Molecular subclassification of medulloblastoma and its utility for disease prognostication

Edward Carl Schwalbe

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Newcastle University

Faculty of Medical Sciences

Northern Institute for Cancer Research

October 2011

Declaration

I certify that no part of the material documented in this thesis has previously been submitted for a degree or other qualification in this or any other university. I declare that this thesis represents my own unaided work, carried out by myself, except where it is acknowledged otherwise in the thesis text.

Ed Schwalbe

October 2011

Acknowledgements

I would like to thank my supervisors, Professor Steven Clifford, and Dr. Simon Bailey for their unstinting support and encouragement through the project. Their expert guidance has helped make this project a challenging, but hugely enjoyable part of my life. Next, I would like to thank the members of the Brain Tumour Research Group, without whom I would have floundered. In particular, I would like to thank Dr. Janet Lindsey for her most helpful advice and critical appraisal of my work over the last four years. In addition, Dr. Dan Williamson has been most helpful in introducing me to the clustering techniques that underpin chapter 4. Within the wider NICR, I would also like to thank Mike Cole, in particular, for his patient and explanatory responses to some of my more obtuse statistical questions. Also within the NICR, Professor Anthony Moorman and Dr. Olaf Heidenreich, my thesis progression overseers, have helped ensure that my project has run smoothly, and have contributed perspicacious insight and criticism that have served to improve the thesis.

Finally, I would like to thank the Katie Trust and their fundraisers, whose hard work in raising money made this project possible.

Dedication

I would like to thank Jo and Eric for their support and encouragement, including uprooting to a different city, which has helped to ensure that the thesis and writing-up period have been as painless as possible.

Abstract

Medulloblastoma is the most common malignant brain tumour of childhood. Transcriptomic classification of the disease has indicated the existence of discrete molecular subgroups of medulloblastoma, although the precise number, nature and clinical significance of these subgroups remains unclear. Two groups, characterised by activation of the WNT and SHH signalling pathways, are common to all published studies. An assay for the rapid diagnosis of medulloblastoma subgroups was therefore designed, using transcriptomic gene signatures of pathway activation for the WNT and SHH signalling pathways. The successful validation of these gene signatures *in vitro* and *in silico* enabled a meta-analysis of 173 new and published cases to be performed, which defined the molecular and clinico-pathological correlates of the disease subgroups more precisely. WNT subgroup cases were associated with *CTNNB1* mutation, chromosome 6 loss and classic histology and were diagnosed > 5 years of age. SHH cases predominated in infants and showed an age-dependent relationship to desmoplastic / nodular histology. WNT / SHH independent tumours showed all histologies, peaked at 3 to 6 years and were associated with chromosome 17p loss.

A novel DNA methylation array-based approach was next applied to disease subclassification. Using consensus clustering, based on non-negative matrix factorisation, four methylomic subgroups were identified in a training cohort (*n* = 100), which were robustly validated in a test cohort (*n* = 130). The subgroups were characterised by significant relationships to specific clinico-pathological and molecular markers. Two subgroups were characterised by activation of the WNT and SHH signalling pathways and showed equivalent clinico-pathological and molecular characteristics to the previously defined transcriptomic subgroups. For the WNT / SHH independent subgroups, group I was associated with a loss of chromosome 17p, whereas group II was enriched for large cell / anaplastic (LCA) histology. The WNT subgroup was associated with a favourable prognosis, while no survival differences were apparent between the remaining subgroups (SHH, group I, group II). Specific methylation biomarkers were identified for the discrimination of all subgroups. Assays of DNA methylation status were robust in derivatives of FFPE tissues, enabling testing in routinely-collected clinical material.

Finally, the prognostic potential of methylomic biomarkers was investigated in a large clinical trials-based cohort (*n* = 191), with particular focus on the non-WNT subgroups (*n* = 163), where subgroup membership was not prognostic. Using the Cox Boost algorithm, which adds high dimensional data to mandatory clinical covariates to form cross-validated prognostic Cox survival models, the methylation status of *MXI1* and *IL8* were each identified as independent prognostic markers. These were incorporated into a novel risk stratification scheme, based on the cumulative assessment of disease risk using clinical (metastatic disease; poor prognosis), pathological (LCA pathology, poor prognosis) and methylomic variables (WNT subgroup, favourable prognosis; *MXI1* and *IL8* status). Importantly, this scheme assigns 46% of cases to a low risk group of patients (>90% survival) who could potentially be treated less intensively, with the aim of reducing therapy-associated late effects. This model out-performed the current clinical and other state-of-the-art medulloblastoma risk classification schemes. These data provide clear precedent for the utility of DNA methylation biomarkers for disease subclassification and prognostication in medulloblastoma, and their clinical application in diagnostic tumour biopsies.

List of abbreviations

3’ UTR	Three prime untranslated region
A	Adenine
AATK	Apopotosis-associated tyrosine kinase
Ab	Antibody
ABTA	American Brain Tumour Association
ANOVA	Analysis of variance
APC	Adenomatous polyposis coli
aPRIMES	Array based profiling of reference-independent methylation status
ASCL2	Achaete-scute complex homologue 2 (Drosophila)
AT / RT	Atypical teratoid / rhabdoid tumours
AUC	Area under the curve
B2M	Beta-2-microglobulin
BASH	Beadarray subversion of harshlight
BCHE	Butyrylcholinesterase
BCL2	B-cell CLL / lymphoma 2
BCR-ABL	Breakpoint cluster region c-abl oncogene 1, non-receptor tyrosine kinase
BLAST	Basic local alignment search tool
BLAT	Basic local alignment tool
BMI1	BMI polycomb ring finger oncogene
BP	Base pair
BRAF	v-raf murine sarcoma viral oncogene homologue B1
BRCA1	Breast cancer 1, early onset
BRCA2	Breast cancer 2, early onset
C	Cytosine
CA	Co-activator
CART	Classification and regression tree
CASP8	Caspase 8, apoptosis-related cysteine peptidase
CBTRUS	Central brain tumour registry of the United States
CCDC46	Coiled-coil domain containing 46
CCKAR	Cholecystokinin A receptor
CCNB1	Cyclin B1
CCRG	Childhood cancer research group
CDH1	Cadherin 1, type 1, E-cadherin (epithelial)
CDH13	Cadherin 13, H cadherin (heart)
CDKN2A	Cyclin dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
cDNA	Complementary DNA
CGH	Comparative genomic hybridisation
CHARM	Comprehensive high throughput arrays for relative methylation
CI	Confidence interval
CIMP	CpG island methylator phenotype
CLAS	Classic type medulloblastoma
CNGA3	Cyclic nucleotide gated channel alpha 3
CNN	Copy number neutral
CNS	Central nervous system
COL1A2	Collagen, type I, alpha 2
CpG	Cytosine-guanine dinucleotide

CREBBP	CREB binding protein
cRNA	Complementary RNA
CRUK	Cancer Research UK
CSF	Cerebro-spinal fluid
CTCF	CCCTC-binding factor (zinc finger protein)
CTTNB1	Catenin (cadherin-associated protein), beta 1
CV	Coefficient of variation
Cy3	Cyanine 3 dye
Cy5	Cyanine 5 dye
DAPK1	Death-associated protein kinase 1
DCC	Deleted in colorectal carcinoma
DHH	Desert hedgehog
dHPLC	Denaturing high performance liquid chromatography
DKK1	Dickkopf homologue 1 (Xenopus laevis)
DKK2	Dickkopf homologue 2 (Xenopus laevis)
DN	Desmoplastic / nodular medulloblastoma
DNA	Deoxyribose nucleic acid
*DNMT*1	DNA methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3B
DSH	Dishevelled
EFS	Event free survival
EGFR	Epidermal growth factor receptor
EGL	External granule layer
ERBB2	v-erb-b2 erythroblastic leukaemia viral oncogene homologue2, neuroma / glioblastoma
ES	Embryonic stem
ESR1	Oestrogen receptor 1
FFPE	Formalin-fixed, paraffin embedded
FGF18	Fibroblast growth factor 18
FGF4	Fibroblast growth factor 4
FISH	Fluorescent, in-situ hybridisation
FP	False positive
FSTL5	Follistatin-like 5
G	Guanine
GAB1	GRB2-associated binding protein 1
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GEO	Gene expression omnibus
GLI1	GLI family zinc finger 1
GLI2	GLI family zinc finger 2
GLI3	GLI family zinc finger 3
GNP	Granule neuron precursor
GRM1	Glutamate receptor, metabotropic 1
GRM8	Glutamate receptor, metabotropic 8
GSK-3β	Glycogen synthase kinase 3 beta
GSTP1	Glutathione-S-transferase pi 1
GTP	Guanine triphosphate
Gy	Gray
H & E	Haematoxylin and eosin

HAT	Histone acetyl transferase
HC	Hierarchical clustering
HDAC	Histone deacetylase
HFE	Hemochromatosis
HGF	Hepatocyte growth factor (hepapoietin A; scatter factor)
Hh	Hedgehog
HIC1	Hypermethylated in cancer 1
HL	Hodgkin's lymphoma
HOMOD	Homozygosity mapping of deletion
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homologue
IDH1	Isocitrate dehydrogenase 1
IHH	Indian hedgehog
IL8	Interleukin 8
ITIH2	Inter-alpha (globulin) inhibitor H2
IVT	In-vitro transcription
KCDT11	Potassium channel tetramerisation domain containing 11
KCNA1	Potassium voltage-gated channel, shaker-related subfamily, member 1
KRAS	v-ki-ras2 Kirsten rat sarcoma viral oncogene homologue
LCA	Large cell / anaplastic medulloblastoma
LDHB	Lactate dehydrogenase B
LEF	Lymphoid enhancer factor
LINE-1	Long interspersed nuclear element 1
LOH	Loss of heterozygosity
LOOCV	Leave-one-out cross-validation
MAD	MAX dimerisation protein
MATH1	Atonal homologue 1 (Drosophila)
MAX	MYC associated factor X
MB	Medulloblastoma
MBEN	Medulloblastoma with extensive nodularity
MBP	Methyl-cytosine binding protein
MC-SVM	Multi-class support vector machine
MDS	Myelodysplastic syndrome
MeDIP	Methylated DNA immunoprecipitation
MET	Met proto-oncogene (hepatocyte growth factor receptor)
µg	Microgram
MGMT	O-6-methylguanine-DNA methyltransferase
MICAL1	Microtubule associated monoxygenase, calponin and LIM domain containing 1
µl	Microlitre
MLH1	MutL homologue 1, colon cancer, nonpolyposis type 2 (E. coli)
MLL2	Myeloid / lymphoid or mixed –lineage leukaemia 2
MLL3	Myeloid / lymphoid or mixed –lineage leukaemia 3
µm	Micrometre
MRI	Magnetic Resonance Imaging
mRNA	Messenger RNA
MSH2	MutS homologue 2, colon cancer, nonpolyposis type 1 (E. coli)
MSP	Methylation specific PCR
MXI1	MAX interactor 1
MYCC	v-myc myelocytomatosis viral oncogene homologue (avian)

MYCL	v-myc myelocytomatosis viral oncogene homologue 1, lung carcinoma derived (avian)
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)
NC	Not classifiable
ng	Nanogram
NGS	Next generation sequencing
NHL	Non-Hodgkin's lymphoma
NMB	Newcastle medulloblastoma
NMF	Non-negative matrix factorisation
NOS2A	Nitric oxide synthase 2, inducible
NPR3	Natriuretic peptide receptor C
OD	Optical density
ONS	Office for National Statistics
OS	Overall survival
OTX2	Orthodenticle homeobox 2
$P14^{ARF}$	Cyclin dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4) ARF
$P16^{INK4A}$	Cyclin dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4) INK4A
PALB2	Partner and localiser of BRCA2
PCA	Principal component analysis
PcG	Polycomb group proteins
PCR	Polymerase chain reaction
PDGF	Platelet derived growth factor
PDGFB	Platelet derived growth factor beta
PDLIM3	PDZ and LIM domain 3
PEFR	Peak expiratory flow rate
PFS	Progression free survival
PNET3	Primitive-neuroectodermal tumour 3 trial
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PTCH1	Patched 1
PTCH2	Patched 2
PTEN	Phosphatase and tensin homologue
PYGL	Phosphorylase, glycogen, liver
QC	Quality control
RAB33A	RAB33A, member RAS oncogene family
RB1	Retinoblastoma 1
REN	Renin
RIN	RNA integrity number
RJG	Richard J Gilbertson
RLGS	Restriction landmark genome scanning
RMA	Robust multi-array average
RNA	Ribonucleic acid
ROC	Receiver Operating Characteristic
rRNA	Ribosomal RNA
RT	Reverse transcriptase
RT-PCR	Reverse transcriptase PCR

SFRP1	secreted frizzled-related protein 1
SHH	Sonic hedgehog
SIOP	International Society of Paediatric Oncology
SMARCA3	Helicase like transcription factor
SMARCA4	SWI / SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
SMO	Smoothened homologue (Drosophila)
SNP	Single nucleotide polymorphism
SOCS1	Suppressor of cytokine signalling 1
SPDEF	SAM pointed domain containing ets transcription factor
SPINT1	Serine peptidase inhibitor, Kunitz type 1
SPINT2	Serine peptidase inhibitor, Kunitz type 2
SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homologue (avian)
SUFU	Suppressor of fused homologue (Drosophila)
SVM	Support vector machine
T	Thymine
TCF	T cell factor
TET	Ten-eleven translocation
TF	Transcription factor
TFPI2	Tissue factor pathway inhibitor 2
TIMP3	TIMP metallopeptidase inhibitor 3
TMEM45A	Transmembrane protein 45A
TNC	Tenascin C
TP	True positive
TP53	Tumour protein 53
TSG	Tumour suppressor gene
TSP-1	Thrombospondin-1
TSS	Transcriptional start site
U	Uracil
UKCCSG	United Kingdom Children's Cancer Study Group
UV	Ultra-violet
VEGF	Vascular endothelial growth factor
VEGFA	Vascular endothelial growth factor A
VHL	Von-Hippel Landau
WCRF	World cancer research fund
WGSBS	Whole genome shotgun bisulfite sequencing
WHO	World Health Organisation
WIF1	WNT inhibitory factor 1
WNT	Wingless
WT1	Wilms' tumour 1
YAP1	Yes-associated protein 1
ZIC2	Zic family member 2 (odd-pairing homologue, Drosophila)

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Cancer

Cancer is an umbrella term for a numerous, heterogeneous collection of diseases, all characterised by uncontrolled growth and invasion of cells that encroach upon or displace surrounding tissue. Under normal conditions, tight regulation of cellular proliferation and apoptosis ensures a homeostasis of cell numbers that maintains normal tissue architecture and function. Any event that disrupts this homeostasis in favour of proliferation, by affecting genes involved in processes such as apoptosis and / or proliferation can lead to aberrant proliferation, which in turn can generate a tumour.

A tumour is simply defined as an abnormal growth of body tissue. It encompasses benign, pre-malignant and malignant forms, with cancer, by definition, a malignant tumour, because of its ability to invade and infiltrate surrounding tissues. Invasion is the process by which cells break away from the primary tumour and move into adjacent tissues. Metastasis is defined as the spread of secondary tumours, which detach from the primary tumour, to other non-adjacent locations in the body via the blood or lymphatic system. The determination of metastatic status is an important prognostic factor in most cancers. More than 200 different cancers have been described, and cancer remains one of the most important disease burdens worldwide (CRUK, 2011). Although many cancers have environmental determinants, ultimately, these environmental determinants have to materially alter the expression of genes to favour inappropriate proliferation of a cell, meaning that cancer is fundamentally a genetic disease.

1.1.1 Cancer incidence

Despite an increased understanding of the determinants of cancer, and the lifestyle choices that can reduce the risk of developing cancer, cancer incidence in the developed world has continued to rise, with a lifetime risk of developing cancer in the UK of 40% in 2008 (Figure 1.1). The reasons for this increase are complex, but must be understood in the context that improvements in the treatment of diseases other than cancer mean that people are living longer and therefore are at an increased risk of developing cancer through increased longevity.

As shown in Figure 1.1, there are marginally more cancers in males than females. Notwithstanding the gender-specific cancers (principally breast, ovarian, uterine and prostate cancers), males have a higher incidence and mortality in almost every major cancer (Figure 1.2), and efforts are being undertaken to encourage lifestyle changes in men which may reduce their excess cancer burden, due to an increased incidence of smoking, higher alcohol intakes and a reluctance to seek medical attention (Peate, 2011).

In 2008, the most recent year with available statistics, 309,500 people were diagnosed in the UK (CRUK, 2011). In the same year, there were 150,000 deaths from cancer in the UK (CRUK, 2011). Interestingly, rates of cancer in the UK are substantially higher than in many other developed countries. The UK has the 22nd highest cancer rate in the world, and estimates suggest that approximately one third of these cancers are preventable through changes in diet, weight loss and an increase in physical activity (WCRF, 2010). Cancer is predominantly a disease that affects the elderly, with childhood cancer (defined here as age at diagnosis less than 15 years of age) representing only 1% of all newly diagnosed cancers in the UK, equivalent to 1500 cases annually (CRUK, 2011) (Figure 1.3).

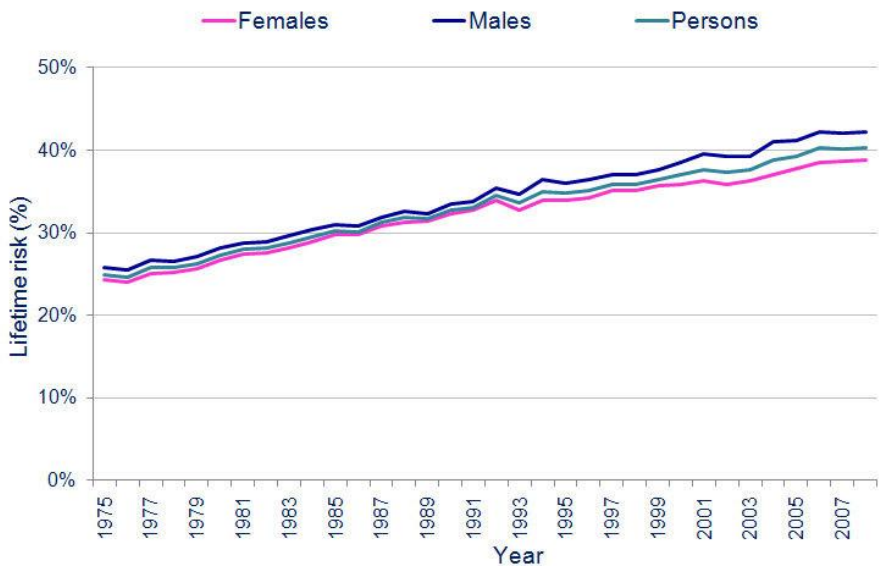
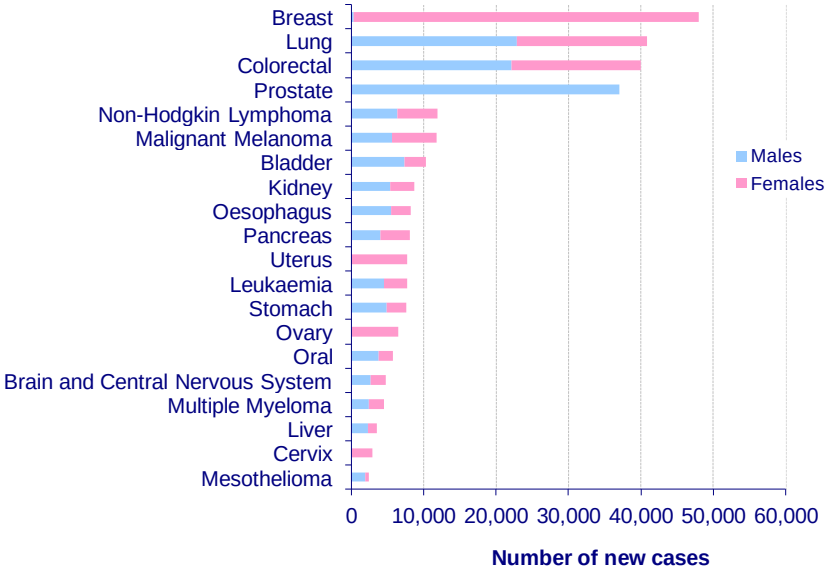


Figure 1.1. Increasing lifetime risk of cancer in Great Britain, 1975-2008. Figure taken from Cancer Research UK cancer stats (CRUK, 2011).

A



B

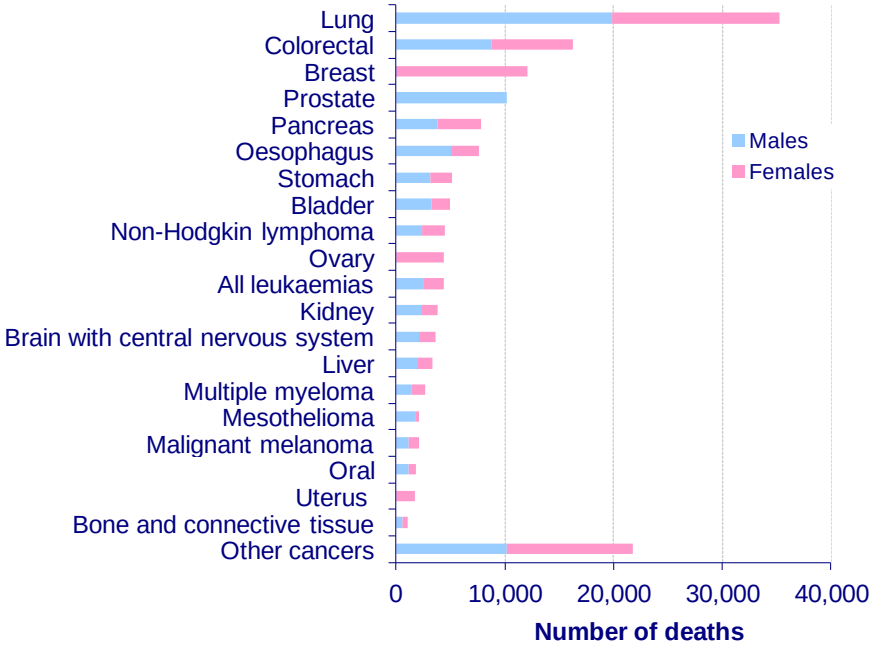


Figure 1.2. Incidence and mortality rates of common cancers in the UK. **A**.The 20 most commonly diagnosed cancers in the UK in 2008. **B**. The 20 most common causes of death from cancer in the UK in 2008. Data supplied from http://info.cancerresearchuk.org/cancerstats.

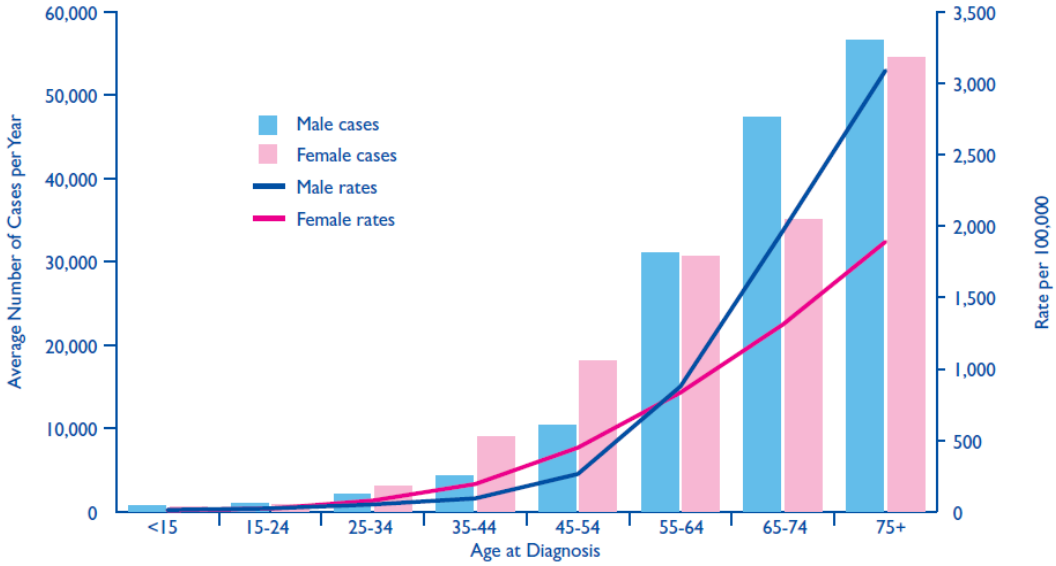


Figure 1.3. Average number of new cancers per year increases with age. Figure taken from Cancer Research UK cancer stats (CRUK, 2011).

1.1.2 Childhood cancer

In contrast to adult cancers, childhood cancers (defined here as age at diagnosis less than 15 years) often arise in locations derived from embryonal tissues, in contrast to adult cancers, which are primarily epithelial in origin. Leukaemias and cancers of the central nervous system (CNS) comprise greater than 50% of childhood cancers (Figure 1.4). Survival rates can vary according to cancer type, with a five year survival rate for all paediatric cancers of 78% in the period from 2001-2005 (CCRG, 2010), although survival rates for aggressive forms of glioma are as low as 5%. Cancer remains one of the leading causes of death in childhood, responsible for over 300 deaths a year in the UK (CRUK, 2011). Moreover, the deleterious side effects that are a consequence of cancer treatments in children confer a heavy burden on survivors and their carers, through the late effects of treatment, which can include neuro-cognitive and neuro-endocrine defects. There is a pressing need to further characterise the nature and behaviour of childhood cancers, both to reduce the death rate, but also to identify methodologies that might mitigate the severity of treatment side effects. This is a problem unique to childhood cancers, due to the long survival times and relatively high cure rates.

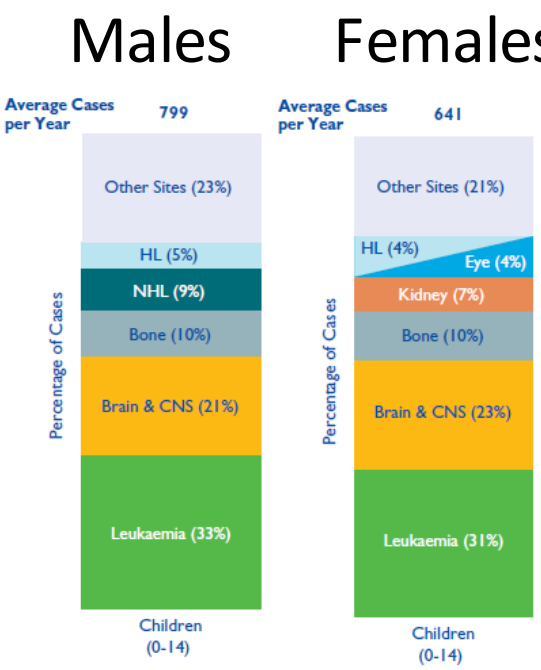


Figure 1.4. Incidence of most common cancers of childhood in the UK. HL – Hodgkin's lymphoma; NHL – Non-Hodgkin's lymphoma. Figure adapted from Cancer Research UK cancer stats (CRUK, 2011)

1.1.3 Cancer is a multi-stage process

Normal physiological and cellular processes have evolved to keep cancer in check, and in order for a normal cell to become malignant, it must acquire various capabilities in order to evade these defences. Histopathological investigations have described a typical progression in solid tumours, initiating with tissue hyperplasia (excessive proliferation of normal cells), which progresses into dysplasia (excessive proliferation of immature cells), and finally into invasive carcinoma. In their landmark papers, Hanahan and Weinberg describe six hallmarks of cancer which a cell must acquire in order to become malignant (Hanahan and Weinberg, 2011; Hanahan and Weinberg, 2000), (Figure 1.5). Although the acquisition of these abilities will be common to nearly all cancers, the mechanisms by which these abilities are acquired are very heterogeneous, both within specific cancers and across all cancers. The specific processes are discussed below.

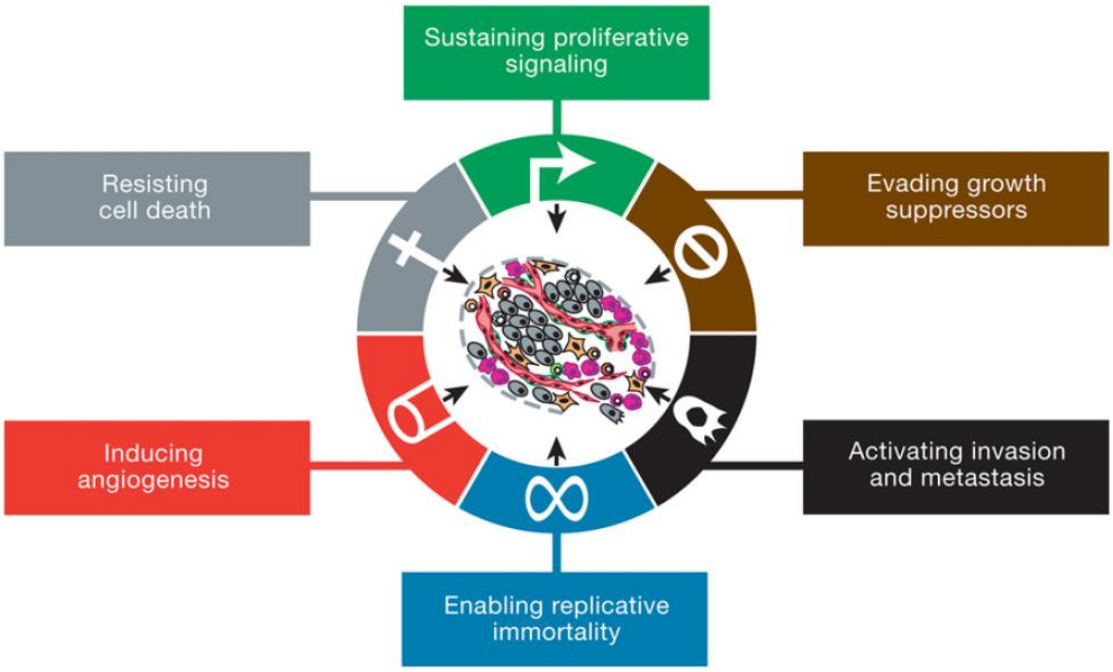


Figure 1.5. The six hallmark capabilities of cancer, necessary for tumour growth. Figure taken from (Hanahan and Weinberg, 2011)

1.1.3.1 Sustained proliferative signalling

In normal tissues, the processes that mediate entry into and progression through the cell cycle are tightly controlled, ensuring a homeostasis of cell numbers, and a maintenance of tissue architecture. Cancer cells acquire the capability to sustain proliferative signalling through various mechanisms: They may produce growth factor ligands themselves, or stimulate normal cells within the adjacent stroma to produce proliferative growth factors. Cancer cells can also become hyper-responsive to growth factors by increasing the number of receptors at the cell surface. Alternatively, the constitutive activation of proliferative signalling pathways can obviate the need for growth factors (Hanahan and Weinberg, 2011).

1.1.3.2 Evasion of growth suppressors

In addition to the stimulatory mechanisms outlined above, cancer cells must also evade negative regulation of cell proliferation. Many examples of tumour suppressor genes (TSGs) (section 1.1.6) have been described that act as control proteins that determine whether the cell is free to proceed along proliferative pathways, or whether cellular senescent and apoptotic pathways should be activated (Hanahan and Weinberg, 2011).

1.1.3.3 Activating invasion and metastasis

The processes which govern invasion and metastasis remain poorly understood, but a series of discrete steps that characterise invasion and metastasis have been described. In this series, local invasion is the first step, which is governed by changes to the mechanisms by which cancer cells interact with adjacent cells and the extracellular matrix. The next step is intravasation by cancer cells into adjacent blood and lymphatic vessels, where they can be transported to distant tissues. After transportation, cancer cells escape into the surround parenchyma and form small nodules of cancer cells (micro-metastases). Finally, the micro-metastases grow into additional macroscopic tumours, a process termed colonisation.

1.1.3.4 Enabling replicative immortality

Cancers are widely accepted as having acquired limitless replicative potential. In normal cells, only a limited number of successive cell divisions are possible before cells become senescent, a non-proliferative but viable state. If division is continued, cells reach a crisis phase and the majority of cells die. Rarely, cells acquire limitless replicative potential, and it has been established that changes in telomere structure are the mechanism by which such cells can become immortalised (Blackburn, 1991). Under normal conditions of cell division, telomere length successively shortens, until the telomeres have largely eroded, have lost their protective functions and trigger senescence. Telomerase, an enzyme that can counter telomere erosion, is activated in ~90% of immortalised cells (including cancer cells) and its expression is correlated with a resistance to senescence and crisis / apoptosis (Hanahan and Weinberg, 2011).

1.1.3.5 Induction of angiogenesis

Invading and growing tumours require nutrients and oxygen, and the capability to evacuate metabolic wastes. This capability is supplied by the stimulation of angiogenic mechanisms that provide a tumour-associated neo-vasculature. The induction of angiogenesis is mediated by various factors that either induce or oppose angiogenesis. The most widely studied examples of these factors are the inducer vascular endothelial growth factor A (VEGF-A) and the inhibitor thrombospondin-1 (TSP-1). Under normal conditions, the *VEGFA* gene is activated during embryonal and post-natal development, as well as in pathological conditions in the adult. In cancer, *VEGFA* expression is induced by hypoxia and oncogene signalling. TSP-1 binds trans-membrane receptors which stimulates suppressive signals that can counter pro-angiogenic stimuli (Hanahan and Weinberg, 2011).

1.1.3.6 Resisting cell death

Cancer cells must overcome apoptotic mechanisms to persist and proliferate. The activation of apoptotic processes can be triggered by stresses resulting from elevated levels of oncogene (section 1.1.5) signalling, and DNA damage due to hyper-proliferation. Two apoptotic regulatory mechanisms exist for determining the activation of apoptosis. One receives and processes extra-cellular triggers of apoptosis,

the other senses intracellular pro-apoptotic signals. Both culminate in the activation of proteases which proteolytically disassemble the cell before phagocytosis. The most commonly observed mechanism for the evasion of apoptosis is the loss of tumour protein 53 (*TP53*) and its tumour suppressor function (Hanahan and Weinberg, 2011).

1.1.3.7 Genetic model for development of cancer

More than one mutation is necessary for development of cancer. A series of mutations to several classes of gene is normally required before a normal cell can transform into a tumour cell. The monoclonal origin of cancer is a model that described the process of tumourigenesis being dependent upon a series of genetic and epigenetic events that sequentially accumulate in a cell population that is derived from a single initiating mutation in a single cell.

Under the model, a mutation in a single cell provides a selective growth advantage over surrounding somatic cells. The mutated cell proliferates, and creates a population of cells with the same genotype, increasing the likelihood of additional tumorigenic mutations and proliferative capability. Each successive event provides an additional growth advantage and is clonally expanded until a tumour is formed which has acquired the capabilities described in section 1.1.3.

In their classic paper from 1990, Fearson and Vogelstein present a model for the clonal evolution of colorectal carcinoma (Fearon and Vogelstein, 1990) that formalised the model for the clonal evolution of cancer. Colorectal tumours almost always arise from pre-existing benign tumours (adenomas), and the process by which normal epithelium forms adenomas, which progress to full carcinoma, is governed by the continued acquisition of genetic and epigenetic defects that drive proliferation and expansion, as shown in Figure 1.6.

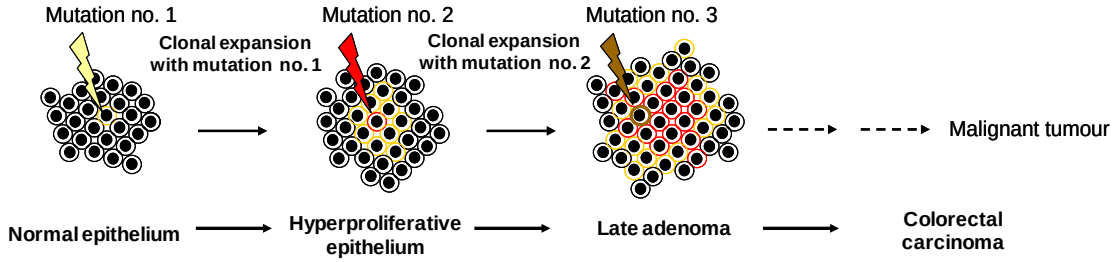


Figure 1.6. A model for the development of colorectal cancer. Although the specific genetic changes that occur through tumour progression will vary, the underlying histological changes are likely to be common for every colorectal carcinoma. Figure adapted from Fearon and Vogelstein, 1990.

The monoclonal origin of cancer has been confirmed in several types of cancer by chromosome X inactivation studies, which have demonstrated that cells within a single tumour, and cells of anatomically distinct tumours from an individual patient, have the same number of inactive X chromosomes and therefore originate from a single cell (Wang et al., 2009b; Jones et al., 2005; Jacobs et al., 1992). However, some cancers have been shown to contain cells with a variable pattern of X chromosome inactivation, suggestive of a polyclonal origin, whereby cancers develop from multiple colonies of genetically distinct cells that interact to initiate tumorigenesis (Parsons, 2008).

1.1.4 Genetic determinants of cancer

As discussed in section 1.1, cancer initiates with an abnormal proliferation of cells with the capability of infiltrating and displacing surrounding tissues. There is a widely accepted view that a cell must acquire multiple genetic lesions in order for this to occur, as described in section 1.1.3.7. The majority of these lesions arise from environmental exposure to mutagenic factors (ultra-violet (UV) light, ionising radiation, chemical mutagens present in the environment). However, there are a number of inherited cancer syndromes where affected individuals are heterozygous for a mutated allele, and are vulnerable to acquiring a further mutation in the second allele, which can act as the initiating step of the multi-stage process of cancer. One example of cancer genes are the *BRCA* family of genes, with female members of affected families harbouring harmful mutations being particularly susceptible to developing breast or ovarian cancers. Women carrying harmful mutations in *BRCA1* or *BRCA2* have a

lifetime risk of breast cancer that is five times normal, and a risk of ovarian cancer that is ten to thirty times normal (Kadouri et al., 2007). Two gene classes are fundamental to the development of cancer: Oncogenes and tumour suppressor genes.

1.1.5 Oncogenes

An oncogene is a gene that, when mutated or expressed at abnormally high levels, has the potential to cause cancer by directly promoting cell growth and proliferation, inhibiting apoptosis, or facilitating other hallmarks of a cancer cell illustrated in section 1.1.3. The first oncogene, *SRC*, was identified in 1970 in a chicken retrovirus (Martin, 1970). Since their initial description in viruses, studies in cancer cell lines identified that oncogenes can be mutated forms of normal genes, termed proto-oncogenes (Stehelin et al., 1976). Proto-oncogenes encode proteins whose normal function is to regulate cell growth and differentiation. They can be converted into oncogenes by acquiring enhanced functionality through dominant gain-of-function mutations, or by an upregulation of expression of the normal product (Strachan and Read, 1999). Many proto-oncogenes have subsequently been identified and, with that, came the realisation that many oncogenes are different components of the same pathway, implicating a relatively small number of pathways in the development of cancer. Since the activation of many oncogenes has been identified in diverse cancers, there is hope that specific pathway inhibitors will be useful against a broad range of cancer types (Vogelstein and Kinzler, 2004), defined by their oncogenic activation profile. Examples of well-studied oncogenes are shown in Table 1.1. The mechanisms by which proto-oncogenes become oncogenes are described below.

Oncogene	Normal Function	Cancer Association
BRAF	Intracellular signalling	Melanoma
EGFR	Growth factor receptor	Glioma, non-small cell lung cancer
ERBB2	Growth factor receptor	Breast
HRAS	Component of signal transduction pathway	Colorectal carcinoma, bladder, rhabdomyosarcoma
MYCN	Transcription factor	Neuroblastoma / medulloblastoma
PDGFB	Growth factor	Dermatofibrosarcoma, fibroblastoma
KRAS	Intracellular signalling	Colorectal
VEGF	Promotes angiogenesis	Metastatic colorectal

Table 1.1. Examples of proto-oncogenes, their normal function and examples of cancer in which they can become activated. Data gathered from the Cancer Genome Project (www.sanger.ac.uk/genetics/CGP/Census/),(Strachan and Read, 1999).

1.1.5.1 Activation of proto-oncogenes

Many cancer cells contain many copies of structurally intact oncogenes, a process named gene amplification. For example, many neuroblastoma cancers are amplified in the oncogene *MYCN* (Schwab, 1990), and some medulloblastomas are amplified for *MYCC* (Bigner et al., 1990). The increase in gene copy number can promote messenger RNA (mRNA) expression of the oncogene which can drive tumourigenesis. The mechanisms that lead to gene amplification are poorly understood, although genomic instability is a hallmark of cancer cells and a loosening of the checkpoints that control genome integrity can permit gene amplification to occur.

An alternative mechanism for proto-oncogene activation is through point mutation. For example, specific point mutations in the *HRAS* gene can abrogate the GTPase

activity of the HRAS protein. In turn, this causes the G-protein-coupled receptor signalling protein GTP-RAS to be inactivated more slowly, leading to excessive cellular response to signalling from G-protein coupled receptors (Khosravi-Far and Der, 1994).

The karyotype of tumour cells is typically abnormal, with structural abnormalities such as chromosomal gains, losses or translocations relatively common. One mechanism for oncogene activation is by generation of fusion genes along well-defined breakpoints. The best known example of this is the Philadelphia chromosome, observed in 90% of patients with chronic myeloid leukaemia (Whang et al., 1963). This involves a balanced reciprocal 9;22 translocation, forming a novel fusion gene, *BCR-ABL* (Strachan and Read, 1999). An alternative translocation mechanism transposes oncogenes into active chromatin domains. The translocation of the *MYCC* gene to actively transcribed chromatin regions proximal to immunoglobulin loci is a hallmark of Burkitt's Lymphoma (Strachan and Read, 1999). In this case, the intact gene is translocated, and, without its normal control, is transcribed at an inappropriately high level.

1.1.6 Tumour suppressor genes

Tumour suppressor genes (TSGs) are a class of genes whose gene products prevent a cell from progressing towards cancer. They can be inhibitory to processes that drive cellular proliferation and growth or can function to promote genome stability, by ensuring fidelity in DNA replication, repairing damaged DNA, or promoting apoptosis if the damage is irreparable (Charles, 2004).

The two-hit hypothesis of TSG inactivation was proposed by Knudsen in 1971, after he observed that the age of onset of retinoblastoma followed second order kinetics, with the implication that two separate genetic events were necessary for tumourigenesis to occur (Knudson, 1971). He recognised that mutations in TSGs were recessive, and inactivation of a TSG would be bi-allelic, in contrast to oncogene mutations, which involve gain-of-function and are dominant. Later experiments with retinoblastoma validated Knudsen's hypothesis. In familial retinoblastoma, one mutated form of the *RB1* gene is inherited, with only one more hit at the affected locus necessary for the initiation of tumourigenesis. As a result, the tumour tends to occur bilaterally in children < 2 years of age. Sporadic retinoblastomas require the independent

inactivation of both *RB1* alleles and tend to occur unilaterally in older children (Cavenee et al., 1983).

Since their initial description, more than 100 TSGs have been identified through their association with sporadic and inherited cancers. However, there are notable exceptions to the two hit rule, with loss of a single copy of the *PTCH1* gene sufficient to initiate medulloblastoma, due to haplo-insufficiency (Goodrich et al., 1997), and the loss of a single copy of *PTEN* can promote progression of prostate cancer by the same mechanism (Kwabi-Addo et al., 2001). Commonly inactivated TSGs are listed in Table 1.2.

Gene	Normal Function	Cancer Association
APC	Scaffold protein	Colorectal
BRCA1	Transcriptional control	Breast and ovarian carcinoma
CDKN2A	Regulates cell division	Melanoma
DCC	Regulates growth processes	Colorectal carcinoma
RB1	Regulates cell division	Retinoblastoma
TP53	Regulates cell division and apoptosis	Lung, medulloblastoma
VHL	Regulates cell division and angiogenesis	Kidney
WT1	Transcriptional control	Wilms' Tumour

Table 1.2. Examples of TSGs, their normal function and examples of cancers in which they are inactivated. Table adapted from Strachan and Reid, 1999.

1.1.6.1 Inactivation of tumour suppressor genes

Many mechanisms for the inactivation of TSGs have been reported. Whole chromosomal loss, chromosomal copy number neutral (CNN) loss of heterozygosity (LOH), mitotic recombination, gene conversion, gene deletion, point mutation and epigenetic inactivation through hypermethylation of the gene promoter (section 1.1.9) are all recognised mechanisms for the inactivation of TSGs (Strachan and Read, 1999). Importantly, bi-allelic loss can occur via the same mechanism, or by any combination of two mechanisms. Chromosomal loss and loss with duplication can lead to LOH compared to constitutional DNA, which is detectable through running genomic arrays such as array CGH (comparative genomic hybridisation) and SNP (single nucleotide polymorphism) arrays and is a well-described method for detecting candidate tumour suppressor loci in cancer.

1.1.7 Epigenetics

Epigenetics is the study of heritable changes in gene expression caused by mechanisms other than changes in the primary DNA sequence, and there is an increasing body of evidence that critical events in tumourigenesis may be driven by epigenetic mechanisms (Baylin and Jones, 2011; Jones and Baylin, 2002; Jones and Laird, 1999). As of 2011, four major classes of epigenetic regulation have been shown to be important in cancer: DNA methylation, chromatin remodelling, post-translational histone modifications and gene regulatory micro RNAs. None of these processes function in isolation, with the permissiveness for gene transcription being the result of a complex interplay of epigenetic processes which determine whether DNA is in its inactive (heterochromatin) or active form (euchromatin), shown in Figure 1.7, and is discussed in detail in section 1.1.8.

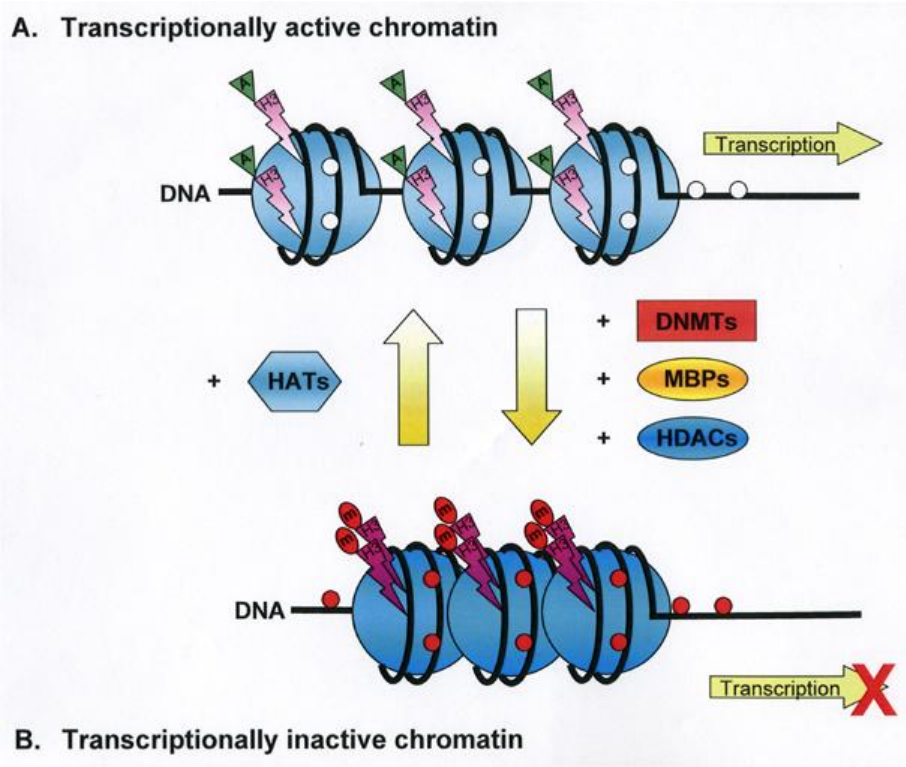


Figure 1.7. Epigenetic control of gene transcription by dynamic changes in DNA methylation status and chromatin modifications. In the upper panel, **A**,DNA is depicted in its active form, euchromatin; nucleosomes (large blue circles) are relaxed, histone H3 (lightning bolt) is acetylated (green triangle) and CpG dinucleotides are unmethylated (white circles). DNA (black line) is now permissive for induction of transcription. Through the interplay of DNMTs (DNA methyltransferases), MBPs (methyl-binding proteins), HDACs (histone deacetylases) and HATs (histone acetyltransferases), DNA can be rendered transcriptionally inactive (lower panel, **B**, which depicts heterochromatin, defined by nucleosome compaction). Nucleosomes have condensed, histone H3 has become methylated (red ovals) and deacetylated, and CpG dinucleotides are now methylated (red circles). Figure taken from (Lindsey et al., 2005).

1.1.8 Regulation of gene expression by epigenetic mechanisms

The architecture of gene expression states has been subject to intense study over the last decade, and the complexity and number of processes that can affect gene transcription have become more fully characterised (Baylin and Jones, 2011). The organisation of gene promoters is dependent on chromatin structure and organisation at regions flanking the promoter, and on the specific interplay of regulatory proteins at the site of the promoter, mediated by dynamic changes to methylation and / or acetylation of regulatory proteins.

1.1.8.1 Histone modifications and nucleosome reorganisation

Active gene promoters are characterised by nucleosome-depleted regions upstream of their transcriptional start sites (TSSs). The nucleosomes that flank these depleted regions are marked by trimethylation of lysine 4 of histone 3 (H3K4me3), have acetylated lysines, and harbour the histone variant, H2.AZ, which may destabilise nucleosomes to facilitate initiation of transcription (Figure 1.8). The transcribed regions of active genes also show enrichment for specific histone modifications, such as histone H3 lysine 36 trimethylation (H3K36me3), which may enable elongation of transcription (Guenther et al., 2007). Although DNA methylation represses initiation of transcription in promoter regions (Guenther et al., 2007), dense methylation of gene bodies may also promote elongation of transcription (Maunakea et al., 2010). Enhancer regions, shown in Figure 1.8, are characterised by nucleosome depletion, the presence of H3K4me1 and acetylation of lysine 27 of histone H3 (H3K27Ac).

DNA methylation, mediated by DNA methyltransferases (DNMTs), stabilises epigenetic gene silencing (Lin et al., 2007) and 'locks' DNA in a silent state (Deaton and Bird, 2011) (Figure 1.8), in promoter regions that lack H2A.Z, have nucleosomes located across the TSS, and which additionally harbour repressive histone modifications such as histone H3 lysine 9 bi- or trimethylation (H3K9me2 or H3K9me3). It is now understood that the maintenance of genome-wide expression states is determined by the balance of transcriptionally permissive and transcriptionally repressive chromatin modifications (Chi and Bernstein, 2009).

In addition to DNA methylation, promoters can also be silenced by polycomb group proteins (PcGs), shown in Figure 1.8. In humans, the protein complex PRC2 mediates histone H3 lysine 27 trimethylation (H3K27me3), which is recognised by the protein complex PRC1, which inhibits transcriptional elongation by ubiquitylation (Stock et al., 2007) and chromatin compaction (Eskeland et al., 2010). While DNA methylation at promoters is thought to remain relatively stable following development, polycomb-mediated repression is a more dynamic repression system (Mohn et al., 2008). In embryonic stem cells, CpG islands (section 1.1.9), silenced by PcG proteins, possess the active transcriptional histone marks H3K4me3 and H3K27me3, existing in a 'bivalent' state, which is poised between active transcription and stable repression. This bivalency may allow regulatory flexibility, for activation when needed during differentiation (Chi and Bernstein, 2009), or conversion to a repressive state for genes whose expression is not required. Bivalent CpG island promoters are observed in one fifth of CpG island promoters in embryonic stem (ES) cells (Ku et al., 2008), but are also found to a lesser extent in other cell types (Roh et al., 2006). It is thought that bivalent gene promoters are prone to acquiring *de novo* DNA methylation in cancer and pre-cancerous cells (Ohm et al., 2007; Widschwendter et al., 2007), providing a mechanism for the DNA hypermethylation observed at specific gene promoters in cancer.

Additional control is provided by global modifications to chromatin architecture. Insulators such as *CTCF* (Guelen et al., 2008), along with PcG occupancy, organise DNA into repressive heterochromatin or active euchromatin loops to enable large-scale repression or to connect distal enhancers and proximal promoters. The major mechanisms and modification that activate and repress DNA transcription are summarised in Table 1.3. A more detailed summary of the proteins involved in controlling transcription at the TSS is given below.

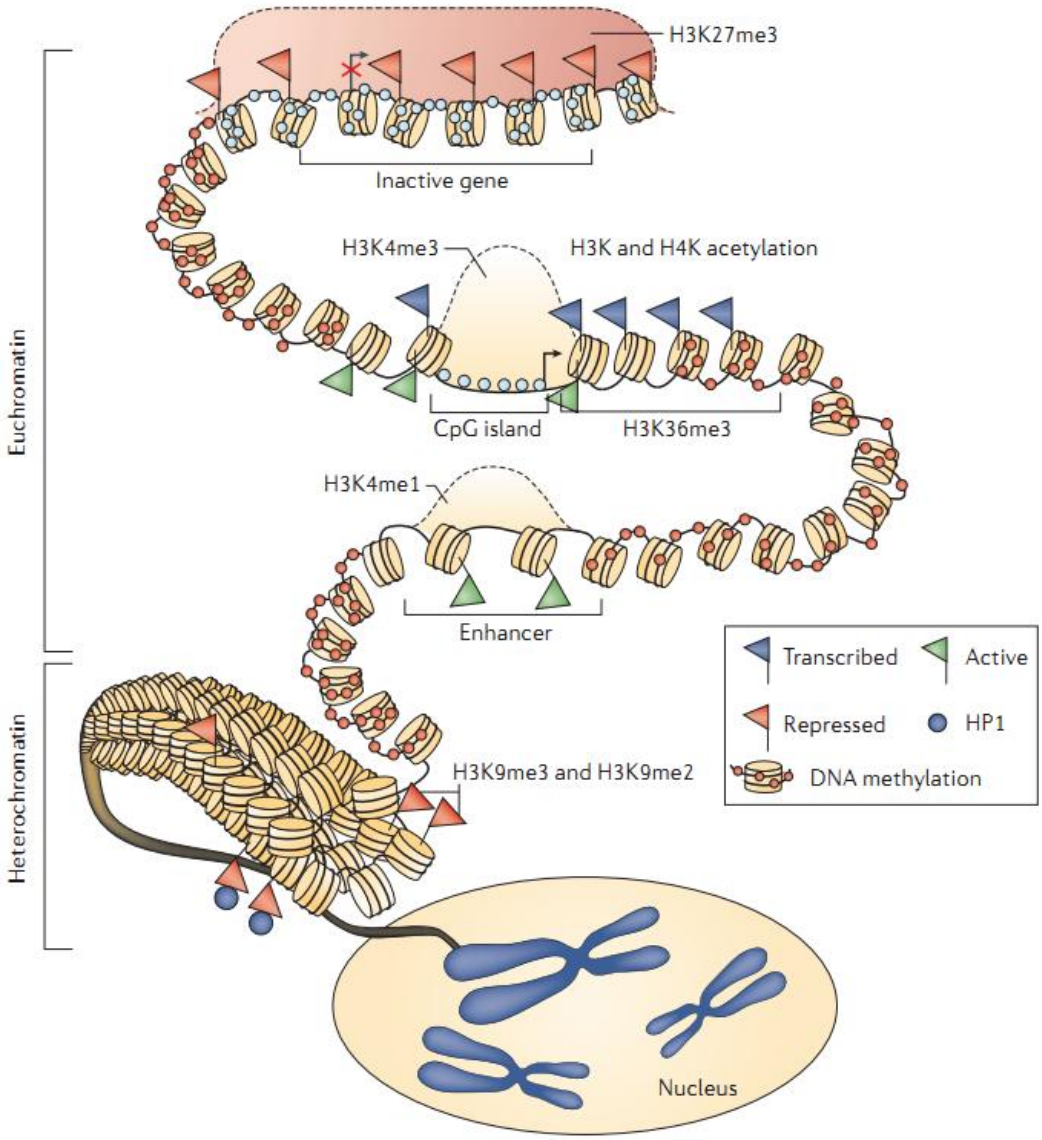


Figure 1.8. Model of the structure of the epigenome in normal human cells. A silenced gene (indicated by an X over the transcriptional start site) at the top of the figure has its promoter region occupied by a Polycomb group (PcG) complex, indicated by a red oval, that mediates repressive changes including the repressive histone modification trimethylation of lysine 27 on histone 3 (H3K27me3). CpG dinucleotides within this region are unmethylated (blue circles) and nucleosomes (yellow reels) are positioned over the transcriptional start sites. The gene promoter beneath the silenced gene is in a permissive state for gene transcription, characterised by active histone H3 lysine 4 trimethylation (H3K4me3) at the promoter. In addition, histones H3 and H4 are acetylated at key lysines, and the trimethylation of histone 3 lysine 36 (H3K36me3) aids transcriptional elongation. The transcriptional start site is not occupied by nucleosomes. A distal enhancer for this gene is also shown, characterised by an active nucleosome configuration and the signature histone modification for enhancers, methylation of histone H3 lysine 4 (H3K4me1). Towards the bottom of the figure, transcriptionally inactive heterochromatin is shown, with compacted nucleosomes, the presence of histone H3 lysine 9 bi- and trimethylation (H3K9me2 and H2K9me3), which are signature repressive marks of constitutive heterochromatin and extensive DNA methylation. Finally, heterochromatin is folded into chromosomes within the nucleus, as shown. Figure taken from Baylin and Jones, 2011.

