F. J., Kieran, M. W., Bradner, J. E., Beroukhim, R., and Cho, Y.-J. BET bromodomain inhibition of MYC-amplified medulloblastoma. *Clin Cancer Res* **20**(4), 912–25 (2014).

[271] Kieran, M. W. Targeted treatment for sonic hedgehog-dependent medulloblastoma. *Neuro Oncol* **16**(8), 1037–47 (2014).

[272] Gajjar, A., Stewart, C. F., Ellison, D. W., Kaste, S., Kun, L. E., Packer, R. J., Goldman, S., Chintagumpala, M., Wallace, D., Takebe, N., Boyett, J. M., Gilbertson, R. J., and Curran, T. Phase I study of vismodegib in children with recurrent or refractory medulloblastoma: a pediatric brain tumor consortium study. *Clin Cancer Res* **19**(22), 6305–12 (2013).

[273] Rodon, J., Tawbi, H. A., Thomas, A. L., Stoller, R. G., Turtschi, C. P., Baselga, J., Sarantopoulos, J., Mahalingam, D., Shou, Y., Moles, M. A., Yang, L., Granvil, C., Hurh, E., Rose, K. L., Amakye, D. D., Dummer, R., and Mita, A. C. A phase I, multicenter, open-label, first-in-human, dose-escalation study of the oral smoothened inhibitor Sonidegib (LDE225) in patients with advanced solid tumors. *Clin Cancer Res* **20**(7), 1900–9 (2014).

[274] Amakye, D., Jagani, Z., and Dorsch, M. Unraveling the therapeutic potential of the Hedgehog pathway in cancer. *Nat Med* **19**(11), 1410–22 (2013).

[275] Brun, S. N., Markant, S. L., Esparza, L. A., Garcia, G., Terry, D., Huang, J.-M., Pavlyukov, M. S., Li, X.-N., Grant, G. A., Crawford, J. R., Levy, M. L., Conway, E. M., Smith, L. H., Nakano, I., Berezov, A., Greene, M. I., Wang, Q., and Wechsler-Reya, R. J. Survivin as a therapeutic target in Sonic hedgehog-driven medulloblastoma. *Oncogene* **0** (2014).

[276] Kadota, R. P., Stewart, C. F., Horn, M., Kuttesch, J. F. J., Burger, P. C., Kepner, J. L., Kun, L. E., Friedman, H. S., and Heideman, R. L. Topotecan for the treatment of recurrent or progressive central nervous system tumors - a pediatric oncology group phase II study. *J Neurooncol* **43**(1), 43–7 (1999).

[277] Blaney, S. M., Phillips, P. C., Packer, R. J., Heideman, R. L., Berg, S. L., Adamson, P. C., Allen, J. C., Sallan, S. E., Jakacki, R. I., Lange, B. J., Reaman, G. H., Horowitz, M. E., Poplack, D. G., and Balis, F. M. Phase II evaluation of topotecan for pediatric central nervous system tumors. *Cancer* **78**(3), 527–31 (1996).

[278] Stewart, C. F., Iacono, L. C., Chintagumpala, M., Kellie, S. J., Ashley, D., Zamboni, W. C., Kirstein, M. N., Fouladi, M., Seele, L. G., Wallace, D., Houghton, P. J., and Gajjar, A. Results of a phase II upfront window of pharmacokinetically guided topotecan in high-risk medulloblastoma and supratentorial primitive neuroectodermal tumor. *J Clin Oncol* **22**(16), 3357–65 (2004).

[279] Kawasaki, H. and Taira, K. Retraction: Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells. *Nature* **426**(6962), 100 (2003).

[280] Guidelines for nomenclature of mouse and rat strains. http://www.informatics.jax.org/mgihome/nomen/strains.shtml. Accessed: 2015-03-19.

[281] Guidelines for nomenclature of genes, genetic markers, alleles, and mutations in mouse and rat. http://www.informatics.jax.org/mgihome/nomen/gene.shtml. Accessed: 2015-03-19.

[282] Pazzaglia, S., Mancuso, M., Rebessi, S., Di Majo, V., Tanori, M., Biozzi, G., Covelli, V., and Saran, A. The genetic control of chemically and radiation-induced skin tumorigenesis: a study with carcinogenesis-susceptible and carcinogenesis-resistant mice. *Radiat Res* **158**(1), 78–83 (2002).

[283] Pazzaglia, S., Mancuso, M., Tanori, M., Atkinson, M. J., Merola, P., Rebessi, S., Di Majo, V., Covelli, V., Hahn, H., and Saran, A. Modulation of patched-associated susceptibility to radiation induced tumorigenesis by genetic background. *Cancer Res* **64**(11), 3798–806 (2004).

[284] Mao, J., Ligon, K. L., Rakhlin, E. Y., Thayer, S. P., Bronson, R. T., Rowitch, D., and McMahon, A. P. A novel somatic mouse model to survey tumorigenic potential applied to the Hedgehog pathway. *Cancer Res* **66**(20), 10171–8 (2006).

[285] Xie, J., Murone, M., Luoh, S. M., Ryan, A., Gu, Q., Zhang, C., Bonifas, J. M., Lam, C. W., Hynes, M., Goddard, A., Rosenthal, A., Epstein, E. H. J., and de Sauvage, F. J. Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**(6662), 90–2 (1998).