1.1.8.2 Regulatory processes at the transcriptional start site

A tetrameric complex that surrounds a nucleosome has been identified as the initiating *complex* for DNA methylation (Ooi et al., 2007). The complex consists of two molecules of the *de novo* methyltransferase DNMT3A and two of DNMT3L, an inactive isomer expressed in embryonic stem cells. Nucleosomes that contain the histone mark H3K4me3 cannot be bound by this complex, targeting DNA methylation to regions such as inactive promoters, intragenic regions and regions not subject to transcriptional control. The enzyme family responsible for methylating cytosine residues are the DNA methyltransferases (DNMTs), with DNMT1 maintaining the *de novo* methylation patterns determined by DNMT3A and DNMT3B. In differentiated somatic cells, DNMT1 shows a strong preference for recognising hemi-methylated DNA intermediate produced during replication of a methylated sequence, perpetuating pre-existing methylation patterns between parent and daughter cells. The interplay between DNMT3A / B and DNMT1 is being defined in normal and cancer cells. DNMT3A and DNMT3B are thought to repair errors made by DNMT1 after DNA synthesis (Jeong et al., 2009). The recent description of enzymes that remove DNA methylation, such as the TET family of proteins, which form 5-hydroxymethylcytosine from methylated cytosine, have been demonstrated to be important during development and tumourigenesis (Williams et al., 2011; Kriaucionis and Heintz, 2009). The contribution and position of the determinants of transcription in a transcriptionally active and repressed gene are shown in Figure 1.9.

In an actively transcribed promoter, widely spaced nucleosomes contain acetylated histone H3 and H4, with additional H3K4me3. Polycomb protein is absent, and a transcription activator complex, consisting of a co-activator protein (CA), transcription factor (TF) and histone acetyl transferase (HAT) can access the TSS. DNMTs are excluded from the promoter, preventing *de novo* methylation of CpG dinucleotides. In its inactive form, the promoter DNA is characterised by methylation of CpG dinucleotides, with the promoter region now accessible to DNMTs. The promoter itself is occupied by transcriptionally repressive proteins, including methyl-cytosine binding proteins (MBPs) and nucleosomes in which the histones have become deacetylated in a process mediated by histone deacetylases (HDACs). The nucleosomes have additionally acquired methylated lysine 9 of histone H3, mediated by a lysine 4

methyltransferase. Finally, the transcription activator complex is no longer able to access the TSS (Figure 1.9).

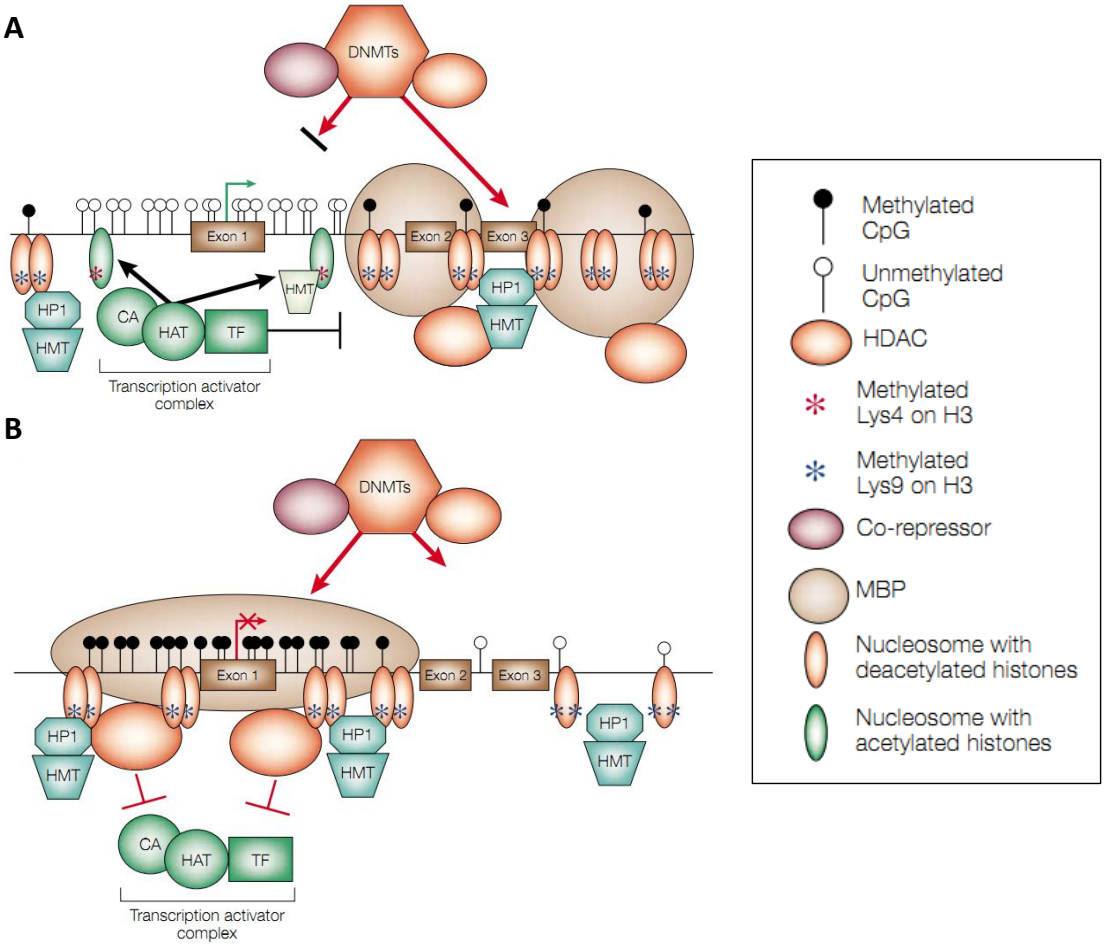


Figure 1.9. A CpG rich promoter in transcriptionally activated and repressed configurations. **A.** The activated promoter is surrounded by widely-spaced, relaxed nucleosomes, which contain H3K4me3, whose methylation is mediated by Lysine 4 methyltransferase, HMT, and acetylation marks, mediated by histone acetyl transferase (HAT). The open chromatin state allows entry of the transcription activator complex, consisting of a co-activator protein (CA), transcription factor (TF) and HAT. DNA methyltransferases (DNMTs) are excluded from the promoter, preventing *de novo* methylation of CpG dinucleotides. **B.** The transcriptionally repressed promoter is characterised by dense promoter CpG methylation. Methyl-cytosine binding proteins (MBPs) bind to the DNA and recruit histone deacetylase (HDAC) complexes, which, in turn, deacetylate the nucleosomes and, in conjunction with trimethylation of lysine 9 of histone H3 (H3K9me3), generate a tightly compacted chromatin structure that prevents access of the transcriptional activator complex. The promoter is also accessible to DNMTs, enabling the maintenance of the repressive CpG methylation and the associated transcriptional silencing. HP1 – chromodomain protein HP1α; HMT – Lysine 4 methyltransferase. Figure adapted from (Jones and Baylin, 2002).

Repressive chromatin	Active chromatin
Nucleosomes across promoter region	Nucleosome depletion at promoter
Histone H3 lysine 9 dimethylation	Histone H3 lysine 4 trimethylation
Histone H3 lysine 9 trimethylation	Histone H3 lysine 36 trimethylation
Deacetylation of histones H3 and H4	Acetylation of histones H3 and H4
Methylated DNA across promoter region	Unmethylated DNA across promoter region
For certain gene promoters: Repression not dependent upon DNA methylation across promoter; dependent upon the binding of PcGs which repress transcription	In enhancer region: Nucleosome depletion Histone H3 lysine 4 methylation Histone H3 lysine 27 acetylation

Table 1.3. Catalogue of repressive and active chromatin marks for the control of gene transcription in humans. PcG – polycomb group proteins.

1.1.9 DNA methylation in cancer

The most widely studied epigenetic change is methylation of the 5' carbon of cytosine at CpG dinucleotides. The dinucleotide CpG is rare within the genome relative to other dinucleotide pairs (Bird, 1980). This is because the methylated cytosine is able to spontaneously deaminate to thymine. It is thought that only those methylated CpGs which have an important role in gene regulation have persisted in the genome through selective pressure, the rest being lost over evolutionary time without detriment to host fecundity. DNA loci with a high GC percentage proximal to gene promoters are known as CpG islands (Gardiner-Garden and Frommer, 1987). Computational identification of CpG islands from the initial draft of the human genome estimated that there were 29,000 CpG islands (Lander et al., 2001). One current definition of a CpG island is that of a sequence with length > 500 bases, GC content > 50% and CpG dinucleotides at an observed : expected ratio > 0.6 (Wang and Leung, 2004), but there is no strict definition. Although CpG islands are generally located in the 5' upstream promoter region of a gene, there are atypical CpG islands located within the gene or 3' to the gene (Nguyen et al., 2001). Approximately 60% of human genes are associated with a CpG island. A recent study demonstrated hypermethylation (inappropriately high levels of methylation) across a 1MB region of colorectal cancer cells, suggesting that this large scale silencing may be a mechanism for LOH across large chromosomal regions (Frigola et al., 2006). It has been suggested that whilst methyl groups have no effect on base pairing, they may affect DNA–protein interactions and inhibit transcription by interfering with its initiation; DNA methylation can reduce the binding affinity of transcription factors and recruit methylation-dependent DNA binding-proteins, which act as transcriptional repressors (Jones and Laird, 1999).

Generally, promoter-associated CpG islands of normal adult somatic cells are thought to be predominantly unmethylated, with their associated genes being transcriptionally active. This is in contrast to most CpG sites outside of CpG islands, which are predominantly methylated. This methylation may be necessary to silence intact retrotransposons (principally, LINE-1 (long interspersed nuclear element-1) and *Alu* elements), preventing genome instability (Walsh et al., 1998). It has been reported that the global hypomethylation (inappropriately low levels of methylation) observed in cancer cells (Goelz et al., 1985) may provide a mechanism for an increase in genome

instability by activating previously silenced retrotransposons, by loss of imprinting and / or chromosomal instability.

Studies in cancer have shown that against a global cancer hypomethylation background, hypermethylation is detectable in specific DNA regions (de Bustros et al., 1988). In 1994, the first demonstration that DNA hypermethylation at a promoter can lead to silencing of a TSG was described in the Von Hippel Landau (VHL) gene (Herman et al., 1994). Subsequent investigations have revealed that CpG island hypermethylation of TSGs varies according to tumour type (Costello et al., 2000). Canonical cancer genes such as the DNA repair genes *MLH1, BRCA1* and *MGMT* have all been shown to undergo silencing by DNA hypermethylation in cancer cells (Esteller, 2006). Current estimates are that 5-10% of genes whose promoter regions are unmethylated after completion of embryonal development acquire methylation in cancer (Bird, 2002). More recently, comprehensive analyses of genome-wide DNA methylation patterns have revealed the existence of CpG island 'shores' (Irizarry et al., 2009), regions of DNA that lie upstream but close to CpG islands. These regions showed more variable methylation patterns than within CpG islands, with a strong relationship to gene expression, and are tightly linked to the tumour tissue of origin.

Epigenetic gene silencing in cancer has also been shown to be important in predisposing to mutational events that can drive tumourigenesis. The mismatch repair gene *MLH1* is frequently methylated in sporadic tumours and is associated with an observed microsatellite instability (Jones and Baylin, 2002). Conversely, mutational events that are associated with epigenomic defects are also being identified. For example, mutations in *IDH1* (Parsons et al., 2008) occurring in gliomas correlate with a CpG island hypermethylator phenotype (Noushmehr et al., 2010). In addition, cytosine methylation can influence tumorigenicity, since 5-methylcytosine can deaminate to thymine; up to 50% of inactivating point mutations of the TSG *TP53* in somatic cells occur at methylated cytosines (Rideout et al., 1990). The role of DNA hypermethylation in enabling pathway disruption through the transcriptional repression of multiple microRNAs is becoming more widely recognised. For example, a hypermethylation signature of microRNA loci, associated with metastasis, has been reported (Lujambio et al., 2008). A selection of genes commonly transcriptionally silenced by promoter hypermethylation in cancer is shown in Table 1.4.

Gene Function	Gene Name	Cancer Type	References
Cell-cycle regulation	*RB1*	Retinoblastoma	(Greger et al., 1994)
	P16^{INK4A}	Colon, lung, many others	(Herman et al., 1995)
	P14^{ARF}	Colon	(Xing et al., 1999)
	RASSF1A	Lung, liver, brain	(Dammann et al., 2000)
Tumour-cell invasion	*CDH1*	Breast, gastric, thyroid, leukaemia, liver	(Wheeler et al., 2001)
	CDH13	Lung, ovarian, pancreas	(Sato et al., 1998)
	TIMP3	Brain, kidney	(Esteller et al., 2001)
	VHL	Renal cell	(Herman et al., 1994)
DNA repair / detoxification	*MLH1*	Colon, endometrial, gastric	(Herman et al., 1998)
	MGMT	Brain, colon, lung, breast	(Esteller et al., 1999)
	GSTP1	Prostate, liver, colon, breast, kidney	(Lee et al., 1994)
	BRCA1	Breast, ovarian	(Dobrovic and Simpfendorfer, 1997)
Chromatin remodelling	*SMARCA3*	Colon, gastric	(Moinova et al., 2002)
Cell signalling	*SOCS1*	Liver, colon, multiple myeloma	(Nagai et al., 2002)
Transcription	*ESR1*	Colon, breast, lung, leukaemia	(Issa et al., 1994)
Apoptosis	*DAPK1*	Lymphoma	(Katzenellenbogen et al., 1999)

Table 1.4. Examples of genes silenced by aberrant DNA hypermethylation in cancer. *RB1*, retinoblastoma 1; *p16^{INK4}*, cyclin dependent kinase inhibitor 2A INK4A; *p14^{ARF}*, cyclin dependent kinase inhibitor 2A ARF; *RASSF1A*, Ras association domain family protein 1 isoform A; *CDH1*, cadherin 1; *CDH13*, cadherin 13; *TIMP3*, tissue-inhibitor of metalloproteinase 3; *VHL*, Von Hippel Lindau; *MLH1*, mutL, E. Coli, homolog of, 1; *MGMT*, methylguanine-DNA methyltransferase; *BRCA1*, breast cancer 1; *GSTP1*, glutathione S-transferase 1; *SMARCA3*, swi / snf-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 3; *SOCS1*, suppressor of cytokine signalling 1; *ESR1*, oestrogen receptor 1; *DAPK1*, death-associated protein kinase. Table adapted from (Robertson, 2005).

1.1.9.1 DNA methylation patterns provide a rich source of biomarkers

DNA methylation patterns hold rich promise as biomarkers for detection of cancer, assessment of prognosis, and predicted response to therapy. The assessment of *TFPI2* methylation in stool and blood samples has been reported as a novel biomarker for the detection of colorectal cancer (Glockner et al., 2009). Similarly, a methylation signature of active ovarian cancer (Teschendorff et al., 2009) may have potential as a non-invasive strategy for early detection in at-risk groups. Finally, using a methylation-sensitive restriction enzyme, comparative hybridisation based approach, a 112 locus methylation signature has been developed that is able to predict progression free survival in ovarian cancer with 95% accuracy (Wei et al., 2006).

An assessment of the DNA methylome of 1054 cancers on the GoldenGate methylation array platform (section 2.7) demonstrated that the tissue type origin of cancers could be determined from its DNA methylation profile (Fernandez et al., 2011), indicating that DNA methylation patterns could be useful for determining the cell of origin of cancer, which might have implications for treatment strategies. More specifically, hypermethylation marks that silence the *MGMT* gene in glioma can predict response to the chemotherapeutic drug, temozolomide (Esteller et al., 2000).

There are an increasing number of cancer drugs which function to reverse the epigenetic abnormalities of cancer. DNA demethylating agents such as azacytidine and decitabine inhibit all DNA methyltransferases and have shown promise in increasing survival in patients with myelodysplastic syndrome (MDS) and increasing the time period between diagnosis of MDS and full-blown acute myeloid leukaemia (Silverman and Mufti, 2005). The precise mode of action for this class of drugs remains poorly understood. It is conceivable that treatment with demethylating agents could reactivate previously silenced TSGs and slow tumour progression. There is also interest in the potential potentiating effects of epigenetic treatments, when allied with conventional chemotherapeutic treatments. The application of chemotherapy to a tumour creates a strong selective pressure for changes in gene expression that reduce the effectiveness of the drug. These changes in gene expression may be mediated by the hypermethylation and silencing of specific genes which bind to the drug. If the

demethylating agent is applied alongside the chemotherapy, development of resistance to chemotherapy may be retarded (Teodoridis et al., 2004).

Concerns remain over the non-specific nature of these drugs, which have the potential to reactivate normally silenced genes in addition to hypermethylated tumour specific genes, and it has been demonstrated that treatment of cancer cell lines with demethylating agents led to the methylation-dependent dysregulation of many hundreds of genes, including classes of genes that promote and inhibit cell growth (Suzuki et al., 2002).

1.2 Paediatric tumours of the central nervous system

Cancers of the CNS are the second most common group of cancers in childhood (Gurney, 1999), responsible for 40% of cancer deaths in children in the UK (Soffietti et al., 2002). Between 1998 and 2002, CNS tumour incidence was 4.40 cases per 100,000 person years for children aged 0 - 19 years in the United States (CBTRUS, 2011). Although survival rates have improved, 40% of children diagnosed with a CNS tumour will die from their disease; moreover, survivors suffer considerable morbidity, usually in the form of neuro-endocrine and neuro-cognitive defects, from the effects of their treatments (Pizer and Clifford, 2009).

The World Health Organisation (WHO) classification of CNS tumours has standardised their classification from a consensus view of neuro-pathologists and geneticists (Louis et al., 2007), enabling worldwide epidemiological studies and clinical trials to be implemented from common diagnostic and classification criteria. The identification and characterisation of brain tumours now depends on genetic and immunohistochemical markers in addition to traditional histopathological features.

1.2.1 Embryonal tumours of the CNS

The majority of childhood CNS tumours arise within the brain. Gliomas, which can arise in both the brain and spine, although the brain is by far the most common origin, comprise 50% of childhood brain tumours. Primitive neuro-ectodermal tumours (PNET) of the CNS comprise ~20% of childhood brain tumours and are the most common malignant paediatric brain tumours, as shown in Figure 1.10. Initially classified as a single entity, the 2000 WHO classification (Kleihues et al., 2002) recognised five distinct embryonal tumours (medulloblastoma, supratentorial PNET, medulloepithelioma, ependymyoblastoma and atypical / rhabdoid tumours, characterised by undifferentiated neuro-epithelial cells (Sarkar et al., 2005)). This was based on the recognition that these five variants have distinct morphological, immunohistochemical, molecular, clinical and biological profiles. Medulloblastoma is by far the most common variant, comprising 95% of PNETs (Sarkar et al., 2005). The 2007 re-evaluation of brain tumour classification (Louis et al., 2007) separated medulloblastoma from other PNETs and recognised five distinct medulloblastoma variants, discussed in section 1.3.1.

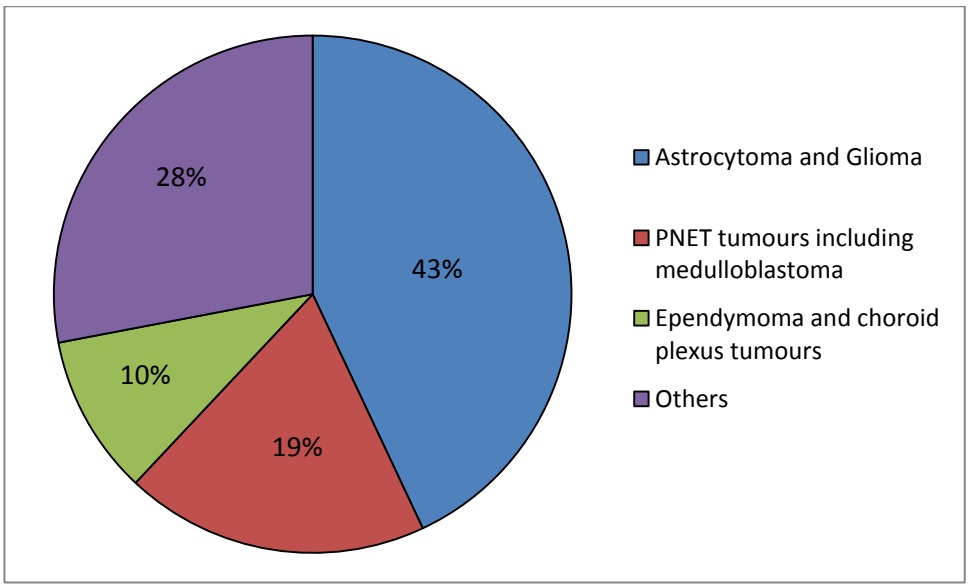


Figure 1.10. Occurrence of childhood brain tumours. Data retrieved from Cancer Stats (CRUK, 2011).

1.3 Medulloblastoma

In 1925, Bailey and Cushing identified a novel tumour, localised to the cerebellum, and distinct from gliomas (Bailey and Cushing, 1925). The authors postulated that this undifferentiated tumour, which they named medulloblastoma, derived from embryonic cells in the ependymal lining of the fourth ventricle of the cerebellum (Figure 1.11) and subsequently pioneered treatment of medulloblastomas using radiotherapy, reviewed by (Ferguson and Lesniak, 2005).

In the intervening years since its initial description, the tumour has been well described and both the means of diagnosis and therapeutic strategies have greatly improved. It arises, by definition, in the posterior fossa (a region of the brain that contains the brainstem and cerebellum), usually from the roof of the fourth ventricle, and presents as a midline tumour. More commonly in older children, the tumours can also arise in a cerebellar hemisphere.

Medulloblastoma is the most common malignant brain tumour of childhood, accounting for 20% of all paediatric brain malignancies, with an incidence of 0.5 / 100,000 children / year (Crawford et al., 2007), corresponding to approximately 90 cases a year in the UK (Pizer and Clifford, 2009). Approximately 70% of cases occur in childhood (3 to 15 years of age), with 10-15% infant cases (< 3 years of age). There is a bi-modal distribution in age of incidence, with peaks at 3-4 years and 8-9 years of age (Crawford et al., 2007; Ellison, 2006; Ellison, 2002). There is a gender imbalance observed with the incidence of medulloblastoma in males 1.5-2 times higher than in females (Crawford et al., 2007).The origin of medulloblastoma is discussed in later sections (1.3.13).

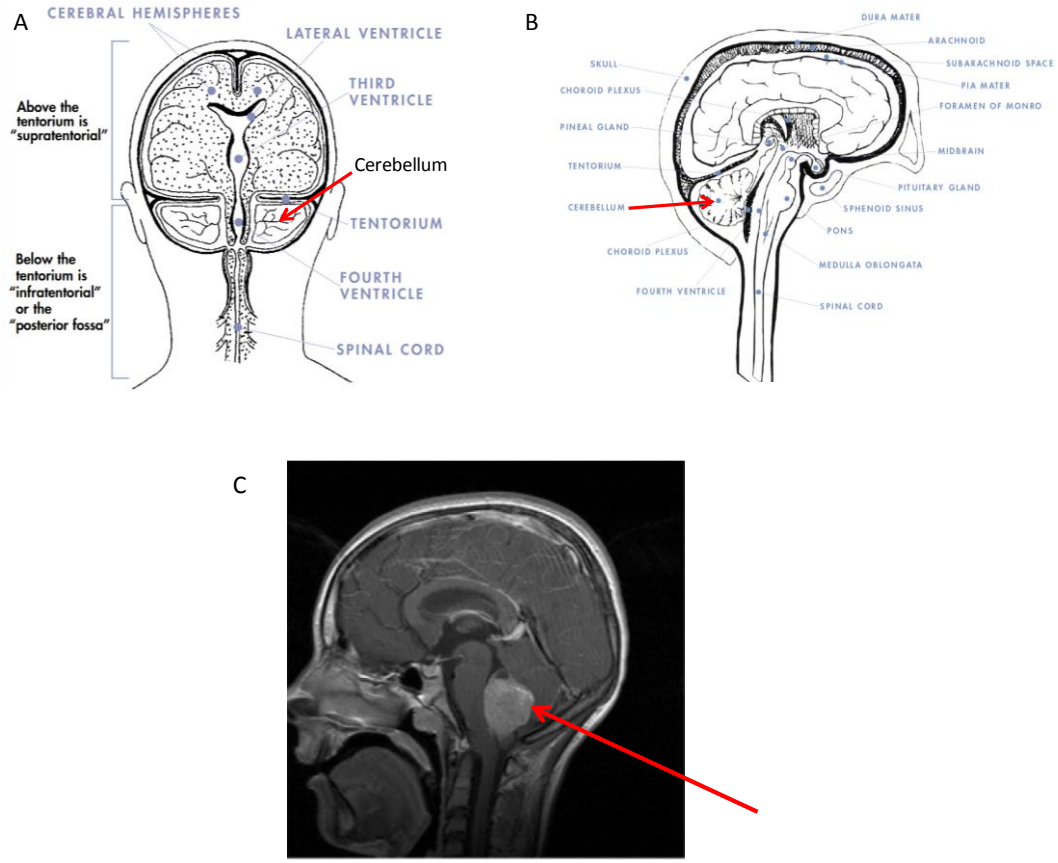


Figure 1.11. Anatomy of brain and location of medulloblastoma. **A.** Dorsal-ventral schematic view of the brain, showing location of cerebellum and fourth ventricle. The location of brain regions is denoted by arrows (cerebellum is shown with a red arrow). **B.** Sagittal section showing regions of the brain, indicated by arrows (cerebellum is shown with a red arrow). Schematic brain sections taken from a primer of brain tumours (ABTA, 2010). **C.** T1- weighted sagittal section magnetic resonance imaging (MRI) scan of child with medulloblastoma. Medulloblastoma is visible as grey staining mass indicated by red arrow. Image kindly provided by Dr. Simon Bailey (Paediatric Brain Tumour Research Group, Newcastle University, UK).

1.3.1 Histology of medulloblastoma

Medulloblastoma is currently classified using WHO criteria (Louis et al., 2007) as a grade IV tumour because of its invasive nature, metastatic capability and primitive cellular appearance. Under WHO criteria, 5 distinct histological disease variants are recognised: classic medulloblastoma, desmoplastic / nodular (DN) medulloblastoma, medulloblastoma with extensive nodularity (MBEN), large cell medulloblastoma and anaplastic medulloblastoma. Typically, the histopathological subtype of the tumour is determined by an experienced neuro-pathologist after surgical excision of the tumour.

1.3.1.1 Classic medulloblastoma

The classic medulloblastoma (Figure 1.12A) is the most common of the histopathological medulloblastoma subtypes, comprising 70-80% of medulloblastomas (Pizer and Clifford, 2009; Gilbertson and Ellison, 2008). This tumour variant is typically composed of small round cells with a high nuclear to cytoplasm ratio that occur in sheet-like areas of densely packed undifferentiated cells, with hyperchromatic nuclei visible after H & E (haematoxylin and eosin) staining (Ellison, 2002).

1.3.1.2 Desmoplastic / nodular medulloblastoma

The desmoplastic / nodular (DN) variant of medulloblastoma is observed in 10-15% of cases (Pizer and Clifford, 2009; Gilbertson and Ellison, 2008; McManamy et al., 2007; Ellison, 2002). It is most often observed in young children (Pizer and Clifford, 2009), representing > 50% of infant cases (<3 years old at diagnosis) and just 5% of cases aged 3 – 15 years (McManamy et al., 2007) and also appears to have a higher prevalence in adult cases (Remke et al., 2011b; Ellison, 2002). The DN subtype is characterised by scattered round or ovoid nodules separated by reticulin-rich desmoplastic inter-nodular regions (Figure 1.12B). Desmoplasia in infant cases has previously been associated with an improved prognosis (Rutkowski et al., 2005), although the prognostic value of desmoplasia in non-infants is unclear. Tumour cells located within inter-nodular / desmoplastic regions are more densely packed and pleomorphic than nodular cells (Ellison, 2002). The extent of nodularity can be variable and an increased nodular density has been associated with an improved prognosis (Verma et al., 2008; McManamy et al., 2007).

1.3.1.3 Medulloblastoma with extensive nodularity

The 2007 classification of CNS tumours identified a new histological subgroup of medulloblastoma, the medulloblastoma with extensive nodularity (MBEN), which is closely related to the DN disease variant (section 1.3.1.2), differing in a markedly expanded lobular architecture and a correspondingly reduced inter-nodular component (Figure 1.12C). It comprises 1-2% of medulloblastomas (McManamy et al., 2007). The nodules contain round cells with uniform nuclei with a high level of neuronal differentiation and a low proliferative index (Louis et al., 2007). This variant

most often presents in infant patients (< 3 years old at diagnosis) and is associated with a favourable prognosis (McManamy et al., 2007; Ellison, 2002).

1.3.1.4 Large cell medulloblastoma

The large cell variant of medulloblastoma comprises ~ 4% of medulloblastomas, and is characterised by large cells with a large, pleomorphic nucleus, prominent nucleoli and an abundant cytoplasm (Figure 1.12D). This subtype has a higher mitotic and apoptotic rate than other medulloblastoma variants, and, as a consequence, regions of necrosis are often apparent (Gilbertson and Ellison, 2008; Ellison, 2002). The large cell variant has been associated with a poor prognosis (Eberhart et al., 2002; Brown et al., 2000).

1.3.1.5 Anaplastic medulloblastoma

Anaplastic medulloblastoma comprises 10-20% of medulloblastomas (Gilbertson and Ellison, 2008). The anaplastic variant is characterised by cells with nuclear pleomorphism and moulding of cells, where cells wrap around each other (Figure 1.12E) (Ellison, 2002). Although areas of anaplasia can be present in all medulloblastoma histopathological subtypes (sections 1.3.1.1, 1.3.1.2, 1.3.1.3 and 1.3.1.4), it is particularly pronounced in this anaplastic subtype. The anaplastic variant has previously been associated with a poor prognosis (Giangaspero et al., 2006; Eberhart and Burger, 2003). Because the large cell and anaplastic variants form a continuum with a shared poor prognosis (Gilbertson and Ellison, 2008; Brown et al., 2000), they are typically grouped as a single large cell / anaplastic (LCA) category in studies of medulloblastoma.

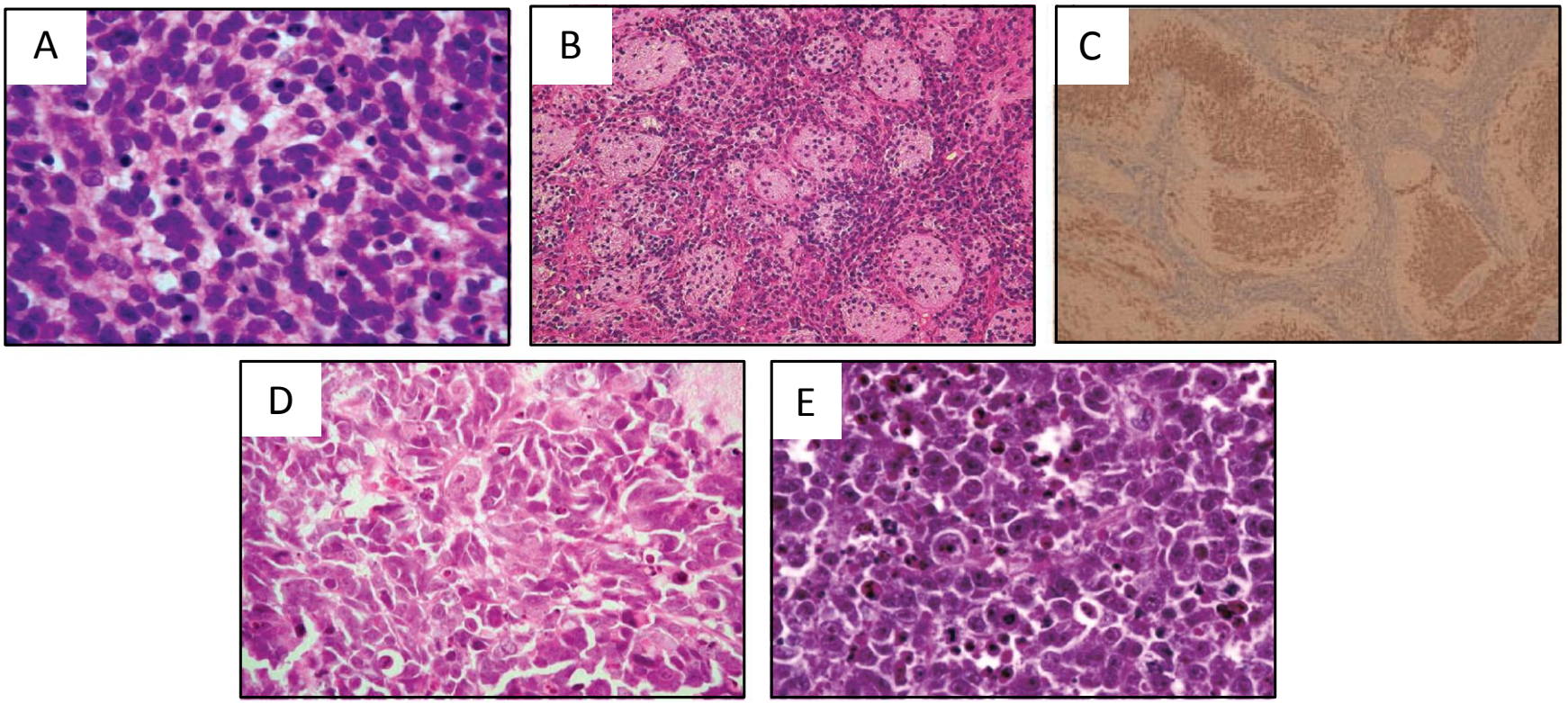


Figure 1.12. Histopathological subtypes of medulloblastoma. Unless stated otherwise, section shown is stained with haematoxylin and eosin. **A.** Classic medulloblastoma. **B.** Desmoplastic / nodular (DN) medulloblastoma. **C.** Medulloblastoma with extensive nodularity (MBEN), stained with Neu-N antibody (specific for neural cells). **D.** Anaplastic medulloblastoma. **E.** Large cell medulloblastoma. Sections **A,B,D,E** taken from (Gilbertson and Ellison, 2008). Section **C** taken from (Ellison, 2002).

1.3.2 Clinical presentation and diagnosis of medulloblastoma

Most medulloblastoma patients present with a raised intracranial pressure due to hydrocephalus, caused by blockage of the cerebrospinal fluid (CSF) pathways at the fourth ventricle. This is manifested clinically by headaches, vomiting, lethargy and drowsiness, which is observed in ~80% of patients at diagnosis. Disruption of the cerebellum may be indicated by ataxia, and pressure on the brain stem can result in cranial nerve palsies and long tract signs (Pizer and Clifford, 2009). In older children, behavioural changes and a declining academic performance may also be present (Wilne et al., 2007). Infants may present with non-specific symptoms such as irritability and vomiting, often leading to delays in the diagnosis of medulloblastoma, although raised intra-cranial pressure in infants can be indicated by an enlarged head.

Patients suspected to have a medulloblastoma are diagnosed using standard neuro-imaging techniques. Computed tomography (CAT) and magnetic resonance imaging (MRI) can both be used in diagnosis, and if a medulloblastoma is indicated, it is usual to scan using MRI, if it hasn't already been applied, due to the higher resolution that this technique affords (Packer et al., 1999). On MRI scans, the tumour is visible as a heterogeneous hypo-intense mass (Halperin et al., 2001), as shown in Figure 1.11C. Since metastasis to other areas of the CNS via CSF is a well validated marker of poor risk, which influences treatment decisions and occurs in 35% of cases (Pizer and Clifford, 2009), typically an MRI scan of the entire cerebrospinal axis is performed to assess disease spread (Packer et al., 1999) (section 1.3.3).

1.3.3 Tumour staging

Tumour staging is an important prognostic factor and its assessment is fundamental to selecting appropriate treatment strategies (section 1.3.5). The extent of dissemination of the tumour is determined pre-operatively by MRI scans of the entire cerebro-spinal axis, in addition to cytological examination of CSF following tumour excision. The determination of metastatic state is assigned with reference to Chang's criteria (Chang et al., 1969), shown in Table 1.5.

Stage	Criteria
M0	No evidence of gross subarachnoid metastasis and no tumour cells observed in CSF (50%)
M1	Microscopic tumour cells observed in CSF (20%)
M2	Gross nodular seeding in cerebellum, cerebral subarachnoid space, or in third or fourth ventricles (5-10%)
M3	Gross nodular seeding in spinal subarachnoid space (20%)
M4	Extra-neural metastasis, uncommon (1%)

Table 1.5. Description of M stage classification system for medulloblastomas. M stage (Chang et al., 1969) is assigned based on pre-operative MRI scans and post-excision cytological examination of the CSF. Approximate percentage incidences of each M stage is shown in parentheses. Table adapted from (Zeltzer et al., 1999).

1.3.4 Prognostic factors in medulloblastoma

A number of well-validated factors have been shown to be important determinants of patient prognosis in medulloblastoma. Current risk stratification schemes consider the age of the patient (infant patients under 3 years of age at diagnosis), metastatic spread and residual disease following surgical excision to confer membership of a poor risk patient group, with the remaining patients being classified as standard risk. Current, well supported prognostic markers, with accompanying evidence are summarised in Table 1.6.

The poorer prognosis of infant cases may partly be a consequence of the avoidance of radiotherapy in these cases, because of the unacceptable neuro-cognitive *sequelae* associated with cranial radiotherapy to a small, developing brain. Therapies in infant cases are often designed to avoid or delay the use of cranio-spinal radiotherapy. It has also been reported that infant medulloblastomas can be genetically and histopathologically distinct, which may relate to their clinical behaviour; the relative contribution of biology and therapy to outcome in infants remains unclear (Crawford et al., 2007; Packer et al., 1999).

Metastatic status (section 1.3.3) is also fundamental to treatment selection. Patients with metastatic disease at presentation to the CSF (M1), brain (M2) or spine (M3) are assigned to poor risk treatment protocols (Zeltzer et al., 1999). Residual disease is identified by post-operative MRI scans, performed within 72 hours of surgery, after which time post-operative changes impede the identification of residual disease. Subtotal resection, defined by any residual disease with a cross-sectional area of greater than or equal to 1.5 cm^2, has been shown to be a significant marker of poor prognosis (Pizer and Clifford, 2009).

In recent years, novel molecular prognostic factors have been identified (section 1.3.16), some of which are now validated in ≥2 clinical trials cohorts (Table 1.6) and are being incorporated into state-of-the-art classification schemes used in upcoming clinical trials. In particular, the good prognosis conferred by WNT (Wingless) pathway activation in medulloblastoma and the poor prognosis associated with a large cell / anaplastic histological subtype are being incorporated into the risk stratification

schemes for the PNET 5 / 6 clinical trials which are due to commence in 2012 (Pizer and Clifford, 2009).

Category	Prognostic factor	Reference
Favourable risk		
Histology	Desmoplasia in patients < 3 years of age	(McManamy et al., 2007; Rutkowski et al., 2005)
Biology	WNT pathway activation (β-catenin nuclear accumulation)	(Clifford et al., 2006; Ellison et al., 2005)
Adverse risk		
Clinical / radiological	Age < 3 years of age	(Rutkowski et al., 2005; Zeltzer et al., 1999)
	Metastatic disease	(Zeltzer et al., 1999; Bailey et al., 1995)
	Post-surgical residual disease ≥ 1.5 cm^2	(Pomeroy et al., 2002; Zeltzer et al., 1999)
Histology	Large cell / anaplastic medulloblastoma	(Eberhart and Burger, 2003; McManamy et al., 2003; Ellison, 2002)
Biology	*MYC* family gene amplification	(Lamont et al., 2004; Eberhart and Burger, 2003)

Table 1.6. Currently accepted prognostic factors in medulloblastoma, validated in ≥2 clinical trials cohorts. Markers of favourable and adverse risk are listed, with supporting references included. Table adapted from (Pizer and Clifford, 2009).

1.3.5 Treatment of medulloblastoma

After diagnosis, the fundamental treatment modality is surgical excision, with the aim of achieving maximal removal, while sparing as much normal brain as possible, followed, if appropriate, by post-operative cranio-spinal radiotherapy and / or chemotherapy. Due to the devastating side effects, cranio-spinal radiotherapy is avoided if possible in infant cases under the age of 3, who are treated with specialised protocols (section 1.3.5.1.3). The treatment decisions are currently based on the prognostic factors age at diagnosis, metastatic stage and residual disease, as outlined in section 1.3.4.

1.3.5.1 Treatment protocols

1.3.5.1.1 Current treatments for standard risk patients

Standard risk patients > 3 years old at diagnosis are currently treated with protocols designed to reduce as far as possible neuro-cognitive and neuro-endocrine *sequelae*, whilst maintaining a low rate of relapse. The first line of treatment following excision is typically cranio-spinal radiation of 24 Gray (Gy), coupled with a posterior fossa boost to 56 Gy, allied to eight doses of concomitant weekly vincristine, followed by up to eight cycles of adjuvant chemotherapy with lomustine, cisplatin and vincristine (Packer et al., 2006). This achieves 5 year progression free survival (PFS) rates of 79%-81% (Packer et al., 2006; Packer et al., 1999). A further reduction of cranio-spinal radiation to 18 Gy is being investigated in younger children (Pizer and Clifford, 2009).

1.3.5.1.2 Current treatments for high risk patients

High-risk non-infant patients are currently defined by positivity for metastatic disease (M1 / M2 / M3, observed in approximately 30% of patients) and / or residual disease. Second-look surgery, in which residual disease is removed in a second operation, is becoming more popular and enables the re-assignment of patients to a standard risk protocol (Pizer and Clifford, 2009). Since 5 year PFS for high risk patients is approximately 60% (Pizer and Clifford, 2009; Gilbertson, 2004), the primary focus for this group of patients is to achieve a cure, rather than avoidance of neuro-cognitive and neuro-endocrine *sequelae*.

Current standard treatments for high risk patients involve an initial dose of 35 – 36 Gy cranio-spinal radiotherapy, with a boost of 18 – 20 Gy to the posterior fossa. Further boosts to sites of metastatic spread are usually administered, if appropriate. Chemotherapy is generally administered, although the precise time of application is variable (Crawford et al., 2007).

1.3.5.1.3 Current treatments for patients under 3 years of age

Treatment of medulloblastoma in very young children represents a very difficult balance between the necessity to achieve cure and not breaching an acceptable level of neuro-cognitive *sequelae*. For this reason, chemotherapeutic treatment following surgical excision is the most widely accepted treatment, with the aim to delay, or prevent altogether, the need to administer radiotherapy. As an alternative to irradiation of the entire cranio-spinal axis, focal irradiation of the tumour is becoming increasingly common (Rutkowski et al., 2010). However, treatment with the same chemotherapeutic regimes used as an adjuvant to radiotherapy in older children leads to a poor outcome, with 5 year PFS of approximately 30% (Crawford et al., 2007). A recent study identified large survival differences by histological subtype in children aged under 5 at diagnosis (Rutkowski et al., 2010). 8 year event free survival (EFS) rates were 55% for children with DN or MBEN histologies, 27% for children with classic histologies and 14% for children with LCA histologies. Novel treatment strategies have been explored. In particular, a German study found that an intensive chemotherapeutic regime in infants with complete tumour resection and no solid metastases led to a 5 year PFS of 82%, although PFS for patients with solid metastases or residual disease was worse (33% and 50% 5 year PFS, respectively). Notably, this study demonstrated that infants with desmoplastic tumours had a 5 year PFS of 85%, demonstrating the prognostic advantage of desmoplasia in the under 3s (Rutkowski et al., 2005).

1.3.5.1.4 Current treatments for relapsing patients

While patients who have not previously been irradiated can be successfully treated after relapse with radiotherapy (Ridola et al., 2007), for the majority of patients whose primary treatment involved radiotherapy, the prognosis is dismal, with survival rates of just 2% (Pizer and Clifford, 2009). Over 75% of relapses in children occur within two

years of diagnosis (Packer et al., 1999), with 80% of relapses at the leptomeninges, alone or in combination with relapse at the primary site (Pizer and Clifford, 2009). High dose chemotherapy may be efficacious in patients with an isolated relapse, although prognosis is dismal, with most treatments being palliative in this group of patients.

1.3.5.2 Treatment sequelae

The improved survival observed for medulloblastoma patients has come at the cost of considerable side effects that are a consequence of treatment. Survivors often have to contend with long term intellectual impairment, due to damage during development from the tumour and from radiotherapy treatments. The extent of intellectual impairment is inversely related to the age of the patient (Fossati et al., 2009). In addition, neuro-endocrine and spinal damage can affect growth, bone development and can lead to early puberty. Chemotherapeutic regimes can cause ototoxicity, infertility and lead to secondary cancers as a result of treatment. In general, the neuropsychological outcome for patients is poor, with survivors exhibiting lifelong social and educational difficulties (Pizer and Clifford, 2009).

1.3.6 Challenges for the improvement of patient outcome

The multi-modal approaches outlined above, allied to advances in surgical technique and post-operative care, have combined to achieve an improvement in 5 year survival rates from 3% to 70% over the last 50 years (Crawford et al., 2007; Newton, 2001). Whilst this progress is encouraging, fundamental challenges still remain; there is a clear need to continue to explore treatment options in the 30% of patients who still experience relapse, whilst for those patients who achieve cure, there is a need to reduce as much as possible the intensity of treatment, so that neuro-cognitive defects are minimised.

A troika of approaches are being employed to achieve these aims: improved treatments, the use of novel biological agents that specifically target the inappropriately activated signalling pathways that drive tumourigenesis (see sections 1.3.9, 1.3.10), and improved risk stratification.

Surgical techniques that minimise damage to healthy surrounding brain tissue continue to be refined, while clinical trials focus on optimising combinations of

chemotherapy and radiotherapy (Pizer and Clifford, 2009; Crawford et al., 2007). It is possible that proton therapy, which allows very tight constraint over the areas of tissue targeted compared to conventional radiotherapy, might further spare healthy brain tissue, and is being investigated as a replacement for cranio-spinal radiotherapy (Fossati et al., 2009).

Improvements in understanding of the molecular determinants of medulloblastoma have demonstrated subgroups of medulloblastoma characterised by activation of the WNT and Sonic hedgehog (SHH) signalling pathways (discussed in detail in sections 1.3.9 and 1.3.10). Drugs which specifically target these pathways may enable targeted control of the tumour without the devastating side effects associated with current treatments for medulloblastoma.

The characterisation of molecular predictors of disease outcome has been accomplished in tandem with an improved understanding of disease biology. These novel molecular predictors, summarised in Table 1.6, are being evaluated for their inclusion into refined risk stratification schemes. However, the proposed diagnostic molecular markers (WNT status and amplification of the *MYC* oncogenes) identify only approximately 20% of patients, and there is a pressing need for additional well-validated molecular markers of prognosis, discussed further in section 1.3.16. A more accurate appraisal of patient risk at diagnosis, that incorporates additional molecular disease determinants, represents a simple and cost effective way of ensuring that patients are treated with appropriately tailored treatment strategies, reducing the severity of neuro-cognitive *sequelae* in good prognosis patients, whilst continuing to aggressively treat patients with poor prognosis, where achieving cure is the sole treatment objective.

1.3.7 Genetics of medulloblastoma

The study of three areas has greatly aided the elucidation of the molecular genetics of medulloblastoma – the study of familial cancer syndromes whose phenotype includes medulloblastoma, the application of whole-genome and molecular cytogenetic techniques and the targeted investigation of specific oncogenes and TSGs. The major findings from these investigations are summarised in the following sections.

1.3.7.1 Familial cancer syndromes

Although the majority of medulloblastomas (~95%) are sporadic, the first insights into the molecular basis of the disease were obtained by studying familial cases of medulloblastoma. Familial linkage analysis enabled the identification of genes involved in the molecular pathogenesis of medulloblastoma. The phenotypes of the familial cancer syndromes basal cell nevus (Gorlin) syndrome (Gorlin et al., 1965), Turcot syndrome (Turcot et al., 1959) and Li-Fraumeni syndrome (Li and Fraumeni, 1969) all can include medulloblastoma.

Germline mutations in the tumour suppressor *PTCH1* have been identified in patients with Gorlin syndrome (Johnson et al., 1996), also known as naevoid basal cell carcinoma syndrome. The rate of medulloblastoma in individuals affected by Gorlin syndrome is 600 times that of the general population. The *PTCH1* gene encodes a protein that is a membrane-bound receptor involved in the developmental SHH pathway (section 1.3.9). Subsequent investigations revealed that 10% of sporadic medulloblastomas displayed *PTCH1* mutations (Raffel et al., 1997), indicating a role for the SHH pathway in medulloblastoma (section 1.3.9.1).

Turcot syndrome is a hereditary disease characterised by colonic polyps, colorectal cancer and neuroepithelial tumours (Hamilton et al., 1995). Type I Turcot syndrome is characterised by hereditary non-polyposis colorectal cancer syndrome, mutations in DNA mismatch repair genes and an early onset of malignant glioma (Paraf et al., 1997); in type II Turcot syndrome, affected individuals are characterised by germline mutations of the TSG *APC*, and an increased incidence of CNS tumours, principally medulloblastoma. APC protein binds β-catenin (Rubinfeld et al., 1993), acting as an antagonist of the WNT signalling pathway (section 1.3.10), and mutation of the *APC* gene can free β-catenin from proteosomal degradation, activating the WNT pathway. Mutations in *APC* have been identified in less than 5% of sporadic medulloblastomas, but the identification of the importance of the WNT pathway for tumourigenesis in individuals affected by type II Turcot syndrome enabled the importance of the WNT pathway in medulloblastoma to be recognised.

Li-Fraumeni syndrome is associated with multiple tumour types, including medulloblastoma and mutations in the *TP53* gene (located at chromosome 17p13.1)

have been identified in affected families (Malkin et al., 1990). Subsequently, it has been reported that the overall frequency of *TP53* mutations in medulloblastoma (both familial and sporadic cases) is from 7-16% (Lindsey et al., 2011; Pfaff et al., 2010; Tabori et al., 2010; Ellison, 2002). The role of *TP53* as an important TSG has been widely discussed (Malkin, 2001), but briefly, *TP53* acts as a transcription factor in response to DNA damage, regulating the cell cycle and apoptosis; inactivated *TP53* serves to bypass cell cycle check-points, avoiding apoptosis, and contributing to immortalisation of the cancer cell.

The gene mutations identified in familial cases pointed towards a role for aberrantly expressed WNT and SHH pathways in medulloblastoma pathogenesis. The roles of the SHH and WNT pathways in cancer are now well characterised (Klaus and Birchmeier, 2008; Marino, 2005) and are discussed in sections 1.3.9 and 1.3.10. Additional familial syndromes have implicated further disease pathways in medulloblastoma (Table 1.7), but the mechanisms by which they drive carcinogenesis are poorly understood.

Syndrome	Gene	Locus	Tumour Susceptibility	Risk	Reference
Turcot Syndrome (Type 2)	*APC*	5q21 – 22	Medulloblastoma Multiple colorectal adenoma	79% 100%	(Paraf et al., 1997; Hamilton et al., 1995)
Gorlin Syndrome	*PTCH1*	9q22.3	Medulloblastoma Basal cell carcinoma	3 - 5% 100%	(Hahn et al., 1996; Johnson et al., 1996)
Li-Fraumeni Syndrome	*TP53*	17p13.1	Medulloblastoma Multiple primary neoplasms	~2% 3-25%	(Kleihues et al., 1997; Malkin et al., 1990)
Fanconi Anaemia subtype 'D1'	*BRCA2*	13q12.3	Medulloblastoma Wilms' tumour Neuroblastoma Haematological malignancies (infants) Breast cancer (adult)	High	(Reid et al., 2007; Hirsch et al., 2004; Offit et al., 2003)
Fanconi Anaemia subtype 'D1'	*PALB2*	16p12.1			
Rubenstein – Taybi Syndrome	*CREBBP*	16p13.3	Medulloblastoma Other CNS tumours Neural crest tumour	Rare Rare Rare	(Taylor et al., 2001)
Coffin – Siris Syndrome	-	-	Medulloblastoma	Rare	(Fleck et al., 2001; Rogers et al., 1988)

Table 1.7. Familial cancer syndromes implicated in medulloblastoma. Table adapted from (Gajjar and Clifford, 2010).

1.3.7.2 Recurrent chromosomal abnormalities

Investigations into the chromosomal abnormalities observed in medulloblastoma have identified a series of non-random chromosomal gains and losses that are characteristic of medulloblastoma. Array CGH and fluorescent *in situ* hybridisation (FISH)-based techniques for assessing copy number have more recently been augmented with SNP arrays and LOH analyses.

Abnormalities of chromosome 17 represent the most frequently observed chromosomal defect in medulloblastoma. Chromosome 17p loss has been identified in 30-40% of medulloblastomas and is commonly associated with a reciprocal gain of 17q, resulting in isochromosome 17q (i(17q)) in 30-40% of medulloblastomas (Northcott et al., 2009; Nicholson et al., 2000; Bigner et al., 1997; Bigner et al., 1988). Approximately 15% of tumours show gain of 17q without loss of 17p and gain of the whole of chromosome 17 is observed in 5-10% of tumours (Cho et al., 2011; Northcott et al., 2009; Lamont et al., 2004). Loss of chromosome 17p in isolation has been observed in 20% of patients and has been associated with a poor prognosis (Gilbertson et al., 2001). Isochromosome 17q has been associated with an unfavourable prognosis (Pfister et al., 2009; Pan et al., 2005), although this relationship has not been observed in other studies (Ellison et al., 2011b).

The common loss of 17p is suggestive that this region harbours a TSG that is inactivated in medulloblastoma and the proposed region has been determined to lie between 17p13.1 and 17p13.3 (Jung et al., 2004). The location of the *TP53* gene at 17p13.1 was of great interest, but the frequency of *TP53* mutations of ~10% in sporadic medulloblastoma indicates that this gene is not the common target gene associated with loss of 17p (Ellison, 2002). The *REN* (*KCDT11*) gene, which is situated at 17p13.2, has been suggested as a TSG, since it inhibits the SHH signalling pathway and is deleted in medulloblastoma (Di Marcotullio et al., 2004). Finally, epigenetic inactivation of the *HIC1* gene, located at 17p13.3, has been linked, in conjunction with loss of 17p, with medulloblastoma (Rood et al., 2002). The precise role for *HIC1* has yet to be determined, but mice heterozygous for *HIC1* develop a range of malignancies (Chen et al., 2003), consistent with a two-hit inactivation of a tumour suppressor.

The ascertainment of samples from recurrent medulloblastoma has been challenging, since surgery is not usually indicated at relapse, and treatments are usually palliative. However, a study of a cohort of 28 paired presentation and relapse cases identified that the acquisition of chromosomal abnormalities associated with a poor risk (17q gain, gain of chromosome 6) is characteristic of recurrent tumours (Korshunov et al., 2008).

Other non-random chromosomal aberrations have been associated with medulloblastoma. Gain of chromosome 7 or isolated 7p / 7q gain is the second most frequent gained region, observed in 20 - 30% of tumours (Northcott et al., 2009; Lo et al., 2007; Avet-Loiseau et al., 1999), and recent work has linked loss of 10q in association with 7q gain as being prognostic for poor outcome in paediatric medulloblastoma (Pezzolo et al., 2011). Interestingly, the genes *HGF* and *MET* are located on chromosome 7q, and HGF / MET signalling plays a role in cerebellar development, with stimulation of Hgf / Met signalling in mice preventing apoptosis in granule neuron precursors (GNPs) within the cerebellum (Kongkham et al., 2008). A comprehensive analysis of 212 medulloblastomas using SNP arrays identified frequent loss of chromosomes 6, 8, 9q, 11, 16q and 17p and gains of chromosomes 1q, 7 and 17q (Northcott et al., 2009), suggesting that these chromosomes harbour genes important in disease pathogenesis which are currently unidentified.

The demonstration that medulloblastoma comprises discrete disease subgroups, two of which characterised by activation of the WNT and SHH pathway (section 1.3.9, 1.3.10, 1.3.12.1), has enabled the identification of specific chromosomal abnormalities associated with subgroups. The WNT subgroup is associated with loss of chromosome 6 in the majority of cases (Clifford et al., 2006) and is independent of chromosome 17 abnormalities (Pizer and Clifford, 2009), although the putative TSG located on chromosome 6 has also not yet been definitively identified.

Alongside gross chromosomal abnormalities, there are locus-specific amplifications of specific oncogenes that have been demonstrated to be important in medulloblastoma, alongside dysregulation of specific genes, which are discussed in the following sections.

1.3.7.3 Dysregulation of specific medulloblastoma-associated genes

The recent whole exome sequencing strategy employed on 22 medulloblastomas, with validation in an 88 member cohort, represented a first attempt to catalogue genes commonly mutated, amplified or deleted in medulloblastoma (Parsons et al., 2011), although this report cannot be regarded as definitive, since genes (for example, *MYCN*) with aberrant alterations known to be associated with medulloblastoma at low frequencies (5% or less) were not identified (Table 1.8). The novel observation that *MLL2* and *MLL3* can be mutated in medulloblastoma raises the possibility that inactivation of these genes, which are involved in chromatin regulation, may disrupt transcriptional regulation of normal brain development and differentiation (Lim et al., 2009).

Aside from *MLL2*, the most frequently mutated genes in medulloblastoma are core members of the SHH and WNT pathways (*PTCH1* and *CTNNB1*, respectively), which are discussed in sections 1.3.9 and 1.3.10.

One mechanism for oncogene activation is through amplification, where an oncogene with non-mutated sequence is transcribed at an inappropriately high level (section 1.1.5.1). TSGs can also be inactivated through epigenetic silencing, and these mechanisms are discussed in the following sections.

Gene	Number of mutations (%)	Number of amplifications (%)	Number of deletions (%)
PTCH1	22 / 88 (25%)	0 / 23 (0%)	0 / 23 (0%)
MLL2	12 / 88 (14%)	0 / 23 (0%)	0 / 23 (0%)
CTNNB1	11 / 88 (13%)	0 / 23 (0%)	0 / 23 (0%)
TP53	6 / 88 (7%)	0 / 23 (0%)	0 / 23 (0%)
MYC	0 / 88 (0%)	3 / 23 (13%)	0 / 23 (0%)
PTEN	3 / 88 (3%)	0 / 23 (0%)	0 / 23 (0%)
OTX2	0 / 88 (0%)	2 /23 (9%)	0 / 23 (0%)
SMARCA4	3 / 88 (3%)	0 / 23 (0%)	0 / 23 (0%)
MLL3	3 / 88 (3%)	0 / 23 (0%)	0 / 23 (0%)

Table 1.8. An exome screen of 22 medulloblastomas identifies genes with potential relevance to medulloblastoma pathogenesis. Genes with aberrant alterations in at least 2 / 22 members of the primary cohort were validated in a larger 88 member cohort and their frequency and incidence is shown for that cohort. A cohort of 23 tumours, which included all members of the primary cohort, was analysed for copy number alterations. Table adapted from Parsons et al., 2011.

1.3.7.3.1 Amplification of oncogenes in medulloblastoma

The amplification of oncogenes is a well-described mechanism for disease progression in medulloblastoma. Multiple studies have reported both the presence and poor prognosis associated with amplification of members of the *MYC* family of oncogenes (*MYCN, MYCC* and *MYCL*) in medulloblastoma (Northcott et al., 2009; Pfister et al., 2009; Rutkowski et al., 2007; Lamont et al., 2004; Aldosari et al., 2002; Brown et al., 2000). The *MYC* genes contribute to many mechanisms of tumourigenesis, including transformation, cell proliferation and differentiation, apoptosis, and genome instability (Secombe et al., 2004). Amplification of *MYCC* is observed in approximately 5% of medulloblastomas, with *MYCN* amplification at a similar frequency (Pfister et al., 2009; Lamont et al., 2004; Aldosari et al., 2002). Increased expression of *MYCC* can promote the proliferation of cerebellar granule cells in mouse models and an anaplastic histology in medulloblastoma cell lines (Stearns et al., 2006; Fults et al., 2002), while over-expression of *MYCN* is sufficient to drive the pathogenesis of metastatic medulloblastoma in mouse models (Swartling et al., 2010). *MYCL* amplification is only rarely observed (1-2%) (Northcott et al., 2009).

The regulation of the MYC proteins is governed by their formation of heterocomplexes with MAX, which promotes transcription at genes containing an E-box binding site. MAX is also able to form repressive complexes with the MAD family of proteins and the interplay between MYC, MAD and MAX governs the transcriptional status of the targets of MYC (Grandori et al., 2000).

The oncogenes *OTX2* (Parsons et al., 2011; Northcott et al., 2009) and the SHH-associated *GLI2* oncogene (Northcott et al., 2011; Rieber et al., 2009) have been consistently reported at a frequency of <5% in medulloblastoma.

1.3.7.3.2 Epigenetic silencing of medulloblastoma genes

Candidate gene approaches applied to the study of CpG island methylation in medulloblastoma have been reviewed (Lindsey et al., 2005). *RASSF1A*, *CASP8* and *HIC1* (Lindsey et al., 2005) show promoter CpG island hypermethylation, consistent across multiple studies, that is associated with transcriptional silencing in significant proportions of cases. Genes hypermethylated in cancer encode proteins that have a

diverse range of functions, including microtubule stabilisation and regulation of mitosis (*RASSF1A*), regulation of transcription (*HIC1*), and apoptosis (*CASP8*).

Recent work has identified possible epigenetic mechanisms for activation of the WNT and SHH signalling pathways. The *PTCH1-C* promoter was identified as hyper-methylated in medulloblastomas, and reactivation of expression of *PTCH1* following treatment with the DNA methyltransferase inhibitor 5'aza-2'-deoxycytidine in cell lines with a methylated *PTCH1-C* promoter indicates that epigenetic silencing of *PTCH1* may be important (Diede et al., 2010). The silencing of the *SFRP* gene family though DNA methylation has been suggested to be a mechanism by which inhibition of the WNT pathway could be reduced, which may contribute to excessive WNT signalling (Kongkham et al., 2010).

In other tumours, evidence has emerged for the presence of hyper-methylator phenotypes (West and Barrett, 1993). It is characterised by tumour-specific hypermethylation at several hundred CpG islands (Costello et al., 2000). Since its initial description in Syrian hamster cells, this CpG island hyper-methylator phenotype (CIMP) has been described primarily in colorectal cancers, but also in neuroblastoma, gastric carcinoma, hepatocellular carcinoma, and biliary and ampullary carcinoma (Teodoridis et al., 2008). More recently, a subtype of glioblastoma was reported to be characterised by a CpG island methylator phenotype (Noushmehr et al., 2010). No evidence for CIMP in medulloblastoma has yet been reported.

Genome-wide techniques for assessing the methylation status of gene promoters in medulloblastoma have become increasingly available and are described in section 1.3.13. Importantly, a genome-wide characterisation of the DNA methylome in medulloblastoma has not previously been undertaken.

1.3.8 Aberrant activation of embryonal signalling pathways

Cell signalling pathways play important roles in developmental processes. Inappropriate activation of these pathways can drive the cell towards inappropriate growth and differentiation. Through the study of familial cancer syndromes, mutations in key components of the WNT and SHH signalling pathway have been implicated in

the pathogenesis of medulloblastoma, pathways fundamental to our understanding of medulloblastoma.

1.3.9 Sonic hedgehog signalling pathway

The Hedgehog pathway is an important regulator of embryonic development, including stem cell maintenance, cell differentiation and proliferation, and tissue polarity. Initially discovered in 1980, the hedgehog gene (*Hh)* is named after the spiked phenotype observed in mutant *Drosophila* larvae (Nusslein-Volhard and Wieschaus, 1980). Subsequently, three mammalian homologues of the *Drosophila* hedgehog gene (*Hh)* have been identified: *Sonic (SHH), Indian (IHH)* and *Desert (DHH);* the most widely studied homologue is the *SHH* gene. The SHH pathway has diverse effects in different cellular contexts. It can act as a morphogen in cell fate determination and a mitogen in the development of organs (Ingham and McMahon, 2001). The crucial role for the SHH pathway in cerebellar development is discussed below, but the pathway is also important in adults, where it plays a role in the proliferation of adult stem cells (in particular, neural stem cells (Gupta et al., 2010)), tissue repair and renewal and tissue homeostasis (Hooper and Scott, 2005).

The importance of inappropriate SHH pathway activation in cancer is now widely recognised, with the link between Gorlin's syndrome (section 1.3.7.1) and basal cell carcinoma providing the first evidence for its involvement. The SHH pathway is summarised in Figure 1.13. During normal SHH pathway signalling, SHH ligand binds to PTCH1, which releases the receptor SMO from inhibition. In turn, this activates a signalling cascade, possibly mediated by G proteins, that activates the GLI family of transcription factors, by forming an activated complex, GLI A, that transcriptionally activates SHH pathway target genes, such as Cyclin D1 and *MYCN* (Kenney et al., 2003). In the absence of SHH ligand, GLI proteins are proteolytically processed to generate the repressive GLI R complex, which prevents SHH target genes from being transcribed.

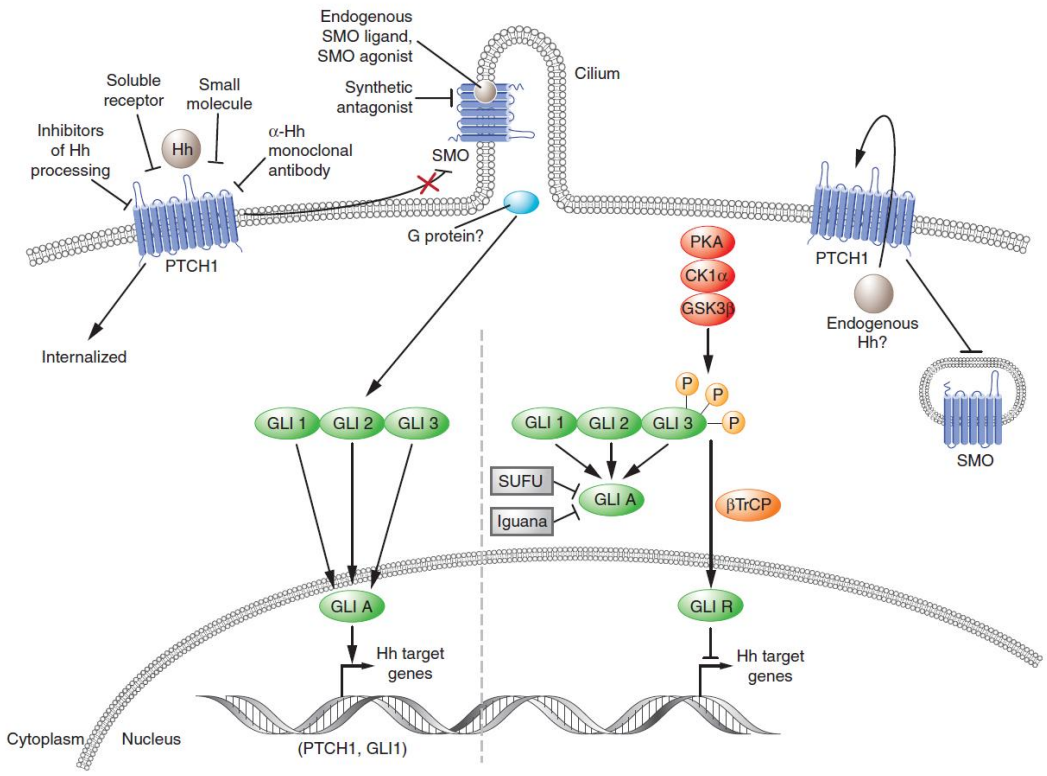


Figure 1.13. The human SHH pathway. In this pathway, signalling is initiated by SHH protein binding to its receptor, PTCH1, which in turn releases SMO from inhibition. SMO activates downstream transcription factors of the GLI gene family, which translocate to the nucleus and activate target genes of the SHH pathway (Carlotti Jr et al., 2008). The mechanisms controlling the GLI proteins are not well understood in humans, although it has been demonstrated that SUFU (suppressor of fused) is a powerful GLI1 antagonist that mediates inhibition of transcription, export from the nucleus and cytoplasm sequestration (Barnfield et al., 2005). GLI1 and GLI2 proteins function as transcriptional activators, whilst GLI3 is a transcriptional repressor (Ruiz i Altaba, 1997). In the absence of Hh ligand, the repressor GLI R is generated shown at the bottom right of the figure, whilst with an activated SMO, the GLI A activator is produced, which drives expression of target genes. The target genes include *PTCH1* itself, bone morphogenic proteins and other genes involved in cell proliferation and differentiation (Ferretti et al., 2005), including Cyclin D1, *MYCN, BMI1, BCL-2* and *VEGF*. Figure taken from (Rubin and de Sauvage, 2006).

1.3.9.1 Role of the SHH pathway in cerebellar development

The development of the cerebellum is still poorly understood, although it is now clear that the developmental processes that occur within the cerebellum are related to the origin and pathogenesis of medulloblastoma (section 1.3.13). There is growing evidence that oncogenic abnormalities in the constituent cells and signalling pathways that govern the complex structures of the cerebellum may drive the formation of different subtypes of medulloblastoma.

At an early stage of cerebellar development, cells are derived from two distinct germinal zones; one from precursor cells in the roof of the fourth ventricle, which gives rise to GABAergic precursors, including Purkinje cells; the other from cells within the rhombic lip, which comprises of GNP cells, which produce the external granule layer (EGL) (Gilbertson and Ellison, 2008). The EGL persists into the second year post-natally.

The distinct precursor cells of the cerebellum are orchestrated by the activation of specific signalling pathways to guide the development of the cerebellum. The SHH pathway plays a crucial role in this development, where its primary role is to control the expansion, differentiation and migration of GNPs in the EGL of the cerebellum (Ruiz i Altaba et al., 2002), as shown in Figure 1.14. The importance of the SHH pathway is further reinforced by experiments that inhibit SHH signalling in mouse models of cerebellar development, where inhibition results in a marked decrease in the proliferation of the EGL (Dahmane and Ruiz i Altaba, 1999); conversely, recombinant SHH drives the proliferation of GNPs (Wechsler-Reya and Scott, 1999).

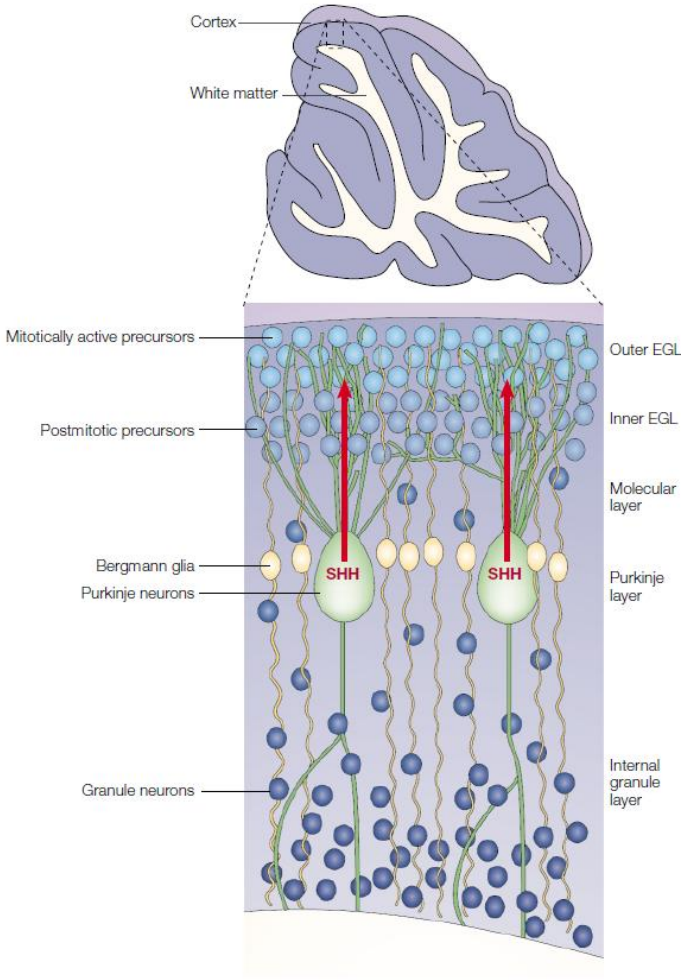


Figure 1.14. Role for SHH pathway in post-natal development of the cerebellum. The secretion of SHH ligand from Purkinje neurons promotes the rapid proliferation of granule cell precursors in the EGL. These cells migrate through the molecular and Purkinje layer, whilst undergoing terminal differentiation, where they become mature granule neurons within the internal granule layer of the cerebellum. Figure taken from Ruiz I Altaba et al., 2002.

1.3.9.2 The SHH pathway in medulloblastoma

Activation of the SHH signalling pathway characterises approximately 25% of medulloblastomas. Mutations in SHH pathway components (*PTCH1, SUFU,* SMO) have been described for subsets of medulloblastomas with activated SHH pathway (Ellison et al., 2003), although there remains a sizeable proportion of SHH medulloblastomas for which a causative mechanism has not been identified, indicating that other mechanisms for pathway activation remain to be discovered.

The mechanisms by which the SHH pathway can drive medulloblastoma have been elucidated in mouse models of the disease. 10-15% of $Ptc^{+/-}$ mice go on to develop medulloblastoma (Goodrich et al., 1997) and express high levels of *Gli1*, consistent with activation of the SHH pathway. More than half show residual populations of GNP cells at the surface of the cerebellum that fail to undergo terminal differentiation and migration to the internal granule layer, suggestive that medulloblastoma arises from these residual populations in this model (Goodrich et al., 1997).

Activation of the SHH pathway is observed in the majority of infant cases (Northcott et al., 2011) and is also associated (although not exclusively) with a desmoplastic phenotype. A similar enrichment has been observed in adult cases, where an incidence of 50% has been reported (Remke et al., 2011b). Multiple studies have identified distinct sets of differentially expressed genes that characterise medulloblastoma with an activated SHH signalling pathway (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006). No apparent differences in survival between SHH and WNT / SHH independent tumours have been identified (Cho et al., 2011; Ellison et al., 2011a; Northcott et al., 2010), although this requires additional validation in clinically controlled trials cohorts. Similarly, no difference in survival has been reported between adult and childhood SHH medulloblastomas (Northcott et al., 2011; Remke et al., 2011b).

1.3.10 WNT signalling pathway

The WNT signalling pathway plays an important role in embryogenesis and cancer, but it is also important in the regulation of normal physiological processes in adults. Initially recognised to be important in developmental biology, its relevance to cancer

later became apparent when the interaction with the tumour suppressor APC and β-catenin was identified (Rubinfeld et al., 1993; Su et al., 1993). The identification of mutations in the *APC* gene in families affected with Turcot syndrome (section 1.3.7.1) linked their syndrome to inappropriate activation of the WNT signalling pathway. Subsequent work identified that up to 85% of sporadic colorectal cancers are associated with truncating mutations in *APC* (Groden et al., 1991). More rarely, gain of function mutations in *CTNNB1* (the β-catenin gene) have been reported in colorectal cancers (Klaus and Birchmeier, 2008), and it is this type of mutation that is most commonly observed in WNT pathway activated medulloblastomas (section 1.3.10.2).

In the absence of WNT ligands, cytoplasmic β-catenin is recruited into a destruction complex, consisting of APC, Axin and Glycogen synthase kinase-3β (GSK-3β), where it is N-terminal phosphorylated by casein kinase-1α and GSK-3β. After phosphorylation, β-catenin is targeted for proteosome mediated degradation, ensuring that cytoplasmic levels of β-catenin remain low.

Activation of the canonical WNT signalling pathway is initiated by secreted WNT proteins binding to Frizzled receptors in the plasma membrane. This interaction can be inhibited by several proteins, including the secreted, frizzled, related proteins (SFRPs), Dickkopfs (DKKs) and WNT-inhibitory factor 1 (*WIF1*) (Klaus and Birchmeier, 2008). In the next step, Dishevelled (*DSH*) is recruited to the plasma membrane, where it interacts with Frizzled to mediate the translocation of Axin to the plasma membrane and inactivation of the destruction complex. This inactivation enables the cytoplasmic stabilisation and subsequent translocation of β-catenin to the nucleus.

In the nucleus, β-catenin forms a transcriptionally active complex with the lymphoid enhancer factor (LEF) and T cell factor (TCF) transcription factors. Target genes of the transcriptional complex include the proto-oncogene, *MYC*, proliferative genes such as *cyclin D1* and cell signalling genes (*VEGF, FGF4* and *FGF18*) (Klaus and Birchmeier, 2008). The canonical WNT signalling pathway is summarised in Figure 1.15.

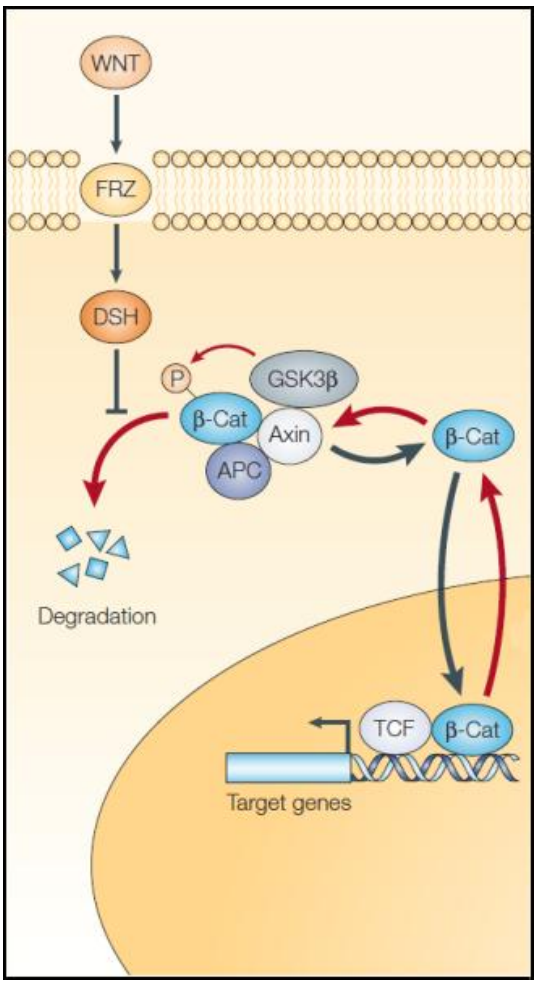


Figure 1.15. The WNT signalling pathway. In the absence of WNT ligand, β-catenin is recruited by the cytoplasmic destruction protein complex, which contains APC, Axin and glycogen synthase kinase -3β (GSK-3β). GSK-3β phosphorylates the N-terminal domain of β-catenin, targeting it for degradation. WNT ligand binds to the trans-membrane ligand frizzled (FRZ). Dishevelled (DSH) has an inhibitory effect on the phosphorylation of β-catenin by GSK3β, enabling β-catenin to translocate to the nucleus, where it forms a transcriptionally active complex with lymphoid enhancer factor (LEF) and T cell factor (TCF), and mediates transcription of proliferative target genes, such as *MYCC, Cyclin D1* and *AXIN2*. Figure adapted from (Scotting et al., 2005).

1.3.10.1 Role of the WNT pathway in cerebellar development

The role for the WNT pathway in cerebellar development is poorly understood, although it has been shown that deletion of the *Wnt-1* gene from mice completely blocks cerebellar development by preventing the formation of the midbrain-hindbrain junction from which the cerebellum is derived (Thomas and Capecchi, 1990). Recent work in mouse models reported the expression of WNT pathway target genes in the lower rhombic lip of the cerebellum at day 11.5 of embryonic development, and the dorsal brainstem at day 15.5 (Gibson et al., 2010), which helped to identify a putative cell of origin for medulloblastomas with an activated WNT pathway (section 1.3.13).

1.3.10.2 Dysregulation of the WNT pathway promotes medulloblastoma

The most commonly observed mutation in medulloblastoma with an activated WNT signalling pathway are activating mutations that affect the phosphorylation domain of the β–catenin protein, affecting 8% of sporadic medulloblastomas (Gilbertson and Ellison, 2008) and ~95% of WNT-activated medulloblastomas (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006). Now, in the absence of WNT ligand, the β-catenin protein can no longer be phosphorylated by the complex of APC, axin and GSK-3β and marked for degradation (Figure 1.15). Instead, unmodified β–catenin is free to translocate to the nucleus and initiate transcription of proliferative WNT pathway target genes.

Mutations in other WNT pathway components have also been identified. Individuals affected by Turcot syndrome harbour mutations in *APC,* one of the proteins responsible for targeting β–catenin for degradation (section 1.3.7.1), and rare mutations of *AXIN* have also been reported (Baeza et al., 2003). The majority of WNT pathway activated tumours are of classic histology (Ellison et al., 2005; Eberhart et al., 2000), occur in non-infants and have dual peaks of incidence at 10 and 20 years. They have also been associated with a favourable prognosis (Cho et al., 2011; Northcott et al., 2010; Fattet et al., 2009; Clifford et al., 2006; Ellison et al., 2005).

WNT activated tumours have distinct molecular and genomic defects that enable robust classification of these tumours. They can be identified by a nucleopositive

immunophenotype for β–catenin (Ellison et al., 2005; Eberhart et al., 2000), and this is the basis for the identification of WNT positive cases in the upcoming PNET5 clinical trials cohorts (Pizer and Clifford, 2009), discussed further in section 1.3.16.

1.3.11 Novel treatment options through inhibition of signalling pathways

The treatment of cancers through the specific inhibition of activated developmental signalling pathways represents an attractive therapeutic goal, since the non-tumour-specific cellular damage caused by intensive chemotherapy and radiotherapy could be reduced with adjuvant inhibitor therapy, enabling a much improved quality of life for the patient, both during and after treatment, whilst still achieving a cure. In recent years, targeting of tumour-specific mutant proteins with specific inhibitors have been reported, including the use of imatinib to target the BCR-ABL fusion protein in patients with chronic myelogeneous leukaemia (Druker et al., 2006), and the use of erlotinib to target EGFR in patients with non-small-cell lung cancer (Jänne et al., 2005). These approaches are attractive also, since they are applicable to any cancer with a specific pathway activation, and in future, it may be possible to define cancers not only by their location, but also their signalling pathway activation status, so that personalised therapies can be administered.

The recognition that up to 25% of medulloblastoma patients have tumours characterised by SHH signalling pathway activation and that WNT pathway activation is observed in a further 10% has prompted investigation into the discovery and utility of specific inhibitors for these pathways that may be useful for treating medulloblastoma. Progress for inhibition of the SHH pathway is more advanced and is reviewed first.

Pre-clinical data in mouse models demonstrated that suppression of the SHH pathway using a small molecule inhibitor of Smoothened could eliminate medulloblastoma in *Ptc1*$^{+/-}$*p53*$^{-/-}$ mice (Romer et al., 2004). Subsequently, the inhibitor has been developed in humans as the drug vismodegib (GDC-0449), which is an orally available inhibitor of the SHH pathway gene *SMO.* A recent case report detailed the dramatic (albeit transient) regression of metastatic medulloblastoma in an adult male (Rudin et al., 2009). After three months, the patient relapsed, and a later publication identified a mutation in SMO that had no effect on SHH signalling but disrupted the ability of GDC-

0449 to bind SMO (Yauch et al., 2009). As of 2011, there is an ongoing phase II trial evaluating the efficacy and safety of GDC-0449 for the treatment of adults with recurrent or refractory medulloblastoma (Gupta et al., 2010). Second-generation inhibitors of SMO will be designed to retain activity to both the wild type and mutant forms of SMO (Dijkgraaf et al., 2011).

The recent discovery that itraconazole, a routinely administered antifungal, was a potent inhibitor of the SHH pathway (Kim et al., 2010), raises hopes that this could be an easily tolerated and clinically robust method for the inhibition of medulloblastoma, although the drug does not pass through the blood / brain barrier and would need to be administered directly to the CSF. Since it inhibits the SHH pathway by preventing the ciliary accumulation of SMO, distinct from the mode of action of SMO inhibitors such as vismodegib, this drug could be evaluated in conjunction with other targeted therapies with the aim to prevent resistance mechanisms from developing.

The inhibition of SHH pathway in young children with medulloblastoma needs to proceed with caution, since in a mouse medulloblastoma model, treatment with a specific SHH pathway inhibitor caused permanent defects in bone structure, by inducing premature fusion of the growth plates (Kimura et al., 2008). It is unclear whether this side effect would be observed in humans, and trials should proceed with care to evaluate optimal toxicity endpoints.

WNT inhibitors are not as fully developed as SHH inhibitors and are yet to enter clinical trials. This was primarily due to a perceived lack of druggable targets, although recent progress has raised hope that inhibitors can be developed. Inhibition of the WNT pathway would have clinical utility outside of medulloblastoma; most importantly, over 80% of colorectal cancers are driven by WNT pathway mutations (Garber, 2009). Cyclo-oxygenase (COX) inhibitors have been shown to down-regulate WNT signalling, through poorly understood mechanisms (Barker and Clevers, 2006); subsequently, specific inhibitors have been identified: the drug XAV939 has been reported to antagonise WNT signalling by stabilising axin, which is the concentration-limiting component of the β-catenin destruction complex (Huang et al., 2009). The application to medulloblastoma is unclear, however, since the most frequently observed activation mechanism for WNT signalling is an activating mutation of β-catenin, which prevents

interaction with the destruction complex. Indirect inhibition of the WNT pathway may also have clinical utility. In mouse medulloblastoma xenograft models, down-regulation of the WNT pathway was achieved by inhibiting the phosphatidylinositol 3-kinase / Akt signalling (Baryawno et al., 2010). The application of WNT inhibitors must be considered with the recognition for the crucial role of the WNT pathway in self-renewal, particularly in the bone marrow and gut, and the serious side-effects that would accompany a systemic inhibition of the pathway.

The treatment of subgroups of medulloblastoma characterised by activated embryonal signalling pathways with specific inhibitors holds great promise for the development of future treatment strategies that minimise side-effects and maximise tumour-specificity. However, progress has been limited, and routine use of these approaches in medulloblastoma remains several years away, although the first SHH inhibitors are entering clinical trials. There also remain unanswered questions about the suitability of the inhibitors in young children, with the potential for devastating side-effects if normal developmental processes that rely on SHH or WNT pathway activation are irreversibly halted.

1.3.12 High throughput genomic and transcriptomic methods in medulloblastoma

The application of high-throughput genomic and transcriptomic techniques has helped characterise the recurrent changes that define subgroups of medulloblastoma, which have subsequently proved useful for the characterisation of disease determinants, clinical behaviours and disease stratification in what is now recognised as a molecularly heterogeneous disease.

1.3.12.1 Transcriptomic subgrouping of medulloblastoma

Since the first application of transcriptomic arrays to medulloblastomas in 2002, which demonstrated that medulloblastoma was molecularly distinct from PNET, atypical teratoid / rhabdoid tumours (AT / RT) and glioma (Pomeroy et al., 2002), further advances have led to an evolving molecular classification of medulloblastoma. More recent studies have focussed on investigating the molecular heterogeneity within medulloblastomas, in order to subclassify the disease, so that subgroup specific

disease features and clinical behaviours can be identified. Many of these studies have also applied genomic arrays in conjunction with the transcriptomic assays, enabling an integrative approach to disease classification, which subclassifies medulloblastoma by its transcriptome, with subgroups associated with enrichments for specific chromosomal aberrations.

Unsupervised microarray experiments by Kool and colleagues (*n* = 62) and Thompson and colleagues (*n* = 46) identified five disease subgroups (Kool et al., 2008; Thompson et al., 2006). Thompson and colleagues used unsupervised hierarchical clustering (HC) (section 2.9.1) to identify five disease subgroups, designated A-E. The B and D subgroups were characterised by expression signatures of the WNT and SHH pathways, respectively. Subsequent investigations identified mutations in *CTNNB1* and *PTCH1 / SUFU* in subgroups B and D, respectively, confirming their designation as WNT and SHH subgroups of medulloblastoma. The WNT subgroup was also strongly associated with monosomy of chromosome 6. A common theme to all transcriptomic subclassifications of medulloblastoma has been the relative difficulty in assigning definitive transcriptomic, genomic and clinico-pathological features that robustly identify the remaining WNT / SHH independent subgroups across different studies, compared to the WNT and SHH disease subgroups. Thompson and colleagues identified that anaplasia was enriched in subgroup E tumours, and isochromosome 17q in subgroup C, but no other associations with these WNT / SHH independent subgroups were identified (Thompson et al., 2006).

Kool and colleagues also identified five disease subgroups (again designated A-E) using HC. For the majority of cases, this transcriptomic classification was paired with array CGH. Again, WNT (A) and SHH (B) disease subgroups were identified, characterised by *CTNNB1* mutation, monosomy 6 and good outcome (WNT cases) and *PTCH1* mutation and 9q loss (SHH cases). Subgroups C and D were associated with expression signatures of neuronal activation and glutamate and gamma-aminobutyric acid receptors, whereas D and E were associated with retinal gene expression patterns. Chromosome 17 abnormalities were more common in groups C and D. WNT / SHH independent subgroups (C-E) were associated with an increased incidence in metastatic disease. There were also notable differences in clinico-pathological features, with WNT cases occurring in older children and associated with a classic histology. By

contrast, SHH cases were strongly associated with infant and DN disease. Their study demonstrated that there was a degree of transcriptomic overlap between the WNT / SHH independent subgroups, with shared increases in expression of genes involved in protein biosynthesis, NOTCH and platelet-derived growth factor (PDGF) pathways. Likewise, group C overlapped with group D, and group D with group E, again emphasising the difficulty in assigning robust markers that define the subgroups (Kool et al., 2008).

In 2010, Northcott and colleagues undertook a larger study on 103 samples, investigating both transcriptomic and genomic signatures of medulloblastoma (Northcott et al., 2010). Again, using an unsupervised HC approach, they identified 4 disease subgroups. They described a WNT subgroup, associated with a WNT expression signature, monosomy of chromosome 6, and a good prognosis, and a SHH subgroup, characterised by a SHH expression signature, desmoplastic histology, and an increased incidence of infant and adult cases; there were two further WNT / SHH independent subgroups, C and D. There was a male preponderance of cases in subgroups C and D, an increased incidence of isochromosome 17q and both shared expression of genes involved in neuronal development. In contrast, WNT and SHH subgroup cases were enriched for expression of genes involved in axonal guidance. Subgroup C tumours were characterised by genes involved in phototransduction, glutamate signalling, gain of 1q, poorer survival and, along with WNT subgroup cases, increased *MYCC* expression. The paradoxical relationship between *MYCC* expression in WNT and non-WNT cases is interesting, since its increased expression is associated with a favourable prognosis in WNT cases, but an unfavourable prognosis in group C tumours, and provides an example of pleiotropic behaviour of prognostic biomarkers across subgroups of medulloblastoma. It is possible that levels of *MYCC* in WNT tumours are not at the supraphysiological levels found in group C, which depend upon gain or amplification of the *MYCC* gene, which was exclusively observed in this group. Subgroup C tumours all occurred in children under 10 years of age, in contrast to subgroup D cases, with a peak incidence from 11-15 years. Subgroup D tumours had an over-representation of genes involved in cyclic-AMP signalling, semaphorin signalling in neurons and p53 signalling.

Uniquely in this study, four immunohistochemical markers, DKK1, SFRP1, NPR3 and KCNA1 specifically identified the WNT, SHH, C and D subgroups, respectively. Survival analyses in a separate tissue microarray cohort classified by these antibodies confirmed a poorer survival in subgroup C tumours, although questions remain whether the antibody-based classification is exactly comparable to the transcriptomic classification, since the age profiles of the subgroups C and D differ between transcriptomic and antibody based classifications.

This approach to classifying medulloblastomas, coupled with the additional recent description of robust WNT and SHH subgroup immunohistological markers (WNT – nuclear accumulation of β-catenin; SHH – immunoreactivity for GAB1 (Ellison et al., 2011a)), requires additional validation, but offers promise for the routine assignment of medulloblastoma subgroup in hospital pathology laboratories (Eberhart, 2011).

Also in 2010, the largest study to date, by Cho and colleagues, tested the gene expression of 194 medulloblastomas, and used a consensus-clustering non-negative matrix factorisation (NMF) approach to identify six subgroups (c1 to c6), an approach similar in concept to the consensus clustering approach described in section 2.9.4. Briefly, in the first step, optimal numbers of metagenes were estimated using the whole dataset. In the second step, partitions of 85% of the dataset were re-iteratively selected and clustered using the partitioning around medoids algorithm (Kaufman and Rousseeuw, 2008). The stability of each solution was assessed for each number of clusters and the most stable solution selected. Once again, WNT (c6) and SHH (c3) disease subgroups were identified, characterised by expression of WNT and TGF-β genes, monosomy of chromosome 6 (WNT), and enrichment for genesets associated with SHH signalling, loss of 9q (SHH) and amplification of *GLI2*. Subgroup c1 was characterised by high levels of *MYCC* expression, coupled with gains and amplifications of *MYCC.* In addition, expression of photoreceptor genes and *GABRA5* were features shared with the closely related subgroup, c5. Subgroup c5 also had an increased number of chromosomal abnormalities compared to subgroup c1. Subgroups c2 and c4 shared similarities in gene expression, with both enriched for expression of markers of neuronal differentiation, including the glutamatergic markers *GRM1* and *GRM8*. Subgroup c4 also showed evidence for expression of photoreceptor markers and a *MYCC*-mediated expression signature, similar to subgroup c1. Immunostaining for

photoreceptor and neuronal / glutamatergic markers in this subgroup demonstrated how the mixed expression phenotype could be due to distinct cellular populations within the tumour (Cho et al., 2011).

It is possible that the increased size of the study (*n* = 194) by Cho and colleagues (nearly double the size of the next biggest study by Northcott and colleagues (*n* = 103)) enabled the subtle differences between closely related subgroups c2 and c4, and c1 and c5 to be discerned, and that the remaining three studies were not adequately powered to detect these differences, which might represent 'subgroups within subgroups'. Recent work demonstrating that adult and infant SHH medulloblastomas are clinically and molecularly distinct (Northcott et al., 2011) and the identification of three molecular subgroups in adult medulloblastoma, with Northcott subgroup C only rarely identified (Remke et al., 2011b) demonstrates that there is evidence for 'subgroups within subgroups', although it remains to be seen what implications this may have for risk stratification.

In summary, consensus for the precise number of medulloblastoma subgroups has yet to be reached. However, in every reported study, distinct WNT and SHH subgroups were apparent, although the nature and number of WNT / SHH independent subgroups remains unclear; this observation underpinned the investigations reported in chapter 3. The reported molecular and clinico-pathological features of each subgroup vary between studies, and because of the differing numbers of subgroups reported, there is necessarily some overlapping of features, but it is possible to extract some commonalities between studies, as summarised in Figure 1.16.

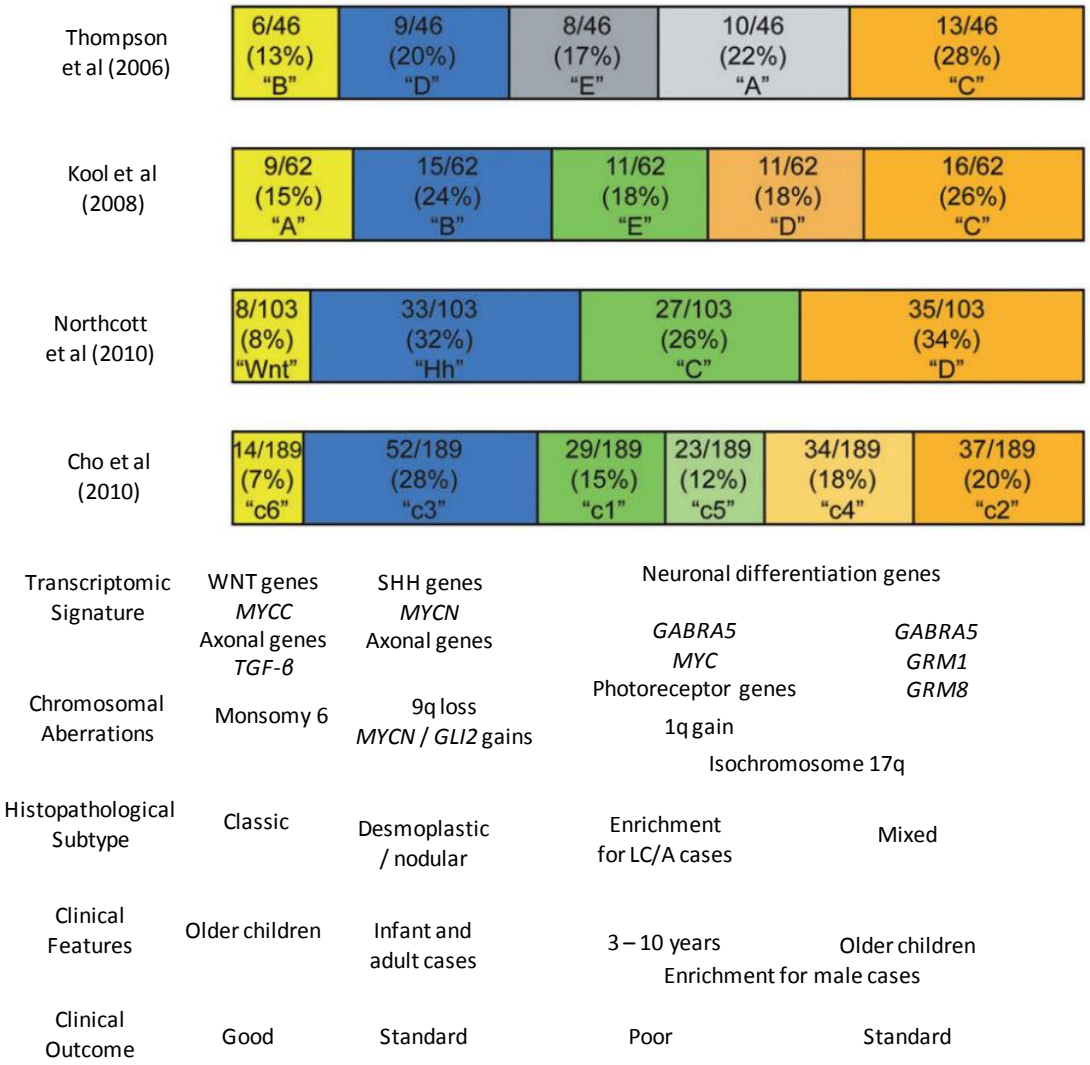


Figure 1.16. Transcriptomic classification of medulloblastoma identifies robust subgroups characterised by activation of the WNT and SHH signalling pathway. For WNT / SHH independent subgroups, reports differ as to the precise number (from two to four WNT / SHH independent subgroups) and nature of these subgroups. The subgroup names, frequency and incidence of subgroup membership and molecular and clinico-pathological features of subgroups are listed for four transcriptomic investigations of medulloblastoma (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006). Figure adapted from Eberhart, 2011.

1.3.13 Cellular origins of medulloblastoma

Originally assumed to develop from embryonic cells in the ependymal lining of the fourth ventricle of the cerebellum (Bailey and Cushing, 1925), our refined understanding of medulloblastoma has led to the proposal that different subtypes of medulloblastoma arise from different types of progenitor cell (Gilbertson and Ellison, 2008). Discussed in sections 1.3.9 and 1.3.10, there are subgroups of medulloblastoma defined by their activation of the WNT and Sonic hedgehog (SHH) embryonal signalling pathways. Recent work in mouse models of medulloblastoma with activated WNT (Gibson et al., 2010) or SHH pathways (Schuller et al., 2008; Yang et al., 2008) has described a cell of origin for tumours activated in these signalling pathways.

One of the most widely studied models of medulloblastoma is the *Ptc* mutant mouse (Goodrich et al., 1997). Homozygous mutation of the *Ptc* gene is embryonic lethal but 10-15% of heterozygous $Ptc^{+/-}$ mice develop medulloblastomas. By creating conditional mutants of *Ptc*, Yang and colleagues were able to inactivate *Ptc* in GNP cells, which led to a phenotype of severe hyperplasia of the external granule layer (Figure 1.17). Most GNP cells continued to differentiate into neurons, but remnant populations remained that continued to divide and formed tumours in 100% of animals (Yang et al., 2008). Interestingly, deleting *Ptc* in neural stem cells also induced medulloblastoma, but only from those stem cells that had committed to the granule cell lineage.

By examining the Brain Explorer 2 (http://www.brain-map.org), Gibson and colleagues catalogued the expression of 24 WNT and 25 SHH signature genes of medulloblastoma in the developing mouse brain (Gibson et al., 2010). These experiments confirmed that SHH signature genes were expressed at the upper rhombic lip at embryonic day (E) 11.5 and within the cerebellum at E15.5. In contrast, WNT signature genes were expressed at the lower rhombic lip at E11.5 and the dorsal brainstem at E15.5. By targeting activating mutations of *CTNNB1* to specific locations, it was determined that, while mutations in the cerebellum had little impact on progenitor cell populations, mutations within the dorsal brainstem caused the abnormal accumulation of cells, which persisted into adulthood and progressed to form medulloblastoma, in conjunction with *Tp53* mutation. Additional evidence came from MRI scans of human WNT and SHH medulloblastoma, where tumours infiltrated the dorsal brainstem and

cerebellum, respectively (Figure 1.17). Finally, 10-15% of mice with an activating mutation in β-catenin and deletion of *Tp53* developed medulloblastoma that recapitulated the molecular features of human WNT-subtype medulloblastoma. The authors then went on to demonstrate that mutant β-catenin prevented the dorsal-ventral migration of progenitor cells that would normally develop into the dorsal brainstem.

In summary, it seems likely that SHH-subtype medulloblastomas arise from mutations within GNP cells of the developing cerebellum, whilst WNT-subtype medulloblastoma arise from the dorsal brainstem; they are distinct diseases that involve different cells of origin (McCarthy, 2011). The cell of origin for the majority of medulloblastomas, that arise independently of activation of the WNT or SHH pathway, is still unclear, although the increasing recognition that medulloblastoma is a heterogeneous disease, with multiple mechanisms of carcinogenesis, underscores the importance of treatment strategies that include consideration of the biological subtype of the tumour.

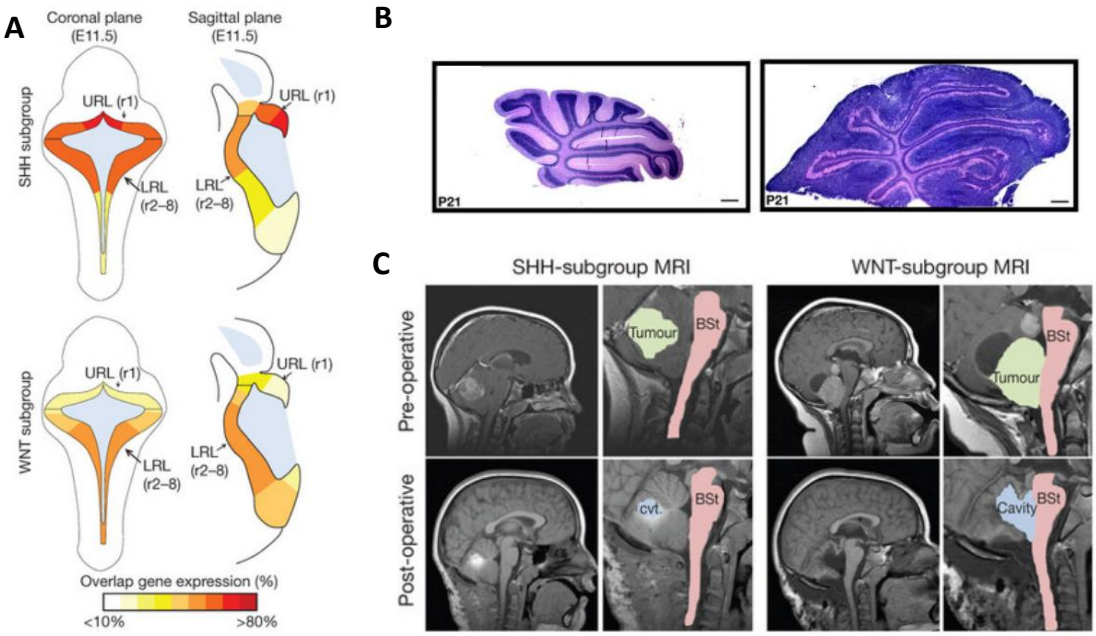


Figure 1.17. The cell of origin for WNT and SHH subtype medulloblastomas. **A**. Expression of WNT and SHH signature genes in mouse embryos at E11.5. SHH subgroup signature genes are expressed most strongly at the upper rhombic lip (URL). WNT signature genes are expressed most strongly at the lower rhombic lip (LRL). **B**. Normal and Math1-Cre / $Ptc^{C/C}$ mouse cerebellum shows at 21 days post-natally that mutant mouse (right hand panel) has severe hyperplasia of the external granular layer, where granule neuron precursors develop. **C**. Human medulloblastomas show different areas of tumour infiltration that is dependent upon subtype. MRI scans show sagittal sections with close-up showing brain stem and tumour. For the SHH medulloblastoma (left), the tumour has arisen within the cerebellum, whereas for the WNT medulloblastoma (right), the tumour has arisen from the dorsal brainstem. Part A and C taken from Gibson et al., 2011. Part B taken from Yang et al., 2008.

1.3.14 Methods for the genome-wide analysis of methylation patterns

Methods for the analysis of DNA methylation patterns have been developed that utilise techniques that rely on a methylation-dependent treatment of the DNA before amplification or hybridisation. There are three main approaches: endonuclease digestion, affinity enrichment and bisulfite conversion (discussed more fully in section 2.6), summarised in Table 1.9.

Pre-Treatment	Analytical Step			
	Locus-specific analysis	Gel-based analysis	Array-based analysis	NGS based analysis
Enzyme Digestion	*HpaII*-PCR (Wahlfors et al., 1992)	RLGS (Hayashizaki et al., 1993)	CHARM (Irizarry et al., 2008)	Methyl-seq (Brunner et al., 2009)
Affinity Enrichment	MeDIP-PCR		MeDIP (Weber et al., 2005) aPRIMES (Pfister et al., 2007)	MeDIP-seq (Down et al., 2008)
Sodium bisulfite		Sanger bisulfite sequencing (Frommer et al., 1992) MSP (Herman et al., 1996)	GoldenGate (Bibikova et al., 2006) Infinium (Bibikova et al., 2009)	WGSBS (Lister et al., 2009)

Table 1.9. Major techniques for DNA methylation analysis. CHARM – comprehensive high-throughput arrays for relative methylation; MeDIP – methylated DNA immunoprecipitation; MSP – methylation sensitive PCR; NGS – next generation sequencing; RLGS – restriction landmark genome scanning. Table adapted from Laird, 2010.

The first techniques for assessing locus specific DNA methylation patterns used methylation-sensitive restriction endonucleases to fragment DNA in a methylation-dependent manner, which could later be identified by gel electrophoresis and / or Southern blotting. However, these techniques are subject to false-positive results, due to incomplete digestion of DNA for reasons other than DNA methylation (Laird, 2010).

Later, restriction landmark genome scanning (RLGS) was the first truly genome-wide technique for assessing DNA methylation patterns (Hayashizaki et al., 1993).

Differences in DNA methylation are detected as difference in the pattern of restriction fragments generated by digestion with a methylation-sensitive restriction enzyme, separated by two-dimensional gel electrophoresis. This technique has been widely used to identify imprinted loci and sites that are methylated in a cancer- or tissue-specific manner (Costello et al., 2000; Plass et al., 1996). The use of RLGS and similar techniques has declined, as more reliable methods, that are less labour intensive, have become available ((Laird, 2010)). An alternative approach is differential methylation hybridisation (Huang et al., 1999), which involves digestion of one pool of genomic DNA with a methylation-sensitive restriction enzyme and mock digestion of a second pool. The two pools are amplified and fluorescently labelled, prior to a competitive, two-colour array hybridisation. The relative intensities of fluorescent signal from the digested and mock-digested pools can be used to infer methylation status at specific loci. An optimised workflow, which uses the methylation-dependent endonuclease, *Mcr*BC, to provide a greater sensitivity to densely methylated regions, followed by array hybridisation, is known as CHARM (comprehensive high-throughput arrays for relative methylation (Irizarry et al., 2008)). The techniques outlined above, which rely on restriction-enzyme mediated enrichment of methylated sequences, are being adapted to replace the array hybridisation step with next generation sequencing (Brunner et al., 2009).

The affinity enrichment of methylated regions of DNA using specific antibodies for methylated-cytosines has proved to be useful for the identification of methylated DNA. One example of this type of technique, MeDIP (methylated DNA immunoprecipitation) (Weber et al., 2007; Keshet et al., 2006; Weber et al., 2005), involves hybridisation to a capture array following affinity-enrichment for methylated DNA. MeDIP-seq, which applies next generation sequencing to affinity-enriched methylated DNA, replacing the array analysis step, has recently been described (Down et al., 2008).

The discovery that sodium bisulfite treatment deaminated unmethylated cytosine residues to uracil, while sparing methylated cytosine residues (section 2.6) prompted the introduction of many techniques that utilise this primary sequence change, which gives a reliable readout of methylation status. Low throughput techniques, such as bisulfite Sanger sequencing (section 2.6.1) (Frommer et al., 1992), quickly became the ‘gold standard’ against which other techniques are compared, since the change in DNA

sequence is unambiguous and is, by definition, locus specific. Later, methylation-sensitive polymerase chain reaction (MS-PCR) (section 2.6.2) (Herman et al., 1996) was reported for the rapid assay of locus-specific methylation status. The reduced sequence complexity of bisulfite-treated DNA (three bases instead of four), has resulted in decreased hybridisation specificity, which has hindered the introduction of array-based assay of bisulfite-treated DNA; however, through careful probe design, Illumina have released three generations of methylation arrays, that assay the methylation status of specific CpG dinucleotides (section 2.7) (Sandoval et al., 2011; Bibikova et al., 2009; Bibikova et al., 2006). The arrays are reported to be robust, reproducible and, importantly, are suitable for use with DNA extracted from formalin-fixed, paraffin embedded (FFPE) tissue (Thirlwell et al., 2010).

It is likely that all these techniques will eventually be superseded by whole genome shotgun bisulfite sequencing (WGSBS), which apply next generation sequencing to bisulfite treated DNA (Lister et al., 2009), although approximately one tenth of the CpG dinucleotides in the mammalian genome are refractory to analysis by this means (Laird, 2010). Nanopore-based sequencing, which can directly read unmodified methylated cytosine residues at single molecule resolution, remains in development, but holds rich promise for true single cell measurements of the DNA methylome in future (Clarke et al., 2009).

This project utilised Illumina's Golden Gate Cancer Panel I methylation microarrays (section 2.7) to characterise the DNA methylome of medulloblastoma. These arrays were chosen due to their low sample requirements relative to the other techniques outlined above (less than 1 µg), their low cost, their suitability for use with FFPE tissues and their high reproducibility.

1.3.15 Methylomic analysis of medulloblastoma

Candidate-gene based approaches for the investigation of DNA methylation mediated silencing of gene expression in medulloblastoma (section 1.3.7.3.2) have been superseded by techniques that enable global assays of DNA methylation status.

RLGS has been used to indicate the levels of global methylation in medulloblastoma, and refinements in techniques allowed the identification of a number of DNA

fragments which are differentially methylated in tumour tissues and the normal cerebellum (Smiraglia et al., 1999). Fruhwald and colleagues identified that 1% of CpG islands were methylated in primary medulloblastomas and identified DNA sequences for which hypermethylation correlated with a poorer prognosis (Fruhwald et al., 2001).

Subsequently, cell line experiments compared mRNA expression before and after treatment with the DNA demethylating agent, 5-aza,2'-deoxycytidine, so that expression under the control of DNA methylation could be identified. Using this approach, *COL1A2* expression was shown to be re-activated following 5-aza,2'-deoxycytidine treatment. Further investigations in primary tumours revealed that dense biallelic methylation associated with transcriptional silencing was observed in the majority of cases tested. In addition, promoter hypomethylation of the *COL1A2* gene was strongly associated with infant desmoplastic cases, illustrating the utility of methylation status as biomarkers for disease phenotypes (Anderton et al., 2008). A similar approach identified an inhibitor of HGF / MET signalling, serine protease inhibitor kunitz-type *2* (*SPINT2*), as a putative tumour suppressor silenced by promoter methylation in medulloblastoma. One third of primary medulloblastomas tested showed evidence for promoter methylation of *SPINT2.* Furthermore, re-expression of *SPINT2* in cell lines, reduced proliferative capacity *in vitro* and increased survival in a xenograft model *in vivo* (Kongkham et al., 2008). Subsequently, this technique also identified the methylation dependent silencing of the *SFRP* family of genes in a minority of primary medulloblastomas, which may release WNT signalling from inhibition, contributing towards excessive WNT signalling in the disease (Kongkham et al., 2010).

A different approach was taken by Pfister and colleagues, who described a method called array-based profiling of reference-independent methylation status (aPRIMES), which enabled the assessment of direct methylation status (Pfister et al., 2007). The method is based on the differential restriction and competitive hybridisation of methylated and unmethylated DNA by methylation-specific and methylation-sensitive restriction enzymes, respectively. The ratio of methylated to unmethylated DNA is determined by the intensity of signal after competitive hybridisation to a microarray containing probes present within CpG islands. Using this approach, *ZIC2* was identified as a frequently methylated gene in paediatric medulloblastomas. This approach is

limited by the demanding technical requirements (primarily, generation of microarrays from CpG island enriched libraries) and the requirement for 100% processivity for the methylation-specific and methylation-sensitive restriction enzymes. Moreover, the assay by definition is limited to only assessing the methylation status of CpG islands, so that genome-wide estimates of methylation are not possible.

Other techniques that utilise bisulfite treatment (section 2.6) to directly assess genome-wide methylation patterns have also been reported, but until now, have not been applied to medulloblastoma. A novel high-throughput technology that directly measures DNA methylation on bisulfite treated DNA has been developed by Illumina (Bibikova et al., 2006). This technology uses an existing bead based microarray platform to assess methylation status using ~1500 probes that map to ~800 cancer related genes (section 2.7), and provided the platform for the investigations reported in chapters 4 and 5. The later introduction of the Infinium and Infinium HD arrays (27,000 and 450,000 CpG sites respectively) (Sandoval et al., 2011; Bibikova et al., 2009) utilise similar technology to provide a truly genome-wide characterisation of the methylome.

At present, the utility of DNA methylation patterns to subclassify medulloblastoma has not been reported. The work reported in chapters 4 and 5 uses a DNA methylation array-based approach to identify whether DNA methylation has utility for disease subclassification and prognostication.

1.3.16 Molecular markers of prognosis in medulloblastoma

The relationship between survival and prognostic molecular markers in medulloblastoma has been reported in more than 200 articles over the last 30 years (Gilbertson, 2011). Despite this, few markers have been validated in ≥2 trials cohorts (Table 1.6), although the prognostic ability of activation of the WNT pathway (assessed by nuclear accumulation of β-catenin) and the poor risk conferred by amplification of the *MYC* oncogenes are to be tested in the upcoming PNET5 and PNET 6 clinical trials (Pizer and Clifford, 2009). The realisation that medulloblastoma comprises a heterogeneous group of diseases with discrete cells of origin and underlying abnormalities may enable a more robust identification of molecular markers that define subtype-specific survival differences.

This is illustrated by conflicting reports concerning the prognostic effect of mutations in *TP53.* An initial report by Tabori and colleagues reported a universally poor survival in children with somatic *TP53* mutation (Tabori et al., 2010). Subsequent investigations revealed that the survival effects of *TP53* mutation were dependent upon the disease context in which they occur; *TP53* mutation frequently occurs alongside activation of the WNT pathway, and is therefore compatible with a favourable prognosis (Lindsey et al., 2011; Pfaff et al., 2010); on the other hand, it also frequently occurred alongside amplification of *MYCN*, a proposed marker of poor prognosis (Pfister et al., 2009), but was not prognostic overall (Pfaff et al., 2010), demonstrating that an understanding of the background molecular heterogeneity is critically important to the investigation of the prognostic value of *TP53*.

Despite this pessimistic view of molecular prognostication in medulloblastoma, there are a handful of examples, validated in multiple studies, which demonstrate that real progress has been made. The improved survival observed in medulloblastomas with an activated WNT signalling pathway has been validated across multiple studies (Fattet et al., 2009; Clifford et al., 2006; Gajjar et al., 2006; Ellison et al., 2005). Likewise, the poor prognosis associated with amplification of *MYCC* has been replicated in multiple studies (Ellison et al., 2011b; Pfister et al., 2009; Eberhart et al., 2004; Lamont et al., 2004). Finally, a recent report from Remke and colleagues focussed on disease prognostication within the WNT / SHH independent subgroups of medulloblastoma (Remke et al., 2011a). Using transcriptomic approaches, they identified four disease subgroups in 64 samples. They noted that *FSTL5* was prognostic across all disease subgroups, with highest expression identified in Northcott group C tumours (see section 1.3.12.1), which were associated with a poor prognosis in their own and previously reported cohorts (Northcott et al., 2010). Importantly, the expression of *FSTL5* in Northcott group D tumours was highly variable, which led the authors to propose that high *FSTL5* expression might identify Northcott group D tumours with a poor prognosis. Next, they tested the prognostic ability of immunostaining for *FSTL5* in a larger, independent cohort, and confirmed that *FSTL5* was prognostic across the whole cohort in univariate and multivariate survival models, and also was prognostic within the WNT / SHH independent Northcott subgroup C and D tumours. Thus, this validated biomarker holds promise as a marker of poor prognosis, with particular value

for identifying high risk Northcott group D or low risk Northcott group C patients, for more effective risk stratification in future clinical trials.

While few reported prognostic molecular markers of medulloblastoma have been robustly validated and progressed into trials, there is an increasing understanding of the absolute importance of verifying the reported prognostic marker in an independent (preferably clinical trials) cohort. Additionally, the survival implications for membership of the non-WNT molecular subgroups are unclear, and the behaviour of currently accepted clinical markers is also unclear, requiring further investigation in trials cohorts. Since the large majority of historical trials cohorts are comprised of tissues that are FFPE, any reported marker should be amenable to testing in such materials, either through standard immunohistochemical techniques, or through techniques that are insensitive to the alterations to nucleic acids that occur as part of the fixing process.

1.4 Summary and aims

Medulloblastoma is the commonest malignant brain tumour of childhood, comprising 20% of all CNS cancers. While 5 year survival rates have reached ~ 80% using current risk-adapted combination therapies, there remain unmet needs for an understanding of how to successfully treat the remaining 20% of patients who relapse, and for the introduction of novel risk stratification schemes that can identify patients with a good prognosis, for whom less aggressive treatments can be administered, reducing the risk of adverse treatment *sequelae*, whilst maintaining PFS. Likewise, if patients with a particularly poor prognosis could be identified, novel combinations of intensive therapies coupled with suitable adjuvant therapies could be applied that maximise their chance of survival.

In recent years, considerable efforts have been applied to molecular subclassification of the disease. Transcriptomic classification of the disease has provided evidence for between 4 to 6 disease subgroups (section 1.3.12.1), with two subgroups, respectively defined by activation of the SHH and WNT pathways, common to all studies. While means for the identification of WNT subgroup cases are available, at the commencement of this project, there were no assays for the identification of SHH subgroup medulloblastomas, and a clear need for a simple, validated assay to identify WNT and SHH subgroup medulloblastomas. The relationships between non-WNT subgroup membership and molecular and clinico-pathological disease correlates are not well defined and have previously been investigated in cohorts of limited size; a meta-analysis of transcriptomic datasets, using data from multiple studies, has not previously been undertaken.

The recent availability of DNA methylation microarrays enables the investigation of any differential patterns of DNA methylation within medulloblastomas, for the identification of discrete methylomic disease subgroups, and has not previously been attempted. This may help to bring clarity to the transcriptomic subclassification of the disease; any agreement between the number of medulloblastoma subgroups, estimated using transcriptomic and methylomic means, could help reach consensus. Additionally, the use of DNA methylation arrays offers a means for the subclassification

of medulloblastoma in historical, FFPE-derived, trials cohorts that would not be possible using transcriptomic techniques.

Finally, the relationship of disease subgroup membership to survival is not fully understood. While the good prognosis conferred by membership of the WNT subgroup has been widely reported (section 1.3.10.2), the prognostic implications of membership of the non-WNT transcriptomic subgroups remain unclear and bear further scrutiny (Cho et al., 2011; Ellison et al., 2011b; Northcott et al., 2010).

In future, for an improved understanding of the disease, it will become necessary to consider medulloblastoma as an umbrella term for a mixed group of heterogeneous cancers, with different disease determinants and clinical behaviours; the behaviour and incidence of previously reported clinico-pathological, molecular and prognostic markers in medulloblastoma needs to be assessed in the context of molecular subgroup membership; it is also possible that some correlates and markers will remain prognostic independent of subgroup ('universal' markers of poor risk in medulloblastoma).

This project aimed to investigate novel methods for the classification of medulloblastoma, and to investigate the utility of these methods for improved disease prognostication, with the following specific aims.

1. To identify minimal gene expression signatures for the identification of WNT and SHH pathway activation in medulloblastoma, to apply those signatures to a test cohort and publically available transcriptomic datasets, and to undertake a meta-analysis of clinical and molecular correlates of medulloblastomas classified by their signalling pathway activation status (chapter 3).

2. Using a novel DNA methylome profiling technology, to investigate whether medulloblastomas display differential patterns of DNA methylation and any ability of such patterns for disease subclassification. If any such patterns were identified, to investigate any relationship to WNT and SHH pathway activation, relationship to molecular and clinico-pathological correlates and survival (chapter 4).

3. To undertake a comprehensive survival analysis of medulloblastoma; to investigate whether previously identified prognostic markers in medulloblastoma display pleiotropic behaviours across disease subgroups and the utility of methylated loci as biomarkers for prognostication. Finally, to investigate the integration of any proposed prognostic methylated loci into improved risk stratification schemes for medulloblastoma (chapter 5).

Chapter 2. Materials and methods

2.1 Cohorts

The cohorts of primary medulloblastomas used for the investigations reported in chapter 3 consisted of an initial 55 member medulloblastoma cohort, whose characteristics are described below.

A representative cohort of 55 medulloblastomas was analysed, detailed in Table 2.1. The cohort comprised of 33 (60%) NMB (medulloblastoma cases collected in Newcastle and 22 (40%) RJG (medulloblastoma cases supplied by Dr. Richard Gilbertson (Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, TN, USA)). Histological subtypes comprised 39 classic (71%), 5 LCA (9%) and 11 DN tumours (20%). The cohort included medulloblastomas regardless of age at diagnosis. There were 11 (20%) infant cases (aged under 3 at diagnosis), 41 (75%) children (>3-15 years) and 3 (5%) adults (≥16 years). The cohort contained 21 (38%) female and 34 (62%) male cases, with a male: female ratio of 1.6:1. M stage classification according to Chang's criteria (Chang et al., 1969) was available for most cases. Since the nature of the reported M stage varied (some cases reported M0 / M1 as a single category), M staging was simplified into two categories; M- cases were defined as being M0, M1 or M0 / 1. M+ status was assigned to cases reported as M2 or M3 (no M4 cases were observed in any cohort studied). The majority of cases for which there was available information were M- (36 / 44 (82%)), with 8 M+ cases (18%).

Study ID	Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	*PTCH1* mutation
1	RJG 112	0.7	F	DN	0	0	0
2	RJG113	2.5	M	LCA	1	0	0
3	RJG114	2.9	M	CLAS	0	0	0
4	RJG115	4.6	M	CLAS	0	NA	NA
5	RJG116	19.0	M	DN	0	0	0
6	RJG 117	12.8	M	CLAS	1	0	0
7	RJG118	2.6	F	CLAS	1	0	0
8	RJG 120	3.6	M	CLAS	NA	0	0
9	RJG121	4.1	M	LCA	0	0	0
10	RJG122	2.5	M	LCA	0	0	0
11	RJG123	15.4	M	CLAS	0	0	0
12	RJG124	6.3	M	DN	1	0	0
13	RJG126	2.6	M	DN	1	0	0
14	RJG127	4.8	F	DN	0	0	1
15	RJG 129	11.5	M	CLAS	NA	0	0
16	RJG131	5.7	F	DN	1	0	0
17	RJG133	16.7	M	CLAS	0	0	1
18	RJG 135	3.6	F	CLAS	0	0	0
19	RJG136	2.9	M	DN	0	0	1
20	RJG140	12.5	M	DN	0	0	0
21	RJG141	9.8	F	CLAS	0	0	0
22	RJG142	1.0	F	DN	0	0	1
23	NMB 20	6.6	M	CLAS	NA	0	0
24	NMB43	10.0	M	CLAS	0	0	0
25	NMB45	12.6	M	CLAS	0	0	0
26	NMB46	5.1	M	CLAS	0	0	0
27	NMB51	6.8	M	CLAS	0	0	0
28	NMB52	8.6	F	CLAS	0	0	0
29	NMB 59	6.6	M	LCA	0	0	0
30	NMB60	5.0	M	CLAS	0	0	0
31	NMB61	10.3	M	CLAS	0	1	0
32	NMB63	11.5	M	CLAS	0	0	0
33	NMB 64	1.4	F	DN	NA	0	0
34	NMB 65	9.3	M	CLAS	NA	0	0
35	NMB 66	4.9	M	CLAS	NA	0	0
36	NMB 68	6.9	F	CLAS	NA	0	0
37	NMB69	7.9	M	CLAS	0	0	1
38	NMB70	3.3	F	CLAS	0	0	0
39	NMB 71	15.8	M	CLAS	NA	0	0
40	NMB76	7.5	M	CLAS	0	0	0

Study ID	Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	*PTCH1* mutation
41	NMB77	8.5	F	CLAS	0	0	0
42	NMB78	5.5	M	CLAS	1	0	0
43	NMB79	3.5	F	CLAS	1	0	1
44	NMB80	10.2	F	CLAS	0	0	0
45	NMB81	14.2	F	DN	0	0	0
46	NMB82	5.4	M	CLAS	0	0	0
47	NMB 83	10.5	M	CLAS	NA	0	0
48	NMB 84	15.1	F	CLAS	NA	0	0
49	NMB 87	1.6	M	CLAS	NA	0	0
50	NMB88	17.0	F	CLAS	0	0	0
51	NMB89	4.6	F	CLAS	0	0	0
52	NMB90	3.0	F	CLAS	0	0	0
53	NMB92	4.3	M	CLAS	0	0	0
54	NMB93	10.0	M	CLAS	0	1	0
55	NMB94	9.0	F	CLAS	0	1	0

Table 2.1. Clinical demographics of the cohort used for chapter 3. The cohort consisted of 55 primary medulloblastomas obtained from across the UK and Europe (*n* = 33), and from Dr. Richard Gilbertson (*n* = 22). Study ID, sample ID, age in years at diagnosis, and gender are shown. Histological subtype is coded as follows: CLAS – classic, DN – desmoplastic / nodular, LCA – large cell / anaplastic. M stage is coded based on Chang's criteria (see section 1.3.3). 0 – M- (M stage 0 / 1), 1 – M + (M stage 2 / 3). *CTNNB1* mutation: 0 – no mutation detected; 1 – mutation detected. Missing data are indicated with NA.

The training and test cohorts that formed the basis of chapters 4 and 5 are common to both and are also described below.

The training cohort initially consisted of 108 medulloblastomas and is listed in Table 2.2. This cohort comprised 86 NMB cases, 15 RJG and 7 PNET 3 (medulloblastoma cases from the PNET3 (Primitive-neuroectodermal tumour 3) clinical trials cohort, detailed below) cases. Histological subtypes (see below) comprised 80 classic type (74%), 17 desmoplastic / nodular (16%), 10 LCA (9%) and 1 MBEN case (1%). The cohort included medulloblastomas regardless of age at diagnosis. The median age at diagnosis was 7.6 years. There were 18 (17%) infant cases (aged under 3 at diagnosis) and 7 (6.5%) adult cases (aged over 16 at diagnosis). The cohort contained 67 male

cases (62%) and 41 female cases (38%), with a male: female ratio of 1.6:1. M stage classification according to Chang's criteria (Chang et al., 1969) was available for 96 cases (89%). The majority of cases for which there was available information were M- (77 / 96 (80%)), with 19 M+ cases (20%). Survival data was available for the majority of cases (95 cases (88%)), with a median follow up time of 4.8 years.

The test cohort consisted of 143 primary medulloblastoma tumours from patients entered onto the SIOP / UKCCSG PNET3 clinical trial (Taylor et al., 2003), hereafter referred to as PNET3. The trial took place from March 1992 to January 2000. Patients defined as average risk, aged from 3 to 16 years, were recruited from across Europe and placed into one of two treatment arms (radiotherapy alone *versus* application of chemotherapy prior to radiotherapy). The study found that there was a significant difference in 3 and 5 year EFS between treatment arms that was not repeated when overall survival (OS) was examined. The cohort details are listed in Table 2.3. The test cohort comprised 122 classic type (85%), 10 DN (7%) and 11 LCA (8%) medulloblastomas. The median age at diagnosis was 8.6 years. The cohort contained 88 male cases (62%) and 55 female cases (38%), with a male: female ratio of 1.6:1. M stage classification was available for 142 / 143 cases (99%), with 114 M- cases (80%) and 28 M+ cases (20%). Survival data were available for every case, with a median follow up time of 9.8 years.

For both cohorts, histological variants were confirmed on review by Professor David Ellison (Chair of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA) according to the 2007 WHO classification criteria (Louis et al., 2007). Additional clinico-pathological correlate data were obtained from a variety of sources. M stage, patient age and survival data were kindly provided by Dr. Simon Bailey (Newcastle University Brain Tumour Research Group). The molecular correlates were kindly provided by other Newcastle University Brain Tumour Research Group members: Chromosome 6 and 17 heterozygosity was assayed by Dr. Meryl Lusher and Dr. Hisham Megahed respectively, using the HOMOD (homozygosity mapping of deletions) methodology (Goldberg et al., 2000), a technique that can identify LOH in tumour samples without the necessity for matched constitutional DNA. Chromosome 6 and 17 heterozygosity statuses were determined by analysis of six polymorphic microsatellite markers spanning each chromosome using previously described methods (Langdon et al., 2006).

Activating mutations of *CTNNB1* were identified by Dr. Meryl Lusher. A mutational analysis of *CTNNB1* was performed by direct sequencing using previously reported primers and methods (Ellison et al., 2005). Data on amplification of the *MYC* family genes *MYCC* and *MYCN* were supplied by Dr. Sarra Ryan, who measured gene amplification using a real-time PCR-based technique. Samples where the haploid copy number for either gene was greater than 5 were classed as amplified.

Information on WNT and SHH pathway activation came from two sources. Firstly, the GeXP assay for the transcriptomic assignment of pathway activation described in sections 2.4 and 4.3.8 was used to assign pathway activation in cases extracted from fresh frozen tissue for which mRNA was available. Secondly, a recent study by Ellison et al. described antibody markers of the WNT and SHH molecular medulloblastoma subgroups (Ellison et al., 2011a). In their study, immunoreactivity for filamin A and YAP1 identified SHH and WNT tumours, whilst GAB1 immunoreactivity characterised SHH medulloblastomas and nuclear immunoreactivity for β-catenin identified WNT tumours (Figure 2.1). Since their study was based on a cohort comprising PNET3 trials cases, there was significant overlap between the subgroup assignments made in Ellison et al's study and the sample comprising the test cohort in this study (data were available for 115 / 143 (80%) test cohort samples and for 5 / 108 (5%) training cohort samples. The clinico-pathological and molecular correlates are shown for the training and test cohorts in Table 2.2 and Table 2.3 respectively. The summarised cohort demographics, after the removal of samples failing quality control, are described in Table 4.3. A summary of the cohorts used in this study, both in primary investigations and for assay development and validation, are listed in Table 2.4.

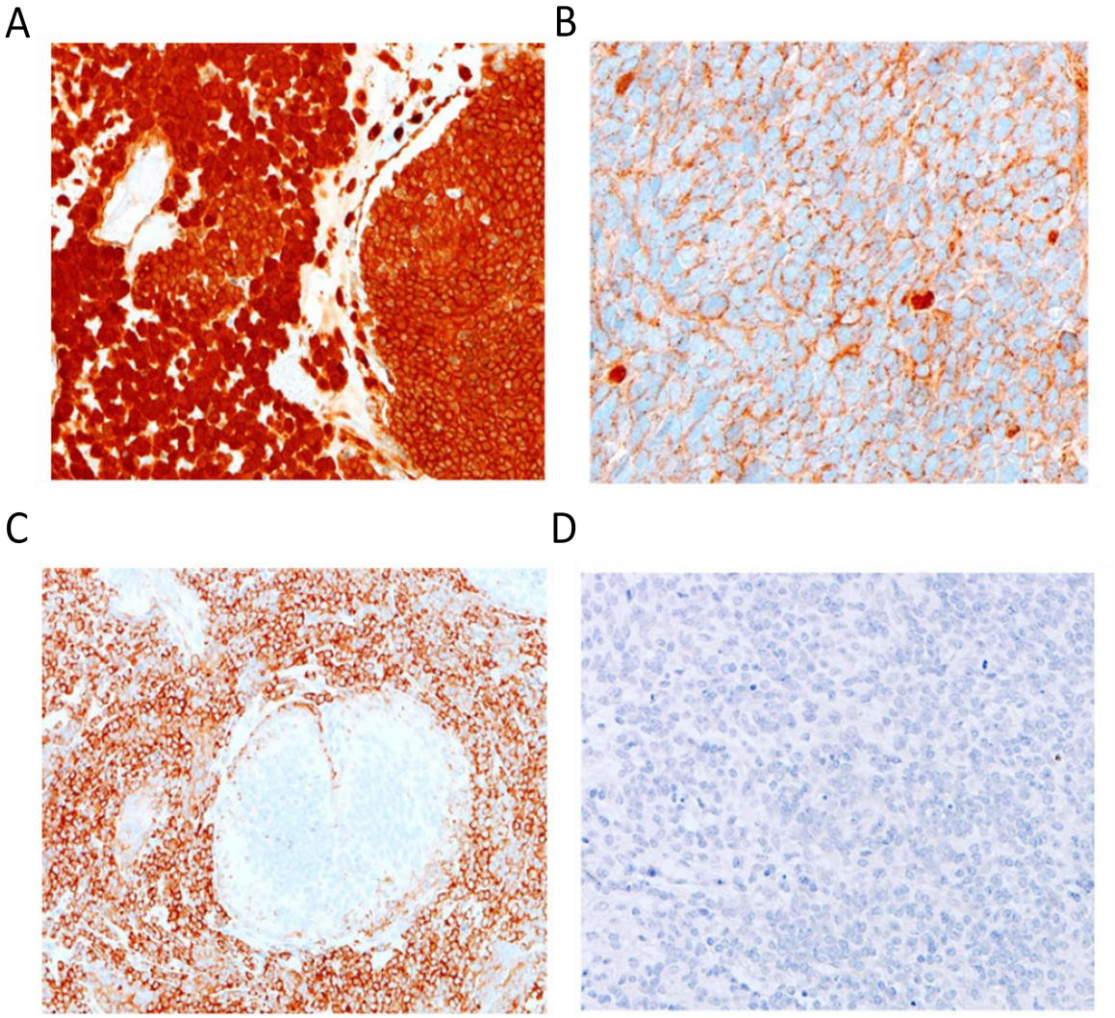


Figure 2.1. Immunohistochemistry for CTNNB1 and GAB1 antibodies using specific antibodies respectively identifies WNT and SHH subgroups of medulloblastoma. Each panel shows a representative medulloblastoma section with H&E counterstaining. Panel A and B show samples positive and negative for nuclear accumulation of β-catenin, a marker of WNT pathway activation in medulloblastoma. The occasional cell with nuclear accumulation in the negative sample is not sufficient to ascribe nuclear accumulation positivity. Panel C shows a sample positive for expression of GAB1 in a DN medulloblastoma. Nodules show lack of staining; desmoplastic regions show dense staining. Panel D shows a case with absence of GAB1 staining, indicating an absence of SHH pathway activation. Figure adapted from Ellison DW et al., 2011a.

Sample ID	Age	Sex	Tumour Type	Centre	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	EFS Status	EFS Time (years)
NMB108	4.6	F	CLAS	BELG	1	0	NA	NA	0	0	1	NA	0	4.6
NMB109	8.0	M	CLAS	BELG	0	0	0	NA	0	0	0	NA	0	2.3
NMB110	10.0	F	CLAS	BELG	0	0	0	NA	0	0	0	NA	0	1.6
NMB111	10.0	M	CLAS	BELG	0	0	0	NA	0	0	NA	NA	0	2.5
NMB112	43.0	F	CLAS	BELG	0	0	0	NA	0	0	0	NA	0	2.3
NMB113	3.0	M	CLAS	BELG	0	0	NA	NA	0	0		NA	0	4.1
NMB125	6.0	F	CLAS	BUD	NA	0	0	NA	0	0	0	NA	0	3.0
NMB126	8.4	F	CLAS	BUD	NA	0	0	NA	0	0	0	NA	NA	NA
NMB128	4.5	M	CLAS	BUD	NA	0	0	NA	0	1	0	NA	NA	NA
NMB129	5.7	M	LCA	BUD	NA	0	0	NA	0	0	0	NA	NA	NA
NMB131	10.3	M	CLAS	BHAM	0	1	1	NA	0	0	2	NA	0	7.4
NMB133	6.9	F	CLAS	BHAM	0	0	0	NA	0	0	NA	NA	0	4.8
NMB134	4.9	M	CLAS	BHAM	1	0	0	NA	0	1	0	NA	NA	NA
NMB135	11.2	F	CLAS	BHAM	0	1	1	NA	0	0	2	NA	0	4.4
NMB136	10.5	M	CLAS	BHAM	1	0	0	NA	0	0	0	NA	NA	NA
NMB137	15.1	F	CLAS	BHAM	1	NA	0	NA	0	0	0	NA	1	1.2
NMB138	3.3	M	CLAS	BHAM	0	NA	0	NA	0	0	1	NA	1	0.8
NMB139	12.7	M	CLAS	BHAM	0	1	1	NA	0	0	2	NA	0	2.8
NMB140	9.4	M	CLAS	BHAM	0	0	0	NA	NA	NA	0	NA	0	2.7
NMB141	11.7	F	CLAS	BHAM	0	0	0	NA	0	0	1	NA	0	2.2
NMB142	12.1	M	CLAS	BHAM	1	0	0	NA	0	0	0	NA	0	2.3

Sample ID	Age	Sex	Tumour Type	Centre	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	EFS Status	EFS Time (years)
NMB143	2.8	M	LCA	BHAM	0	0	0	NA	0	0	0	NA	1	0.8
NMB144	5.1	M	CLAS	BHAM	1	0	0	NA	0	0	0	NA	1	2.0
NMB147	19.6	M	CLAS	SOTON	0	0	0	NA	0	0	0	NA	NA	NA
NMB148	1.6	M	CLAS	SOTON	0	0	0	NA	0	0	1	NA	1	0.3
NMB149	6.3	M	LCA	SOTON	NA	0	0	NA	0	0	0	NA	1	1.3
NMB151	9.8	M	CLAS	SOTON	1	0	0	NA	0	0	0	NA	1	1.6
NMB152	0.5	M	CLAS	SOTON	0	0	0	NA	0	0	0	NA	1	5.1
NMB153	3.3	M	CLAS	SOTON	NA	0	0	NA	0	0	0	NA	0	13.2
NMB154	4.7	M	CLAS	SOTON	NA	0	0	NA	0	0	0	NA	NA	NA
NMB155	19.8	M	CLAS	SOTON	NA	0	0	NA	0	0	0	NA	NA	NA
NMB156	7.3	M	DN	SOTON	0	0	0	NA	0	0	0	NA	0	13.1
NMB157	5.4	M	DN	SOTON	1	0	0	NA	0	0	0	NA	0	10.6
NMB159	22.8	M	DN	NCL	NA	0	0	NA	0	0	1	NA	NA	NA
NMB16	0.1	M	CLAS	NCL	0	0	0	NA	NA	NA	0	NA	1	0.0
NMB162	5.1	M	CLAS	NCL	0	0	0	NA	0	0	0	NA	0	7.5
NMB164	7.1	F	LCA	NCL	0	0	0	NA	0	0	0	NA	0	2.6
NMB165	12.7	M	CLAS	NCL	0	0	0	NA	NA	NA	0	NA	0	4.3
NMB166	9.7	F	CLAS	NCL	0	0	0	NA	1	0	0	NA	0	5.6
NMB167	6.5	F	CLAS	NCL	0	0	0	NA	0	0	0	NA	0	4.0
NMB168	9.8	F	LCA	NCL	0	0	0	NA	0	0	0	NA	1	1.8
NMB169	8.9	M	LCA	NCL	0	0	0	NA	1	0	0	NA	1	0.2

Sample ID	Age	Sex	Tumour Type	Centre	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	EFS Status	EFS Time (years)
NMB17	4.0	M	CLAS	NCL	0	0	0	NA	NA	NA	0	NA	0	5.7
NMB171	7.7	M	CLAS	NOTT	NA	0	0	NA	0	0	0	NA	NA	NA
NMB180	10.7	M	CLAS	LEEDS	0	0	0	NA	0	0	0	NA	0	5.1
NMB181	8.4	M	DN	LEEDS	0	0	0	NA	0	1	1	NA	1	0.1
NMB182	9.6	M	CLAS	LEEDS	0	0	0	NA	0	0	0	NA	0	4.6
NMB184	8.6	M	CLAS	LEEDS	0	0	0	NA	0	0	0	NA	0	3.6
NMB185	14.0	M	CLAS	CARD	1	0	0	NA	0	0	0	NA	0	4.7
NMB186	8.0	M	CLAS	CARD	0	0	0	NA	0	0	0	NA	1	3.1
NMB187	3.7	M	CLAS	NOTT	0	0	0	NA	0	0	0	NA	0	4.8
NMB188	8.6	F	CLAS	NOTT	NA	0	0	NA	0	0	0	NA	0	0.2
NMB189	8.6	M	CLAS	NOTT	0	0	0	NA	NA	NA	0	NA	0	3.6
NMB190	11.7	M	CLAS	NOTT	1	0	0	NA	NA	NA	0	NA	1	6.6
NMB199	9.8	M	CLAS	NCL	1	0	0	NA	NA	NA	0	NA	1	1.3
NMB200	1.3	F	DN	NCL	0	0	0	NA	0	0	1	NA	0	7.0
NMB202	40.0	F	DN	NCL	0	0	0	NA	0	0	1	NA	0	5.9
NMB203	6.2	M	CLAS	NCL	0	0	0	NA	NA	NA	0	NA	0	4.2
NMB227	4.0	M	CLAS	NCL	1	NA	0	NA	NA	NA	0	NA	1	1.3
NMB250	4.8	M	CLAS	NCL	0	0	0	NA	NA	NA	0	NA	0	2.0
NMB251	3.3	F	CLAS	NCL	NA	0	NA	NA	NA	NA	NA	NA	NA	NA
NMB252	13.1	F	CLAS	NCL	0	1	1	NA	NA	NA	2	NA	0	1.1
NMB253	0.4	M	DN	POL	0	0	0	NA	NA	NA	1	NA	0	1.2

Sample ID	Age	Sex	Tumour Type	Centre	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	EFS Status	EFS Time (years)
NMB254	1.5	F	DN	POL	0	0	0	NA	NA	NA	1	NA	0	0.9
NMB33	1.5	F	DN	NCL	0	0	0	NA	NA	NA	1	NA	NA	NA
NMB43	9.0	M	CLAS	NCL	0	0	0	1	0	0	0	NA	1	4.2
NMB45	12.6	M	CLAS	NCL	0	0	0	1	0	0	0	NA	0	8.4
NMB51	6.8	M	CLAS	NCL	0	0	0	0	0	0	0	NA	0	8.0
NMB52	8.6	F	CLAS	NCL	0	0	0	1	0	0	0	NA	0	7.2
NMB60	5.0	M	CLAS	BHAM	0	0	0	0	NA	NA	0	NA	1	1.0
NMB63	11.5	M	CLAS	NCL	0	0	0	0	0	0	1	NA	0	7.4
NMB64	1.5	F	MBEN	BHAM	0	0	0	NA	0	0	NA	NA	1	1.6
NMB65	9.3	M	CLAS	BHAM	0	0	0	NA	0	0	NA	NA	0	5.5
NMB69B	7.8	M	CLAS	CARD	0	0	NA	NA	NA	NA	0	NA	0	8.7
NMB76	7.5	M	CLAS	BHAM	0	0	0	0	0	0	0	NA	1	2.4
NMB77	8.5	F	CLAS	BHAM	0	0	0	1	0	0	0	NA	0	6.4
NMB78	5.5	M	CLAS	BHAM	1	0	0	0	0	0	0	NA	1	0.3
NMB79	3.5	F	CLAS	CAM	1	0	0	0	0	0	1	NA	0	6.5
NMB80	10.2	F	CLAS	CAM	0	0	0	0	0	0	0	NA	0	1.9
NMB81	14.2	F	DN	CAM	0	0	0	0	0	0	1	NA	0	3.8
NMB82	5.4	M	CLAS	NCL	0	0	0	1	1	0	0	NA	0	4.8
NMB88	17.0	F	CLAS	BRI	0	0	0	1	0	0	0	NA	0	2.8
NMB89	4.6	F	CLAS	BRI	0	0	0	1	0	1	0	NA	0	7.5
NMB90	3.0	F	CLAS	BRI	0	0	0	0	0	0	0	NA	0	10.4

Sample ID	Age	Sex	Tumour Type	Centre	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	EFS Status	EFS Time (years)
NMB93	10.0	M	CLAS	BRI	0	1	1	0	0	0	2	NA	0	1.5
NMB94	9.0	F	CLAS	BRI	0	1	1	0	0	0	2	NA	0	5.0
PNET30119	15.8	M	CLAS	PNET3	0	1	1	0	NA	NA	NA	2	0	5.3
PNET30131	5.4	F	LCA	PNET3	0	1	0	1	NA	NA	NA	2	0	5.0
PNET30139	7.6	F	CLAS	PNET3	0	1	1	0	NA	NA	NA	2	1	1.8
PNET30146	9.1	F	CLAS	PNET3	0	0	1	NA	NA	NA	NA	NA	1	2.8
PNET30147	10.8	F	CLAS	PNET3	0	1	1	0	NA	NA	NA	2	0	12.3
PNET30209	10.4	M	CLAS	PNET3	0	0	NA	NA	0	0	NA	NA	0	7.5
PNET350129	9.9	M	CLAS	PNET3	0	0	1	0	NA	NA	NA	2	0	11.2
RJG112	0.7	F	DN	RJG	0	0	0	NA	0	0	NA	NA	1	1.3
RJG113	2.5	M	LCA	RJG	1	0	0	0	NA	NA	0	NA	0	15.5
RJG114	2.9	M	CLAS	RJG	0	0	0	NA	0	0	NA	NA	1	1.1
RJG115	4.6	M	CLAS	RJG	0	0	0	NA	0	0	NA	NA	0	14.4
RJG116	19.0	M	DN	RJG	0	0	0	0	0	0	1	NA	0	14.0
RJG118	2.6	F	CLAS	RJG	1	0	0	NA	0	0	NA	NA	0	13.4
RJG121	4.1	M	LCA	RJG	0	0	0	1	NA	NA	0	NA	0	12.6
RJG122	2.5	M	LCA	RJG	0	0	0	0	0	0	0	NA	1	0.9
RJG124	6.3	M	DN	RJG	1	0	0	0	0	0	0	NA	1	0.3
RJG126	2.6	M	DN	RJG	1	0	0	0	0	0	1	NA	0	0.1
RJG127	4.8	F	DN	RJG	0	0	0	0	0	0	1	NA	0	8.3
RJG131	5.7	F	DN	RJG	1	0	0	0	0	0	0	NA	1	0.2

Sample ID	Age	Sex	Tumour Type	Centre	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	EFS Status	EFS Time (years)
RJG135	3.6	F	CLAS	RJG	0	0	0	NA	0	0	NA	NA	NA	NA
RJG141	9.8	F	CLAS	RJG	0	0	0	0	0	0	0	NA	0	4.7
RJG142	1.0	F	DN	RJG	0	0	0	0	0	0	1	NA	1	0.3

Table 2.2. Clinical demographics of the training cohort. The cohort consisted of 108 primary medulloblastomas obtained from across the UK and Europe (*n* = 86), from Dr. Richard Gilbertson (*n* = 15) and from the PNET3 clinical trial (*n* = 7). Patient ID, age in years at diagnosis, and gender are shown. Histological subtype is coded as follows: CLAS – classic, DN – desmoplastic / nodular, LCA – large cell / anaplastic, MBEN – medulloblastoma with extensive nodularity). The centre from which the sample was received is shown: BELG – Belgium; BHAM – Birmingham, UK; BRIS – Bristol, UK; BUD – Budapest, Hungary; CAM – Cambridge, UK; CARD – Cardiff, UK; LEEDS – Leeds, UK; NCL – Newcastle, UK; NOTT – Nottingham, UK; PNET3 – PNET3 trials cohort; POL – Poland; RJG – Memphis, TN, USA; SOTON – Southampton, UK. M stage is coded based on Chang's criteria (see section 1.3.3). 0 – M- (M stage 0 / 1), 1 – M + (M stage 2 / 3). *CTNNB1* mutation: 0 – no mutation detected; 1 – mutation detected. Chromosome 6 LOH: 0 – no loss of heterozygosity detected; 1 – loss of heterozygosity detected. Chromosome 17 LOH: 0 – no loss of heterozygosity detected; 1 – loss of heterozygosity detected. *MYCC* amplification: 0 – no amplification detected; 1 – amplification detected. *MYCN* amplification: 0 – no amplification detected; 1 – amplification detected. Signalling pathway activation: 0 – WNT / SHH independent; 1 – SHH pathway activated; 2 – WNT pathway activated. Signalling pathway antibody status: 0 – WNT / SHH independent; 1 – SHH antibody positivity; 2 – WNT antibody positivity. Event free survival status (0 - no event, 1 – event) and time in years are given. Missing data are indicated with NA.

Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	Treatment	EFS Status	EFS Time (Years)
PNET30002	12.8	F	CLAS	0	1	0	0	0	0	NA	2	1	0	13.0
PNET30009	6.3	F	CLAS	0	1	1	0	0	0	NA	2	1	0	12.5
PNET30012	4.7	M	CLAS	0	0	0	0	0	0	NA	0	1	0	9.8
PNET30013	9.9	M	CLAS	0	0	0	1	0	0	NA	0	1	1	3.5
PNET30015	12.9	M	CLAS	0	0	0	1	0	0	NA	0	2	1	0.7
PNET30018	3.4	M	CLAS	0	0	0	0	0	0	NA	0	1	0	12.6
PNET30019	15.4	M	CLAS	0	0	0	0	0	1	NA	0	2	1	0.6
PNET30028	7.2	F	CLAS	1	0	0	1	0	0	NA	2	1	0	10.7
PNET30030	14.6	M	DN	0	0	0	0	0	0	NA	NA	1	1	5.7
PNET30031	7.2	M	CLAS	0	0	0	0	0	0	NA	0	2	1	2.1
PNET30032	9.7	M	CLAS	0	NA	NA	NA	0	1	NA	NA	1	0	10.0
PNET30033	8.8	F	CLAS	0	0	0	0	0	1	NA	0	2	0	12.1
PNET30035	11	M	CLAS	0	0	0	0	0	0	NA	1	2	1	0.8
PNET30038	10.8	F	LCA	0	0	0	1	0	0	NA	1	1	1	2.0
PNET30039	8.5	F	CLAS	0	0	1	0	0	0	NA	2	2	0	8.6
PNET30041	7.6	M	CLAS	0	NA	NA	NA	0	0	NA	NA	1	0	11.9
PNET30043	8.6	F	CLAS	0	NA	NA	NA	0	0	NA	NA	1	1	2.2
PNET30044	9.2	M	DN	0	0	0	0	0	0	NA	1	2	0	10.2
PNET30047	7.3	M	DN	0	0	0	0	0	0	NA	1	2	0	10.8
PNET30048	14.4	M	CLAS	0	0	0	0	0	0	NA	0	1	0	10.4

Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	Treatment	EFS Status	EFS Time (Years)
PNET30051	7.5	F	CLAS	0	1	1	0	0	0	NA	2	2	0	9.7
PNET30052	9.4	M	CLAS	0	1	0	0	0	0	NA	2	1	0	10.2
PNET30054	6.2	M	CLAS	0	0	0	0	0	0	NA	0	1	1	3.3
PNET30062	6.6	M	CLAS	0	0	0	0	0	0	NA	0	1	0	8.0
PNET30065	5.7	F	CLAS	0	0	0	0	0	0	NA	0	1	0	11.7
PNET30066	5.9	M	CLAS	0	0	0	0	0	0	NA	0	2	1	1.5
PNET30072	7.6	F	CLAS	0	0	0	0	0	0	NA	0	1	1	1.5
PNET30075	6.5	F	DN	0	0	0	0	0	0	NA	1	2	0	2.1
PNET30083	4.5	F	CLAS	0	0	0	0	0	0	NA	0	2	0	10.6
PNET30105	8.3	M	CLAS	0	0	0	0	0	0	NA	0	2	0	11.1
PNET30106	6.4	M	CLAS	0	0	0	0	0	0	NA	0	1	1	1.3
PNET30107	5.9	F	CLAS	0	0	0	1	0	0	NA	0	2	0	11.2
PNET30112	7.1	F	CLAS	0	1	1	0	0	0	NA	2	1	0	11.3
PNET30113	4.1	F	CLAS	0	0	0	1	1	0	NA	0	2	1	0.3
PNET30116	9.5	M	CLAS	0	0	0	0	0	0	NA	0	2	1	1.5
PNET30120	6.3	M	CLAS	0	0	0	0	0	0	NA	0	1	0	10.1
PNET30121	13.2	M	CLAS	0	0	0	1	0	0	NA	0	2	0	11.0
PNET30124	6.2	F	LCA	0	0	0	0	0	0	NA	1	2	0	10.5
PNET30126	8.4	M	CLAS	0	0	0	1	0	0	NA	NA	1	1	2.2
PNET30129	7.9	M	LCA	1	0	0	1	0	0	NA	0	1	1	1.1

Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	Treatment	EFS Status	EFS Time (Years)
PNET30132	5.1	F	CLAS	0	0	0	0	0	0	NA	0	2	0	10.4
PNET30134	8.4	M	LCA	0	0	0	1	0	0	NA	0	2	1	2.6
PNET30137	13.4	M	LCA	0	1	1	0	0	0	NA	2	1	0	11.3
PNET30141	6.4	M	CLAS	0	0	0	0	0	0	NA	0	1	0	12.3
PNET30145	5	M	CLAS	0	0	0	0	0	0	NA	NA	2	1	1.3
PNET30148	4.2	M	CLAS	0	0	0	0	0	0	NA	NA	1	0	10.2
PNET30150	8.9	F	CLAS	0	0	0	1	0	0	NA	0	1	0	12.2
PNET30152	10.2	F	CLAS	0	0	0	0	0	0	NA	2	1	0	0.1
PNET30160	8.8	F	DN	0	0	0	0	0	0	NA	NA	1	0	9.0
PNET30161	8.1	M	CLAS	0	0	0	0	0	0	NA	0	2	0	11.8
PNET30164	8.3	F	CLAS	0	0	0	0	0	0	NA	0	1	1	6.2
PNET30165	12.4	M	CLAS	0	0	0	0	0	0	NA	1	2	0	7.1
PNET30166	8.6	M	CLAS	0	0	0	0	0	0	NA	NA	1	1	7.1
PNET30172	8.4	M	CLAS	0	0	0	0	0	0	NA	2	1	0	8.8
PNET30175	5.4	F	CLAS	0	0	0	0	0	0	NA	1	2	0	6.3
PNET30178	7.7	M	CLAS	0	NA	NA	NA	0	0	NA	NA	2	0	6.0
PNET30179	3.5	F	CLAS	0	0	0	0	0	0	NA	NA	1	0	11.2
PNET30180	9.9	M	CLAS	0	0	0	1	0	0	NA	2	2	0	9.1
PNET30185	10.3	M	CLAS	0	0	0	0	0	0	NA	NA	1	0	11.1
PNET30186	11.6	F	CLAS	0	0	0	1	0	0	NA	0	2	0	11.1

Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	Treatment	EFS Status	EFS Time (Years)
PNET30191	7.8	M	LCA	0	0	0	0	0	0	NA	0	2	1	0.5
PNET30193	7.3	F	LCA	0	0	0	0	0	0	NA	1	1	0	8.6
PNET30195	6	F	DN	0	0	0	0	0	0	NA	1	2	0	8.8
PNET30199	15.5	F	DN	0	0	0	0	0	0	NA	1	2	0	9.3
PNET30201	6.6	M	LCA	0	0	0	0	0	1	NA	1	2	1	0.5
PNET30202	9.7	F	CLAS	0	NA	0	0	0	0	NA	1	1	0	8.7
PNET30205	9.2	F	CLAS	0	NA	NA	NA	0	1	NA	NA	1	0	7.0
PNET30209	10.4	M	CLAS	0	NA	NA	NA	0	0	NA	NA	1	0	7.5
PNET30210	4.6	M	CLAS	0	0	0	0	0	0	NA	NA	1	0	7.8
PNET350001	10.1	M	CLAS	0	NA	NA	NA	0	0	NA	NA	1	0	8.1
PNET350010	5.9	M	CLAS	0	0	0	0	0	0	NA	NA	2	0	11.7
PNET350011	13.4	M	CLAS	0	0	0	0	0	0	NA	1	2	1	5.3
PNET350012	4.2	F	CLAS	1	0	0	0	1	0	NA	0	1	0	10.8
PNET350015	4.6	F	CLAS	1	0	0	0	0	0	NA	0	2	1	0.6
PNET350019	15.4	M	CLAS	0	0	0	0	0	0	NA	1	2	1	2.1
PNET350021	14	M	CLAS	1	1	0	0	0	0	NA	2	1	0	11.0
PNET350034	14.3	M	CLAS	0	0	0	0	0	1	NA	0	2	0	7.0
PNET350035	9.9	F	CLAS	0	0	0	0	0	0	NA	0	2	0	7.1
PNET350040	15.6	M	CLAS	0	0	0	1	0	0	NA	0	2	1	5.1
PNET350041	10.8	M	CLAS	0	0	0	0	0	0	NA	0	1	0	10.0

Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	Treatment	EFS Status	EFS Time (Years)
PNET350044	8.1	M	CLAS	0	0	0	1	0	0	NA	0	2	0	10.1
PNET350045	6.8	M	CLAS	0	0	0	0	0	0	NA	2	2	0	12.1
PNET350049	13.1	F	LCA	1	0	0	0	0	0	NA	0	1	1	2.2
PNET350056	10.8	M	CLAS	0	1	1	0	0	0	NA	2	2	0	11.8
PNET350057	14	M	CLAS	0	0	0	0	0	0	NA	0	2	0	10.1
PNET350058	11.6	M	CLAS	1	0	0	0	0	0	NA	0	1	0	10.3
PNET350060	10.3	M	DN	NA	NA	NA	NA	0	1	NA	NA	1	1	1.3
PNET350063	5.3	F	CLAS	0	0	0	0	0	0	NA	1	2	1	0.5
PNET350068	6.3	F	CLAS	1	0	0	0	0	0	NA	NA	1	0	11.7
PNET350075	8	F	CLAS	0	NA	1	0	0	0	NA	NA	2	0	11.7
PNET350080	10.3	M	CLAS	0	1	1	0	0	0	NA	2	1	0	10.2
PNET350086	10.4	F	CLAS	1	NA	NA	NA	0	0	NA	NA	1	0	11.8
PNET350088	15.2	F	CLAS	0	0	0	0	0	0	NA	0	2	0	10.3
PNET350090	10.3	F	CLAS	0	0	0	0	0	0	NA	2	1	0	1.3
PNET350091	4.4	M	CLAS	1	0	0	0	0	0	NA	0	1	1	1.6
PNET350099	13	F	CLAS	0	0	0	1	0	0	NA	0	1	1	4.4
PNET350104	5.4	M	LCA	1	0	0	0	0	1	NA	NA	1	1	0.0
PNET350106	7.9	F	CLAS	1	0	NA	NA	0	0	NA	0	1	1	0.7
PNET350116	3.8	F	CLAS	0	0	0	0	0	0	NA	1	2	0	7.0
PNET350120	13.5	M	CLAS	0	0	0	0	0	0	NA	1	2	0	14.8

Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	Treatment	EFS Status	EFS Time (Years)
PNET350124	7.1	M	CLAS	1	0	0	1	0	0	NA	0	1	1	2.2
PNET350128	15.4	M	CLAS	0	NA	NA	NA	0	1	NA	NA	1	0	9.2
PNET350132	14.1	M	CLAS	0	0	0	1	0	0	NA	0	2	0	9.3
PNET350133	9.5	F	CLAS	0	0	0	1	0	0	NA	0	1	0	9.3
PNET350136	12.1	M	CLAS	0	0	0	1	0	0	NA	0	2	0	8.8
PNET350137	6	F	CLAS	0	0	0	0	0	0	NA	0	1	1	5.4
PNET350142	9.4	F	CLAS	1	0	0	1	0	0	NA	0	1	1	1.0
PNET350147	5.3	M	CLAS	0	0	0	0	0	0	NA	NA	2	0	4.8
PNET350150	5.1	M	CLAS	1	0	0	0	0	0	NA	0	1	1	1.1
PNET350161	6.8	F	CLAS	0	0	0	1	0	0	NA	0	2	1	2.5
PNET350163	6.8	F	CLAS	0	0	0	1	0	0	NA	0	2	0	10.1
PNET350165	14	F	CLAS	0	0	0	0	0	0	NA	NA	2	0	5.8
PNET350166	10.6	M	CLAS	0	0	0	0	0	0	NA	0	2	1	3.2
PNET350167	6.9	F	CLAS	0	0	0	0	0	0	NA	0	2	0	10.4
PNET350169	3.1	M	LCA	1	0	0	0	0	0	NA	1	1	1	1.2
PNET350170	10.9	F	CLAS	1	0	0	0	0	0	NA	1	1	0	8.6
PNET350172	5.4	M	CLAS	0	0	0	1	0	0	NA	0	2	0	10.3
PNET350174	12.5	M	CLAS	1	0	0	0	0	0	NA	0	1	0	7.6
PNET350176	7.8	F	DN	0	0	0	0	0	0	NA	1	2	0	9.8
PNET350184	11.8	M	CLAS	0	0	0	0	0	0	NA	0	2	0	10.2

Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	Treatment	EFS Status	EFS Time (Years)
PNET350189	14.3	M	CLAS	0	0	0	1	0	0	NA	0	2	0	10.0
PNET350193	14.1	F	CLAS	0	0	0	0	0	0	NA	0	1	0	8.1
PNET350197	5.3	M	CLAS	1	NA	NA	NA	0	1	NA	NA	1	1	0.8
PNET350198	9.5	M	CLAS	1	0	0	0	0	0	NA	0	1	1	4.2
PNET350204	10.7	M	CLAS	1	NA	NA	NA	0	1	NA	NA	1	0	10.1
PNET350208	9.7	F	CLAS	1	0	0	0	0	0	NA	0	1	0	9.6
PNET350209	9.1	M	CLAS	0	0	0	0	0	0	NA	0	1	1	4.1
PNET350212	10.8	M	CLAS	1	0	0	0	0	0	NA	0	1	0	5.4
PNET350217	9.5	M	CLAS	1	NA	0	0	0	0	NA	0	1	1	4.8
PNET350218	3.4	M	DN	0	0	0	0	0	0	NA	1	1	1	1.5
PNET350224	14.9	M	CLAS	0	0	0	0	0	0	NA	0	2	0	5.0
PNET350241	13.1	M	CLAS	1	0	0	1	0	0	NA	0	1	1	2.2
PNET350244	14.1	M	CLAS	0	0	0	1	0	0	NA	0	2	0	4.6
PNET350248	3.9	M	CLAS	0	0	0	0	0	0	NA	0	1	0	6.0
PNET350250	15.8	M	CLAS	0	NA	NA	NA	0	0	NA	NA	1	1	2.0
PNET350253	9.6	M	CLAS	0	0	0	0	0	0	NA	0	2	1	2.3
PNET350254	5.6	F	CLAS	0	0	0	0	0	0	NA	0	2	0	9.3
PNET350256	4.2	M	CLAS	1	0	0	0	0	0	NA	NA	1	1	0.3
PNET350259	9.1	M	CLAS	1	0	0	0	0	0	NA	0	1	1	0.8
PNET350284	11.2	M	CLAS	0	0	0	0	0	0	NA	0	2	0	5.4

Sample ID	Age	Sex	Tumour Type	M stage	*CTNNB1* mutation	Chromosome 6 LOH	Chromosome 17 LOH	*MYCC* amplification	*MYCN* amplification	Signalling pathway signature	Signalling pathway antibody status	Treatment	EFS Status	EFS Time (Years)
PNET350290	9.7	F	CLAS	0	0	0	0	0	0	NA	1	1	0	8.3
PNET350291	6.7	M	CLAS	0	0	0	0	0	0	NA	NA	1	0	11.0
PNET350292	3.3	M	CLAS	1	0	0	0	0	0	NA	0	1	0	7.8

Table 2.3. Clinical demographics of the test cohort. The cohort consisted of 143 primary medulloblastomas obtained from the PNET3 clinical trial. Patient ID, age in years at diagnosis, and gender are shown. Histological subtype is coded as follows: CLAS – classic; DN – desmoplastic / nodular; LCA – large cell / anaplastic. The centre from which the sample was received is shown: M stage is coded based on Chang's criteria (see section 1.3.3): 0 – M- (M stage 0 / 1), 1 – M + (M stage 2 / 3). *CTNNB1* mutation: 0 – no mutation detected; 1 – mutation detected. Chromosome 6 LOH: 0 – no loss of heterozygosity detected; 1 – loss of heterozygosity detected. Chromosome 17 LOH: 0 – no loss of heterozygosity detected; 1 – loss of heterozygosity detected. *MYCC* amplification: 0 – no amplification detected; 1 – amplification detected. *MYCN* amplification: 0 – no amplification detected; 1 – amplification detected. Signalling pathway activation: 0 – WNT / SHH independent; 1 – SHH pathway activated; 2 – WNT pathway activated. Signalling pathway antibody status: 0 – WNT / SHH independent; 1 – SHH antibody positivity; 2 – WNT antibody positivity. Treatment status is indicated (1 – chemotherapy plus radiotherapy; 2 – radiotherapy only). Event free survival status (0 - no event; 1 – event) and time in years are given. Missing data are indicated with NA.

Cohort Name	Size	Source	Description	Chapter
Chapter 3 primary investigation cohort	55	Primary medulloblastomas from UK and Europe (*n* = 33) and USA (*n* = 22), with DNA and available clinico-pathological data	This cohort was assessed for mutations in *PTCH1*, *SMO*, *SUFU* and for *PTCH1* and *COL1A2* promoter methylation status. Chromosome 6 and chromosome 17p were also tested for LOH	3
Chapter 3 GeXP investigation cohort	39	Primary medulloblastomas from UK, Europe (*n* = 25) and USA (*n* = 14), with DNA, RNA and clinico-pathological data available	39 / 55 samples from chapter 3 primary investigation cohort had available RNA and were tested for WNT / SHH pathway activation using GeXP assay described in chapter 3	3
Kool et al. (Kool et al., 2008)	62	60 primary medulloblastomas, 2 local relapsed biopsies	Transcriptomic data from Affymetrix U133 plus 2 arrays used to select WNT / SHH signature genes for GeXP assay, to validate GeXP assay (chapter 3) and assess expression patterns (chapter 5)	3, 5
Thompson et al. (Thompson et al., 2006)	46	Primary medulloblastomas	Transcriptomic data from Affymetrix U133aV2 arrays used to select WNT / SHH signature genes for GeXP assay, to validate GeXP assay (chapter 3) and assess expression patterns (chapter 5)	3, 5
Fattet et al. (Fattet et al., 2009)	40	Primary medulloblastomas from single institution	Transcriptomic data from Affymetrix U133 plus 2 arrays used to validate GeXP assay (chapter 3) and assess expression patterns (chapter 5)	3, 5
Primary methylation array training cohort	108	Primary medulloblastomas from UK, Europe and USA	The cohort was run on Golden Gate methylation arrays and, following QC, disease subgroups were identified	4
PNET3 methylation array test cohort	143	Primary medulloblastomas from PNET3 clinical trial (see section 2.1, 5.3.1)	The cohort was run on Golden Gate methylation arrays and, following QC, used to validate subgroup patterns discovered using the primary methylation array training cohort	4
Age-matched survival cohort	191	Comprised PNET3 methylation array test cohort cases passing methylation array QC (*n* = 136), plus age matched (3-16 years) cases passing methylation array QC (*n* = 55) from primary methylation array cohort	This cohort was used to identify potentially prognostic methylomic biomarkers in a cohort with a defined age distribution (aged 3 – 16 years at presentation)	5
Cho et al. (Cho et al., 2011)	194	Primary medulloblastomas from 5 USA oncology centres	Transcriptomic data from Affymetrix HT HG U133A used to assess expression of prognostic methylation biomarkers	5
Northcott et al. (Northcott et al., 2010)	103	Primary medulloblastomas	Transcriptomic data from Affymetrix Human Exon 1.0 ST array used to assess expression of prognostic methylation biomarkers	5

Table 2.4. Summary of cohorts utilised in this study. Sample number, sample source, description and chapter in which its use is described are indicated for each cohort.

2.2 Extraction of nucleic acids

DNA and RNA extractions were performed by members of the Paediatric Brain Tumour Research group (Northern Institute for Cancer Research, University of Newcastle, UK). Dr. Meryl Lusher, Mr. Kieran O'Toole and Dr. Janet Lindsey kindly carried out the extractions. Additional DNA and RNA samples of medulloblastomas were kind gifts of Dr. Richard Gilbertson and were extracted by Twala Hogg (both Department of Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, TN, USA).

2.2.1 DNA extraction

DNA was extracted from FFPE medulloblastoma samples using a Qiagen DNeasy kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions. DNA was extracted from frozen tumour samples using Trizol (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions. Concentrated DNA stocks were stored at -80°C.

2.2.2 RNA extraction

RNA was extracted from frozen material using Trizol (Invitrogen) according to manufacturer's instructions. RNA was resuspended in RNase free water (Invitrogen) and treated with DNase I (Ambion) to eliminate residual contaminating DNA. Concentrated stocks were stored at -80°C.

2.3 Assessment of nucleic acid concentration and quality

2.3.1 Nanodrop

DNA concentrations were assessed with the Nanodrop spectrophotometer (Thermo Scientific), which is able to measure DNA concentration in a sample volume of 1 µl. DNA and RNA efficiently absorb UV light, with an absorption maximum at 260 nm. In contrast to nucleic acids, proteins have an absorption maximum of 280 nm. The machine measures the optical density (OD) of the sample at wavelengths of 260 and 280 nm, and reports a nucleic acid concentration as well as an OD_{260} / OD_{280} ratio. This provides a measure of the amount of contaminating protein within the sample.

2.3.2 Bioanalyzer

RNA quality was an important metric in the analyses performed with this material, so a more sensitive method was used to assess RNA quantity and quality. The Bioanalyzer 2100 platform (Agilent, Stockport, UK) is a chip-based micro-fluidics based system for sizing, quantification and quality control of DNA, RNA and protein from a 1 µl sample volume. Micro-fluidics form the basis of an electrophoresis based method. The RNA nano II assay is accurate for the assessment of RNA concentration from 5-500 ng / µl and was performed using the manufacturer's protocol. Samples with concentrations above 500 ng / µl were diluted appropriately so that an accurate reading could be taken. For each sample, the assay reported RNA concentration, 28S:18S ratio and an RNA integrity number (RIN), which is a measure of RNA degradation. High quality RNA would have a RIN above 9 and a 28S:18S ratio close to 2, with degraded samples having a RIN closer to 1 and a 28S:18S ratio less than 1. Two example traces are shown in Figure 2.2.

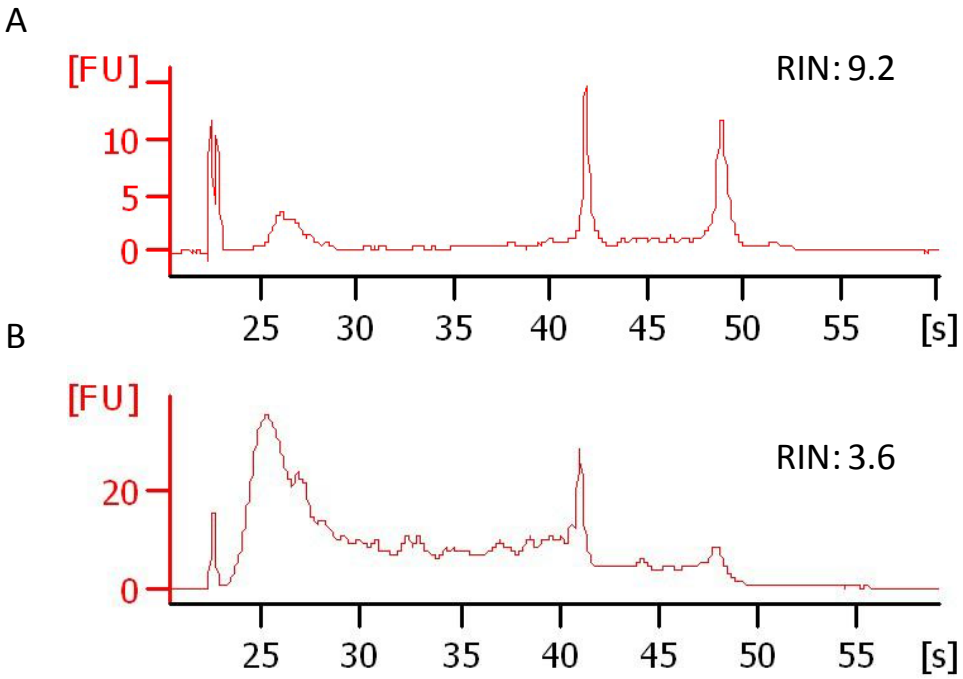


Figure 2.2. Example electropherogram traces from Bioanalyzer. The x axis shows time in seconds, the y axis shows fluorescent units. Panel A shows a trace from a good quality RNA. The 18S and 28S peaks are visible at 42 and 49 seconds respectively. The peak at 27 seconds contains small RNAs, including the 5.8S and 5S ribosomal peaks and transfer RNAs. Panel B shows a poor quality RNA. The amount of degraded RNA has significantly increased, as shown by the large peak at 25 to 30s. The 18S peak is still visible at 42 seconds, but the 28S fraction at 48 seconds is much reduced. For both traces, the RNA integrity number (RIN) is indicated.

2.4 The GeXP assay for assessing expression signatures

The GenomeLab™ GeXP Analysis System Multiplex RT-PCR (reverse-transcriptase – polymerase chain reaction) assay (Beckman Coulter, Fullerton, CA, USA) is a PCR-based assay for the multiplex, simultaneous assessment of gene expression for up to 30 loci from as little as 20 ng of total RNA. Initially, the assay uses gene-specific priming for the quantitative amplification of target loci, with a subsequent universal priming to quantitatively amplify cDNA. It is particularly useful for the testing of larger cohorts for disease or mutation-specific expression signatures, derived from microarray expression array experiments.

Specific chimeric primers, consisting of the locus specific sequence and the universal primer sequence, are constructed for each target locus. Pairs of chimeric primers, specific for each locus, are designed so that each detected amplicon is separated in size from other amplicons. The resultant fragments are separated in a capillary sequencer and visualised by applying a laser to detect the fluorescent tags attached to the PCR products.

The GeXP assay consists of three steps: (i) cDNA synthesis by reverse transcription, (ii) PCR and (iii) fragment analysis on a capillary sequencer, detailed below. In this study, the GeXP assay has been used to design gene expression signatures for the activation of the WNT and SHH signalling pathways (see chapter 3).

In the first step, a multiplex mix of reverse primers (each of which has a sequence complementary to the locus of interest joined to a universal reverse primer) is used to linearly synthesise cDNA from the mRNA target sequence (Figure 2.3).

In the second step, PCR, the first few cycles are driven by the locus specific parts of the chimeric primers. Over the remaining cycles, the fluorescently labelled universal forward primers (contained within the PCR buffer), which are in a large excess, drive amplification from universal primers, with each designed amplicon being equivalently amplified (Figure 2.3).

The final step, fragment analysis, is the application of the PCR products to a capillary electrophoresis system. Each designed amplicon will produce a peak of a specific size, whose area is proportional to the abundance of the target locus. The resultant

electropherogram (Figure 2.4) is used to determine relative abundance of gene transcripts by examining peak areas relative to a control gene.

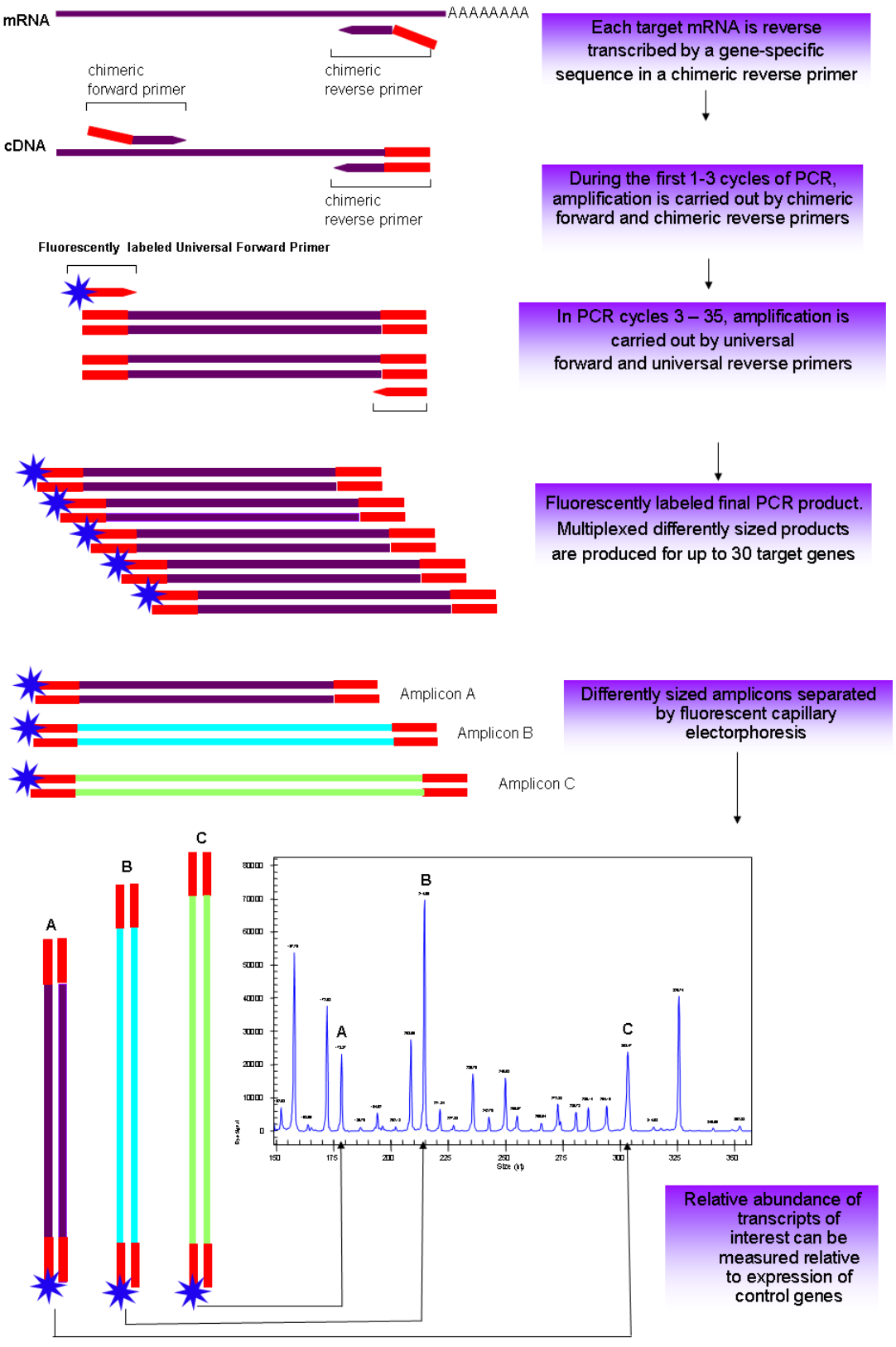


Figure 2.3. Principles of the GeXP assay. Schematic shows how reverse transcription with the reverse primer set only creates assay specific cDNA, During the PCR reaction, after initial amplification by specific chimeric amplicons, the excess of fluorescently labelled universal primers ensures that all amplicons, regardless of size are priming from the same primer pair and are equivalently amplified. Figure adapted from literature supplied by Dr. Jo Craggs (Beckman Coulter, High Wycombe, UK).

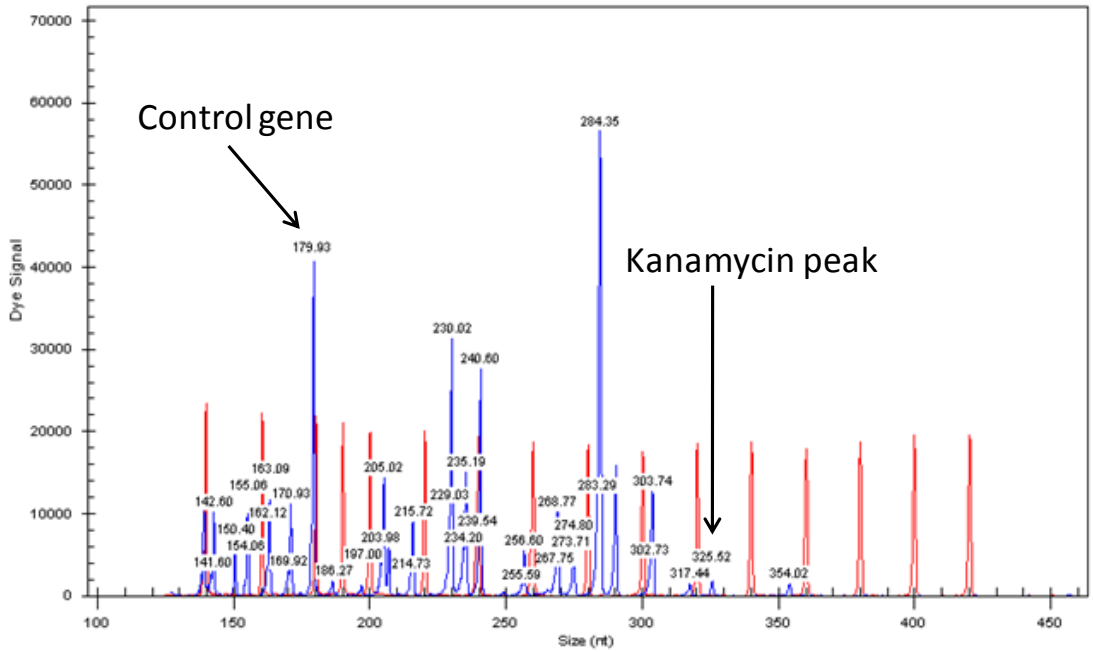


Figure 2.4. The GeXP assay is able to simultaneously assay gene expression for up to 30 loci. In this example trace, the red peaks correspond to ladder peaks used to accurately size the designed fragments. The blue peaks represent the genes being assayed in the experiment. Quantitative assessment of peak areas relative to control genes enable specific gene transcription to be measured. In this case, the control gene is at 179 base pairs (bp) and is indicated by an arrow. The Kanamycin peak at 325bp demonstrates that the reaction was capable of producing detectable product from input RNA and serves as a control for determining that the reagents and experimental procedures have performed adequately. The x axis shows fragment size in nucleotides, the y axis shows relative peak intensity.

2.4.1 Primer Design for the GeXP assay system

The forward and reverse primers are chimeric, with the reverse primers having a 5' end containing a 19 base pair universal sequence followed by a transcript-specific sequence (19 or 20 nucleotides); the forward primers have an 18 base pair universal sequence followed by a transcript-specific sequence (19 or 20 nucleotides). Primer design and multiplexing for the selected loci was undertaken using the GeXP Express Profiler, Gene Expression module (Beckman Coulter). The primer pairs were designed to produce products optimally separated by 7-10 base pairs, with a T_m for the locus-specific primer from 57°C to 63°C. Basic local alignment search tool (BLAST) (Altschul et al., 1990) and basic local alignment tool (BLAT) (Kent, 2002) searches, as well as the e-PCR feature of the UCSC genome browser (Hinrichs et al., 2006) were performed to ensure primer specificity. To eliminate the possibility of detecting genomic DNA, PCR products were designed across exon boundaries; products were also designed where

possible to overlap the Affymetrix probes from which they were derived. For genes with multiple transcripts, amplicons were designed that detected all known transcripts. In addition to the selected loci, endogenous control genes were selected that were equivalently expressed across the cohort of interest. The relative abundance of each locus of interest is calculated by comparison to the expression of a control gene.

An additional manufacturer-specified control primer set that amplifies spiked-in Kanamycin mRNA is included in the Reverse Transcription (reverse primer) and PCR buffers (forward primer) provided. Primers were ordered from VH Bioscience (Newcastle, UK) and were reverse-phase purified following synthesis. Primers were diluted in 18.2MΩ Purelab water (ELGA, High Wycombe, UK) and concentrated stocks of 100µM were stored at -80°C, with working aliquots at 20µM (R primers) and 10µM (F primers) stored at -20°C. The nucleic acid sequences of the designed primers are shown in section 3.3.6.1.

Before commencement of GeXP optimisation experiments, primer multiplexes were prepared. The reverse (R) primer multiplex was prepared so that each primer was at a concentration of 500nM. The forward (F) multiplex was prepared so that each primer was at a concentration of 200nM.

2.4.2 Assessment of RNA concentration and quality

RNA quantity and quality of medulloblastoma samples were assessed using the RNA nano II assay on the Agilent Bioanalyzer 2100 (Agilent, Stockport, UK), using the manufacturer's protocol. For each sample, the assay reported RNA concentration, 28S:18S ratio and a RNA integrity number (RIN), which is a measure of RNA degradation (see section 2.3.2). RNA aliquots at a concentration of 20 ng / µl were prepared as templates for the GeXP assay.

2.4.3 Assay optimisation

For successful optimisation of the GeXP assay, it was first necessary to define a suitable positive control that detected all known transcripts, to ensure that, in every run, all designed peaks are detectable. This is discussed in section 3.3.6.3.

In addition, after completion of assay optimisation, this serves to calibrate the detected peaks. For example, if a peak is designed at 179bp, for certain runs, due to minor variations in the experimental procedure and / or reagents (for example, the poly-acrylamide substrate through which the products are separated), the actual peak could be detected at an estimated distance from 178 to 180bp (this is within normal bounds for the assay). Therefore, the positive control also provides a common reference for where a peak is detected, that may change between different runs.

2.4.3.1 Primer specificity

It was necessary to demonstrate that each primer was generating the desired amplicon of the correct size and that unintended interactions between any single primer and the other primers in the multiplex were not occurring. Undesigned amplicons are undesirable, since they could overlap with designed amplicons of other transcripts. To test for this possibility, singlet 200nM F primers were used to amplify complementary DNA (cDNA) produced from a standard R primer multiplex containing all reverse primers. The defined positive control was used as a template. Results showing more than one peak (in addition to the Kanamycin mRNA control peak) were identified. Loci which showed more than one peak, indicating undesired interaction with other primers, were removed from the multiplex for subsequent experiments.

2.4.3.2 Primer attenuation

The Beckman CEQ system has a dynamic range of detection from 2000 to 130,000 intensity units, so for accurate quantitation, it was important that each expressed transcript was detected within this range. Using a reverse primer multiplex where each primer was at a concentration of 500nM, genes whose peaks that exceeded 130,000 units were identified. A reverse primer multiplex was prepared that omitted these genes. Serial dilutions of these genes' primers were added to this multiplex and the standard GeXP procedure was carried out to determine which dilutions produced peaks within the dynamic range. From these investigations, the optimal primer dilutions were derived and a finalised reverse primer mix, with attenuated primers at their optimal dilutions, was prepared.

2.4.4 GeXP assay procedure

The GeXP assay was performed in triplicate for each sample. Positive controls as well as no-template and reverse-transcriptase-negative controls were included for each experiment.

2.4.4.1 Reverse Transcription Step

The reverse transcription reaction was performed according to the manufacturer's standard protocol. RNA was extracted from medulloblastoma primary tumours and cell lines, as described in section 2.2.2. In a 10 μl reaction volume, 2 μl of 5 X RT (reverse transcriptase) Buffer (containing a Kanamycin RNA reverse primer) (Beckman Coulter), 0.5 μl of reverse transcriptase (RT) (20 units / μl) (Beckman), 2.5 μl of Kanamycin RNA (Beckman) and 1 μl of attenuated reverse primer mix (concentrations of each primer varying from 500 nM to 10 nM) was added to 2.5 μl of RNA template (20 ng / μl). The reverse transcriptase and subsequent PCR reactions were carried out on a GeneAmp® 9700 thermocycler (Applied Biosystems, Foster City, CA, USA). The reaction was heated for 1 minute at 48°C to relax the RNA, then incubated at 37°C for 5 minutes, 42°C for 60 minutes and finally reverse transcriptase enzyme was denatured by heating to 95°C for 4 minutes.

2.4.4.2 PCR step

PCR was carried out at a 10 μl reaction according to manufacturer's instructions (Beckman Coulter). Each reaction contained 2 μl of 5 X PCR Buffer (containing Forward primer for Kanamycin RNA) (Beckman Coulter), 5 mM $MgCl_2$ (Thermo-Hybaid, Ashford, UK), 1.75 units of ThermoStart Taq DNA polymerase (Thermo-Hybaid) and 1 μl of the forward primer multiplex mix (each primer was at a final concentration of 20 nM). 5.35 μl of this mix was added to 4.65 μl of cDNA from the reverse transcription reaction. The PCR reaction was carried out by heating the sample mixture for 10 minutes at 95°C (denaturation and Taq activation, followed by 35 cycles of 94°C (denaturation) for 30s, 55°C for 30s (annealing) and 70°C for 1 minute (extension). PCR products were stored at 4°C until required.

2.4.4.3 Fragment analysis

PCR products were diluted 1:5 with 18.2MΩ pure water prior to loading. 1 μl of each dilution was added to a CEQ sequencing plate (Beckman Coulter) containing 38.5 μl of Beckman sample loading solution (Beckman Coulter) and 0.5μl of size-standard 400 ladder (Beckman Coulter). After mixing, fragments were separated on a CEQ 8000 (Beckman Coulter) using the Frag-3 protocol. Fragments were exported to the GenomeLab GeXP genetic analysis software for analysis.

2.5 Expression Microarray

Since their first description (Schena et al., 1995), expression microarrays have revolutionised the measurements of gene expression. Prior to their discovery, studies of gene expression were limited to single candidate gene-based investigations. They have enabled truly genome-wide assessments of gene expression that have greatly enhanced the ability to detect differentially expressed transcripts between samples.

2.5.1 3′ IVT Affymetrix Arrays

The Affymetrix (Affymetrix, Santa Clara, CA, USA) 3' IVT (*in vitro* transcription) arrays that form the basis for the derivation of expression signatures for signalling pathway activation (see chapter 3) are a high-density oligonucleotide array for the genome-wide measurement of gene expression. Each detected transcript is represented by 11 oligonucleotide probes, each 25 bases in length. For each probe, in addition to the perfect match probe, there is a mis-matched probe at position 13, which can act as a control for non-specific hybridisation. The probe location and a workflow for running the arrays are shown in Figure 2.5. Briefly, cDNA is first prepared from a RNA sample. cDNA is then 3' *in-vitro* transcribed and biotinylated to make biotinylated complementary RNA (cRNA). After fragmentation, the cRNA is hybridised to the microarray, washed to remove non-specific hybridised fragments and stained with PE (phycoerythrin)-conjugated streptavidin. Arrays are read with a laser scanner and intensity scores are summarised for downstream analyses.

The summarised expression data are outputted in .CEL format files that can be read into software and subjected to background correction, normalisation and probe summarisation (section 2.5.2), using an appropriate technique, before examining for differential expression.

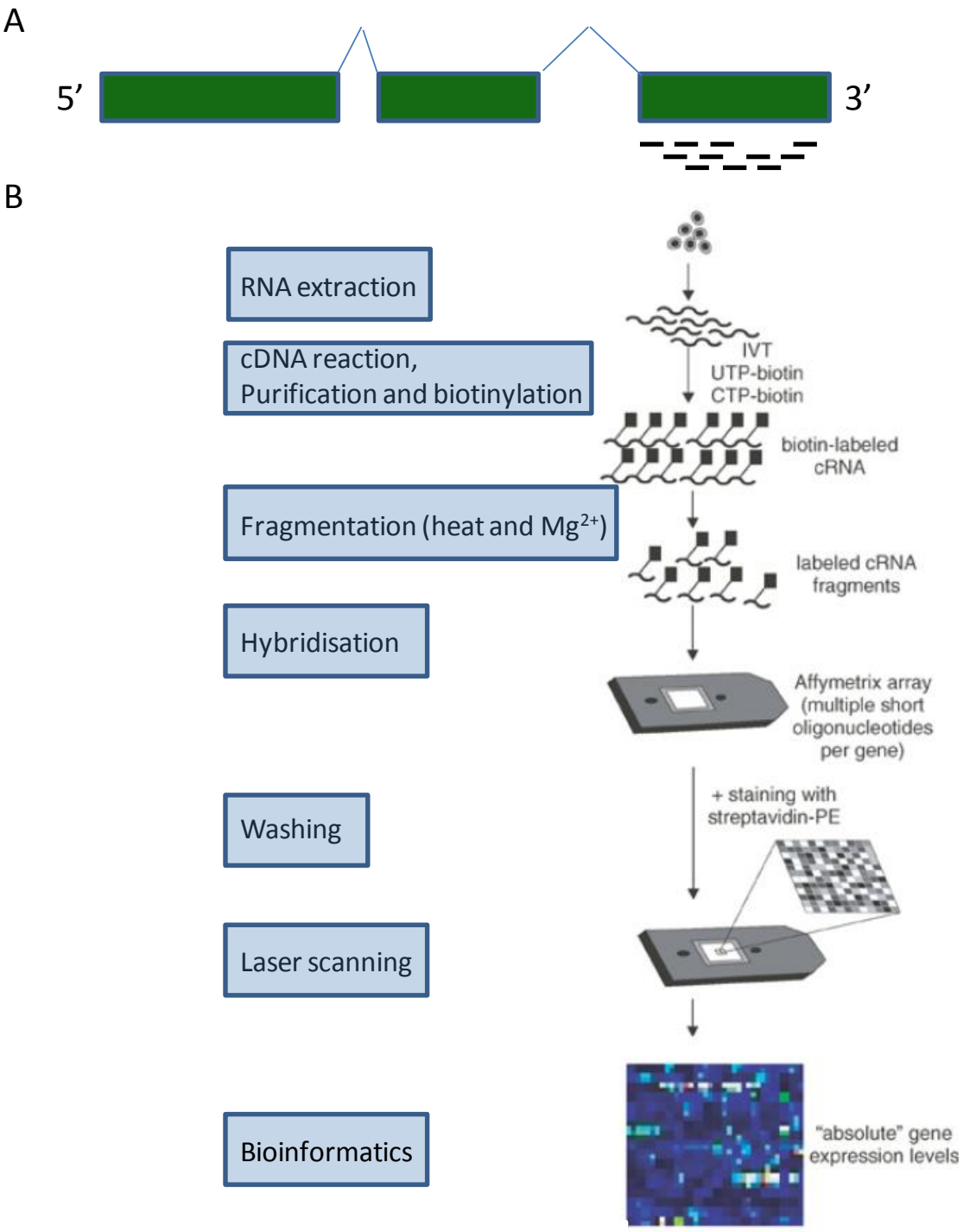


Figure 2.5. Affymetrix 3’-IVT arrays. **A.** The location of 3’-IVT array probes. Eleven probes are shown which span the 3’ end of a given transcript. The 5’ contents of the gene are not assayed. **B.** An Affymetrix 3’-IVT array workflow, adapted from a previously published report (Staal et al., 2003).

2.5.2 Robust multi-array average (RMA)

The robust multi-chip average (RMA) algorithm (Irizarry et al., 2003) is the most widely-used method for background correction and normalisation and summarisation of Affymetrix microarray data. In the background correction step, the mis-matched probe data are discarded, and instead, a background correction is applied which is calculated from the perfect match probes. The background correction works on the assumption that the observed signal (*O*) consists of a normally distributed background noise (*N*), defined by its mean, μ, and its standard deviation, σ, and an exponentially distributed true signal (*S*) defined by the exponent of a constant α, with

$$O = N + S,\ N \sim N(\mu,\sigma 2),\ S \sim Exp(\alpha).$$

The parameters α, μ and σ are assumed to be equal for all perfect match probes on a chip and are estimated for each chip. The estimated noise for each probe is subtracted from the observed intensity of expression to give a corrected measure of expression.

In the next step, the arrays are subjected to quantile normalisation, which transforms each array so that its distribution of intensities is the same and scaled so that the mean is the same. To achieve this, the expression data are ranked for each array, taking the average value at each rank across all arrays and then replacing for each array the actual expression scores with the averaged rank value across all arrays.

In the final step, the 11 probes that hybridise to each detected transcript are summarised into a single value of gene expression using median polishing, a technique that is robust to the effects of outlier probes, since it uses median values, but also because it estimates values based on the entire set of arrays being tested, rather than from a single array.

RMA has been shown to outperform other normalisation and expression summarisation techniques (Millenaar et al., 2006) and has become the most widely used method for processing Affymetrix arrays.

2.6 Bisulfite modification of DNA

Sodium bisulfite treatment modifies all unmethylated cytosine residues to uracil whilst leaving 5-methylcytosines unchanged. This treatment causes a change in the DNA

sequence at the single nucleotide level that can subsequently be detected using sequencing or using PCR based techniques (Herman et al., 1996; Frommer et al., 1992). For all analyses involving bisulfite treated DNA, bisulfite conversion was undertaken using an EpiTect bisulfite kit (Qiagen), according to manufacturer's instructions.

The 'gold standard' for the ascertainment of DNA methylation is through bisulfite sequencing and it is the benchmark against which any alternative means for estimating DNA methylation is measured.

2.6.1 Bisulfite treated DNA primer design and sequencing

For this project, MethPrimer (Li and Dahiya, 2002), was used to design appropriate primers for sequencing of bisulfite treated DNA. Primers are designed to be strand specific as well as bisulfite-specific (so that primers contain non-CpG cytosines which are not complementary to non-bisulfite treated DNA). Since the bisulfite treated DNA is of a lower complexity than untreated DNA, extra constraints are needed for primer selection in addition to those required for standard PCR.

- Primers should not contain any CpG sites within their sequence to avoid preferential selection of methylated or unmethylated DNA.
- To bias against detecting DNA that has not been completely modified, primer should hybridise to primary DNA sequence containing non-CpG cytosine residues (which will be converted into uracil residues if bisulfite conversion has occurred). This will ensure preferential hybridisation to completely modified DNA.

Amplified bisulfite treated DNA can then be sequenced using conventional methods. Direct sequencing enables the average methylation level of the sample from which the DNA was extracted to be determined (Figure 2.6). Samples in which the methylated peak represented >25% of the total peak height in greater than 25% of the analysed CpG sites were classed as showing evidence of methylation (i.e. methylated).

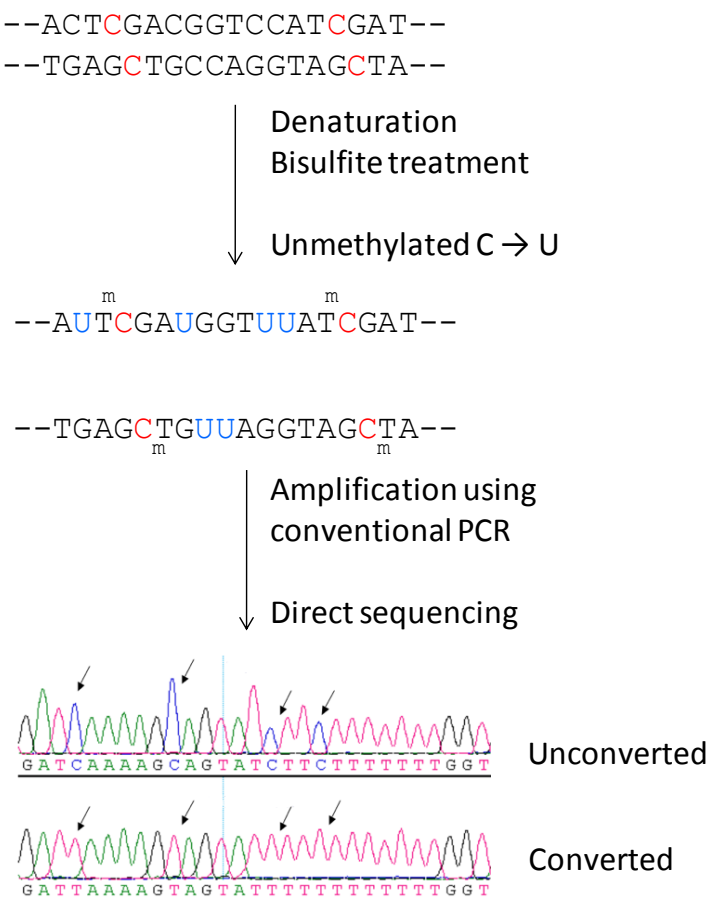


Figure 2.6. Principles of bisulfite conversion and sequencing to identify methylated cytosine residues. In the first step, DNA is denatured and treated with sodium bisulfite. This deaminates unmethylated cytosine residues, while leaving methylated cytosine residues unchanged. In the next step, specific primers, designed to avoid CpG islands, are used to amplify the region of interest. Conventional capillary sequencing can then discriminate between methylated and unmethylated cytosines. In the bottom panel, cytosine residues are indicated with arrows. Methylated cytosine residues will be read as cytosines, as they are unchanged following bisulfite treatment. Unmethylated cytosine residues will be read as thymine, since they were converted to uracil residues after bisulfite treatment. To measure the methylation at discrete CpG dinucleotides, the height of the cytosine peak is divided by the sum of the height of the cytosine and thymine peaks.

2.6.2 Methylation–specific PCR

Methylation-specific PCR (MSP), is an alternative method for analysing methylation status of bisulfite treated DNA, which obviates the need for DNA sequencing (Herman et al., 1996). MSP takes advantage of the sequence differences introduced by bisulfite treatment. Two sets of primers are designed, one specific for methylated (the M pair) and one specific for unmethylated loci (the U pair). For each tested sample, two PCR reactions are performed, one for each set of primers. Amplification with the M pair indicates the presence of methylated DNA; amplification with the U pair indicates the presence of unmethylated residues, and if both primer pairs produce bands, the sample has both methylated and unmethylated CpG dinucleotides (Figure 2.7). MSP has the advantages that it can be performed on small amounts of DNA, is much less labour intensive than bisulfite sequencing and can detect methylated loci at an abundance as low as 0.1% (Herman et al., 1996). The biggest disadvantages are that the results of MSP are qualitative rather than quantitative, and that careful optimisation is necessary to ensure that the PCRs for both U and M are performing equivalently.

Primers for MSP were designed using MethPrimer (Li and Dahiya, 2002), using the following design criteria:

- Primer sequences to contain at least one CpG site at the 3' end, to ensure optimal discrimination between methylated and unmethylated alleles.
- Primer sequences to contain as many CpG sites as possible, to further enhance discriminative power.
- Primers in the M and U pairs should contain the same CpG sites within their sequence, to ensure that any differential methylation in proximal CpG sites outside the region of interest is not affecting the amplification efficiencies.
- Primer sets U and M should have similar melting temperatures so that they can be amplified using the same reaction conditions on the same thermocycler. For this reason, the U set, which will be less complex and contain fewer cytosine residues, is longer than the M set.

For loci assessed by MSP, any sample showing a visible PCR product using primers specific for the methylated sequence, was classed as showing evidence of methylation (i.e. methylated).

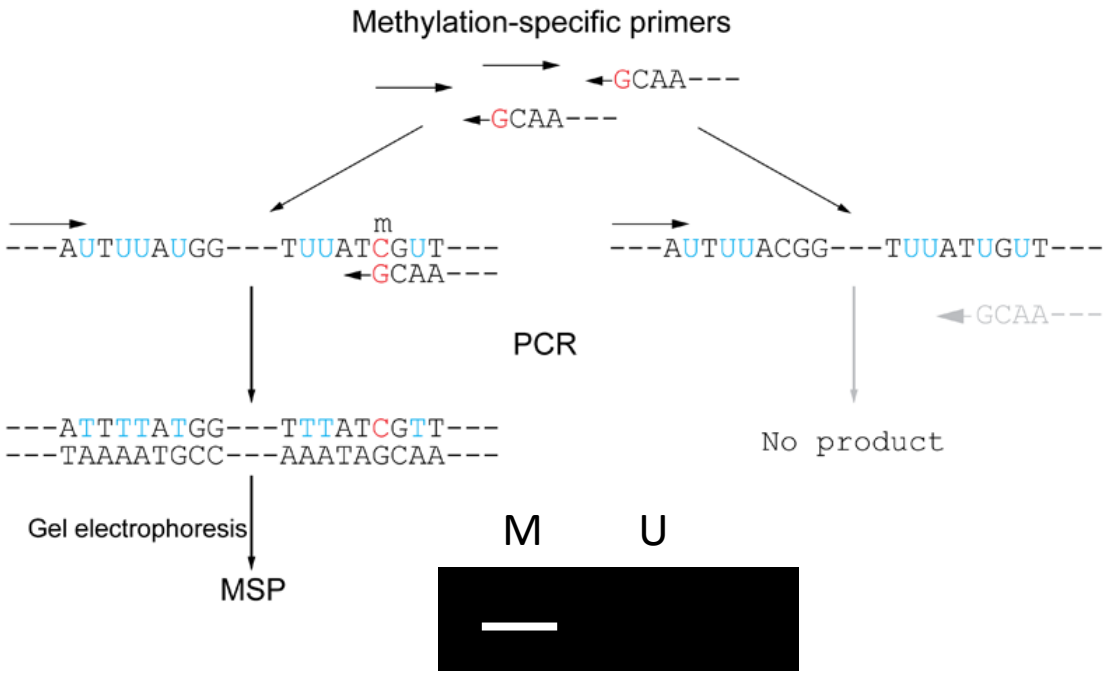


Figure 2.7. Methylation-specific PCR is able to sensitively identify DNA methylation. In this simplified example, a single PCR is set up for two DNA samples, one with a methylated CpG dinucleotide that is unchanged following bisulfite treatment, the other with an unmethylated DNA sample, where unmethylated cytosines have been converted to uracil residues. The methylation-specific reverse primer is able to bind to the target locus, which has a methylated cytosine residue, enabling amplification of the locus to occur. For the unmethylated sample, the conversion of the unmethylated cytosine prevents hybridisation from occurring, which in turn, prevents amplification. When the PCR products for the two samples are electrophoresed on an agarose gel, a band will be apparent for the methylated DNA sample, M, but not for the unmethylated sample, U. In a paired analysis (not shown), a primer pair specific for unmethylated DNA is used to amplify both samples and run in parallel to detect the presence of unmethylated DNA. Figure adapted from Wikipedia (http://en.wikipeida.org/wiki/Bisulfite_sequencing).

2.7 Golden Gate Cancer Panel I Methylation Array

This project utilised a high-throughput microarray technology that directly measures DNA methylation levels on bisulfite treated DNA (Bibikova et al., 2006). Briefly, bisulfite conversion of genomic DNA converts unmethylated cytosine to uracil residues, whilst methylated cytosines are unchanged. Pairs of probes, one pair specific to a converted uracil residue, the other pair specific to an unchanged (methylated) cytosine are hybridised to target DNA. Allele specific extension followed by ligation creates joined probe pairs that can be PCR amplified from common primer sequences. The PCR products contain address sequences that hybridise to a bead-based array microarray format. The ratio of hybridisation intensities from the unmethylated probe : methylated probe is used to determine the methylation status at specific CpG dinucleotides. The assumption is made that this accurately reflects the status of adjacent CpG dinucleotides. This assumption has been validated (Bibikova et al., 2006) using bisulfite sequencing. The principle of the technology is outlined in Figure 2.8. Using this technology, the GoldenGate Cancer Panel I methylation array (Illumina, San Diego, CA) is able to measure methylation status at 1505 loci mapping to 807 genes. The probes have been selected to represent TSGs, oncogenes, imprinted genes and genes involved in DNA repair, cell cycle control, differentiation and apoptosis. In addition, previously reported methylated genes were included (Bibikova et al., 2006). Probes situated both within (*n* = 1044) and outside of CpG islands (CpG islands were defined by Illumina) (*n* = 461), are represented. Probes are represented by a gene name, then the distance of the assayed CpG dinucleotide relative to the TSS (loci upstream of the start site are labelled with P (promoter), downstream loci with E (exon)), then the strand assayed (forward strand (F), reverse strand (R)). Thus an example probe, AATK_E63_R, is measuring methylation at the *AATK* gene, 63 bases downstream of the TSS, with the reverse strand being assayed.

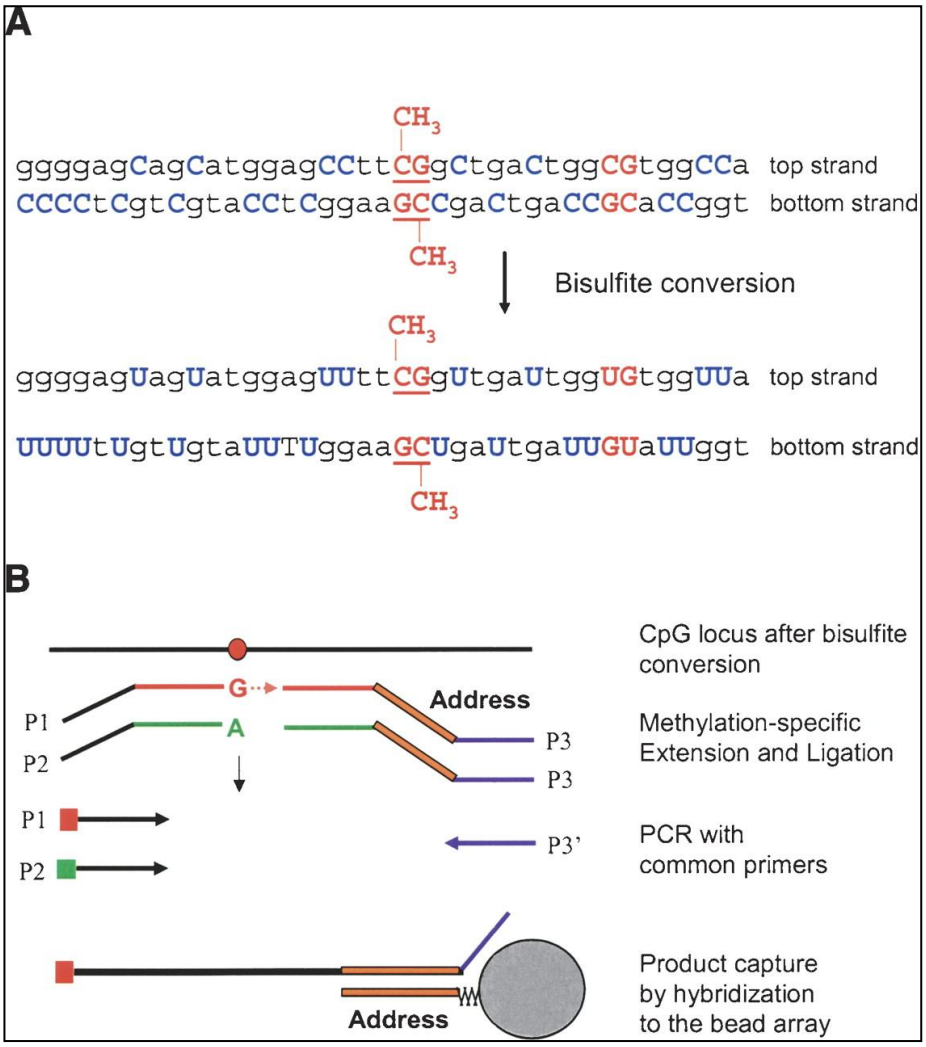


Figure 2.8. The principles of the Illumina GoldenGate methylation array. Figure taken from Bibikova et al., 2006. A. After bisulfite conversion, specific probes can detect methylated and unmethylated CpG residues. B. Allele-specific extension and ligation, followed by PCR from common primer sequences enables quantitative comparative measurement of Cy3 (unmethylated) and Cy5 (methylated)-labelled PCR products when hybridised to a bead-based microarray.

2.8 Methylation microarray Assay

Microarray analysis was performed on the Illumina Golden Gate Cancer Panel I methylation array at the Wellcome Trust Centre for Human Genetics, Oxford, UK according to manufacturer's protocols (Illumina, San Diego, CA, USA).

For each locus to be assayed, four oligonucleotides (two allele-specific oligonucleotides and two locus-specific oligonucleotides) are designed (Figure 2.8). Pooled query oligonucleotides corresponding to the 1505 loci assayed on the array are annealed to

bisulfite treated genomic DNA and washed to reduce non-specific hybridisation. Hybridised oligonucleotides are then extended and ligated to produce amplifiable fragments. Finally, specific fluorescently labelled universal PCR primers (Cyanine 3 (Cy3) – unmethylated; Cyanine 5 (Cy5) – methylated) selectively amplify unmethylated and methylated fragments.

In order to read the signal, pools of 3 µm diameter glass beads bound to specific detection oligonucleotides were randomly assembled into a bead array format. Because the beads are randomly positioned, a decoding process is used to determine the location and identity of each bead at every array location (Gunderson et al., 2004). The fluorescently-labelled amplified PCR products are hybridised to the beads. 96 well arrays containing the bead mixes were scanned in the Bead Array reader, which simultaneously scans at two different wavelengths (532 and 658 nm) to identify levels of Cy3 and Cy5 fluorescence.

Individual files are created for the red and green channels for each sample. Data and associated images are exported ready for downstream analysis by the end user in Bead Studio v3.2 (Illumina, San Diego, CA, USA).

Bead Studio is a general purpose program suitable for the analysis of all Illumina array data. The methylation module is specifically designed for the analysis of methylation array data. After reading in the data, background intensity, calculated from a panel of negative controls was subtracted from each data point and a level of methylation, the β score was calculated as in the following equation:

$$\beta = \frac{\max(\text{Cy5 intensity}, 0)}{(\max(\text{Cy3 intensity}, 0) + \max(\text{Cy5 intensity}, 0) + 100)}$$

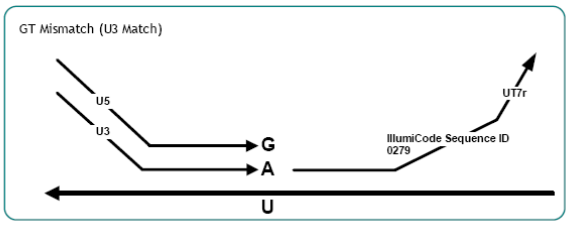
The β-value provides a continuous measure of levels of DNA methylation in samples, ranging from 0 in the case of completely unmethylated sites to 1 in completely methylated sites.

2.8.1 Methylation Microarray Quality Control

Bead Studio contains quality control measures to identify whether the array has run successfully. These are summarised in Figure 2.9. Subsequently, an improved method for the identification of spatial artefacts, BASH (**B**ead**A**rray **S**ubversion of **H**arshlight) (Cairns et al., 2008) became available. This software enabled the identification of spatial artefacts, enabling these areas to be excluded when calculating methylation intensities (Figure 2.10). Since the distribution of beads on the array is random, excluding areas due to anomalous intensities will not exclude specific probes due to their fixed position on the array. Additionally, since for each probe there are on average 60 beads on the array (approximately 30 each of unmethylated / methylated probe sequences), the removal of up to 20% of the beads on the array can be implemented without detriment to the estimation of the β score (Illumina, personal communication).

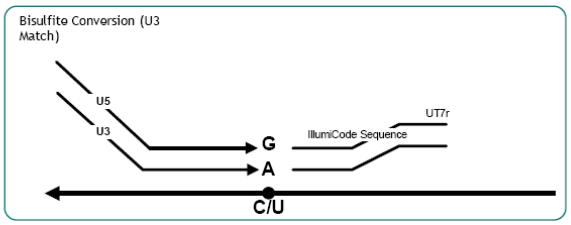
When running the test (PNET3) cohort, although previous FFPE samples had achieved satisfactory results, 2 inter-array replicates were additionally included to test for reproducibility. After masking of the anomalous intensity regions, β scores were calculated in the same way as Illumina's Bead Studio program (section 2.8), and exported for further analysis.

Sample intensities were assessed using the R package, *beadarray* (Dunning et al., 2007). Poorly performing samples were identified by examining scatter plots of signal intensity for both red and green channels. Examples of such plots are shown in Figure 2.11. Finally, quality control (QC) data from Bead Studio were collated with the signal intensity plots shown in Figure 2.11. Samples were deemed to have failed QC if they failed more than one QC test.



Allele specific extension

Two probes, mismatched by one base. Only one probe should ligate and therefore produce a signal after PCR amplification

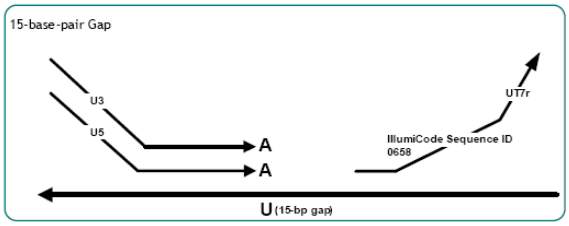


Bisulfite conversion

Primers are designed for the same locus, targeting either converted or unconverted DNA sequences. Presence of both signals indicates incomplete conversion

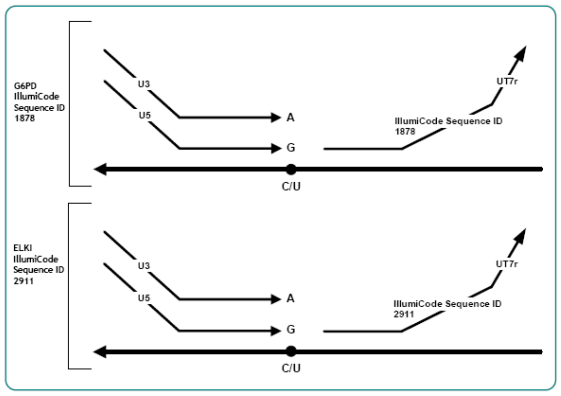
Contamination

For each run on the array reader, 1 out of 4 PCR contamination detection primers are added to each oligo pool. During each run, only this primer should give a signal. A high signal for any of the other three contamination primers indicates contamination from previous runs



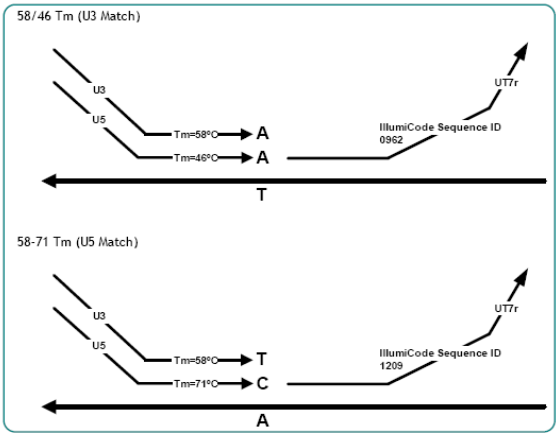
Extension gap

Tests the efficiency of extending 15 bases from the 3' end of the allele specific oligo to the 5' end of the locus-specific oligo



Gender Control

In males, control loci should be unmethylated, whereas females should be hemi-methylated. For cancer samples, this control is unreliable, since loss of X is a frequent occurrence.



Hybridisation Control

Two allele specific oligonucleotides (ASOs) with differing melting temperatures are hybridised to the same non-polymorphic locus. Only one ASO should extend and produce a signal

Figure 2.9. Bead Studio QC measures. Figure panels are adapted from Illumina's literature. Boxes describe each control step.

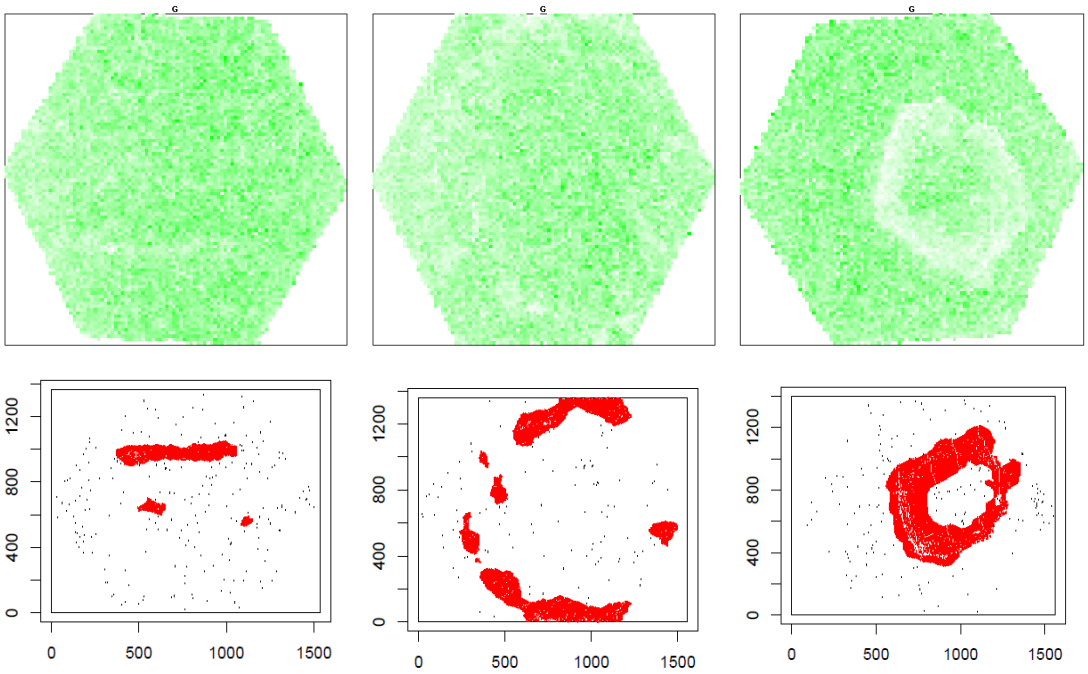


Figure 2.10. Identification of spatial anomalies on bead arrays using BASH (Cairns et al., 2008). The first row shows reconstructions of intensities across the array. The second row shows areas of anomalous intensities identified using the BASH algorithm, marked in red. These areas were ignored during subsequent calculation of the β scores. Columns show three types of spatial anomalies detected. In the first column, there is a small area of lower intensity scores, possibly caused by a fibre on the surface of the array. The second column shows an array where there were low intensities around the edges of the array. The third column shows the effects of an unintended bubble forming on the array surface. It should be noted that the angle of rotation for the BASH masked plots does not match the angle of rotation for the image reconstructions.

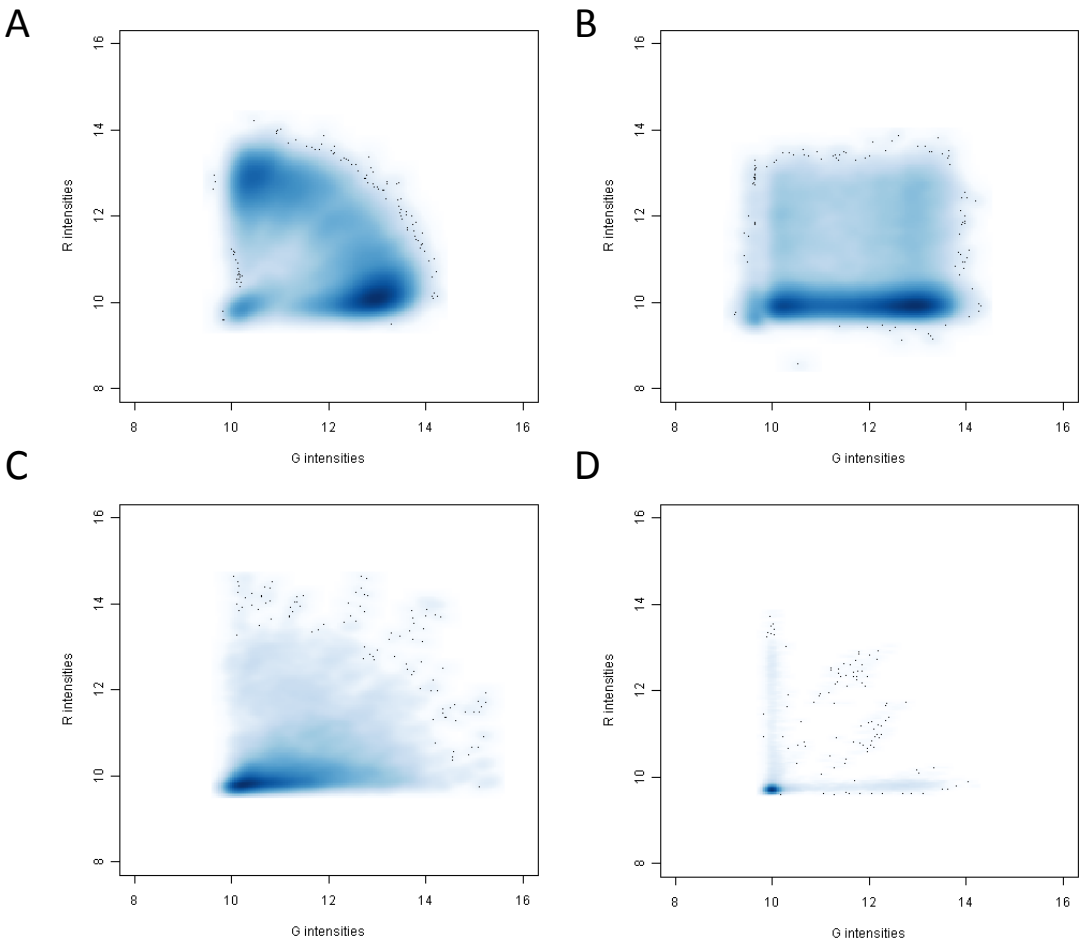


Figure 2.11. Red / green intensity density-scatterplots additionally identify poorly performing samples. Scatterplots display red (R intensities, y axis) and green (G intensities, x axis) for all data points from an individual array. Due to the large number of data points, their occurrence is indicated by shades of blue. High density areas of the plot are shaded dark blue, low density in light blue. Where density is low, individual data points are plotted. Panel A shows a well-performing sample; densest part of scatterplot is at a high green / low red intensity, corresponding to unmethylated probes. The next densest part is at a low green / high red intensity, corresponding to methylated probes. The next three panels (B-D) show intensity scatterplots of failed arrays. Panel B shows an array where the red channel intensities are much reduced. Panels C and D show, with increasing severity, arrays where sample intensities are much reduced at both intensities, indicating failed arrays.

2.8.2 Bland-Altman plot

The Bland-Altman plot (Bland and Altman, 1986) is a method for analysing the agreement between two different assays. It plots the average of the two assays along the x axis and the difference between the two estimates on the y axis (Figure 2.12).

It is designed to overcome the limitation that a simple linear regression between the two estimates of a parameter cannot explain whether there is any agreement between the two estimates. It can only describe any linear relationship between them. If the line of best fit lies along the line of equality, then there is agreement between the two estimates, but Pearson's correlation coefficient, r, does not estimate this agreement. It is also sensitive to the range of the parameter being tested. If the parameter has a wide range, the correlation coefficient will increase. For this project, this type of plot was used to compare estimates of β score estimated by methylation array and bisulfite sequencing and provided a method for assessing agreement between two different methods.

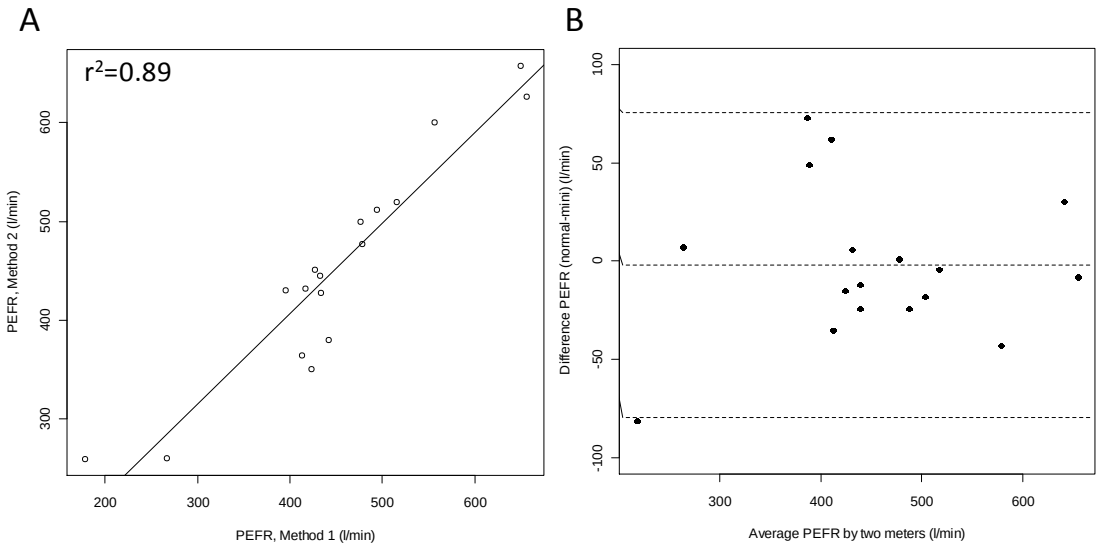


Figure 2.12. Bland-Altman plots are superior to linear regression for analysing agreement between methods. A. Scatter plot showing linear relationship between two methods for measuring peak expiratory flow rate (PEFR), with data from Bland and Altman's original paper (Bland and Altman, 1986). There is a strong linear relationship, as demonstrated by Pearson's r^2 coefficient. B. Bland-Altman example plot using same data. The x axis shows the average PEFR from the two assays and the y axis shows the difference in estimates between the two methods. Horizontal dotted lines show the mean difference and two standard deviations from the mean difference. It can be seen that there is a large difference in estimates of PEFR between the two methods, and that they are not equivalent, despite the strong linear relationship between the two methods.

2.9 Methods for unsupervised clustering

Cluster analysis is a broad term that encompasses a number of techniques for dividing a dataset into clusters of objects that are similar to one another, and formed an integral part of the analyses for this thesis. The different techniques used are briefly discussed below.

2.9.1 Hierarchical clustering

Hierarchical clustering (HC) is a statistical method for the grouping of related samples into clusters which form part of a wider hierarchy. It can be agglomerative, which is an iterative process in which clusters are merged until a single cluster which contains all samples is obtained. At each step, the two most similar clusters are grouped to form a merged cluster (Figure 2.13). Less commonly, divisive HC can be used, in which all samples begin in one cluster. In this case, the most disparate sample (i.e. the most dissimilar to all other observations) initiates a 'splinter group'. In subsequent steps, the algorithm reassigns samples that are more similar to the splinter group than the original unifying group. This is repeated iteratively until each sample has been placed into a discrete group, with the samples in a hierarchy.

The measure of cluster distance is also important. The most commonly utilised measures are average agglomeration (in which the distance between two clusters is the average of the dissimilarities between the samples in one cluster and the samples in the other cluster) and complete (in which the distance between two clusters is given by the largest dissimilarity observed between a sample in the first cluster and a sample in the second cluster) (Figure 2.13).

To perform HC, it is first necessary to measure the similarity between samples. This takes the form of a dissimilarity matrix, which quantifies the dissimilarity (or distance) between each sample. For numerical data, the most commonly used functions for calculating this dissimilarity are Euclidean and Manhattan distance (Figure 2.13). Once a hierarchy of clusters has been derived, by cutting the tree at an appropriate level, samples can be placed into discrete clusters. While this is one of the strengths of HC, since the cluster derivation can be guided by the person doing the analysis, this is also

a weakness, since there is a degree of subjectivity in sample assignment, which is absent for other techniques, such as *k* means clustering (section 2.9.5).

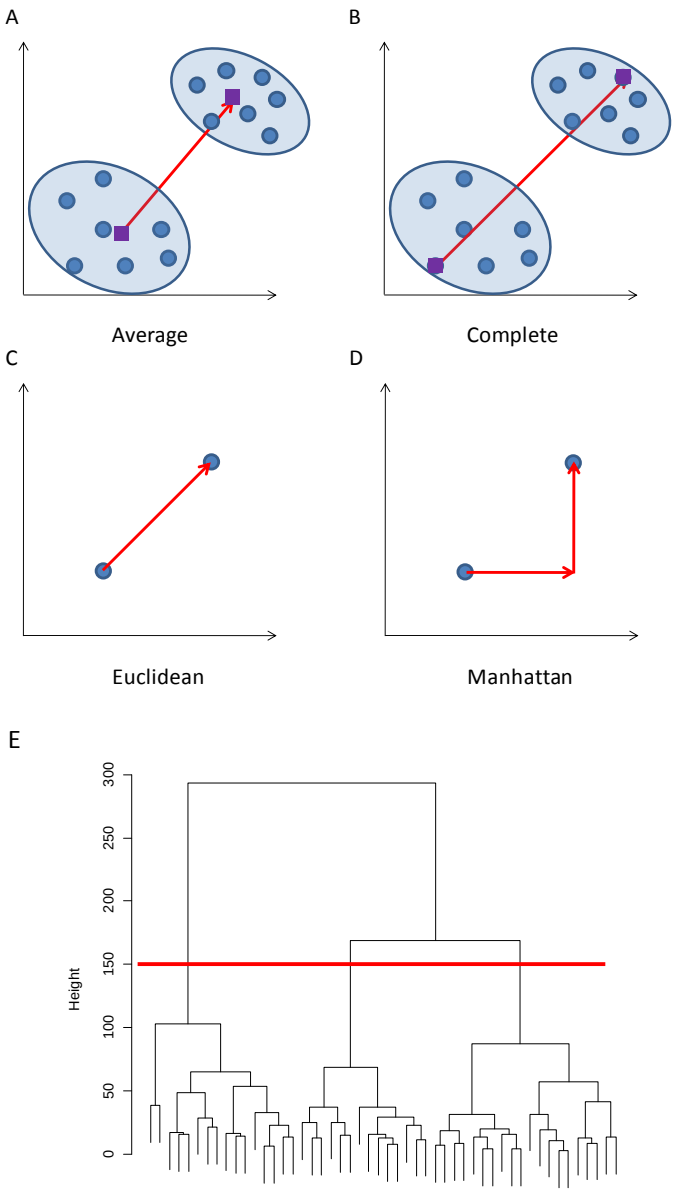


Figure 2.13. Construction of a hierarchical cluster dendrogram. A and B**.** The type of agglomeration method used can affect the structure of the dendrogram. The most commonly used agglomeration methods are shown. In average linkage (A), the average distance between two clusters determines their dissimilarity. For complete linkage (B), the dissimilarity between two clusters is determined by the distance between the most dissimilar members of a cluster. C and D. The most common methods for initially determining the dissimilarity between the samples comprising the dataset is shown. The Euclidean function specifies the distance between two samples to be the root-sum-of-squares difference, or less formally, as the crow flies. The Manhattan function specifies the distance between two samples to be the sum of the absolute difference between their positions, or less formally as the city block difference. E**.** An example dendrogram, which is the graphical output of a hierarchical clustering, generated from average agglomeration (part A) and Euclidean distance (part C) is shown for a hypothetical dataset. In this case, three clear clusters are visible, and if the tree is cut at a height of 150, shown with a red line, cluster membership can be assigned on this basis.

2.9.2 Principal component analysis

While not formally a clustering method, principal component analysis (PCA) is a useful method for the visualisation and interpretation of the variation within a complex dataset. It is fundamentally a dimension reduction procedure, reducing the complexity of a dataset to a smaller number of dimensions that can be used to reveal underlying structure. In a high-dimensional setting, as is the case for genome-wide arrays, there is likely to be a high degree of redundancy within the dataset, i.e. some variables are highly correlated, enabling the reduction of the observed variables into a smaller number of principal components, which are artificial variables that account for as much of the variation in the dataset as possible. A simple two dimensional example illustrating the first two principal components is shown in Figure 2.14.

2.9.2.1 Biplots

A biplot is an exploratory graph that enables information from both samples and variables to be displayed on the same plot. This has been extended to representing the results of PCA, and entails the plotting of the two most important principal components along the x and y axes, with the samples being shown as a scatterplot along these axes and the projection of the variables from which the loadings have been calculated to be drawn as arrows, enabling the contribution of the variables, measured within the dataset, to the principal components, to be determined. An example biplot is illustrated in Figure 2.14.

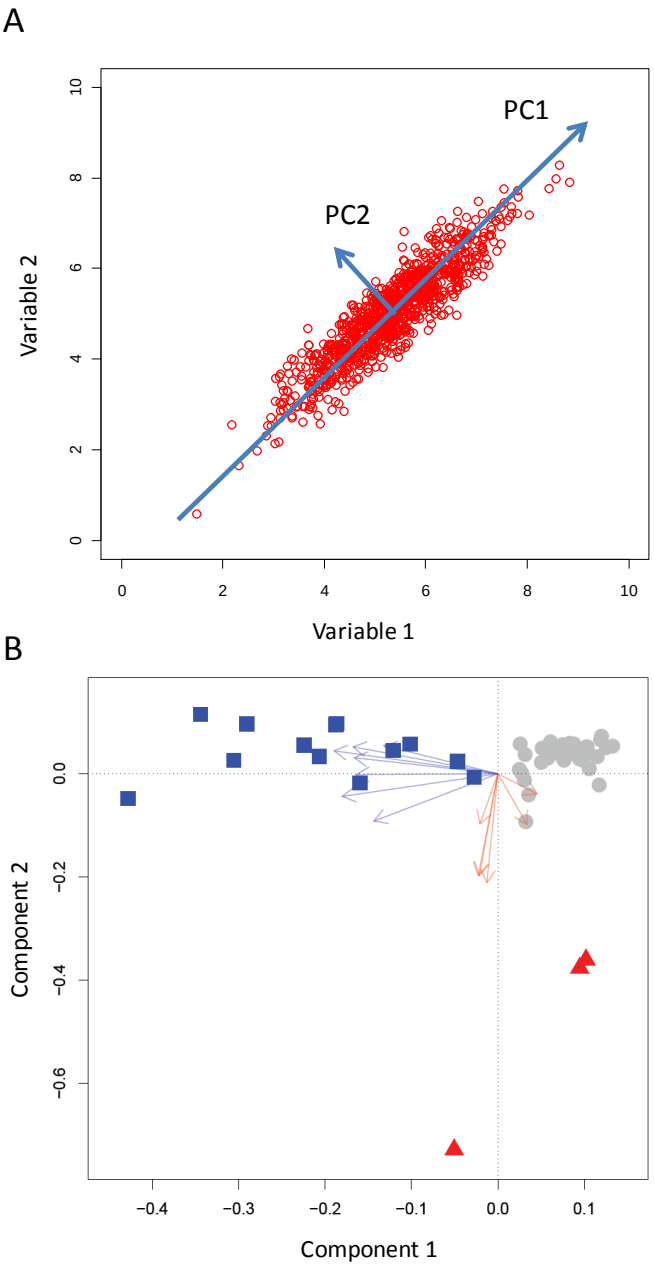


Figure 2.14. Principal component analysis and biplots. A. A two dimensional example of a dataset comprising two variables is shown. The first principal component, which accounts for as much variation in the dataset as possible, is shown as a blue line, labelled PC1. The second principal component, PC2, is calculated orthogonally to the first and represents the next largest amount of variation in the dataset. B. An example biplot is shown for a hypothetical dataset. The samples are represented as points on a scatterplot, showing the values for the first and second principal components, and have been grouped into three (grey circles, red triangles and blue squares). The arrows represent the projections of the variables that comprise the dataset and show how the red arrows represent variables that identify the red triangle group, and the blue arrows represent variables that identify the blue square group. The grey circles appear to be negative for all variables, since they are unaffected by the projections of the variables.

2.9.3 Non-negative matrix factorisation

Like PCA, NMF (non-negative matrix factorisation) (Brunet et al., 2004; Lee and Seung, 1999) is a methodology for complexity reduction in high-dimensional data. Whilst initially applied to image recognition and textual analysis (Lee and Seung, 1999), its utility for class discovery with gene expression data was subsequently recognised (Brunet et al., 2004; Kim and Tidor, 2003) and NMF is now a widely used technique for class discovery in biological research.

Essentially, the application of NMF to biological data involves taking high dimensional data and describing the variation within the dataset by reducing the variation to a small number of metagenes, a process conceptually similar to principal component analysis (section 2.9.2). NMF differs from PCA in that each metagene consists of a positive linear combination of variables, so that metagene expression is more biologically meaningful than for principal components, which can take negative values and are not intuitive. A single metagene, however, can be defined in terms of its correlative (and anti-correlative) genes in order to uncover biological meaning. After factorisation into a small number of metagenes, the dataset can then be clustered to reveal class memberships. The goal of NMF is, from an input matrix A, consisting of N probes and M samples, to identify a small number (k) of metagenes that best approximate the patterns of variation across the input matrix. This is equivalent mathematically to factorising the matrix A into W and H matrices. Matrix W is of size N probes by k metagenes and represents the correlation of each input probe with each metagene k. Matrix H is of size k metagenes by M samples and represents the expression of each metagene for each sample. A simple factorisation on a small dataset where k = 2 is illustrated in Figure 2.15.

Although the convention is to talk in terms of the 'expression' of metagenes, the application of NMF to methylation array datasets will nevertheless identify metagenes defined by methylation, whose 'expression' levels are related to the input methylation probes.

NMF has been reported to out-perform other commonly used clustering methods that had previously been applied to high-dimensional biological data, such as HC or self-organising maps (Gao and Church, 2011). These techniques are disadvantaged by their

sensitivity to the metric used to assess sample similarity and, for HC, there is a degree of subjectivity in the choice of clusters. However, the stability and reproducibility of HC can be improved, by bootstrapping the data and performing multiple clustering runs (Suzuki and Shimodaira, 2006) .

NMF also provides a powerful way to assist in classification across datasets. Let an input matrix A_1 be factorised into W_1 and H_1 matrices during the NMF process. Given a second input matrix A_2, after matching features between A_1 and A_2, a matrix W_2 can be calculated based on the correlation of the features of A_2 with the metagenes calculated from the initial dataset. Since NMF is defined by $A \sim W \times H$, and A_2 and W_2 are now known, H_2 can be inferred and represents the metagene expression profile in a second dataset, enabling comparisons between datasets that share the same NMF-based dimension-reduction procedure and therefore the same metagenes (Tamayo et al., 2007).

2.9.3.1 Methods to assess optimal numbers of metagenes

The R package NMF, contains methods useful for deciding upon the optimal number of metagenes, so that the underlying structure of the data can be identified. In particular, the cophenetic correlation coefficient is a measure of the fidelity with which the assigned metagenes represent the underlying structure of the data and can be compared across different numbers of metagenes to identify an optimal number.

Alternatively, a consensus clustering approach can be employed. By iteratively carrying out NMF on a subset of the dataset, extracting H values and then projecting them onto the whole dataset to calculate H values for the whole dataset and performing a classification on the projected H matrix, sample stability can be assessed at a range of metagene and cluster numbers to identify the optimal numbers of both. This is the basis for the approach described in detail in section 2.9.4.

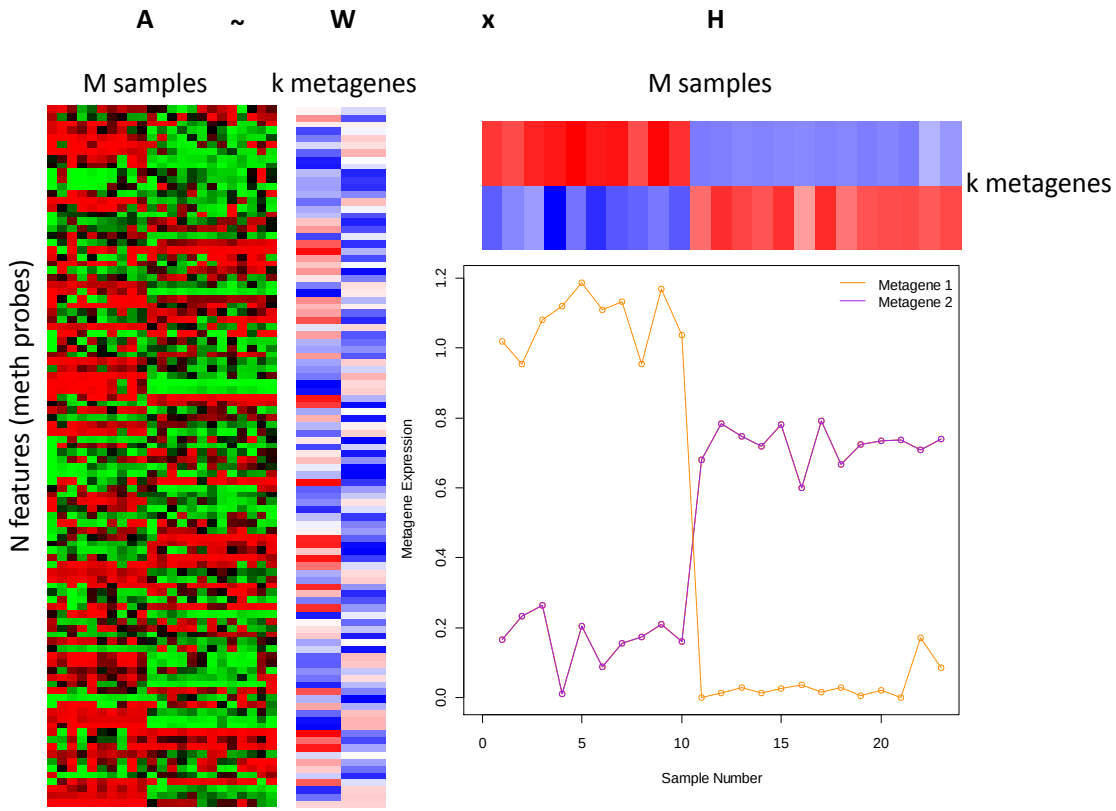


Figure 2.15. NMF can decompose high-dimensional data into discrete metagenes to reveal underlying structure. In this truncated methylation dataset, the heatmap A shows methylation probes (*n* = 100) for medulloblastoma samples (*n* = 23). Unmethylated probes are shown green, methylated probes in red. Hemi-methylated probes are black. The NMF process factorises the matrix A into two matrices W and H. Matrix W has size N rows by k columns, where k is the number of metagenes. Each data point for matrix W represents the coefficient of a methylation probe with that particular metagene. Probes with low correlation to the metagene are shown in blue, high correlation in red. Matrix H has size k rows and M columns, where each column represents the metagene expression profile for a given sample. For matrix H, high metagene expression is shown red, low metagene expression in blue. In this example, there are two metagenes, corresponding to two discrete groups. The matrix A is factorised into matrices W and H, revealing two classes characterised by their metagene expression. The line graph numerically shows the expression of metagenes, using the same data shown in the heatmap above, to illustrate the clear difference in metagene expression between the two classes.

2.9.4 Consensus-based unsupervised clustering approach

Unsupervised cluster analysis was undertaken in order to investigate whether there were any subgroups of medulloblastoma, defined by differences in their methylation patterns. A consensus-clustering based approach was applied. The approach, described below in detail, was initially used to determine whether using the whole dataset or a dataset where non-informative probes were removed, produced the most stable clusters. Maximum stability was achieved using filtered data. Retained probes had at least 5% of samples with a β score over 0.17 *or* at least 5% of samples with a β score less than 0.83. These figures (0.17 / 0.83) were chosen since the limits of detection of this assay are reported to be 0.17 (Bibikova et al., 2006).

The second step involved selecting the optimal number of clusters within the training cohort using NMF (see section 2.9.3). For every combination of 2:6 (*x)* metagenes and 2:6 *(y)* clusters, the following steps were repeated 100 times:

1: Randomly select 80% of the dataset
2: Perform NMF for *x* metagenes on this randomly selected subset
3: Project H values from this factorisation onto whole dataset
4: Classify by *k* means algorithm into *y* clusters using projected H values
5: Store classes for each sample

For each combination of metagenes and clusters, the most stable clustering solution was identified by examining the average sample assignment consistency *i.e.* the average modal cluster score across all samples. After selecting the optimal number of metagenes and clusters, non-classifiable (NC) samples were defined as samples which were assigned to the same cluster less than 80% of the time.

After selecting the optimal number of metagenes and clusters, the consensus clustering approach was applied to the test dataset to pre-emptively identify NC samples within that cohort. This analysis was repeated using datasets with both intact X chromosome probes and with X chromosome probes removed. The consideration of the X chromosome is important, since females with diploid X chromosomes will have one X chromosome silenced through methylation (Wutz and Gribnau, 2007), whilst the single X chromosome of males will be predominantly unmethylated in euchromatic regions. Retaining the X chromosome probes might induce uncertainty in cluster

assignment by confounding with gender specific differences. Since removing X chromosome probes produced a more stable solution, subsequent analyses were carried out on datasets in which the X chromosome probes had been removed.

The third step involved taking the training dataset from which the NC samples were removed and implementing NMF with the optimal number of metagenes. This process was repeated 200 times and the most stable solution selected. The H values for the training dataset were extracted and used to project H values for the test dataset (see section 2.9.3).

The *k* means algorithm (see section 2.9.5) was used to assign samples to fixed classes using the optimal number of clusters identified in the second step, using 1000 iterations and 50 starts. These class assignments formed the basis for subsequent construction of metagene classifiers.

2.9.5 K means clustering

K means clustering is a method of cluster analysis that partitions samples into a pre-specified number of clusters, so that each sample belongs to a cluster with the nearest mean. It is an iterative algorithm, repeating until convergence is reached. It is limited because the number of clusters has to be pre-specified and also that the cluster solution depends on the initial conditions, so that the algorithm has to repeated many hundreds of times to reach a stable solution.

The algorithm consists of two parts. In the first step, each sample is randomly assigned to one of *k* classes. In the second step, the initial cluster means are calculated to be the centroid of the cluster's randomly assigned members. *K* clusters are then created by associating each sample to its nearest cluster mean. The centroid of each cluster is recalculated and becomes the new mean. Samples are reassigned to one of *k* clusters by associating each to its nearest cluster mean. Step two is iteratively repeated until convergence is reached. The process is illustrated in Figure 2.16.

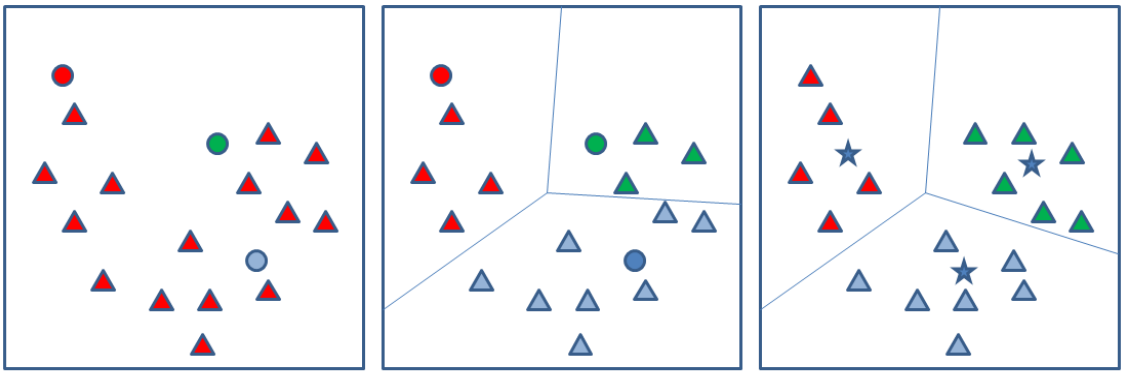


Figure 2.16. The *k* means algorithm for sample classification. The first steps of the algorithm are illustrated. The first step is spread across two panels. Three samples are randomly assigned to be cluster means in the first step (shown as differently coloured circles). In the second panel, samples are assigned to groups on the basis of their nearest mean. In the third panel, the sample means are re-calculated (means no longer correspond to samples and are shown as stars) and the samples reassigned based on the new mean values for the cluster centres. The process is repeated until stability is reached (i.e. there is no change in class assignment).

2.10 Assessment of clustering assignment

One powerful way to assess cluster assignment is also one of the simplest. Plotting the loadings of principal components after PCA (see section 2.9.2) can quickly reveal underlying patterns within the data. If desired, covariance spheroids (see below) can be plotted over putative class assignments to reveal whether there is evidence for separation of the classes. This is illustrated in Figure 2.17.

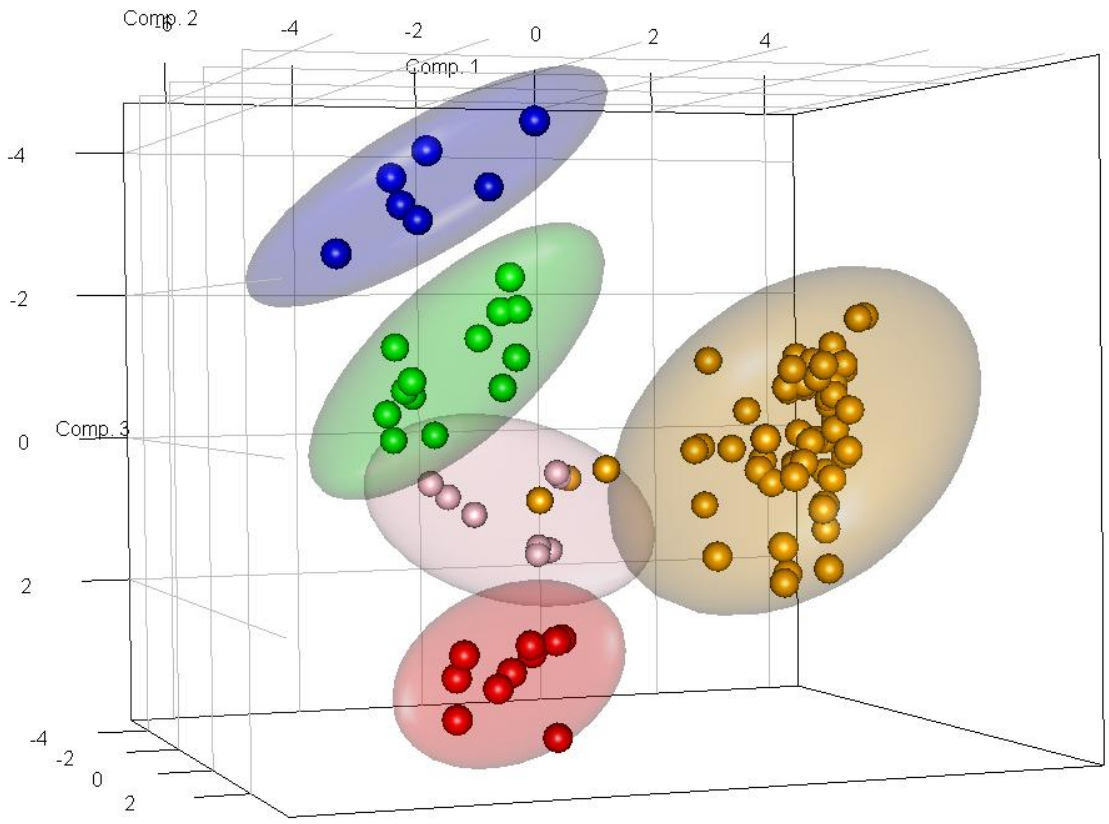


Figure 2.17. Assessment of cluster assignment by PCA. 3D Plot shows subgroup membership for illustrative data plotted against the first three principal components. Component axes are labelled (Comp. 1, 2 and 3). Samples have been assigned to 5 subgroups (red, pink, orange, green, blue). Covariance spheroids have been calculated and the 95% confidence intervals plotted for each subgroup, shown in translucent colours. In this case, three orange cases are overlapping the pink group and are potentially misclassified.

2.10.1 Silhouette Plots

Silhouette plots are a useful, easily interpretable graphical measure of cluster assignment (Rousseeuw, 1987). First, a distance measure is calculated from the dataset in question. This distance measure is used to calculate the average dissimilarity between every sample and the other members of the cluster to which the sample belongs (*a*). For each cluster for which the sample is not a member, the average dissimilarity between the sample and the cluster mean is calculated. The smallest dissimilarity indicates the cluster neighbour of the sample in question and is designated as *b*. Finally, the silhouette score is calculated by subtracting *b* from *a* and dividing by the maximum of b or a, in order to ensure minimum and maximum silhouette scores of -1 and 1 respectively. Observations with a silhouette score close to 1 are very well clustered, since they are not near neighbour clusters. A score of around zero indicates samples that may lie between two clusters and negative scores indicate samples which are potentially placed into the wrong cluster. An example silhouette plot is shown in Figure 2.18.

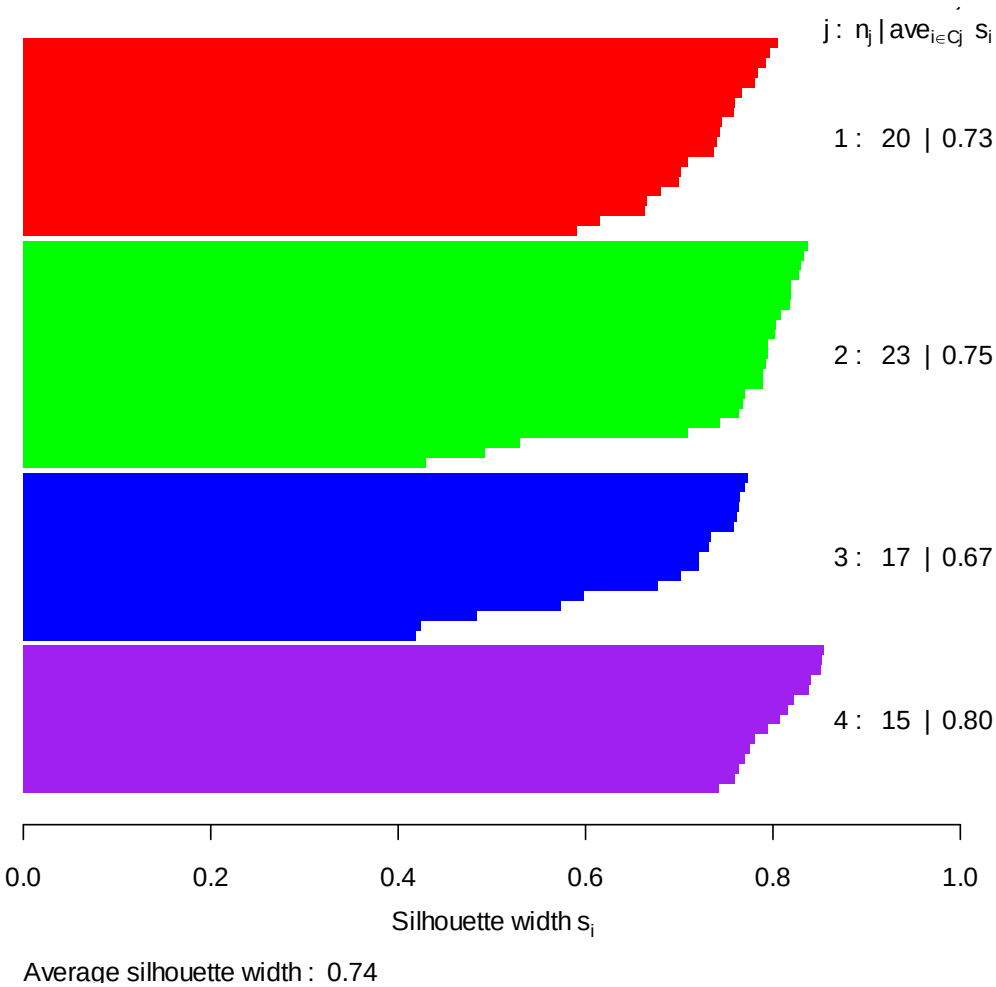


Figure 2.18. Silhouette plots are useful for the assessment of cluster assignment. Each sample is represented by a horizontal bar. Clusters are shown in different colours. For each cluster, samples are ranked in descending order of silhouette score. In this example, there are four clusters, with all samples having a positive silhouette score indicative of a satisfactory clustering. Alongside each cluster, information is included. J is the cluster number; n_j is the number of samples within each cluster; $ave_{i \in cj}$ s_i is the average silhouette width per cluster.

2.11 Over-fitting

The reason for carrying out class discovery in cancer is to identify underlying structures within the data that might ultimately be useful for disease classification, prognostication and / or informing treatment choices.

Due to the high dimensionality of genome-wide assays, the classification of samples using this type of data is very prone to over-fitting, and may lead to the identification of biologically irrelevant subgroups. Classifiers can be constructed that, while performing well on the original dataset, are not reproducible in future studies (Ransohoff, 2004). In particular, for cancers with relatively low incidence such as medulloblastoma, the number of samples in genome-wide studies will be vastly outweighed by the number of variables assessed. One approach to ameliorate this problem is to split the dataset into training and test cohorts, carrying out class discovery in the training cohort and validating any findings in the test cohort. The absolute separation of test data, which are used for validation only, from training data should prevent over-fitting when used to validate classes identified in the training dataset, and is an absolute requirement for any model that may ultimately influence clinical decisions.

2.12 Support Vector Machines

Once class membership has been confidently assigned, classifiers can be used to predict class membership both in the training and test cohort. Predicting membership within the training cohort enables assessment of classifier performance on the dataset from which it was derived. Subject to satisfactory performance on the training dataset, subsequent classification of the test dataset can be undertaken. Class uncertainty can be calculated and samples whose classification is less robust can be identified, and, if necessary, removed from further analyses. Next, clinico-pathological correlates of the assigned groupings can be tested across the training and test datasets to ensure that patterns of correlation are maintained across class, regardless of whether the sample arose from training or test datasets.

The most widely used method of classification in genome-wide studies is support vector machines (SVMs) (Chang, 2001), and has been reported to be the most effective

classifier when applied to the classification of cancer from transcriptomic data (Statnikov et al., 2005). SVMs appear to be insensitive to the high-dimensionality of genome-wide datasets and perform well in a cancer classification context using small numbers of samples (Irgon et al., 2010). Initially applicable only to binary classification, subsequent work demonstrated the utility of multi-category SVMs (MC-SVMs) in cancer (Lee and Lee, 2003). The construction of an SVM is outlined in Figure 2.19. Essentially, training vectors (i.e. variables) are mapped into higher dimensional space by a function ϕ. The SVM then identifies a linear separating hyperplane that separated classes by a maximal margin in this higher-dimensional space. The function ϕ is known as the kernel function and is most commonly linear or radial (see section 2.12.1). Every kernel has a cost parameter, which is the penalty parameter of the error.

MC-SVMs are an extension of the binary SVM described above. For this scenario, the multi-class problem is split into multiple binary classifications, comparing every combination of pairs of classes. Classification is then carried out by a voting strategy, where each combination of class-pair classifiers assigns each sample to one of two classes. All the comparisons are tallied and for each sample, the class with the most votes is returned.

2.12.1 SVM Kernels

The choice of kernel for constructing a SVM depends upon the complexity of data being considered. For high-dimensional data, such as genomic or transcriptomic data, it has been reported that the linear kernel performs well (Furey et al., 2000). For less complex data, a radial kernel is often a more suitable choice. A radial kernel is illustrated in Figure 2.19, where data are shown that are not linearly separable, but become so when transformed into higher dimensional space. When using a radial kernel, it is necessary to specify an additional parameter, gamma, that is optimised during the tuning process (see section 2.12.2).

2.12.2 Tuning SVM model

In order to construct a well-performing SVM, it is necessary to tune the model to identify optimal parameters. At the same time, it is important to avoid over-fitting to the training dataset. This can be avoided by employing cross-validation (section 2.12.3).

During the tuning process, different values of the cost and gamma parameters, if using a radial kernel, and cost only if using a linear kernel are used to construct a SVM, and the accuracy on the training set checked through cross-validation. The optimal value(s) for the parameter(s) is / are returned.

2.12.3 Cross-validation

During the tuning process, cross-validation can be employed to try to avoid overfitting the model to the training data. In this project, leave-one-out cross validation (LOOCV) was used to ameliorate the effects of over-fitting to training datasets. This involves taking in turn a single sample from the training dataset (*n* samples in total) as the validation dataset and all remaining samples (*n*-1) as the training dataset. The SVM is constructed on the training set and used to predict class membership for the single validation sample. This is iteratively repeated until each sample will have acted as a validation set once. The cross validation accuracy is the percentage of data which are correctly classified compared to the SVM constructed using the whole dataset.

2.12.4 Confusion Matrices

A confusion matrix is a simple way to assess classifier performance. If the classifier is applied to a dataset in which the class is already known, the classes assigned by the classifier can be directly compared, and any discordant classifications can quickly be identified. An example confusion matrix is shown in Table 2.5.

		True Class			
		A	B	C	D
Classifier-assigned class	A	25	0	0	1
	B	0	30	0	0
	C	0	0	57	0
	D	0	0	0	18

Table 2.5. Confusion matrices and their interpretation. In this case, true class (A to D) is shown in columns, and the class assigned by the classifier (A to D) is shown in rows. The number of class assignment matches can be assessed. In this example, 1 case with a true class of D has been incorrectly classified as A by the classifier.

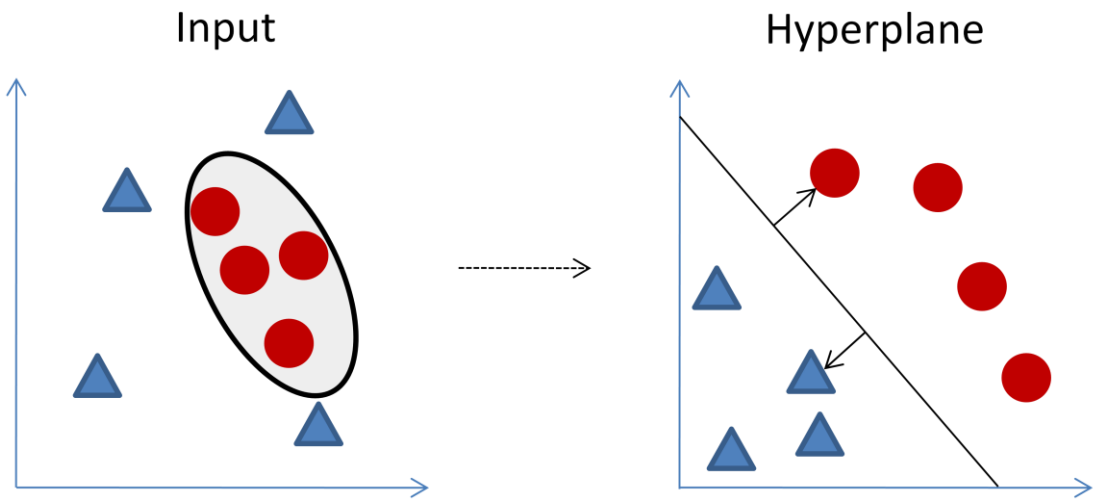


Figure 2.19. Principles of SVM. The figure shows samples from two classes (blue triangles and red circles). Input data are impossible to separate linearly. By transforming data using the specified kernel function, a transformation ϕ is used to construct a higher-dimensional hyperplane, so that the separation between the classes is maximised. Samples that lie closest to the margin, indicated with arrows, are known as support vectors. The line that separates the classes is known as a maximum margin classifier. If it is not possible to separate the classes without error, soft margins are used that split the classes optimally, maximising distance between cleanly split samples.

2.13 Survival analysis in medulloblastoma

2.13.1 Event free and overall survival in medulloblastoma

For survival studies of medulloblastoma, two outcomes are commonly measured: Event free survival (EFS) and overall survival (OS). EFS and OS are defined as the time periods between diagnosis of disease (or commencement of treatment) to relapse or progression / death, respectively. Within medulloblastoma research, EFS has been favoured over OS as a measure of survival since, after relapse, patient survival is exceedingly rare, but time to death may be influenced by second-line therapeutic strategies. The decision to commence further therapies following relapse is variable, with the majority of further treatment being palliative. EFS is a more accurate reflection of survival risk, since the period between relapse and death can be variable due to factors unrelated to disease severity.

2.13.2 Kaplan-Meier plots and log-rank test

The construction and analysis of survival curves is fundamental to the identification of prognostic factors in cancer. The Kaplan-Meier estimator (Kaplan and Meier, 1958) is used to measure the proportion of patients living with or without an event occurring for a set time following diagnosis or treatment. When a large enough sample size is studied, the survival curves will approximate the survival in the whole population. For medulloblastoma, the subdivision of the disease according to clinico-pathological features and / or molecular markers is desirable, so that treatment options can be tailored to patient risk, ensuring that patients with a good prognosis are treated with protocols that, while maintaining cure rates, will reduce therapy-related side effects in later life, whilst continuing to aggressively treat patients with a poor prognosis. An example Kaplan-Meier plot is shown in Figure 2.20.

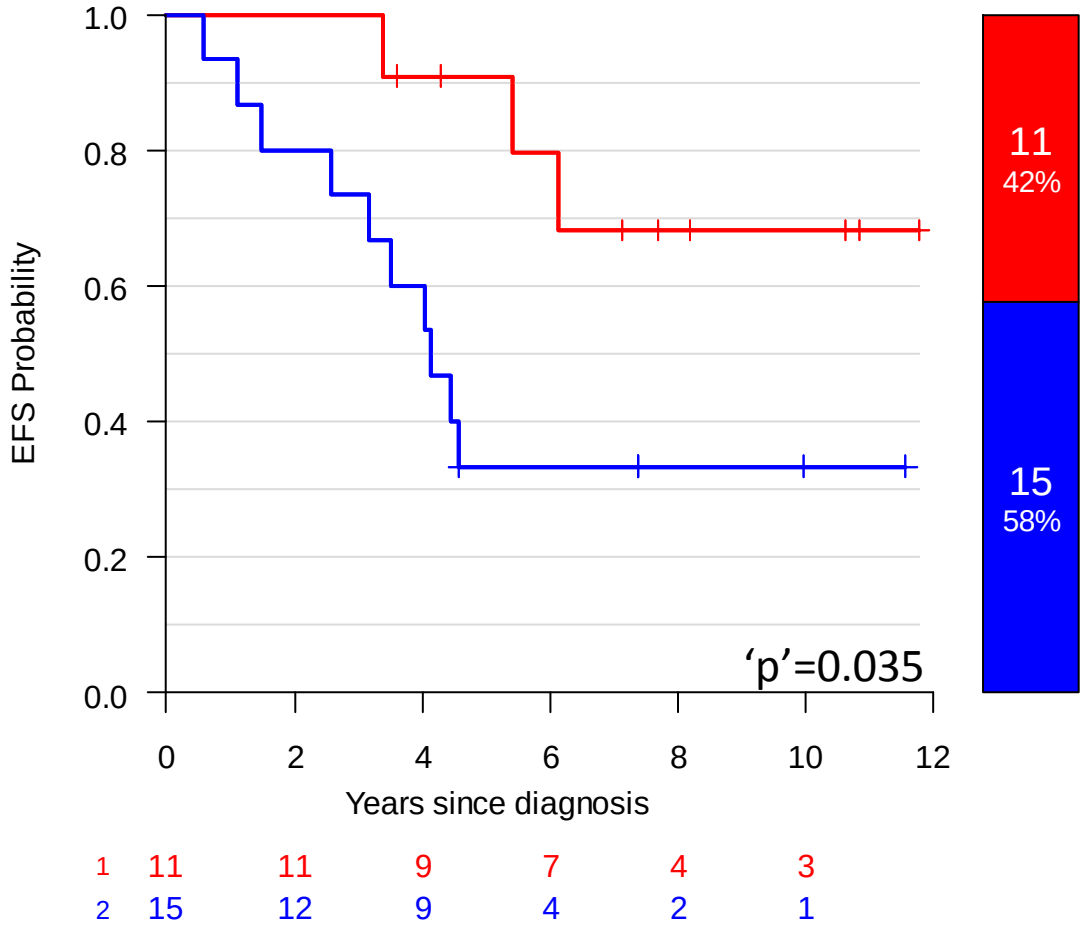


Figure 2.20. Example of a Kaplan Meier curve. Graph shows a small hypothetical cancer cohort that is assessing the difference (if any), between two treatments (1 (red) and 2 (blue)). The two lines represent the observed proportion of relapses in the cohort. Since patients will be recruited to the trial at different time points, the follow up time will be different between patients. If a patient has not yet relapsed at the end of the observation period, the patient is said to be censored at that point. These censor points are marked with pluses on the curve. Relapsing patients will be identifiable where the curve steps downwards. In this plot, the bar graph on the right indicates the absolute number and proportion of patients assigned to each treatment. The table below the graph is an at-risk table and is used to assess the effective population size for each treatment at two-yearly intervals. The number will reduce over time due to patient relapse or through censoring. On the bottom right, the log-rank test assesses whether the difference in the curves observed is statistically significant (see next paragraph).

For statistical interpretation of survival data, the log-rank test (Bland and Altman, 2004; Mantel, 1966) is applicable to censored data. It tests the null hypothesis that there is no difference in survival between two or more groups. It calculates the difference between the observed number of events and the expected number of events if there was no difference between the populations at each timepoint where an event occurs. A chi-squared test of the null hypothesis, at the appropriate degrees of freedom, can be used to calculate a 'p' value. Cox proportional hazards models, which additionally identify the severity of associations, can assess the effects of more than one covariate and are applicable to continuous covariates, will be discussed in section 2.14.

2.14 Cox proportional hazards model

The Cox proportional hazards model (Bradburn et al., 2003; Cox, 1972) is the most widely used multivariate approach for analysing survival data in a clinical setting. At its most simple, it models survival times (or, more specifically, the hazard function) using explanatory variables.

The hazard function is the probability that any individual will experience an event during a given time interval, given that the individual has survived to the beginning of the interval. The Cox proportional hazards model is a form of (multiple) regression, where the relationship between the dependent variable, the hazard function, is compared against explanatory variables.

For each considered explanatory variable, a hazard ratio and 'p' value is reported. The 'p 'value is calculated by dividing the natural log of the hazard ratio by its standard error, and comparing the result against the normal distribution. Values greater than 1.96 are significant at a 5% level. 95% confidence intervals can be calculated for the estimated hazard ratio to determine whether or not this interval includes a hazard ratio of 1, indicating a non-significant covariate. An example output from a Cox proportional hazards model and its interpretation is shown in Table 2.6.

Covariate	Hazard Ratio	95% Confidence Interval (CI)	'P' value
Age	0.91	0.74 – 1.13	0.40
Histology	1.56	1.05 – 2.33	0.029
Treatment	0.73	0.59 – 0.91	0.0052

Table 2.6. An example Cox model from a hypothetical cancer dataset. In this multivariate analysis, each tested covariate is included in a multiple regression against survival. For each covariate, the calculated hazard ratio, 95% confidence interval and 'p' value are given. For continuous variables such as age, the hazard ratio refers to the increase in hazard for an increase of 1 in the value of the covariate. In this example, the risk of an event decreases by 9% if the patient is a year older, after adjustment for other covariates in the model. However, this is not significant, and the 95% confidence intervals span 1, suggesting that age has no bearing on survival. This is in contrast to categorical variables such as histology, where the presence of the histology being tested confers a 56% increase in the risk of an event. Since the 95% confidence intervals do not include 1, this is interpreted as significant at the 5% level. The application of the treatment being tested confers a 27% decrease in the risk of an event, and this is significant at the 1% level.

2.15 Selection of potentially testable putative prognostic methylation probes

Before the identification of prognostic covariates can begin (section 2.16), probes with low variability or low range can be filtered out, since they are unsuitable for consideration as predictive biomarkers, because they are non-testable. Probes with at least 5% of cases highly methylated (β score > 0.8) *and* at least 5% of cases unmethylated (β score < 0.2) were selected as being potentially prognostic. The 5% cutoff was chosen since this is approximately the frequency with which amplifications of the prognostic oncogenes *MYCC* and *MYCN* are observed in the disease (Pfister et al., 2009; Rutkowski et al., 2007).

2.16 Methods of measuring bimodality

An idealised prognostic methylation marker would show a sharply bimodal distribution with modes at 0 and 1 that correlated well with survival. This bimodal difference would be easier to distinguish in subsequent tests than a methylated probe

with a unimodal distribution. To ensure that probes potentially testable in future assays of prognostic probes would be easily testable, the bimodality index (Wang et al., 2009a) can be applied to the pre-filtered methylation probes (described in section 2.15). For each probe tested, the bimodality index is calculated, a measure of the probe bimodality. Since the bimodality index is returned as a continuous variable, probe lists can be ranked in order of bimodality. While previously applied to high-throughput transcriptomic data, histograms of the β score distribution of the top three most and least bimodal probes (Figure 2.21), selected from a pre-filtered selection of methylated probes (section 2.15), demonstrate that this technique is also applicable to high-throughput methylation data.

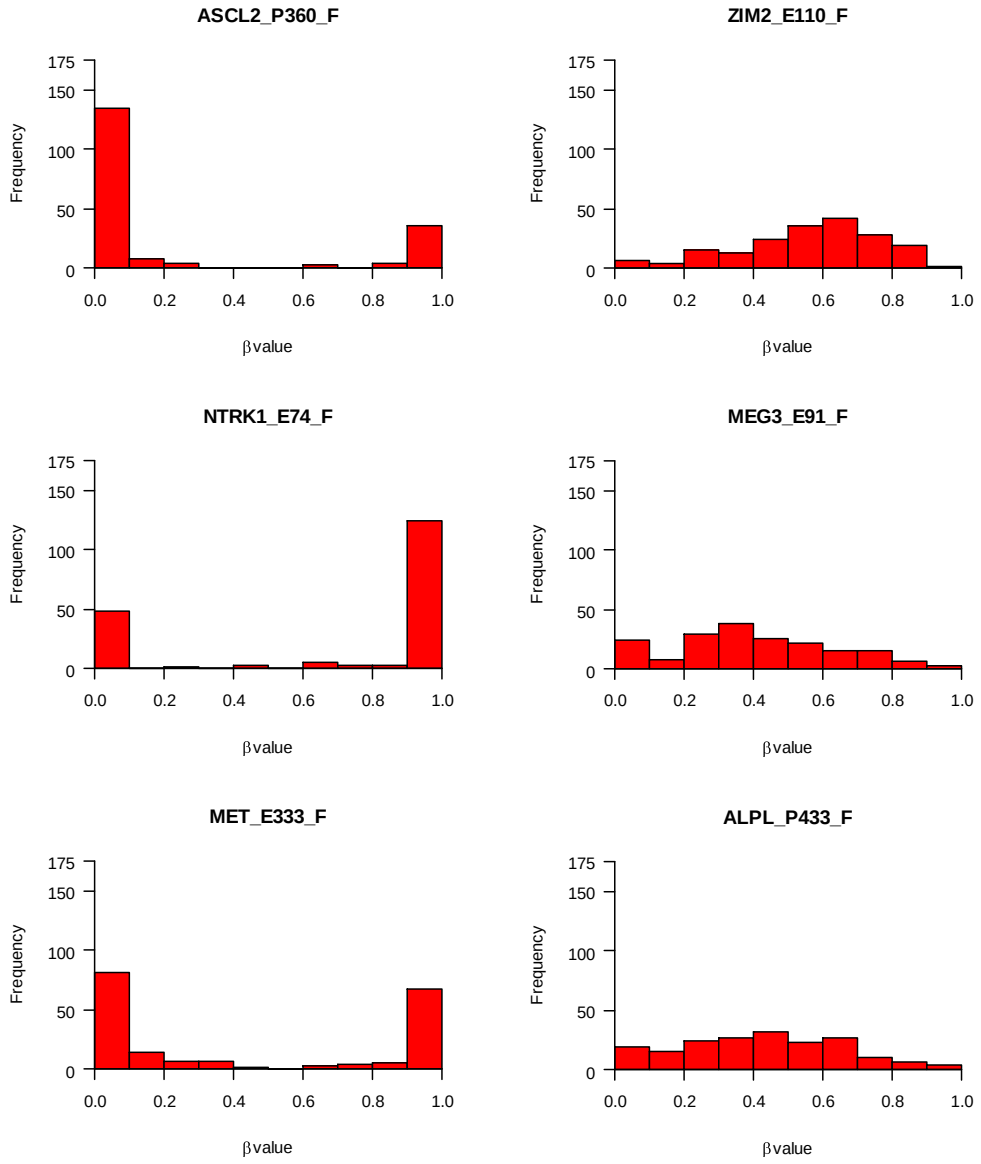


Figure 2.21. Bimodal index can identify probes compatible with subsequent use as prognostic biomarkers. Most (left hand column) and least (right hand column) bimodal methylation probes are shown as barplots. Methylation β score, classed into ten bins is shown on the x axis, with frequency per bin on the x axis.

2.17 Cox boost algorithm

The application of high-dimensional data, typical of transcriptomic or methylomic array-based experiments, in a survival setting is prone to over-fitting (section 2.11), since the number of potential covariates will be much larger than the number of samples on which the study is based. Strategies, all utilising cross-validation-based methods (section 2.12.3), have been explored to overcome this limitation. Supervised principal component survival analysis (Bair and Tibshirani, 2004) combines the effects of many weak predictors of survival to identify stronger predictors of survival. This approach is limited since it reports multi-component survival markers without taking into account any existing clinical variables. This raises the possibility that any reported significant associations are simply due to correlations with previously identified clinical survival correlates.

A more desirable approach would select covariates from high-dimensional data that are complementary to existing clinical survival covariates, avoiding the identification of covariates that are simply correlating with established clinical covariates. This is the approach taken by the Cox boost algorithm (Binder et al., 2009), which tests covariates for prognostic potential alongside mandatory clinical variables. It results in a sparse classifier, i.e. the algorithm preferentially chooses a smaller number of high-dimensional covariates with additional predictive value, with the majority of covariates having no predictive power. To ensure sparseness, a penalty score is chosen which balances the predictive power of each potentially included covariate.

Desired clinical covariates are forced into the model and subsequently, a boosting approach is used to identify potentially prognostic optional covariates. Boosting is a general method of formulating a very accurate predictor by combining a number of less accurate predictive covariates. In this context, a set of predictive covariates from high-dimensional data can be combined with mandatory clinical covariates to produce an optimal survival model. In each boosting step, the estimated prognostic potential for each optional covariate is recalculated, based on the penalty score for including extra covariates and offset by previous boosting steps. After the optimal number of boosting steps, a final model is reported.

For this project, the initial Cox boost model was specified to include the appropriate mandatory clinical variables. Subsequently, the optimal parameters for picking the optional covariates could be selected. The parameter for the optimal number of boosting steps was chosen using LOOCV with respect to the partial log-likelihood. Next, the optimal penalty for including optional covariates was chosen at the optimal number of boosting steps. By optimising the number of boosting steps through cross-validation, the effects of over-fitting can be reduced (Binder et al., 2009).

Finally, the optimal model was calculated, which identified any optional covariates with added prognostic potential. Permuted 'p' values were calculated for the selected optional covariates using LOOCV.

2.18 ROC curves to assess predictive performance

Originally formulated by Allied forces during World War II to help quantitatively assess the verisimilitudinous identification of enemy objects on radar traces, the **R**eceiver **O**perating **C**haracteristic (ROC) plot is a visualisation of the sensitivity (or true positive (TP) rate) versus false positive (FP) rate (which is defined as 1 – specificity).

ROC curves are most often used for the assessment of classification. A perfect classification would have 100% sensitivity and 100% specificity, and a random classifier would have 50% sensitivity and 50% specificity (Figure 2.22).

They have been applied in a clinical setting, where they have several useful features. Firstly, the power to discriminate TPs can be visually identified without resorting to a specific threshold of detection. They are also useful for assessing the discriminatory power from several different classification schemes. Finally, the area under the ROC curve (AUC) can be interpreted as the probability that a test result from a randomly selected individual positive for the outcome being assayed exceeds that of a randomly selected individual negative for outcome, and is a useful summary of the predictive power of the classifier being considered.

Initially applied to assaying disease status and other binary outcomes in a disease setting, the application of ROC curves to survival analysis has been described (Heagerty et al., 2000), and is a useful measure of the predictive power of a disease classifier in a

survival context. An example ROC curve applied to a hypothetical survival cohort and its interpretation is shown in Figure 2.22.

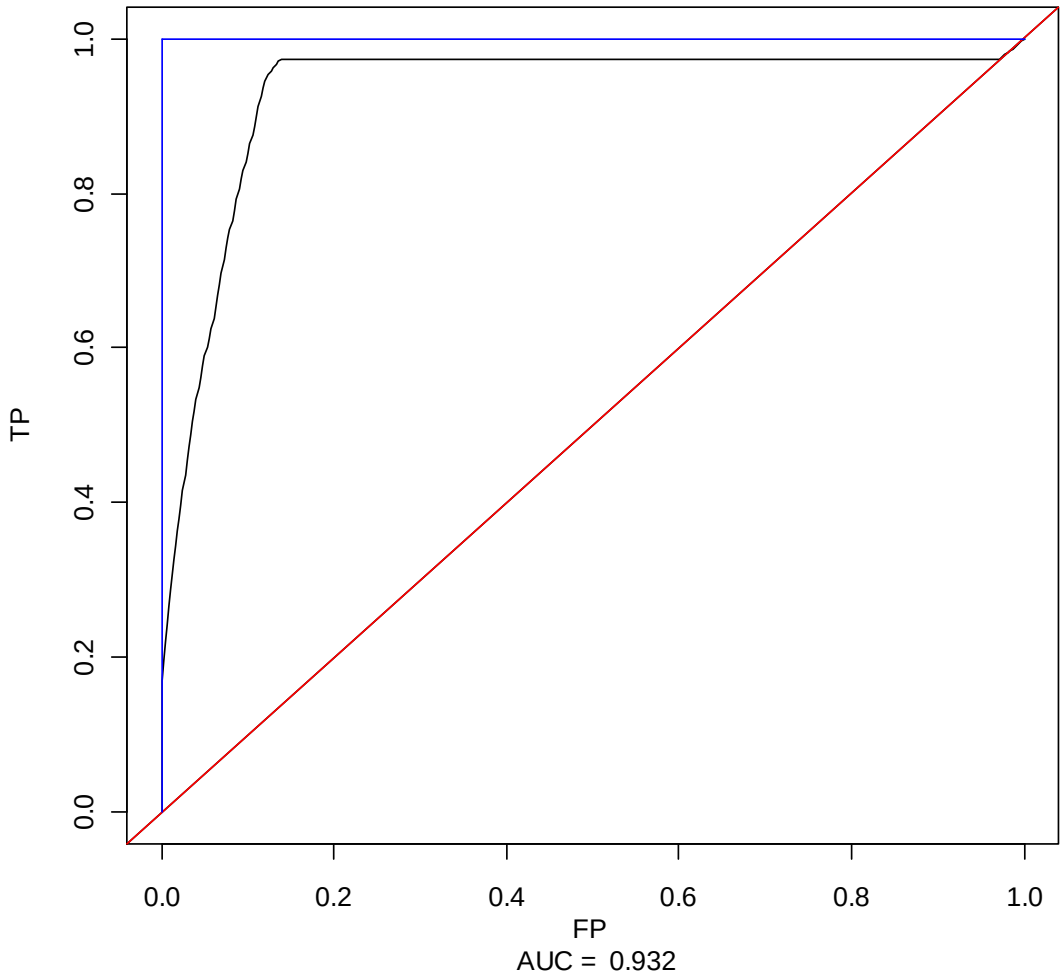


Figure 2.22. The receiver operating characteristic (ROC) curve is useful for assessing classification performance. The ROC curve shown in black is based on a hypothetical cancer survival cohort. For a given time period of follow-up (in this case, 1 year), the classifier (black) demonstrates how the true positive rate (TP – y axis) varies with the false positive rate (FP – x axis). The classifier is performing very well, since for a given true positive rate, the false positive rate is much smaller; in the plot, this is shown by the steeply ascending ROC curve. This is also shown by the area under the curve (AUC) of 0.932. By plotting ROC curves at different time points, the utility of the classifier across different follow up periods can be assessed.

For comparison, a classifier with no predictive power (red line) is also shown along the unit line, demonstrating a situation where it is impossible to tell a true positive from a false positive. A classifier with perfect predictive power (blue line), has an AUC equal to 1.

2.19 Nomograms

Nomograms are a useful method for the graphical representation of the strength of risk factors in a multi-variate model of survival. Although they have fallen out of favour with the widespread adoption of computers (a nomogram can easily be replaced by a simple function), they remain useful for visual inspection of effect sizes in regression models, such as the Cox proportional hazards model.

In this project, nomograms were used to visually compare the risk factors from a Cox proportional hazards model (section 2.14). To assess patient risk, the points score for each risk factor is calculated and the total risk score is then used to read off the hazard score, which is a linear predictor of risk. This is related to the hazard ratio returned from the Cox proportional hazards model, which is the exponent of the linear predictor of risk. Nomograms were plotted using the R package *Design*. An example nomogram is shown in Figure 2.23.

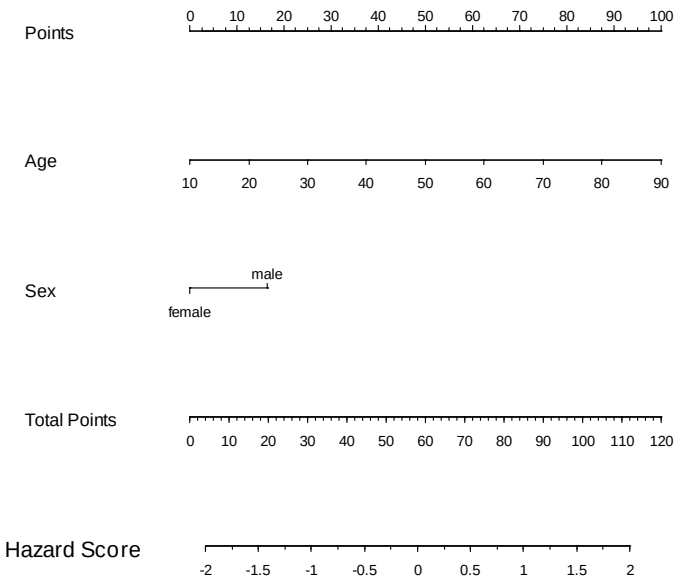


Figure 2.23. Nomograms are useful for visual inspection of the strength of the effects of selected risk factors. In this hypothetical example, two factors (age and sex) are considered for their effect on survival. The points score for each case is calculated based on its age and gender. For example, a 50 year old male would receive 50 points for age and 16 points for being male, giving a total of 66 points, equating to a hazard score (linear predictor) of 0.25. This is equivalent to a hazard ratio of 1.28 relative to the null model.

2.20 Classification and Regression Trees

The classification and regression tree (CART) method for risk stratification was used to identify variables that best separate patients into differential risk groups. The technique is used to create a model that predicts the target variable (in this case, survival) based on several risk factor covariates. The cohort is successively divided into groups of cases with similar response patterns, at each step splitting a node into two subgroups using the covariate that best discriminates survival outcomes using the likelihood ratio test. This continues until subgroups have reached a minimum size or when no covariate can further split subgroups.

It has the advantage that it is simple to understand and interpret, but can be subject to over-fitting (section 2.11). The tree can be cross-validated to avoid this limitation. The R package *rpart* was used to generate the decision tree shown, using a minimum subgroup size of 5 cases and LOOCV (section 2.12.3) to avoid over-fitting. An example decision tree is shown in Figure 2.24.

2.21 Statistical Analyses

Statistical analyses were performed using the R program, version 2.12 (R foundation). Statistical significance was tested using *t* tests, analysis of variance (ANOVA), Fisher's exact test, Tukey's honest significant difference and Chi-squared test. Survival associations were tested using log-rank tests (section 2.13.2) and Cox proportional hazards test (section 2.14). 'P' values were uses to assess significance. Further details of specific analyses are shown where appropriate.

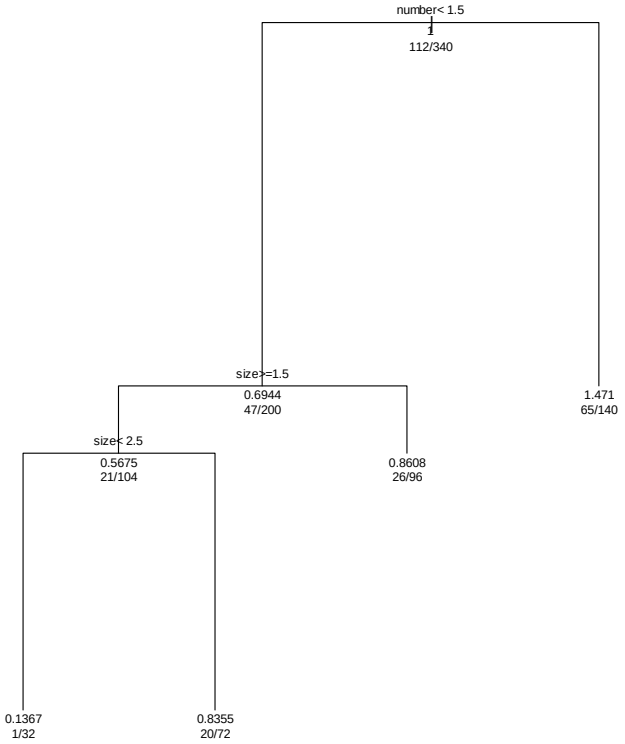


Figure 2.24. CART analysis provides an easily interpretable means of assigning risk stratification. In this hypothetical dataset, two factors, number and size of tumours are considered. The decision tree can be read according to the instruction at the top of each node. For example, the first node considers the number of tumours. In this case, 200 cases have less than 1.5 tumours and are assigned to the left hand branch. The remaining 140 cases are assigned to the right hand branch, which ends in a terminal node, with a hazard ratio of 1.47. Below each hazard ratio, the total number of cases falling within that node is shown as the denominator, with the number of cases experiencing relapse is shown as the numerator. So for the right hand terminal node, 65 / 140 cases experienced a relapse. In this way, the tree identifies four terminal nodes, with hazard ratios of 0.137, 0.836, 0.861 and 1.471. Clearly, in this dataset, a patient with one tumour of between size 1.5 and 2.5 is at a significant survival advantage compared to the remaining cases,

Chapter 3. Design and implementation of an assay for rapid diagnosis of medulloblastoma subgroups

3.1 Introduction

The constitutive activation of WNT and SHH signalling pathways (Clifford et al., 2006; Ellison et al., 2005; Raffel et al., 1997) has been shown to play important roles in the pathogenesis of medulloblastoma (section 1.3.9.2, 1.3.10.2). Moreover, the mutation of specific pathway components represents the majority of mutations associated with medulloblastoma to date.

The SHH pathway is activated by mutation of *PTCH1* in approximately 10% of medulloblastomas (Ellison et al., 2003) and loss of *Ptch1* has been shown to drive tumourigenesis in mouse models of the disease (Kim et al., 2003; Goodrich et al., 1997), where it stimulates inappropriate proliferation of GNP cells (section 1.3.9.1). Similarly, mutation in WNT pathway components have been described in up to 20% of cases (Dahmen et al., 2001; Eberhart et al., 2000; Huang et al., 2000).

Activation of these pathways in medulloblastomas is an important clinical indicator. WNT pathway activated medulloblastomas have a favourable prognosis (>90% OS) (Gajjar et al., 2006; Ellison et al., 2005) and small-molecule inhibitors of the WNT pathway are in the early stages of development (Lu et al., 2011; Baryawno et al., 2010). Small molecule inhibitors of SHH pathway activated tumours show pre-clinical and early clinical activity against the disease (Rudin et al., 2009; Romer et al., 2004).

Transcriptomic investigations of the disease (section 1.3.12.1) have revealed that medulloblastoma comprises distinct molecular disease subgroups (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006). The precise number of subgroups reported varies from 4 to 6, however two subgroups characterised by activation of the WNT and SHH pathway are consistently reported by these studies.

The WNT subgroup is the best described molecular disease subgroup, and is characterised by mutation of *CTNNB1*, nuclear accumulation of β-catenin and chromosome 6 loss (Fattet et al., 2009; Clifford et al., 2006; Thompson et al., 2006; Eberhart et al., 2000). The determinants of SHH subgroup membership are less well defined. Mutations in *PTCH1* are only reported in a subset of SHH cases, indicating a role for other mechanisms and correlates. A number of putative determinants of SHH pathway activation (e.g. *PTCH1* hypermethylation, mutation of *SUFU* / *SMO* and *REN* (*KCTD11*) genetic loss) have been reported in medulloblastoma (Diede et al., 2010; Di

Marcotullio et al., 2004; Taylor et al., 2002; Reifenberger et al., 1998), but their relationship to the SHH subgroup remains to be established. Moreover, the clinical features of the SHH subgroup require clarification. SHH activation has been associated with the DN disease histology in some studies, but not others, and associations with infant cases have also been reported (Kool et al., 2008; Taylor et al., 2002; Pietsch et al., 1997; Wolter et al., 1997).

In the near future, it is likely that identifying WNT and SHH tumours will become an important part of medulloblastoma diagnostics, for the selection of subtype appropriate therapeutic stratification and application of specific biological therapies that target specific signalling pathways. Until recently, robust subgroup identification of non WNT cases has relied on advanced genomic techniques (i.e. microarray) technologies, which are expensive and not practicable for the routine diagnosis of medulloblastoma subgroups. Subsequent to the completion of work reported in this chapter, GAB1 and SFRP1 immunopositivity have been reported to characterise SHH tumours (Ellison et al., 2011b; Northcott et al., 2010) and antibody markers for the remaining WNT / SHH independent tumours have been described (Northcott group C – NPR3; Northcott group D – KCNA1) (Northcott et al., 2010).

The identification of robust biomarkers for subgroups and the development of assays to test them, which can be routinely used across multiple treatment centres, using small amounts of biopsied material, will be essential for any clinical application. A number of new approaches suitable for clinical testing of gene signatures have recently become available. Examples include the GeXP system (Beckman Coulter, Fullerton, CA, USA) and the Nanostring platform (Nanostring Technologies, Seattle, WA, USA). These technologies are a cost-effective and high-throughput means of simultaneously assaying gene expression in multiple genes (GeXP – 30 genes; Nanostring – 50 -70 genes). In this chapter, a novel method for testing gene signatures of WNT and SHH pathway activation, that can be rapidly applied to small amounts of clinical material using an array-independent methodology, was developed for the identification of WNT and SHH subgroup membership in medulloblastomas, which formed the basis for investigations of pathway correlates in wider cohorts. This was coupled with a comprehensive investigation into the determinants of pathway activation to define the clinical and molecular characteristics of each subgroup.

3.2 Aims

The work reported in this chapter aimed to:

- Develop and validate minimal mRNA expression signatures for the diagnosis of WNT and SHH pathway activated subgroups in clinical material.
- Apply the signatures to a cohort of medulloblastoma RNA samples and assign samples to a signature subgroup (WNT, SHH, WNT / SHH independent).
- Apply the signatures *in silico* to available transcriptomic datasets and classify them into signature subgroups.
- Implement a comprehensive investigation of the determinants of SHH pathway activation and investigate their correlation with subgroup membership.
- In a combined cohort, consisting of the primary cohort and the available transcriptomic datasets, to undertake a meta-analysis to investigate the clinical and molecular characteristics of each subgroup and their utility for improved disease management.

3.3 Materials and methods

3.3.1 Description of cohort

A representative, 55 member cohort, comprising RNA extracted from snap-frozen, primary medulloblastoma samples was selected, as detailed in section 2.1. A panel of constitutional DNA samples from 100 normal individuals was obtained from the North Cumbria Community Genetics Project (UK). Research Ethics Committee approval had been obtained for the collection, storage and biological study of all material.

3.3.2 Extraction of nucleic acids

DNA and RNA were extracted from snap-frozen tumour samples and medulloblastoma cell lines MED1, MED8A, D283, D425, D458, D384, D556, D341, UW228 and DAOY using Trizol (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions, as detailed in section 2.2.1. RNA quality and quantity of primary medulloblastoma samples was assessed using the RNA nano II assay on the 2100 Bioanalyzer (Agilent, USA), as described in section 2.3.2.

3.3.3 A mutational screen to identify known genetic defects of medulloblastoma

All coding exons of the *PTCH1* and *SMO* genes were PCR amplified using the primers and conditions shown in Table 3.1. The coding exons of the *SUFU* gene were PCR amplified using previously published primers (Scott et al., 2006). Mutation screening was performed using denaturing high-performance liquid chromatography (dHPLC) on the Transgenomic WAVE system (Transgenomic, Cheshire, UK). To identify mutations, heteroduplex formation, before and after spiking with equal amount of control, wild-type DNA was assessed according to manufacturer's instructions. Samples whose elution profile differed from the wild type controls were directly sequenced on an ABI 377 sequencer (Applied Biosystems). In reported studies, dHPLC has been shown to identify >90% of sequence variants (Scott et al., 2006). Mutation analysis of the *CTNNB1* and *APC* genes was performed as previously described (Ellison et al., 2005). This work was conceived and implemented by Dr. Debbie Scott and Dr. Debbie Straughton (both Newcastle University Brain Tumour Research Group).

Exon	Primer Forward (5'-3')	Primer Reverse (3'-5')	Primer Annealing Temp (°C)	Product Size (bp)
1B	CAATGGAAGGCGCAGGGTCTGA	TTGGGATCGCCCCCGCCAC	56	513
2	GCAAGCGGGCGTGGGCGCGCGG	ATAGAGATATTCGCCAGCGCA	60	337
3	GCAGGTAGTCAGATAACAGA	TGGGAGAAGGCTGCTGGTTTA	60	335
4	GAATTTGCACTAATTTCTTAT	TAGTGTGCCTTAACCTAACGC	61	198
5	AATTTCTCAGGAACACCCCA	AATGTTGCTTATCATTGTTTG	61	239
6	CGATGCGTTTAGAAGGCTCTT	GAAGGGTGGAGAGCAAAACAC	59	299
7	CTTATGCACTGGATTTTAACA	AAGCCACTTCCCTCCTGACA	59	274
8	GAGGCAGTGGAAACTGCTTCCT	TGGTTATGCAATTCTTCATTT	59	266
9	AGGCTTGTGGAAGTGTTCAT	TTCCATGACTGCTCCTGCTTC	61	271
10	GAAACATTAGAATCACAACAC	GGCCCCCAGGCTTGGCCCAGC	61	292
11	CATCTGAATTGCATCTCGCAT	GACATCACAGCTTCCTCTGTTC	63	207
12	TCTGTGCTCCAAGGGGACCAT	TCCCATGTCTGCGGTCTCTGCT	63	266
13	AAGCAGTCCTCTGATTGGGCG	TGGTGCTGTGAGAGAACTCTT	63	244
14a	TCCTAAGTCAGAGCTGTGTAA	GTACTACACCACCGCTGAGCC	60	311
14b	CCAGCTCCGCACGGAGTACG	GAACCTTTGGAGTTCATCAGAT	65	334
15	GAAGAGTCAGTGGTGCTCCCTGG	GTCTTTTGCGTGTTCAGCTTCA	63	416
16	CCACCCCGCTAGGACCAGGGTC	GGTGGAGGCTGCTTGACAGCG	70	288
17	CAAGGCAGAAGTGTGTTTACC	TGCCTTCCCTGGTTCTCCAAA	63	311
18	GAAACTGTGATGCTCTTCTAC	TTCCTTTCAGCATAGCTCTTTC	58	396
19	CCTTGTGGGCAGCCCCAGAGG	GGCTCTGTTTGTCTCCAAGTGG	61	275
20	CAGAGCTGAGCATTTACCAGGT	TAGAAATCATTGTGATTGGG	61	257
21	GTCGGGCACACGGAGGGTGGC	CGGCTGAATGCTGTGTTTCC	63	226
22	TGAGTGTGGCCAGCGGTAAA	GTGGGCCTGGAGGTCCAGCGG	64	383
23a	GTGGGAGCTGCGGGGACCAT	AGCGTCCACTGCCATGGGCA	66	338
23b	CGTTCTCACAACCCTCGGAA	GAGGAGGACAAGGTTCTCCTG	66	381

Exon	Primer Forward (5'-3')	Primer Reverse (3'-5')	Primer Annealing Temp (°C)	Product Size (bp)
1	GCCGAGGTCGTGCGTGTGGCCG	ACCCTTGGAAAGAACAGGCTCA	64	490
2	ACACTGTACCTGCCAGGTCTGA	GGAGCCCAGAGCTTTGTCTCCT	64	437
3	CCATGCTACCTAGATACCTTTC	CTCAGCCTGGAACGTGGGAAGA	64	347
4	GCTCAGTTAAGGGTGTCTGTGT	GCAGGATCTGGCTCTGCCCTAC	64	364
5	GAGACCAGGTAGAGGGAGTACA	ACCTTCTGTCCCACCCCTTCCT	64	368
6	GTATAGTGACTGGTAGGAACGG	CTTGCTGCCAGTACTGGGAGCT	64	290
7	TAATCAGACTTGGGACTCCAGA	CTATGCGACCGGCAGGATGCAG	67	287
8	CCAGCTGGGTGAACTTTGAGGC	GGCAGCAGGGATTTGCCAGGTC	67	310
9	GTGACAGAGCAAGATCCTATCT	TCCTGCCTCTAGCACAGCCTTG	67	367
10	TCCTTCTCTGGAAAGAATGGCA	ATGTGCTAGTCTCTCCCAGCCT	67	315
11	TGGCACTGACTATGGGAGGCAC	TGGAGTACAGGGGCTGTCGGAG	67	298
12	GAGCCAGGGCCCCAGGCTCG	AGGACCTGGGACAGGAAAGA	67	500

Table 3.1. PCR primers and amplification conditions for the mutational analysis of *PTCH1* (upper table) and *SMO* (lower table). Primers were designed according to build version 36 of the NCBI Genomic Sequence Assembly (Genbank Gene ID 5727; http://www.ncbi.nlm.nih.gov/). The addition of 10 x Q solution (Qiagen) was required to amplify exon 1 of SMO.

3.3.4 Analysis of promoter methylation status

Two promoter-associated CpG islands of the *PTCH1* gene, spanning exons 1a and 1c (Diede et al., 2010) were identified and characterised using the Emboss CpGPlot website (http://www.ebi.ac.uk/emboss/cpgplot/). 1a methylation status was determined by MSP (Herman et al., 1996) (section 2.6.2) using primers shown in Table 3.2 and 1c by bisulfite sequencing (Frommer et al., 1992) (section 2.6) using previously published primers (Scott et al., 2006).

The methylation status of the *COL1A2* promoter was assessed by bisulfite sequencing (section 2.6) using previously published primers (Anderton et al., 2008). For loci assessed by bisulfite sequencing, the relative peak intensities at each CpG residue were determined. Methylation status was assigned as described in section 2.6.1. This work was conceived and implemented by Dr Janet Lindsey and Dr Debbie Straughton (both Newcastle University Brain Tumour Research Group).

	Methylated Primers	**Unmethylated Primers**
Product Size (bp)	114	115
Annealing Temp (°C)	62	62
Reverse Primer (5'-3')	ACTACTACTCACACGACGAACGCT	ACTACTCACACACAACAAACACT
Forward Primer (5'-3')	GGTCGTAGAGATTTCGGGATTTTC	GGTTGTAGAGATTTTGGGATTTTTG

Table 3.2. Methylation Specific PCR (MSP) primers for the analysis of the *PTCH1* promoter-associated CpG island 1a.

3.3.5 Loss of heterozygosity analysis

Loss of heterozygosity of chromosome regions 9p22.3 and the p-arm of chromosome 17 were analysed using the HOMOD method (section 2.1), as previously published (Langdon et al., 2006). LOH of chromosome 6 was analysed as previously described (section 2.1). This work was carried out by Dr. Meryl Lusher and Dr. Hisham Megahed (both Newcastle University Brain Tumour Research Group).

3.3.6 Development of GeXP assay to identify signalling pathway activated samples

A multiplex mRNA expression assay (section 2.4) (Schwalbe et al., 2011; Rai et al., 2009) was developed to test tumours for membership of the WNT or SHH medulloblastoma subgroups in the primary medulloblastoma RNA cohort (*n* = 39). Two previously reported independent medulloblastoma expression microarray datasets (Kool et al., 2008; Thompson et al., 2006), were used to design signatures for WNT and SHH pathway activation. Data were respectively downloaded from the St. Jude Research website (http://www.stjuderesearch.org/site/data/medulloblastoma) and the Gene Expression Omnibus (GEO) (Barrett et al., 2007). Data were processed and normalised using the rma algorithm (see section 2.5.2) using R version 2.09 (R core foundation) and Bioconductor (Gentleman et al., 2004).

To ensure homogeneity between datasets, Kool et al's dataset (performed using the Affymetrix U133 plus 2.0 array) was filtered to include only those probe sets present in Thompson et al's dataset (performed on the Affymetrix U133A array). In cases where more than one probe mapped to a gene, the most significant probe was retained, leaving one probe per gene (12,679 probes). Probes differential for the WNT or SHH subgroups, defined by Kool and colleagues (Kool et al., 2008) were identified using *t* tests. These *t* tests were then corrected for multiple testing using Benjamini-Hochberg's false discovery rate correction (Benjamini and Hochberg, 1995). Finally, only those significant genes that showed up-regulation in the subgroup of interest were considered. Previous work had validated 3 SHH-associated genes (*GLI1, PTCH2* and *SFRP1*) and 2 WNT-associated genes (*DKK2* and *WIF1*) by quantitative reverse transcriptase real-time PCR (Thompson et al., 2006) and these genes were also considered for inclusion in the signatures.

Genes highly significant (i.e. within 50 most significant transcripts, WNT probes; within 175 most significant transcripts, SHH probes), in both datasets, in addition to previously validated genes, regardless of rank, were considered for signature inclusion. Genes were prioritised which fulfilled as many of the following criteria as possible; (i) previously validated as markers of signature activation; (ii) 3' untranslated region (UTR) <750 bp, so that the final exon-exon boundary is proximal to the 3' end of transcript. The length of the 3' untranslated region (UTR) was retrieved from BioMart (Smedley et al., 2009); (iii) amenable to primer design subject to gene structure and strict GeXP PCR product size criteria; (iv) amplicon to overlap corresponding Affymetrix probe set; (v) successful PCR amplification using designed primers.

3.3.6.1 Amplicon design

The eXpress designer software (Beckman Coulter) was used to design multiplex primer sets for the GeXP assay. The multiplex was designed in accordance with manufacturer instructions, following the criteria described in section 2.4.1. Briefly, primer sequences were checked for specificity using BLAST (Altschul et al., 1990) and BLAT (Kent, 2002) searches. In addition, primer sequences were checked against dbSNP v129 (Sherry et al., 2001) to ensure that they did not harbour any known SNPs. The control genes *GAPDH*, *B2M* and *28S* ribosomal RNA (*rRNA*) were selected as potential reference genes, on the basis that they were located on chromosomes not frequently associated with copy number alterations in medulloblastoma. The initial multiplex primers are shown in Table 3.3.

Gene Type	Gene Name	Affymetrix Probe	Amplicon Size (bp)	F primer (5' - 3')	R primer (5' - 3')
SHH	*MICAL1*	218376_s_at	141	GACCAGGAGCTACGAGGCTA	TTGACCAAATCCACCAGCTT
CTRL	*B2M*	201891_s_at	149	AGGCTATCCAGCGTACTCCA	TCAATGTCGGATGGATGAAA
SHH	*SFRP1*	202037_s_at	155	TCTGAGGCCATCATTGAACA	TCAGGGGCTTCTTCTTCTTG
SHH	*RAB33A*	206039_at	162	AGGGAGAAGACCGTGGAAAT	AAGACCACGGCATGTACGTT
SHH	*GLI1*	206646_at	170	TTGAGAACCTCAGGCTGGAC	CGGCGTTCAAGAGAGACTG
CTRL	*28S RNA*	N/A	178	GGGGGAGAGGGTGTAAATCT	GCCAATCCTTATCCCGAAGT
WNT	*TMEM45A*	219410_at	194	TGCATGGGATCTGATGGAT	TCCAACTTCTGAGGAGCAGAG
SHH	*BCHE*	205433_at	204	AGCTGGCCTGTCTTCAAAAG	TTCCACTCCCATTCTGCTTC
SHH	*PTCH2*	221292_at	214	TGCTGGTTCCCACTTTGACT	GCTGGTGGACTCAGGATCTC
SHH	*PDLIM3*	209621_s_at	228	TTCAGAGTGCTCCAGGGAAT	TCAGGGTGCCGGTACTTATC
WNT	*DKK2*	219908_at	234	TTGGGATGGCAGAATCTAGG	AGCCCATGAGAACCCTTCTT
WNT	*WIF1*	204712_at	241	CCAGGGAGACCTCTGTTCAA	TCCTCGGCCTTTTTAAGTGA
WNT	*CNGA3*	207261_at	248	TTATTTGCAGGGCCTGTTTC	AGGGACAACACATCCAGCTT
WNT	*PYGL*	202990_at	268	AGCAGCCTGACCTCTTCAAA	AAGGTTCCACGTTCCAGATG
SHH	*ITIH2*	204987_at	275	CACCTAAAGCCCACGGACTA	TTGTAATGCCCGTCAATGAA
WNT	*TNC*	201645_at	283	ACCACAATGGCAGATCCTTC	GCCTGCCTTCAAGATTTCTG
WNT	*CCDC46*	213644_at	318	GTGCGATGTGCAGAGAAAAA	TGGAAGAAGTCCACAGCACA
CTRL	*GAPDH*	217398_x_at	354	CCTGACCTGCCGTCTAGAAA	TTCCTCTTGTGCTCTTGCTG

Table 3.3. SHH and WNT subgroup signature genes selected for expression analysis. The gene type, amplicon size, and forward and reverse primer pairs used for multiplex GeXP assays are listed. Universal tags (Forward: AGGTGACACTATAGAATA, Reverse: GTACGACTCACTATAGGGA) were added to each primer sequence for GeXP analysis. CTRL – control gene. SHH – marker of SHH subgroup tumours. WNT – marker of WNT subgroup tumours. Corresponding Affymetrix probes used in microarray-based expression analysis are shown.

3.3.6.2 Preparation of forward and reverse primer multiplexes

Initial forward and reverse primer multiplex mixes were prepared at 500nM and 200nM respectively, as described in section 2.4.1.

3.3.6.3 Selection of a positive control

The medulloblastoma cell lines MED1, MED8A, D283, D425, D458, D384, D556, D341, UW228 and DAOY were assessed for their suitability as positive controls (see section 2.4.3). Since no cell line could detect all signature genes, it was necessary to mix the selected cell line with mRNA from known WNT and SHH cases, whose positive status

had been previously determined by direct sequencing of the *CTNNB1* and *PTCH1* genes, respectively. The *CTNNB1* and *PTCH1* mutated cases with the most abundant stocks of RNA were selected for inclusion in a hybrid positive control consisting of cell line and primary tumour RNA.

3.3.6.4 Assay optimisation

The assay was designed so that there was redundancy in the signature genes, so that any genes not performing satisfactorily could simply be removed from the primer mix without further optimisation. As discussed in section 2.4.3.1, singlet forward primer reactions were carried out to identify unintended interactions between primers. Genes whose primers were inappropriately interacting with other primers to form undesigned amplicons were removed from the multiplex. Attenuation of primers was applied to reverse primers that initially produced too high a signal (section 2.4.3.2). Reverse primers for *GAPDH, SFRP1, B2M and 28S rRNA* were subsequently diluted 1:20, 1:20, 1:50 and 1:20000 respectively.

3.3.7 GeXP reaction conditions

The GeXP reaction was carried out as described in section 2.4.4, using 50 ng RNA as template per replicate. Reactions were carried out in triplicate.

3.3.8 Comparison of GeXP gene expression with estimates derived from real-time reverse transcriptase PCR

For 5 signature genes, there were matched quantitative real time RT-PCR data available for 23 cases (Thompson et al., 2006), to compare reproducibility between gene expression data estimated by GeXP and real-time RT-PCR. Data were first $\log_2$ transformed, and scaled to a mean of zero and a variance of one to make them comparable. Bland-Altman plots (section 2.8.2), were plotted to assess concordance.

3.3.9 Integration and analysis of combined gene expression data sets

The optimised WNT and SHH signalling pathway signatures (Kool et al., 2008; Thompson et al., 2006) (n = 62 and 46, respectively), were selected from the two microarray datasets used to derive the expression signatures. In addition, a third medulloblastoma microarray dataset, reported by Fattet and colleagues (Fattet et al.,

2009) (*n* = 40), became publicly available in 2009. It was downloaded from GEO and normalised as described in section 3.3.6 and the gene signature probes selected. All three publicly available datasets have linked clinical (age at diagnosis, histological subtype and gender) and *PTCH1* / *CTNNB1* mutational data available (except *PTCH1* mutation data not available from Fattet and colleagues' dataset (Fattet et al., 2009)). Expression data of signature genes were integrated from all three datasets (*n* = 148), together with the primary investigation cohort (*n* = 39).

First, the GeXP data were normalised relative to the selected control gene, *28S rRNA*. Mean expression scores were calculated from independent runs in triplicate. Next, the GeXP data were $\log_2$ transformed to match the expression microarray data. Prior to joining, to ensure that highly expressed signature genes contributed equally towards assigning signature status as lowly expressed signature genes, datasets were separately scaled, on a per-gene basis, to have a mean of zero and a variance of 1.

For each dataset, HC of samples by their signature genes was carried out using Euclidean distance and average agglomeration (section 2.9.1). Biplots (section 2.9.2.1) were plotted for the selected signature genes to support identification of pathway expressing samples. Stacked barplots of signature genes were plotted to identify the extent to which each signature was being expressed. Samples were assigned to a subgroup with reference to HC, supported by biplots of PCA and stacked barplots.

There was some overlap between studies: 11 samples were assessed by expression array by Thompson and colleagues (Thompson et al., 2006) and GeXP; 3 samples were assessed by expression array by Kool and colleagues (Kool et al., 2008) and GeXP, which enabled the comparison of subgroup assignment in individual samples using our signatures, when evaluated using different gene expression assays. Clinical and genomic correlates were combined for the combined, non-overlapping cohort (*n* = 173) (see Table 3.5).

3.3.10 Assessment of subgroup assignment

To assess subgroup assignment, silhouette plots (Rousseeuw, 1987) (section 2.10.1) were constructed for the primary investigation cohort (*n* = 39) and three transcriptomic datasets (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006)

(*n* = 40, 62 and 46, respectively). In addition, the correlation between *CTNNB1* mutation and WNT subgroup membership, and *PTCH1* mutation and SHH subgroup membership was assessed in both the primary GeXP cohort and microarray datasets.

3.3.11 Additional statistical analysis

Fisher's exact and chi-squared tests were used to identify relationships between expression signature status and selected molecular and clinical disease features in the primary investigative cohort (*n* = 39) and in the combined cohort (*n* = 173). Bonferroni corrections for multiple hypothesis testing were applied where appropriate. Additional patient age data were kindly provided by Dr. Marcel Kool (Academic Medical Centre, Amsterdam, the Netherlands).

3.4 Results

In this chapter, a multi-gene expression assay was designed and validated for the routine identification of WNT and SHH medulloblastoma subgroups. A comprehensive investigation of proposed molecular determinants of pathway activation was undertaken in the primary investigation cohort. The assay was then applied to publicly available transcriptomic datasets (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006), and subgroup assignments were pooled for a meta-analysis of the molecular basis, associated biomarkers and clinical relevance of the disease subgroups.

3.4.1 Selection of signature genes

15 genes (eight potential SHH signature genes (*BCHE, GLI1, ITIH2, MICAL1, PDLIM3, PTCH2, RAB33A* and SFRP1) and seven potential WNT signature genes (*CCDC46, CNGA3, DKK2, PYGL, TMEM45A, TNC* and *WIF1*)) were selected subject to the criteria outlined in section 3.3.6. The selected genes and their identification is summarised in Table 3.4, and their expression in the subgroups (WNT, SHH, WNT / SHH independent), defined by Kool and colleagues (Kool et al., 2008), shown in Figure 3.1. Each selected signature gene was highly differential for subgroup membership in both Kool and colleagues' (Kool et al., 2008) and Thompson and colleagues' transcriptomic datasets (Thompson et al., 2006).

3.4.2 Selection of positive control

Cell line D556 detected 11 / 15 target genes and was the best performing cell line. By mixing NMB93 (WNT-positive, *CTNNB1* mutant), NMB81 (SHH-positive, *PTCH1* mutant) and D556 cell line RNA, it was possible to detect all 15 target genes. The positive control for subsequent experiments was defined as a mix of RNAs from D556, NMB81 and NMB93, each at 20 ng / μl in the ratio 2:1:1 respectively

Kool Rank	Gene Name	probe Name	Corrected p value	3'UTR Length	Previously Validated Marker	Thompson Rank	Consideration for signature inclusion?	Successful amplicon design?	Designed amplicon overlaps Affymetrix Probe	Successful PCR amplification
1	*PDLIM3*	210170_at	5.45E-26	545		10	Y	Y	Y	Y
2	*EYA1*	214608_s_at	1.75E-17	1987		21	N	NA	NA	NA
3	*NDP*	206022_at	2.93E-17	1022		3	N	NA	NA	NA
4	*SPHK1*	219257_s_at	6.36E-16	273		24	Y	N - Final Exon (6) too long - 1055bp	NA	NA
5	*NGFR*	205858_at	3.54E-15	2007		53	N	NA	NA	NA
6	*ATOH1*	221336_at	1.02E-14	78	Y	8	Y	N - Only one exon - no exon-exon boundaries	NA	NA
7	*SFRP1*	202037_s_at	3.07E-14	3204	Y	1	Y	Y	N	Y
8	*KIAA0922*	209760_at	7.46E-14	134		9	Y	N - multiple transcripts, final exon	NA	NA
9	*GLI1*	206646_at	1.65E-13	200	Y	27	Y	Y	N	Y
10	*SATB2*	213435_at	1.15E-12	2710		329	N	NA	NA	NA
11	*GAB1*	214987_at	2.14E-12	5378		682	N	NA	NA	NA
12	*FAM198B*	219872_at	2.31E-12	577		3485	N	NA	NA	NA
13	*ANKRD6*	204671_s_at	3.42E-12	2829		22	N	NA	NA	NA
14	*ZNF516*	203604_at	9.52E-12	2833		56	N	NA	NA	NA
15	*SHROOM2*	204967_at	1.30E-11	2505		13	N	NA	NA	NA
16	*BCHE*	205433_at	1.39E-11	601		19	Y	Y	Y	Y
17	*CA14*	219464_at	2.06E-11	418		54	Y	Y	Y	N
18	*MICAL1*	218376_s_at	3.11E-11	276		7	Y	Y	Y	Y
19	*SEPT10*	212698_s_at	8.44E-11	1521		28	N	NA	NA	NA
20	*CRB1*	220522_at	1.31E-10	6139		227	N	NA	NA	NA
21	*METRN*	219051_x_at	2.20E-10	92		16	Y	N	NA	NA
22	*ABCA8*	204719_at	2.96E-10	792		102	N	NA	NA	NA
23	*ABCB4*	207819_s_at	4.41E-10	46		31	Y	N - multiple transcripts, final exon	NA	NA
24	*CTSC*	201487_at	4.66E-10	400		167	Y	N - multiple transcripts, final exon	NA	NA
25	*IRAK3*	220034_at	5.72E-10	6440		492	N	NA	NA	NA
26	*ALDH1A3*	203180_at	1.45E-09	1920		4	N	NA	NA	NA
27	*GCK*	211167_s_at	1.66E-09	864		33	N	NA	NA	NA
28	*PCNT*	203660_s_at	1.98E-09	441		18	Y	N - multiple transcripts, final exon	NA	NA
29	*KLHL4*	214591_at	3.74E-09	3515		1122	N	NA	NA	NA
30	*RGS10*	204319_s_at	3.81E-09	284		2172	N	NA	NA	NA
31	*ANKRD57*	219496_at	3.95E-09	2918		43	N	NA	NA	NA
32	*NRIP2*	215104_at	5.17E-09	1860		2	N	NA	NA	NA
33	*GRIA4*	208464_at	9.13E-09	2517		1214	N	NA	NA	NA
34	*SYPL1*	201259_s_at	9.42E-09	1301		25	N	NA	NA	NA
35	*TEX15*	221448_s_at	9.71E-09	1741		4431	N	NA	NA	NA
36	*LRP5L*	214873_at	1.00E-08	377		334	N	NA	NA	NA
37	*C1orf54*	219506_at	1.12E-08	81		221	N	NA	NA	NA
38	*ARHGEF26*	222121_at	1.26E-08	2353		319	N	NA	NA	NA
39	*NDST3*	220429_at	1.46E-08	3039		82	N	NA	NA	NA
40	*GLO1*	200681_at	1.53E-08	1351		472	N	NA	NA	NA
41	*COL21A1*	208096_s_at	1.85E-08	1066		70	N	NA	NA	NA
42	*CXCR4*	211919_s_at	1.88E-08	519		6	Y	N - two isoforms	NA	NA
43	*IGBP1*	202105_at	2.09E-08	342		995	N	NA	NA	NA
44	*C10orf72*	213381_at	2.27E-08	1205		87	N	NA	NA	NA
45	*PDE1A*	208396_s_at	2.34E-08	310		166	Y	N - multiple transcripts, final exon	NA	NA
46	*RAB33A*	206039_at	2.34E-08	129		12	Y	Y	Y	Y
47	*PLXND1*	38671_at	2.78E-08	1253		145	N	NA	NA	NA
48	*MAP7D3*	219626_at	2.86E-08	1682		3637	N	NA	NA	NA
49	*PNRC1*	209034_at	3.03E-08	963		90	N	NA	NA	NA
50	*EP400*	212375_at	3.03E-08	2786		81	N	NA	NA	NA
60	*PTCH2*	221292_at	7.33E-08	844	Y	157	Y	Y	Y	Y
75	*ITIH2*	204987_at	3.94E-07	332		113	Y	Y	Y	Y

Thompson Rank	Gene Name	Probe Name	Corrected p value	3'UTR Length	Previously Validated Marker	Kool Rank	Consideration for signature inclusion?	Successful amplicon design?	Designed amplicon overlaps Affymetrix Probe	Successful PCR amplification
1	*SFRP1*	202036_s_at	4.02E-15	3204	Y	7	Y	Y	N	Y
2	*NRIP2*	215104_at	1.17E-14	1860		32	N	NA	NA	NA
3	*NDP*	206022_at	1.28E-13	1022		3	N	NA	NA	NA
4	*ALDH1A3*	203180_at	1.23E-12	1920		26	N	NA	NA	NA
5	*PGM1*	201968_s_at	2.11E-11	569		57	Y	N - size constraints	NA	NA
6	*CXCR4*	211919_s_at	4.49E-11	519		42	Y	N - two isoforms	NA	NA
7	*MICAL1*	218376_s_at	7.40E-11	276		18	Y	Y	Y	Y
8	*ATOH1*	221336_at	3.43E-10	78	Y	6	Y	N - Only one exon - no exon-exon boundaries	NA	NA
9	*KIAA0922*	209760_at	3.43E-10	134		8	Y	N - multiple transcripts, final exon	NA	NA
10	*PDLIM3*	210170_at	3.48E-10	545		1	Y	Y	Y	Y
11	*GAS1*	204457_s_at	4.50E-10	1377		164	N	NA	NA	NA
12	*RAB33A*	206039_at	7.45E-10	129		46	Y	Y	Y	Y
13	*SHROOM2*	204967_at	1.34E-09	2505		15	N	NA	NA	NA
14	*ARHGEF7*	202548_s_at	1.35E-09	2623		132	N	NA	NA	NA
15	*ABLIM3*	205730_s_at	2.18E-09	2148		60	N	NA	NA	NA
16	*METRN*	219051_x_at	2.31E-09	92		21	Y	N	NA	NA
17	*PGAM2*	205736_at	2.36E-09	36		147	Y	N - multiple transcripts, final exon		
18	*PCNT*	203660_s_at	3.17E-09	441		28	Y	N - multiple transcripts, final exon	NA	NA
19	*BCHE*	205433_at	3.26E-09	601		16	Y	Y	Y	Y
20	*BTBD3*	202946_s_at	3.35E-09	2942		68	N	NA	NA	NA
21	*EYA1*	214608_s_at	5.35E-09	1987		2	N	NA	NA	NA
22	*ANKRD6*	204671_s_at	9.36E-09	2829		13	N	NA	NA	NA
23	*CABC1*	218168_s_at	1.05E-08	807		138	N	NA	NA	NA
24	*SPHK1*	219257_s_at	1.73E-08	273		4	Y	N - Final Exon (6) too long - 1055bp	NA	NA
25	*SYPL1*	201260_s_at	5.03E-08	1301		34	N	NA	NA	NA
26	*FAM46A*	221766_s_at	7.01E-08	3964		129	N	NA	NA	NA
27	*GLI1*	206646_at	7.64E-08	200	Y	9	Y	Y	N	Y
28	*SEPT10*	214720_x_at	1.18E-07	1521		19	N	NA	NA	NA
29	*SOX9*	202935_s_at	1.73E-07	2023		656	N	NA	NA	NA
30	*GNG3*	222005_s_at	1.73E-07	396		137	Y	Gene overlaps with *BSCL2* exon.	NA	NA
31	*ABCB4*	207819_s_at	1.77E-07	46		23	Y	N - multiple transcripts, final exon	NA	NA
32	*SUN2*	212144_at	2.85E-07	1647		1666	N	NA	NA	NA
33	*GCK*	211167_s_at	3.06E-07	864		27	N	NA	NA	NA
34	*NLGN1*	205893_at	3.80E-07	5340		223	N	NA	NA	NA
35	*C7orf16*	220231_at	3.90E-07	1171		118	N	NA	NA	NA
36	*APBA2*	209870_s_at	3.91E-07	1181		834	N	NA	NA	NA
37	*RBM38*	212430_at	3.91E-07	1790		574	N	NA	NA	NA
38	*NHLH1*	214628_at	4.01E-07	1714		106	N	NA	NA	NA
39	*PPM1H*	212686_at	4.34E-07	4502		167	N	NA	NA	NA
40	*PDLIM2*	219165_at	4.57E-07	555		330	N	NA	NA	NA
41	*TMEM184B*	202027_at	5.85E-07	2398		173	N	NA	NA	NA
42	*SCG5*	203889_at	6.63E-07	488		190	N	NA	NA	NA
43	*ANKRD57*	219496_at	6.67E-07	2918		31	N	NA	NA	NA
44	*FSD1*	219170_at	9.75E-07	192		273	N	NA	NA	NA
45	*PACSIN2*	201651_s_at	1.16E-06	1585		141	N	NA	NA	NA
46	*NLGN4X*	221933_at	1.27E-06	1819		652	N	NA	NA	NA
47	*CIDEB*	221188_s_at	1.27E-06	437		361	N	NA	NA	NA
48	*ABTB2*	213497_at	1.91E-06	1398		58	N	NA	NA	NA
49	*SDC2*	212158_at	2.29E-06	2223		156	N	NA	NA	NA
50	*SCHIP1*	204030_s_at	2.29E-06	591		105	N	NA	NA	NA
113	*ITIH2*	204987_at	0.00011605	332			Y	Y	Y	Y
157	*PTCH2*	221292_at	0.00041743	844	Y		N	Y	Y	Y

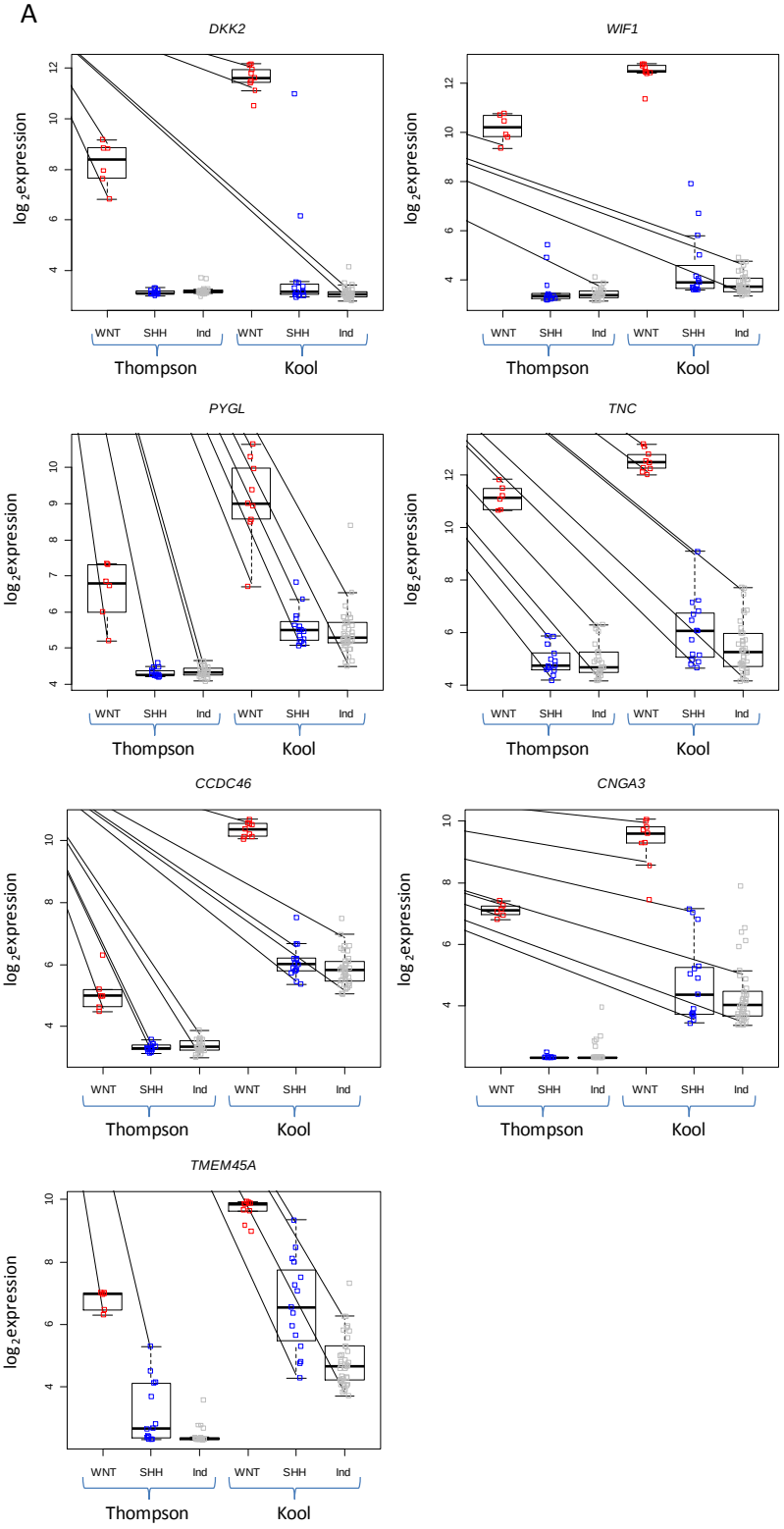
Kool Rank	Gene Name	Probe Name	Corrected 'p' value	3'UTR Length	3'UTR length <750 bp	Previously Validated Marker	Thompson Rank	Consideration for signature inclusion?	Successful amplicon design?	Designed amplicon overlaps Affymetrix Probe	Successful PCR amplification
1	*TMEM51*	218815_s_at	1.38E-35	734	Y		10	Y	N - multiple transcripts, final exon		
2	*GAD1*	206670_s_at	1.38E-35	1435	N		8	N	N - multiple transcripts, final exon		
3	*TNC*	216005_at	1.38E-35	630	Y	Y	1	Y	Y	Y	Y
4	*WIF1*	204712_at	2.35E-34	721	Y		3	Y	Y	Y	Y
5	*LEF1*	210948_s_at	5.49E-32	1206	N	Y	762	N	NA	NA	NA
6	*CCDC46*	213644_at	3.70E-29	429	Y		18	Y	Y	Y	Y
7	*FZD10*	219764_at	8.16E-27	1051	N		NA DOWN-REG	N	NA	NA	NA
8	*DKK2*	219908_at	9.53E-26	2151	N	Y	2	Y	Y	N	Y
9	*BMP4*	211518_s_at	3.48E-25	295	Y		63	Y	Not Attempted - Thompson Rank low	NA	NA
10	*OSR2*	213568_at	3.70E-25	424	Y		11	Y	N - multiple transcripts, final exon	NA	NA
11	*EPHA7*	206852_at	3.14E-24	3405	N		6	N	NA	NA	NA
12	*GABRE*	204537_s_at	1.43E-23	2090	N		34	N	NA	NA	NA
13	*RAI2*	219440_at	3.55E-23	366	Y		5	Y	N - unusual gene structure	NA	NA
14	*DKK4*	206619_at	1.38E-22	31	Y		4	Y	N - multiple transcripts, final exon	NA	NA
15	*FOXA1*	204667_at	1.38E-22	1381	N		55	N	NA	NA	NA
16	*ALK*	208212_s_at	1.51E-22	449	Y		NA DOWN-REG	N	NA	NA	NA
17	*PDE11A*	221110_x_at	1.88E-22	6155	N		1493	N	NA	NA	NA
18	*TNFRSF11B*	204933_s_at	7.11E-22	816	N		2891	N	NA	NA	NA
19	*STC1*	204597_x_at	8.29E-22	2854	N		939	N	NA	NA	NA
20	*TSPAN9*	205665_at	8.29E-22	3442	N		330	N	NA	NA	NA
21	*C22orf31*	206839_at	1.55E-21	69	Y		425	N	NA	NA	NA
22	*WNT11*	206737_at	1.57E-21	738	Y		NA DOWN-REG	N	NA	NA	NA
23	*DLX4*	208216_at	2.47E-21	1013	N		2394	N	NA	NA	NA
24	*AMHR2*	206892_at	3.15E-21	54	Y		NA DOWN-REG	N	NA	NA	NA
25	*RASL11B*	219142_at	4.16E-21	1027	N		12	N	NA	NA	NA
26	*IFT57*	218100_s_at	4.91E-21	1664	N		9	N	NA	NA	NA
27	*PART1*	205833_s_at	6.23E-21	NA*	N		1100	N	NA	NA	NA
28	*CYB5R2*	220230_s_at	8.92E-21	278	Y		NA DOWN-REG	N	NA	NA	NA
29	*TNNC1*	209904_at	2.00E-20	172	Y		985	N	NA	NA	NA
30	*C20orf103*	219463_at	3.68E-20	703	Y		30	Y	Not attempted - better candidates available	NA	NA
31	*TRBC1*	210915_x_at	4.99E-20	219	Y		39	Y	Not attempted - better candidates available	NA	NA
32	*AIMP1*	202541_at	5.23E-20	1527	N		58	N	NA	NA	NA
33	*PTGS1*	215813_s_at	5.23E-20	3158	N		51	N	NA	NA	NA
34	*BAMBI*	203304_at	1.32E-19	537	Y		33	Y	Not attempted - better candidates available	NA	NA
35	*PYGL*	202990_at	3.47E-18	172	Y		15	Y	Y	Y	Y
36	*P4HA2*	202733_at	1.43E-17	1171	N		20	N	NA	NA	NA
37	*EMID1*	213779_at	2.39E-17	780	N		54	N	NA	NA	NA
38	*EMX2*	221950_at	5.67E-17	1313	N		21	N	NA	NA	NA
39	*QPRT*	204044_at	7.26E-17	592	Y		190	N	NA	NA	NA
40	*FGF20*	220394_at	1.01E-16	920	N		32	N	NA	NA	NA
41	*LRP4*	212850_s_at	1.07E-16	2114	N		36	N	NA	NA	NA
42	*LHX6*	219884_at	1.88E-16	2182	N		NA DOWN-REG	N	NA	NA	NA
43	*MYOT*	219728_at	2.08E-16	465	Y		16	Y	Not attempted - better candidates available	NA	NA
44	*C9orf3*	212848_s_at	2.33E-16	477	Y		22	Y	Not attempted - better candidates available	NA	NA
45	*ECEL1*	219914_at	2.73E-16	325	Y		NA DOWN-REG	N	NA	NA	NA
46	*RAB11FIP2*	203883_s_at	2.85E-16	4011	N		29	N	NA	NA	NA
47	*PAPSS1*	209043_at	4.19E-16	582	Y		27	Y	Not attempted - better candidates available	NA	NA
48	*LDLRAP1*	57082_at	6.25E-16	1893	N		95	N	NA	NA	NA
49	*CNGA3*	207261_at	6.26E-16	1345	N		7	Y	Y	Y	N
50	*TMEM2*	218113_at	8.95E-16	1830	N		13	N	NA	NA	NA

Thompson Rank	Gene Name	Probe Name	Corrected 'p' value	3'UTR Length	3'UTR length <750 bp	Previously Validated Marker	Kool Rank	Consideration for signature inclusion?	Successful amplicon design?	Designed amplicon overlaps Affymetrix Probe	Successful PCR amplification
1	*TNC*	216005_at	5.52E-34	630	Y	Y	3	Y	Y	Y	Y
2	*DKK2*	219908_at	6.15E-30	2151	N	Y	8	Y	Y	N	Y
3	*WIF1*	204712_at	3.78E-29	721	Y		4	Y	Y	Y	Y
4	*DKK4*	206619_at	7.60E-27	31	Y		14	Y	N - size constraints	NA	NA
5	*RAI2*	219440_at	1.65E-24	366	Y		13	Y	N - unusual gene structure	NA	NA
6	*EPHA7*	206852_at	1.75E-23	3405	N		11	N	NA	NA	NA
7	*CNGA3*	207261_at	2.19E-22	1345	N		49	Y	Y	Y	N
8	*GAD1*	205278_at	2.39E-21	1435	N		2	N	N - multiple transcripts, final exon	NA	NA
9	*IFT57*	218100_s_at	4.54E-21	1664	N		26	N	NA	NA	NA
10	*TMEM51*	218815_s_at	5.78E-21	734	Y		1	Y	N - multiple transcripts, final exon	NA	NA
11	*OSR2*	213568_at	5.62E-20	424	Y		10	Y	N - multiple transcripts, final exon	NA	NA
12	*RASL11B*	219142_at	8.76E-19	1027	N		25	N	NA	NA	NA
13	*TMEM2*	218113_at	2.31E-18	1830	N		50	N	NA	NA	NA
14	*ALCAM*	201951_at	1.41E-17	2423	N		57	N	NA	NA	NA
15	*PYGL*	202990_at	1.64E-17	172	Y		35	Y	Y	Y	Y
16	*MYOT*	219728_at	2.41E-17	465	Y		43	Y	Not attempted - better candidates available	NA	NA
17	*GPR64*	206002_at	4.03E-17	1554	N		74	N	NA	NA	NA
18	*CCDC46*	213644_at	1.56E-15	429	Y		6	Y	Y	Y	Y
19	*WNT16*	221113_s_at	1.74E-15	1758	N		105	N	NA	NA	NA
20	*P4HA2*	202733_at	1.97E-15	1171	N		36	N	NA	NA	NA
21	*EMX2*	221950_at	1.48E-14	1313	N		38	N	NA	NA	NA
22	*C9orf3*	212848_s_at	2.65E-14	477	Y		44	Y	Not attempted - better candidates available	NA	NA
23	*QPCT*	205174_s_at	3.38E-14	494	Y		1985	N	NA	NA	NA
24	*NET1*	201829_at	6.47E-14	1850	N		52	N	NA	NA	NA
25	*SOCS2*	203372_s_at	7.21E-14	1022	N		146	N	NA	NA	NA
26	*TMEM45A*	219410_at	7.21E-14	425	Y		94	Y	Y	Y	N
27	*PAPSS1*	209043_at	1.87E-13	582	Y		47	Y	Not attempted - better candidates available	NA	NA
28	*CRYZ*	202950_at	3.26E-13	809	N		145	N	NA	NA	NA
29	*RAB11FIP2*	203884_s_at	4.63E-13	4011	N		46	N	NA	NA	NA
30	*C20orf103*	219463_at	6.08E-13	703	Y		30	Y	Not attempted - better candidates available	NA	NA
31	*LPIN1*	212274_at	1.41E-12	2617	N		283	N	NA	NA	NA
32	*FGF20*	220394_at	1.49E-12	920	N		40	N	NA	NA	NA
33	*BAMBI*	203304_at	1.69E-12	537	Y		34	Y	Not attempted - better candidates available	NA	NA
34	*GABRE*	204537_s_at	2.88E-12	2090	N		12	N	NA	NA	NA
35	*FZD6*	203987_at	7.49E-12	1525	N		203	N	NA	NA	NA
36	*LRP4*	212850_s_at	1.62E-11	2114	N		41	N	NA	NA	NA
37	*SPRY2*	204011_at	1.81E-11	781	N		259	N	NA	NA	NA
38	*DSE*	218854_at	6.32E-11	936	N		75	N	NA	NA	NA
39	*TRBC1*	210915_x_at	7.62E-11	219	Y		31	Y	Not attempted - better candidates available	NA	NA
40	*PLEKHF2*	218640_s_at	9.69E-11	1889	N		186	N	NA	NA	NA
41	*SLIT2*	209897_s_at	3.14E-10	1547	N		308	N	NA	NA	NA
42	*FBXL7*	213249_at	3.28E-10	2604	N		115	N	NA	NA	NA
43	*NEFH*	33767_at	5.38E-10	686	Y		172	N	NA	NA	NA
44	*COL5A1*	212489_at	5.46E-10	2536	N		64	N	NA	NA	NA
45	*DEFA5*	207529_at	5.99E-10	121	Y		272	N	NA	NA	NA
46	*DKK1*	204602_at	5.99E-10	849	N	Y	62	Y	Not attempted - better candidates available	NA	NA
47	*ODZ3*	219523_s_at	6.17E-10	2672	N		173	N	NA	NA	NA
48	*ATP8A1*	213106_at	7.59E-10	4640	N		71	N	NA	NA	NA
49	*NRN1*	218625_at	8.85E-10	976	N		NA DOWN-REG	N	NA	NA	NA
50	*MPP1*	202974_at	2.07E-09	660	Y		275	N	NA	NA	NA

*PART1 - processed transcript, not translated

Table 3.4. Selection of WNT and SHH signature genes from two expression array datasets (Kool et al., 2008; Thompson et al., 2006). Table shows the top 50 up-regulated genes within Kool's (left-hand table) and Thompson's (right-hand table) datasets for both SHH (upper table) and WNT (lower table) subgroups, ranked by 'p' value. Selected signature genes are highlighted green. Selected signature genes that failed during assay optimisation are shown in red. Whether or not the gene is a candidate for primer design, subject to criteria listed in section 3.3.6, is shown. For potential candidates, whether or not primer design and subsequent PCR amplification was successful is indicated, including comments if not. Y, yes; N, no; NA, not applicable or not attempted; DOWN-REG, gene is within top 50 most differentially expressed up-regulated genes, although is down-regulated in second dataset.

A


DKK2
WIF1
PYGL
TNC
CCDC46
CNGA3
TMEM45A
log2expression
WNT
SHH
Ind
Thompson
Kool

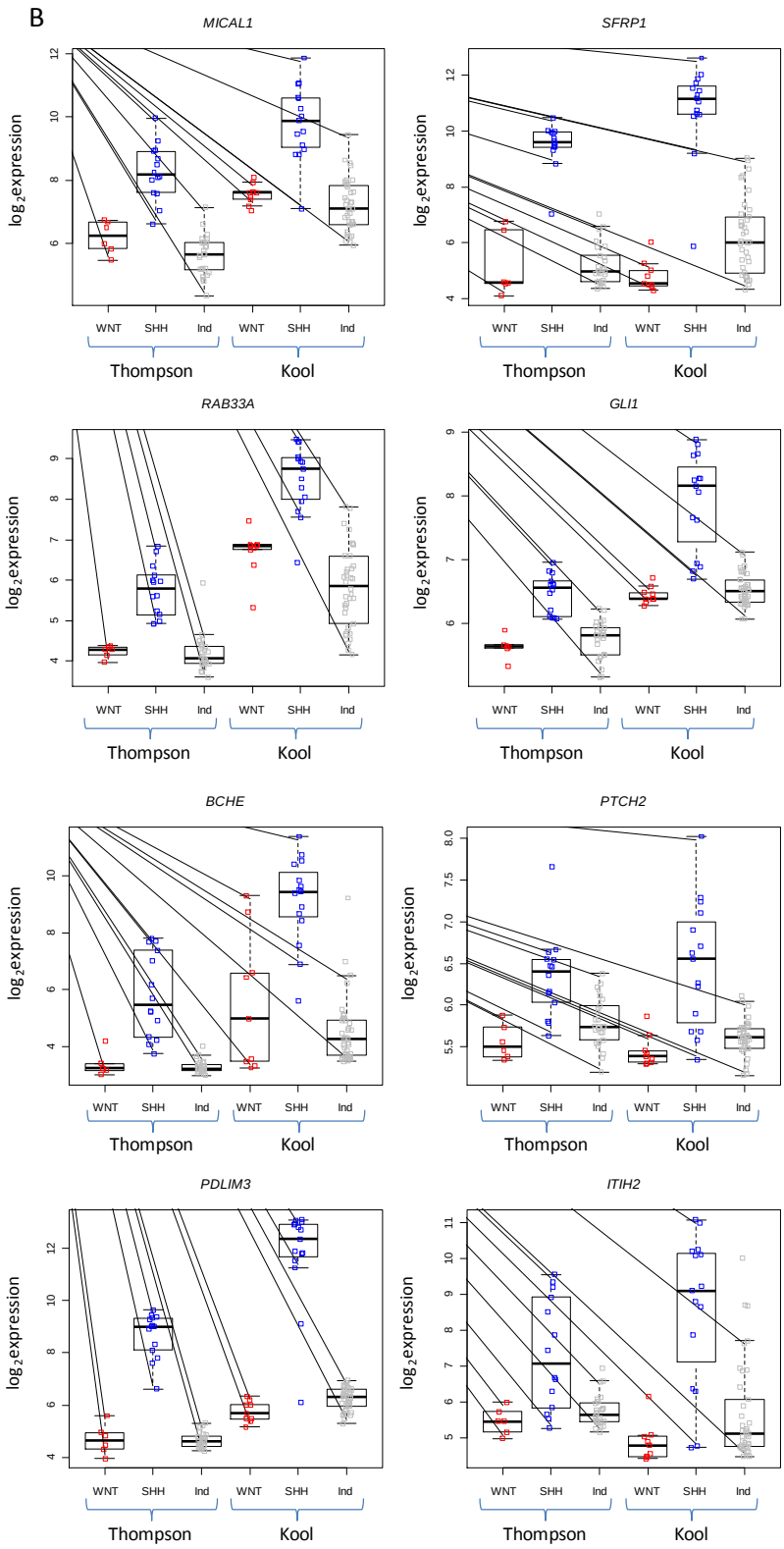


Figure 3.1. Signature genes are differentially expressed in both Kool et al. and Thompson et al. datasets. Box-dotplots are shown for A. Seven WNT subgroup candidate signature genes and B. Eight SHH subgroup candidate signature genes in both datasets (Kool et al., 2008; Thompson et al., 2006). The expression for each subgroup (WNT, SHH and Ind (WNT / SHH independent)) as described by Kool et al. is shown.

3.4.3 Optimisation and assessment of gene signatures for pathway activation

During the assay optimisation (section 2.4.3), it became clear that primers for the gene *CNGA3* were inappropriately interacting with other primers, and this primer set was subsequently removed from the primer multiplex. By running against RNA from known *CTNNB1* and *PTCH1* mutants, the gene *TMEM45A* was shown to not reliably identify WNT subgroup cases. Instead, it recognised both WNT and SHH pathway activated cases, with a higher expression observed in SHH cases. For this reason, the primers for *TMEM45A* were also removed from the multiplex, giving a total of eight SHH signature genes (*BCHE, GLI1, ITIH2, MICAL1, PDLIM3, PTCH2, RAB33A* and *SFRP1*) and five WNT signature genes (*CCDC46, DKK2, PYGL, TNC* and *WIF1*) in the final, optimised gene signature.

3.4.3.1 GeXP assay reproducibility

Data were normalised relative to the expression of the control gene, *28S rRNA.* After excluding gene expression data where each triplicate measurement was equal to zero, since these will introduce bias by artificially reducing the amount of variability observed between triplicate measurements, the coefficient of variation (CV) was calculated, defined by the ratio of the standard deviation to the mean, expressed as a percentage. The average CV for the genes measured in the GeXP assay was 13.29% (average mean 0.19, with average standard deviation 0.024).

3.4.3.2 Comparison of GeXP assay expression score with real time reverse transcriptase PCR estimates of gene expression

For 5 signature genes (*SFPR1, PTCH2, GLI1, DKK2* and *WIF1*), there was matched quantitative real time RT-PCR data available for 23 cases (Thompson et al., 2006), to compare reproducibility between gene expression data estimated by GeXP and real-time RT-PCR. The Bland-Altman plot (section 2.8.2) shown in Figure 3.2 demonstrates that reproducibility is good, with samples with a relatively high level of expression correlating especially well. The two techniques diverge the most when low to intermediate values of expression are encountered.

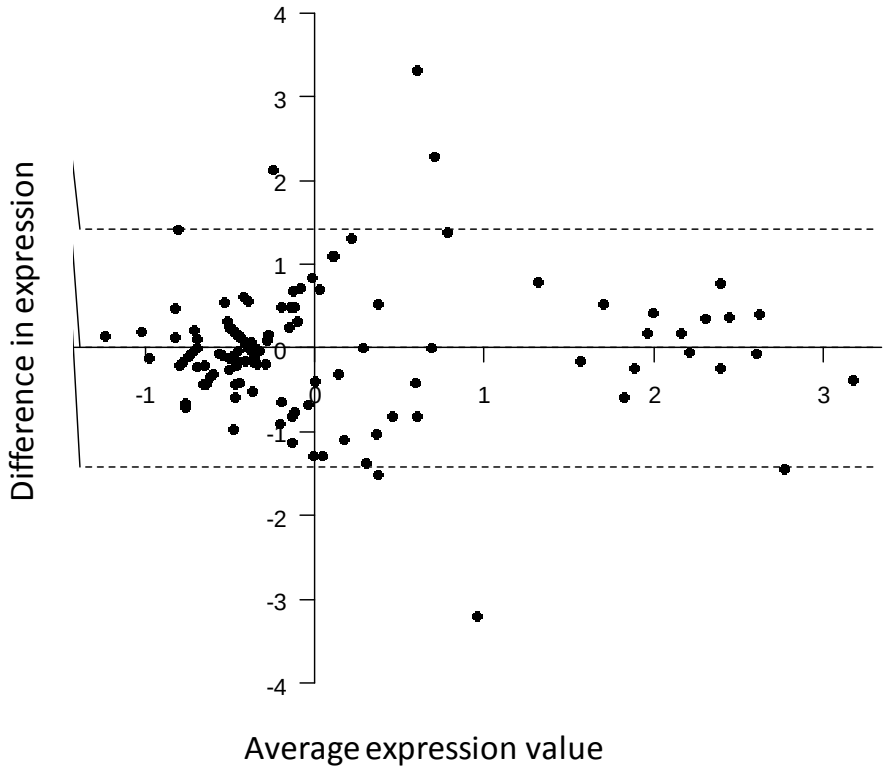


Figure 3.2. Bland-Altman plot shows comparison between paired estimates of gene expression for GeXP and real-time RT-PCR for genes *SFPR1, PTCH2, GLI1, DKK2* and *WIF1* from 23 cases. The x axis shows the average score from the two estimations of relative gene expression and the y axis shows the difference between GeXP and real-time RT-PCR estimates of gene expression. Horizontal dotted lines are plotted at 2 standard deviations of the difference and at the mean difference. Since the mean difference is 5×10^{-11}, this dotted line is not visible as it is obscured by the horizontal axis.

3.4.4 Assignment of pathway activation by WNT and SHH expression signature

In the primary investigation cohort (*n* = 39), the assignment of subgroup was unequivocal for all samples, independent of data analysis method used (unsupervised HC, PCA (Figure 3.3), stacked barplots (Figure 3.4)). Validation of these signatures in three independent medulloblastoma expression microarray datasets (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006) showed that the signatures could be successfully interrogated by hierarchical cluster analysis and PCA, were diagnostic in all cases, independent of cohort or gene expression assay used (Figure 3.3) and showed close consistency with stacked barplot data (Figure 3.4). Signature positivity was concordant with the disease subgroups apparent after independent clustering of the

1500 most variable probes within each entire array dataset (Figure 3.5), and correctly classified disease subgroup (SHH, WNT or WNT / SHH independent) in 99% (146 / 148) of cases overall (Figure 3.3). There was complete concordance in subgroup assignment, using all analytical methods, for the 14 cases analysed in parallel by GeXP and array (section 3.3.9). Silhouette plots (section 2.10.1) of the subgroups assigned by the signature, showed well clustered samples with a positive silhouette score for every member of the primary investigation cohort (*n* = 39) and for 145 / 148 of the transcriptomic datasets. Three samples showed a negative silhouette score and could be considered to be incorrectly classified.

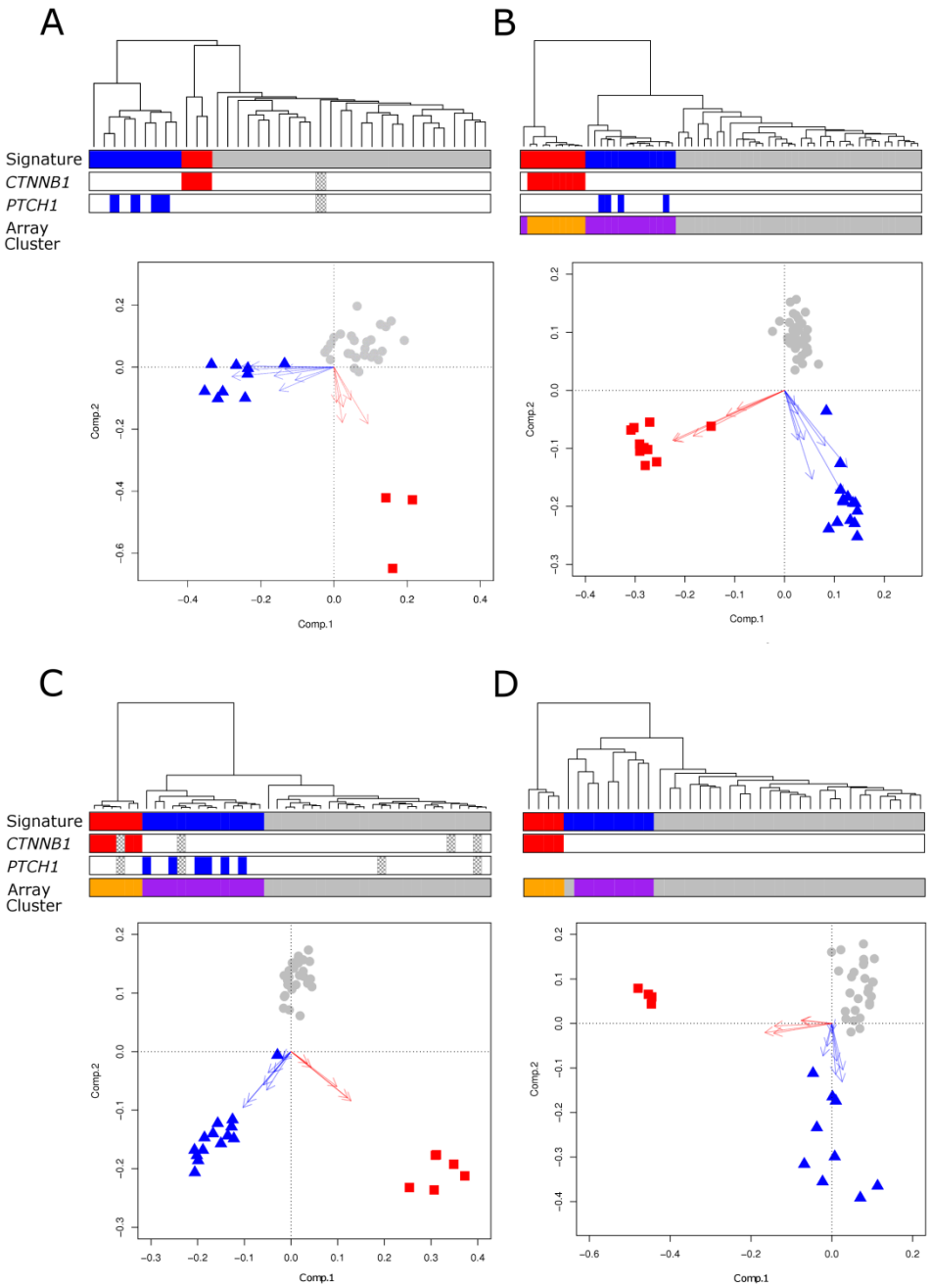


Figure 3.3. Diagnostic WNT and SHH subgroup gene expression signatures recognize equivalent molecular subgroups across multiple medulloblastoma cohorts. Data from 4 independent data sets are shown (A, primary investigation cohort; B, Kool and colleagues (*n* = 62; (Kool et al., 2008)); C, Thompson and colleagues (*n* = 46; (Thompson et al., 2006)); and D, Fattet and colleagues (*n* = 40; (Fattet et al., 2009))). WNT and SHH subgroup expression signature positivity demonstrates close concordance with (i) underlying molecular defects in the respective pathways and (ii) discrete sample clusters identified on independent clustering of the most variable probes in each array data set (Figure 3.5). Each panel (A–D) shows hierarchical clustering of signature genes (WNT subgroup, red; SHH subgroup, blue; WNT / SHH independent, grey). Mutational information for *CTNNB1* and *PTCH1* is shown (coloured boxes, mutation; grey checked boxes, missing data). Array clusters which show concordance with the detected SHH and WNT subgroup signatures derived by clustering the most variable probes of each whole array data set are shown (purple, SHH subgroup; orange, WNT subgroup (Figure 3.5)). Biplots show PCA for each signature gene set. Arrows show projections of expression axes for each gene (SHH signature genes, blue; WNT signature genes, red; WNT / SHH independent cases, grey).

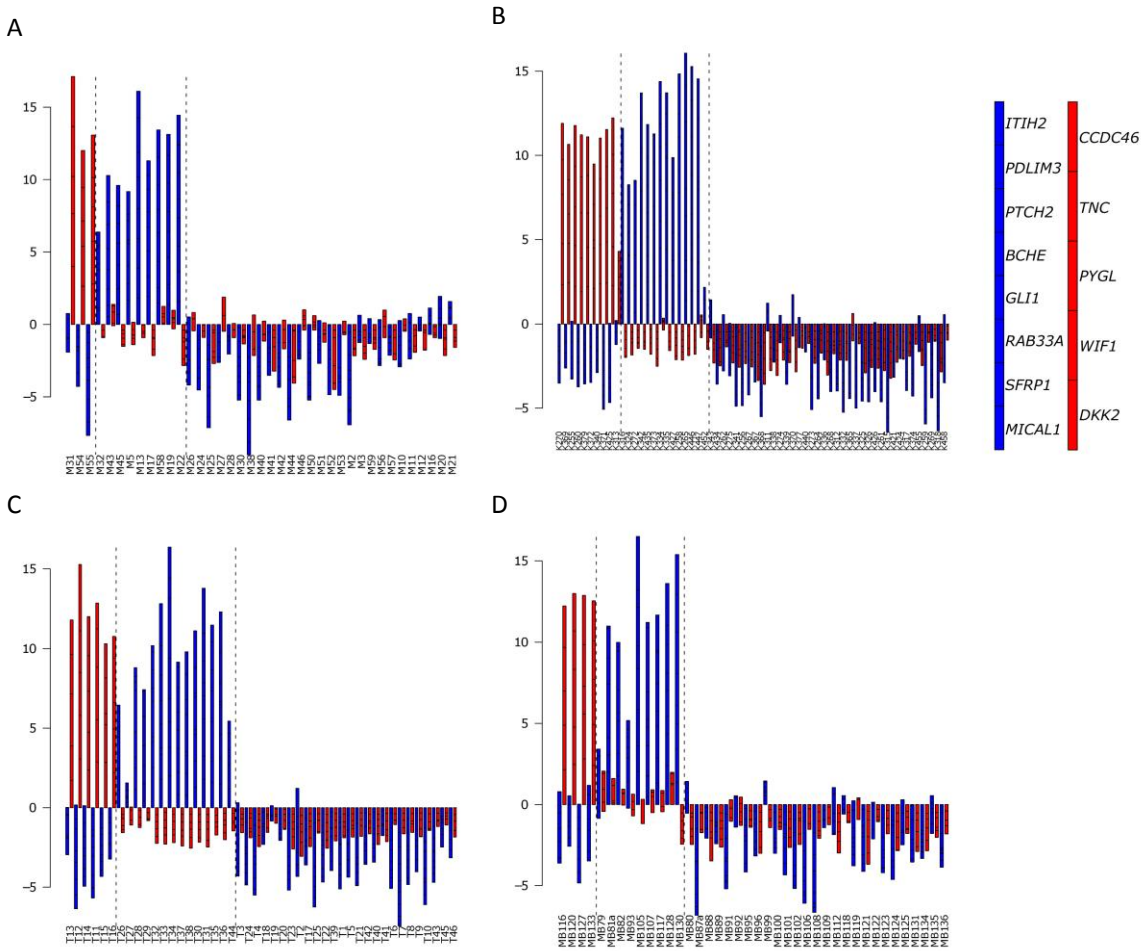


Figure 3.4. Identification of SHH and WNT subgroup medulloblastomas using diagnostic expression signatures. Data are illustrated from the 4 independent data sets using stacked bar plots (A, primary investigation cohort; B, Kool and colleagues (*n* = 62; (Kool et al., 2008)); C, Thompson and colleagues (*n* = 46; (Thompson et al., 2006)); and D, Fattet and colleagues (*n* = 40; (Fattet et al., 2009))). WNT signature genes (red) and SHH signature genes (blue) in combination clearly define subgroup membership. Vertical dashed lines delineate sample groups positive for WNT and SHH signatures by hierarchical clustering and PCA analysis in Figure 3.3. Right-hand panel indicates stacked order of genes for each signature. Before generation of bar plots, expression data from each cohort were scaled on a per gene basis to a mean of zero and a variance of 1. Samples expressing all or most signature genes at above average levels will show bars of greater positive magnitude.

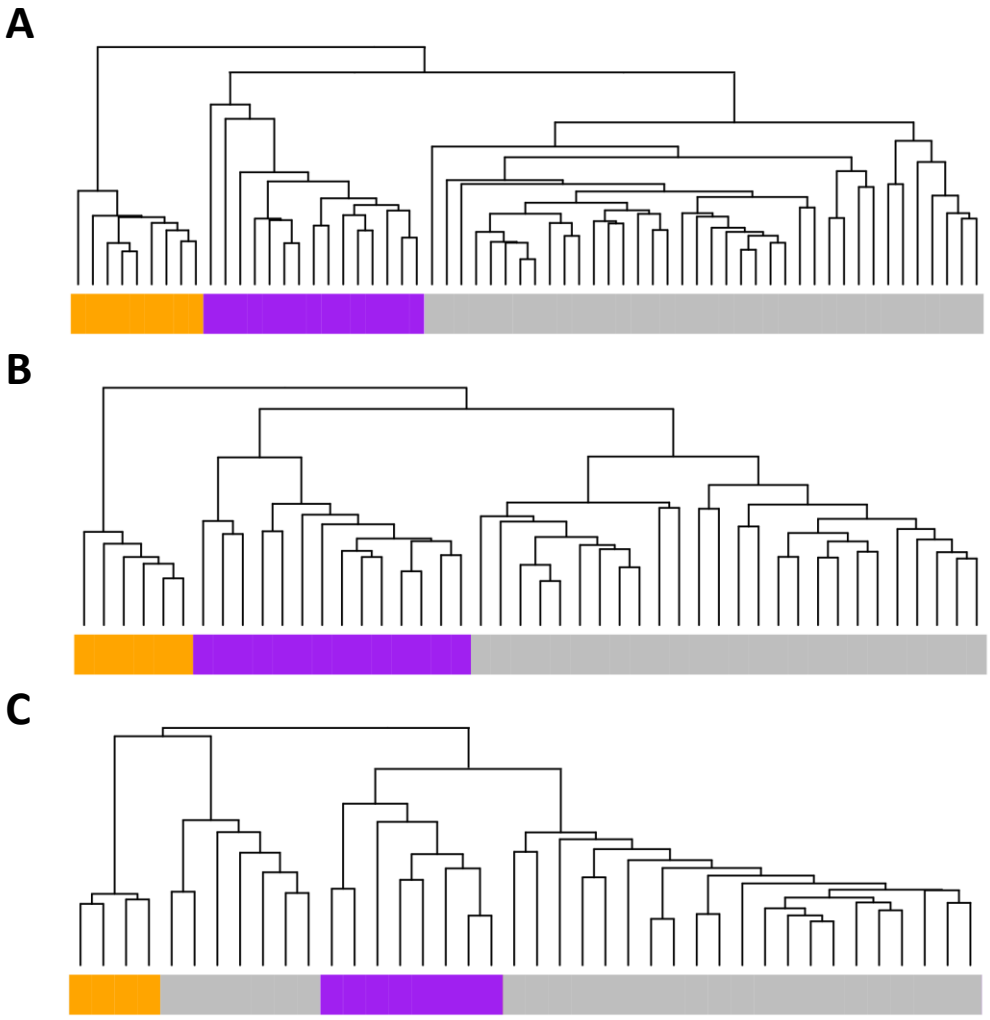


Figure 3.5. Independent hierarchical clustering of the 1500 most differentially expressed genes within transcriptomic datasets from independent medulloblastoma cohorts described by Kool and colleagues (A; (Kool et al., 2008)), Thompson and colleagues (B; (Thompson et al., 2006)) and Fattet and colleagues (C; (Fattet et al., 2009)). Discrete subgroups within each data set, which correlate with the WNT and SHH subgroups recognised by expression signatures described in Figure 2, are marked (orange, WNT; purple , SHH).

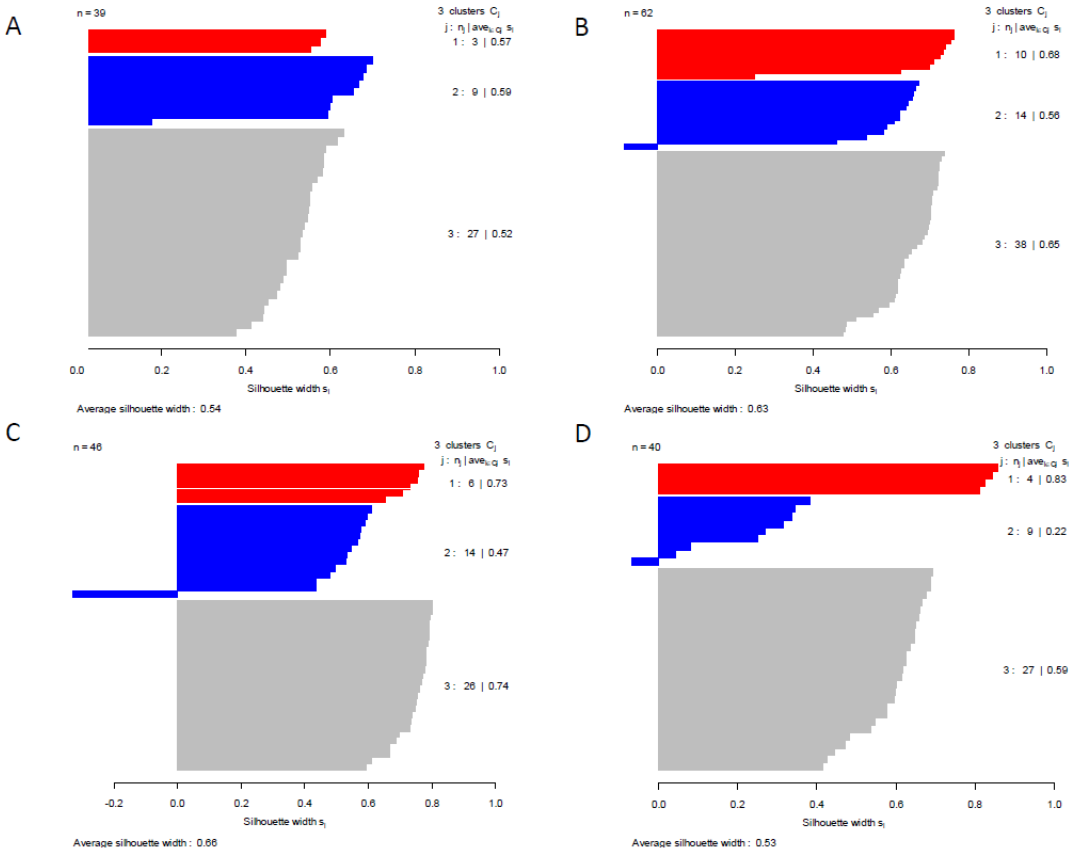


Figure 3.6. Silhouette plots of subgroups assigned by the 13-gene transcriptomic signature demonstrate correctly-clustered samples for 39 / 39 (100%) of the primary investigation cohort and 145 / 148 (98%) of array-based datasets: A, primary investigation cohort; B, Kool et al.; C, Thompson et al.; D, Fattet et al. Discrete subgroups within each data set, which correlate with the WNT and SHH and WNT / SHH-independent subgroups as described in Figures X and Y, are marked (red, WNT; blue , SHH; grey, WNT / SHH independent). A silhouette width (s_i) close to one indicates a very well clustered sample, whereas samples with close to zero width lie between two clusters; samples with negative silhouette widths are potentially placed into the wrong cluster. For each cluster, numbered 1 to 3, the number of cluster members (n_j) and average silhouette width ($ave_{i \varepsilon Cj} S_i$) are given.

3.4.5 Signature positive cases show specific clinical disease features and behaviours

Since the gene expression signatures had been shown to perform consistently well across both GeXP-based and array-based measurements of gene expression, it enabled the molecular subgroup data from all four cohorts to be combined for meta-analysis with the mutational and clinical data which were consistently reported for all studies.

A cohort comprising 173 cases was assessed in this analysis (Table 3.5). Clinical demographics were consistent with previously reported estimates for medulloblastoma (Pizer and Clifford, 2009). There were 115 (68%) classic, 39 (23%) DN and 16 (9%) LCA type medulloblastomas, and 3 cases with data not available. Histological classifications were those reported in the original publications (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006). The gender ratio of the combined cohort was 1.42:1. There were 34 (20%) infant cases (≤3 years old at diagnosis) and 138 (80%) non-infant cases, one case with data not available (overall, median age at diagnosis 6.45, range 0.3 – 35.3 years). There were 30 (21%) M+ disease cases, and 113 (79%) M- disease cases and 44 cases with data not available. Overall, there were 21 (12%) WNT and 42 (24%) SHH subgroup cases observed. The subclassification of the remaining cases has not produced consistent findings in previous studies (Cho et al., 2011; Northcott et al., 2010; Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006). Remaining cases were therefore assigned to the WNT / SHH independent classification as a single group, representing 110 (64%) of cases.

Study	Study ID	Histology	Gender	M Stage	Age	Age (years)	*CTNNB1* mutation	*PTCH1* mutation	Signature	17p loss	9 q loss	Chromosome 6 loss	Exclude
This study	2	LCA	M	1	1	2.5	0	0	WNT/SHH-ind	0	1	0	0
This study	3	CLAS	M	0	1	2.9	0	0	WNT/SHH-ind	1	0	0	0
This study	4	CLAS	M	0	0	4.6	NA	NA	WNT/SHH-ind	NA	NA	NA	0
This study	5	DN	M	0	0	19	0	0	SHH	0	0	0	0
This study	7	CLAS	F	1	1	2.6	0	0	WNT/SHH-ind	0	NA	0	0
This study	9	LCA	M	0	0	4.1	0	0	WNT/SHH-ind	1	NA	0	0
This study	10	LCA	M	0	1	2.5	0	0	WNT/SHH-ind	0	0	0	0
This study	11	CLAS	F	0	0	15.4	0	0	WNT/SHH-ind	0	1	0	0
This study	12	DN	M	1	0	6.3	0	0	WNT/SHH-ind	0	0	0	0
This study	13	DN	M	1	1	2.6	0	0	SHH	0	1	0	0
This study	14	DN	F	0	0	4.8	0	1	SHH	0	NA	0	0
This study	16	DN	F	1	0	5.7	0	0	WNT/SHH-ind	0	0	0	0
This study	17	LCA	M	1	0	16.7	0	1	SHH	0	1	0	0
This study	19	DN	M	0	1	2.9	0	1	SHH	0	0	0	0
This study	20	DN	M	0	0	12.5	0	0	WNT/SHH-ind	NA	0	0	0
This study	21	CLAS	F	0	0	9.8	0	0	WNT/SHH-ind	0	0	0	0
This study	22	DN	F	0	1	1	0	1	SHH	0	0	0	0
This study	24	CLAS	M	0	0	10	0	0	WNT/SHH-ind	1	0	0	0
This study	25	CLAS	M	0	0	12.6	0	0	WNT/SHH-ind	1	0	0	0
This study	26	CLAS	M	0	0	5.1	0	0	WNT/SHH-ind	0	0	0	0
This study	27	CLAS	M	0	0	6.8	0	0	WNT/SHH-ind	0	1	0	0
This study	28	CLAS	F	0	0	8.6	0	0	WNT/SHH-ind	1	0	0	0
This study	30	CLAS	M	0	0	5	0	0	WNT/SHH-ind	0	0	0	0
This study	31	CLAS	M	0	0	10.3	1	0	WNT	0	0	1	0
This study	32	CLAS	M	0	0	11.5	0	0	SHH	0	1	0	0
This study	38	CLAS	F	0	0	3.3	0	0	WNT/SHH-ind	0	1	0	0
This study	40	CLAS	M	0	0	7.5	0	0	WNT/SHH-ind	0	0	0	0
This study	41	CLAS	F	0	0	8.6	0	0	WNT/SHH-ind	1	0	0	0
This study	42	CLAS	M	1	0	5.5	0	0	WNT/SHH-ind	0	0	0	0
This study	43	CLAS	F	1	0	3.5	0	1	SHH	0	1	0	0
This study	44	CLAS	F	0	0	10.2	0	0	WNT/SHH-ind	0	0	0	0
This study	45	DN	F	0	0	14.2	0	0	SHH	0	0	0	0
This study	46	CLAS	M	0	0	5.4	0	0	WNT/SHH-ind	1	0	0	0
This study	50	CLAS	F	0	0	17	0	0	WNT/SHH-ind	1	0	0	0
This study	51	CLAS	F	0	0	4.6	0	0	WNT/SHH-ind	1	0	0	0

Study	Study ID	Histology	Gender	M Stage	Age	Age (years)	*CTNNB1* mutation	*PTCH1* mutation	Signature	17p loss	9 q loss	Chromosome 6 loss	Exclude
This study	52	CLAS	F	0	1	3	0	0	WNT/SHH-ind	0	0	0	0
This study	53	CLAS	M	0	0	4.3	0	0	WNT/SHH-ind	1	0	0	0
This study	54	CLAS	M	0	0	10	1	0	WNT	0	0	1	0
This study	55	CLAS	F	0	0	9	1	0	WNT	0	0	1	0
Kool et al., 2008	255	CLAS	M	NA	0	19.0	1	0	WNT	0	0	1	0
Kool et al., 2008	256	CLAS	M	0	0	7.0	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	258	DN	M	0	1	3.0	0	0	SHH	0	1	0	0
Kool et al., 2008	259	CLAS	M	0	0	15.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	260	CLAS	M	0	0	8.0	1	0	WNT	NA	NA	NA	0
Kool et al., 2008	261	CLAS	M	0	0	4.0	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	262	CLAS	M	0	0	14.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	264	CLAS	F	0	0	7.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	265	DN	M	0	1	3.0	0	1	SHH	0	1	0	0
Kool et al., 2008	267	CLAS	F	0	0	10.0	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	268	CLAS	F	0	0	5.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	269	CLAS	M	0	1	2.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	270	NA	F	0	0	20.0	1	0	WNT	1	1	1	0
Kool et al., 2008	272	DN	M	0	1	2.0	0	0	SHH	0	1	0	0
Kool et al., 2008	273	CLAS	M	0	1	3.0	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	274	CLAS	M	0	0	4.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	275	CLAS	F	0	0	11.0	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	311	CLAS	M	1	0	13.5	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	312	CLAS	M	0	0	7.3	0	0	WNT/SHH-ind	1	1	0	0
Kool et al., 2008	313	CLAS	M	1	0	8.0	0	0	WNT	0	0	0	0
Kool et al., 2008	315	CLAS	M	1	0	6.4	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	316	DN	M	0	1	1.8	0	0	SHH	1	1	0	0
Kool et al., 2008	317	CLAS	M	1	0	3.3	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	318	CLAS	M	1	1	2.8	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	324	CLAS	M	0	1	2.2	0	0	SHH	0	0	0	0
Kool et al., 2008	325	CLAS	M	0	0	3.1	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	326	CLAS	M	0	0	5.9	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	332	DN	F	0	0	4.8	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	334	DN	M	0	0	27.1	0	0	SHH	0	0	0	0
Kool et al., 2008	335	CLAS	F	0	0	35.3	0	0	SHH	1	1	0	0
Kool et al., 2008	336	CLAS	M	0	0	10.3	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	337	CLAS	M	1	0	16.6	0	0	WNT/SHH-ind	0	0	0	0

Study	Study ID	Histology	Gender	M Stage	Age	Age (years)	*CTNNB1* mutation	*PTCH1* mutation	Signature	17p loss	9 q loss	Chromosome 6 loss	Exclude
Kool et al., 2008	338	DN	F	0	0	5.0	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	339	DN	M	0	0	5.3	0	0	WNT/SHH-ind	1	1	0	0
Kool et al., 2008	340	CLAS	F	0	0	7.8	1	0	WNT	0	0	1	0
Kool et al., 2008	341	CLAS	M	0	0	12.2	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	342	CLAS	F	0	1	1.5	0	1	SHH	1	1	0	0
Kool et al., 2008	343	CLAS	F	0	0	25.6	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	365	CLAS	M	0	0	10.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	367	CLAS	F	0	0	6.0	0	0	WNT/SHH-ind	NA	NA	NA	0
Kool et al., 2008	368	CLAS	M	0	0	7.4	0	0	WNT/SHH-ind	NA	NA	NA	0
Kool et al., 2008	369	CLAS	M	0	0	10.4	1	0	WNT	0	0	1	1
Kool et al., 2008	370	CLAS	M	1	0	4.9	0	0	WNT/SHH-ind	NA	NA	NA	0
Kool et al., 2008	371	CLAS	F	0	0	11.2	1	0	WNT	NA	NA	NA	0
Kool et al., 2008	372	CLAS	M	0	0	12.7	1	0	WNT	NA	NA	NA	0
Kool et al., 2008	373	LCA	M	0	1	2.8	0	1	SHH	NA	NA	NA	0
Kool et al., 2008	374	NA	F	NA	NA	NA	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	377	CLAS	M	0	0	5.4	0	0	WNT/SHH-ind	1	0	0	1
Kool et al., 2008	379	CLAS	M	NA	0	10.0	1	0	WNT	0	0	1	1
Kool et al., 2008	421	CLAS	M	0	0	6.0	0	0	WNT/SHH-ind	NA	NA	NA	0
Kool et al., 2008	424	CLAS	M	1	0	5.0	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	425	CLAS	F	0	0	6.0	1	0	WNT	NA	NA	NA	0
Kool et al., 2008	426	CLAS	M	1	1	3.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	427	DN	M	0	0	7.0	0	1	SHH	0	1	0	0
Kool et al., 2008	434	DN	F	0	0	12.0	0	0	WNT/SHH-ind	0	0	0	0
Kool et al., 2008	435	DN	F	0	1	3.0	0	0	SHH	0	0	0	0
Kool et al., 2008	440	DN	F	0	0	3.7	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	446	CLAS	F	0	0	6.3	0	0	SHH	1	1	0	0
Kool et al., 2008	447	CLAS	F	0	0	31.0	0	0	SHH	0	0	0	0
Kool et al., 2008	452	CLAS	M	NA	1	2.5	0	0	SHH	0	0	0	0
Kool et al., 2008	455	CLAS	F	0	1	2.4	0	0	WNT/SHH-ind	1	0	0	0
Kool et al., 2008	458	DN	M	0	0	3.75	0	0	WNT/SHH-ind	0	0	0	0
Thompson et al., 2006	1	LCA	M	0	0	4.1	0	0	WNT/SHH-ind	1	NA	0	1
Thompson et al., 2006	2	DN	M	0	0	12.5	0	0	WNT/SHH-ind	1	NA	0	1
Thompson et al., 2006	3	CLAS	F	0	0	15.4	0	0	WNT/SHH-ind	0	NA	0	1
Thompson et al., 2006	4	DN	F	0	0	9	0	0	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	5	DN	F	1	0	5.7	0	0	WNT/SHH-ind	0	0	0	1
Thompson et al., 2006	6	CLAS	F	1	1	2.6	0	0	WNT/SHH-ind	0	NA	0	1

Study	Study ID	Histology	Gender	M Stage	Age	Age (years)	*CTNNB1* mutation	*PTCH1* mutation	Signature	17p loss	9 q loss	Chromosome 6 loss	Exclude
Thompson et al., 2006	7	CLAS	M	0	0	9.2	0	0	WNT/SHH-ind	0	NA	0	0
Thompson et al., 2006	8	LCA	M	0	1	2.5	0	0	WNT/SHH-ind	0	NA	0	1
Thompson et al., 2006	9	CLAS	F	0	0	3.6	0	0	WNT/SHH-ind	0	NA	0	0
Thompson et al., 2006	10	CLAS	F	1	1	2.8	0	0	WNT/SHH-ind	0	NA	0	0
Thompson et al., 2006	11	CLAS	F	0	0	10.7	NA	NA	WNT	0	NA	1	0
Thompson et al., 2006	12	CLAS	M	0	0	7.3	1	0	WNT	0	NA	1	0
Thompson et al., 2006	13	CLAS	M	0	0	16.2	1	0	WNT	0	NA	1	0
Thompson et al., 2006	14	CLAS	F	0	0	9.2	1	0	WNT	0	NA	0	0
Thompson et al., 2006	15	CLAS	F	0	0	10.8	1	0	WNT	0	NA	1	0
Thompson et al., 2006	16	CLAS	M	0	0	7.3	1	0	WNT	0	NA	NA	0
Thompson et al., 2006	17	CLAS	M	0	0	5.3	NA	NA	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	18	DN	M	1	0	7.5	NA	0	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	19	CLAS	F	0	0	7.3	0	NA	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	20	CLAS	F	0	0	9.8	0	0	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	21	CLAS	M	1	0	4.3	0	0	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	22	CLAS	M	0	0	11.5	0	0	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	23	DN	F	0	0	12.5	0	0	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	24	CLAS	M	0	0	10.7	0	0	WNT/SHH-ind	NA	NA	0	0
Thompson et al., 2006	25	CLAS	F	0	0	10.4	0	0	WNT/SHH-ind	1	NA	0	0
Thompson et al., 2006	26	DN	F	1	1	1.8	0	0	SHH	0	NA	1	0
Thompson et al., 2006	27	LCA	M	0	0	12.5	0	1	SHH	1	NA	0	0
Thompson et al., 2006	28	LCA	F	0	0	11.6	0	0	SHH	1	NA	0	0
Thompson et al., 2006	29	LCA	M	0	0	10.9	0	0	SHH	1	NA	0	0
Thompson et al., 2006	30	LCA	M	0	0	5.6	0	0	SHH	0	NA	NA	0
Thompson et al., 2006	31	DN	F	0	1	1	0	1	SHH	0	0	0	1
Thompson et al., 2006	32	LCA	M	1	0	16.7	0	1	SHH	0	1	0	1
Thompson et al., 2006	33	DN	F	0	0	4.8	0	1	SHH	0	NA	0	1
Thompson et al., 2006	34	DN	M	1	1	2.5	0	1	SHH	0	1	0	1
Thompson et al., 2006	35	DN	F	0	1	2.2	0	1	SHH	0	NA	0	0
Thompson et al., 2006	36	DN	F	0	1	1.2	0	0	SHH	0	NA	0	0
Thompson et al., 2006	37	DN	M	0	1	1.5	0	0	SHH	0	NA	0	0
Thompson et al., 2006	38	DN	F	1	1	1.2	NA	NA	SHH	0	NA	1	0
Thompson et al., 2006	39	LCA	F	0	0	6.3	0	0	WNT/SHH-ind	0	NA	0	0

Study	Study ID	Histology	Gender	M Stage	Age	Age (years)	*CTNNB1* mutation	*PTCH1* mutation	Signature	17p loss	9 q loss	Chromosome 6 loss	Exclude
Thompson et al., 2006	40	LCA	M	1	0	5.8	0	0	WNT/SHH-ind	0	NA	0	0
Thompson et al., 2006	41	LCA	M	1	0	6.6	0	0	WNT/SHH-ind	0	NA	0	0
Thompson et al., 2006	42	CLAS	M	0	1	2.9	0	0	WNT/SHH-ind	1	0	0	1
Thompson et al., 2006	43	LCA	F	1	0	3.3	0	0	WNT/SHH-ind	0	NA	0	0
Thompson et al., 2006	44	DN	M	0	0	5	0	0	SHH	0	NA	0	0
Thompson et al., 2006	45	LCA	M	1	1	3	0	0	WNT/SHH-ind	0	NA	0	0
Thompson et al., 2006	46	CLAS	F	0	0	8.8	0	0	WNT/SHH-ind	0	NA	0	0
Fattet et al., 2009	MB79	CLAS	NA	NA	1	3.0	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB80	DN	NA	NA	0	8.0	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB81a	LCA	NA	NA	0	9.1	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB82	DN	NA	NA	0	3.4	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB87a	CLAS	NA	NA	0	7.3	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB88	CLAS	NA	NA	0	9.5	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB89	CLAS	NA	NA	0	11.2	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB91	CLAS	NA	NA	0	4.9	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB92	CLAS	NA	NA	0	8.9	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB93	DN	NA	NA	0	5.8	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB95	LCA	NA	NA	0	3.4	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB96	CLAS	NA	NA	0	3.7	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB99	NA	NA	NA	0	8.2	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB100	CLAS	NA	NA	0	7.6	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB101	CLAS	NA	NA	0	5.6	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB102	CLAS	NA	NA	0	9.0	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB105	DN	NA	NA	1	1.0	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB106	CLAS	NA	NA	0	4.2	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB107	CLAS	NA	NA	0	10.5	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB108	CLAS	NA	NA	0	8.0	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB109	CLAS	NA	NA	0	5.3	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB112	CLAS	NA	NA	0	11.5	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB116	CLAS	NA	NA	0	12.3	1	NA	WNT	NA	NA	1	0
Fattet et al., 2009	MB117	DN	NA	NA	1	0.3	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB118	CLAS	NA	NA	0	9.6	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB119	CLAS	NA	NA	0	7.8	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB120	CLAS	NA	NA	0	9.0	1	NA	WNT	NA	NA	1	0
Fattet et al., 2009	MB121	CLAS	NA	NA	0	5.8	0	NA	WNT/SHH-ind	NA	NA	0	0

Study	Study ID	Histology	Gender	M Stage	Age	Age (years)	*CTNNB1* mutation	*PTCH1* mutation	Signature	17p loss	9 q loss	Chromosome 6 loss	Exclude
Fattet et al., 2009	MB122	CLAS	NA	NA	0	7.1	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB123	DN	NA	NA	0	3.1	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB124	CLAS	NA	NA	0	10.2	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB125	CLAS	NA	NA	0	3.5	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB127	CLAS	NA	NA	0	5.6	1	NA	WNT	NA	NA	1	0
Fattet et al., 2009	MB128	DN	NA	NA	0	6.5	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB130	DN	NA	NA	1	2.0	0	NA	SHH	NA	NA	0	0
Fattet et al., 2009	MB131	CLAS	NA	NA	0	10.9	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB133	CLAS	NA	NA	0	13.2	1	NA	WNT	NA	NA	1	0
Fattet et al., 2009	MB134	CLAS	NA	NA	0	11.0	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB135	CLAS	NA	NA	1	2.8	0	NA	WNT/SHH-ind	NA	NA	0	0
Fattet et al., 2009	MB136	CLAS	NA	NA	0	5.0	0	NA	WNT/SHH-ind	NA	NA	0	0

Table 3.5. Clinical and molecular features of the combined cohort used for correlative analyses. Reported clinical, pathological and molecular features for three transcriptomic cohorts (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006) together with the cohort described in this study, are shown. Histology (CLAS, classic; LCA, large cell / anaplastic; DN, desmoplastic / nodular). Gender (M, male; F, female). M stage (0, M0 / 1; 1, M2 / 3). Age (>3 years, 0; ≤3 years, 1). *CTNNB1* and *PTCH1* mutation status (0, no mutation detected; 1 – mutation detected). Chromosome 17p and 9q loss, assessed by LOH analysis (this study), array CGH (Kool et al., 2008), FISH (Thompson et al., 2006) (0, no loss detected; 1, loss detected). Chromosome 6 loss, assessed using LOH analysis (this study), array CGH (Fattet et al., 2009; Kool et al., 2008) and FISH (Thompson et al., 2006) (0, no loss detected; 1, loss detected). Duplicate samples analysed in both the current study cohort and other cohorts are marked (exclude column; 0, include; 1, exclude) and were removed from correlative analyses. NA – not available.

3.4.6 Genetic and epigenetic mechanisms of SHH and WNT pathway activation

The patterns of mutation observed for *CTNNB1* and *PTCH1* served to further validate the fidelity of the gene expression signatures developed, since they were exclusively detected in WNT and SHH subgroup cases, respectively, where data was available (Figure 3.3) (combined cohort 'p' = 5.3 x 10^{-8} and 'p' = 0 (chi-squared test), respectively).

Mutation of *CTNNB1* was the primary mechanism and correlate of WNT pathway activation, observed in 19 / 20 (95%) cases where mutational and expression signature data were available. Consistent with this, no *APC* mutations were observed within the 55 cases tested in the primary investigation DNA cohort.

PTCH1 mutation was a major mechanism of SHH pathway activation. 11 / 32 (34%) SHH subgroup cases investigated harboured a *PTCH1* mutation (Figure 3.3). Therefore, a systematic investigation of alternative genetic mechanisms of pathway activation was undertaken in our primary investigation cohort, to identify possible causative mechanisms. This mutational analysis incorporated investigations into all known pathway genes in which mutations have previously been reported; *PTCH1* (Kogerman et al., 2002; Pietsch et al., 1997; Raffel et al., 1997; Wolter et al., 1997), *SUFU* (Taylor et al., 2002) and *SMO* (Reifenberger et al., 1998).

In addition to 5 truncating mutations of *PTCH1* identified, only one further potentially pathogenic missense *SUFU* mutation was identified, with no evidence for *SMO* mutation found (Table 3.6). Additional non-pathogenic variants discovered (e.g. polymorphic variants or intronic changes) are summarised in Table 3.7. Allelic loss of chromosome 17p, targeting *REN* (*KCTD11*) at 17p13.2 has also been previously associated with SHH pathway activation in medulloblastoma (De Smaele et al., 2004; Di Marcotullio et al., 2004) and was observed in 9 / 37 (24%) of cases tested.

			Molecular characteristics				Clinical characteristics		
Case no.	**Gene**	**Exon**	**Nucleotide**	**Amino acid**	**Nature of mutation**	**9q LOH**	**Age (yrs)**	**Sex**	**Pathology**
43	*PTCH1*	2	239delGA	Leu79FS	Truncating	Y	3.5	F	Classic
19	*PTCH1*	3	572insTA	Tyr191FS	Truncating	N	2.9	M	DN
37	*PTCH1*	9	1341delACTC	Leu447FS	Truncating	Y	7.9	M	Classic
17	*PTCH1*	10	1350delC	Leu450FS	Truncating	Y	16.7	M	LCA
14	*PTCH1*	14	2011insC	His671FS	Truncating	Y	4.8	F	DN
22	*PTCH1*	19	3205delGGCATGATGGG	Gly1069FS	Truncating	N	1.0	F	DN
10	*SUFU*	9	C1084T	Arg362Cys	Missense[a]	N	2.5	M	LCA

Table 3.6. Genetic mechanisms of SHH pathway activation in medulloblastoma. Coding sequence mutations detected in *PTCH1*, *SUFU* and *SMO* in 55 primary medulloblastomas. LOH, loss of heterozygosity; M, male; F, female. Nucleotide and amino acid positions are shown relative to Build 36 of the NCBI genomic sequence assembly (Genbank gene IDs 5727 (*PTCH1*), 51684 (*SUFU*) and 6608 (*SMO*); http://www.ncbi.nlm.nih.gov/). [a]variation not identified in 100 non-neoplastic DNA samples. No mutations were detected in *SMO*.

3.4.6.1 Epigenetic mechanisms of SHH pathway activation

Epigenetic mechanisms of SHH pathway activation were additionally investigated, as an alternative to genetic determinants of pathway activation. Two predicted promoter-associated CpG islands, spanning exons 1a and 1c of *PTCH1,* and a promoter-associated CpG island within *SUFU,* were identified and investigated for any evidence of DNA hypermethylation, which may help determine transcriptional silencing (section 1.1.8).

There was no evidence for DNA methylation of the *PTCH1* exon 1a-associated or the *SUFU* CpG island observed in any tumour tested (*n* = 39; Table 3.8), indicating that any transcriptional silencing at these loci through DNA hypermethylation is unlikely to be important in the activation of the SHH pathway in medulloblastoma. In contrast, 12 / 27 (44%) of tumours investigated showed evidence of DNA hypermethylation of *PTCH1* exon 1c, suggesting a potential role for this epigenetic mechanism. Unlike *PTCH1* mutations, however, this hypermethylation was exclusively observed in non-SHH subgroup medulloblastomas, indicating that methylation-dependent silencing at this locus is unlikely to be important in the pathogenesis of *SHH* medulloblastomas. Similarly, no significant associations between any other identified defects associated

with SHH pathway activation (*SUFU* missense mutation and 17p allelic loss (*REN; KCTD11*)) and SHH subgroup status, were identified (Table 3.8).

Gene	Exon	Nucleotide variation	Amino acid variation	Frequency	Nature of Variation
SMO	6	G1203C	Ala401Ala	1/55 (1.8%)	Missense mutation (synonymous)[a]
PTCH1	22	A3583T	Thr1195Ser	1/55 (1.8%)	Polymorphism (non-synonymous)[b]
PTCH1	23a	C3845T	Pro1282Leu	1/55 (1.8%)	Polymorphism (non-synonymous)[b]
PTCH1	23a	C3944T	Pro1315Leu	29/55 (53%)	Polymorphism (non-synonymous)[a,b]
SUFU	8	G1018T	Ala340Ser	2/55 (3.6%)	Polymorphism (non-synonymous)[a,b]
SMO	1	49ins3	17InsLeu	5/55 (9.1%)	Polymorphism (non-synonymous)[c]
SMO	4	G808A	Val269Ile	2/55 (3.6%)	Polymorphism (non-synonymous)[c]
PTCH1	2	C318T	Leu106Leu	3/55 (5.5%)	Polymorphism (synonymous)[b,c]
PTCH1	5	A735G	Thr245Thr	2/55 (3.6%)	Polymorphism (synonymous)[a,b]
PTCH1	12	T1665C	Asn555Asn	17/55 (31%)	Polymorphism (synonymous)[b,c]
PTCH1	12	C1686T	Ala562Ala	14/55 (25.5%)	Polymorphism (synonymous)[b,c]
PTCH1	14b	A2199G	Ser733Ser	13/55 (23.6%)	Polymorphism (synonymous)[b]
PTCH1	23a	G3954A	Pro1318Pro	1/55 (1.8%)	Polymorphism (synonymous)[b]
SMO	4	G852A	Gln284Gln	2/55 (3.6%)	Polymorphism (synonymous)[c]
SMO	5	G1137A	Ala379Ala	1/55 (1.8%)	Polymorphism (synonymous)[c]
SMO	6	G1164C	Gly388Gly	5/55 (9.1%)	Polymorphism (synonymous)[b]
SMO	10	T1722C	Ser574Ser	1/55 (1.8%)	Polymorphism (synonymous)[b]
SMO	12	G2052A	Pro684Pro	4/55 (7.3%)	Polymorphism (synonymous)[c]
SMO	12	A1989G	Pro663Pro	1/55 (1.8%)	Polymorphism (synonymous)[a]
PTCH1	1B	1IVS1B-10 del6	-	1/55 (1.8%)	Intronic variation[d]
PTCH1	1B	1IVS1B-10 ins3	-	11/55 (20%)	Intronic variation[d]
PTCH1	10	1503IVS10+12 g>a	-	1/55 (1.8%)	Intronic variation[d]
PTCH1	11	1504IVS11-12 t>c	-	1/55 (1.8%)	Intronic variation[d]
PTCH1	14	2250IVS14+25 t>c	-	2/55 (3.6%)	Intronic variation[d]
PTCH1	15	2560IVS15+9 g>c	-	30/55 (55%)	Intronic variation[d]
PTCH1	15	2560IVS15+8 g>c	-	1/55 (1.8%)	Intronic variation[d]
PTCH1	17	2887IVS17+21 a>g	-	31/55 (56%)	Intronic variation[d]
SUFU	8	1022IVS8+49 t>c	-	3/55 (5.5%)	Intronic variation[d]
SUFU	9	1023IVS9-23 g>c	-	9/55 (16.4%)	Intronic variation[d]
SMO	1	1IVS1-35 g>t	-	1/55 (1.8%)	Intronic variation[d]
SMO	2	332IVS2-56 t>c	-	41/55 (74.5%)	Intronic variation[d]
SMO	3	537IVS3-26 t>c	-	16/55 (29.1%)	Intronic variation[d]
SMO	3	747IVS3+24 c>g	-	16/55 (29.1%)	Intronic variation[d]
SMO	4	920IVS4+79 t>c	-	14/55 (25.5%)	Intronic variation[d]
SMO	4	920IVS4+55 c>a	-	1/55 (1.8%)	Intronic variation[d]
SMO	4	920IVS4+68 g>a	-	2/55 (3.6%)	Intronic variation[d]
SMO	5	1140IVS5+13 g>a	-	2/55 (3.6%)	Intronic variation[d]
SMO	6	1264IVS6+41 a>g	-	5/55 (9.1%)	Intronic variation[d]
SMO	8	1357IVS8-33 g>c	-	1/55 (1.8%)	Intronic variation[d]
SMO	10	1652IVS10-72 g>a	-	1/55 (1.8%)	Intronic variation[d]

Table 3.7. Additional DNA sequence variations detected in *PTCH1*, *SUFU* and *SMO* in 55 primary medulloblastomas. Nucleotide and amino acid positions are shown relative to Build 36 of the NCBI genomic sequence assembly (Genbank gene IDs 5727 (*PTCH1*), 51684 (*SUFU*) and 6608 (SMO); http://www.ncbi.nlm.nih.gov/). [a]variation not identified in 100 non-neoplastic DNA samples, [b]variation previously reported in the literature or dbSNP (http://www.ncbi.nlm.nih.gov/), [c]variation identified in paired constitutional (blood) DNA sample, [d]intronic variations were not predicted to affect splice sites and were not characterised further.

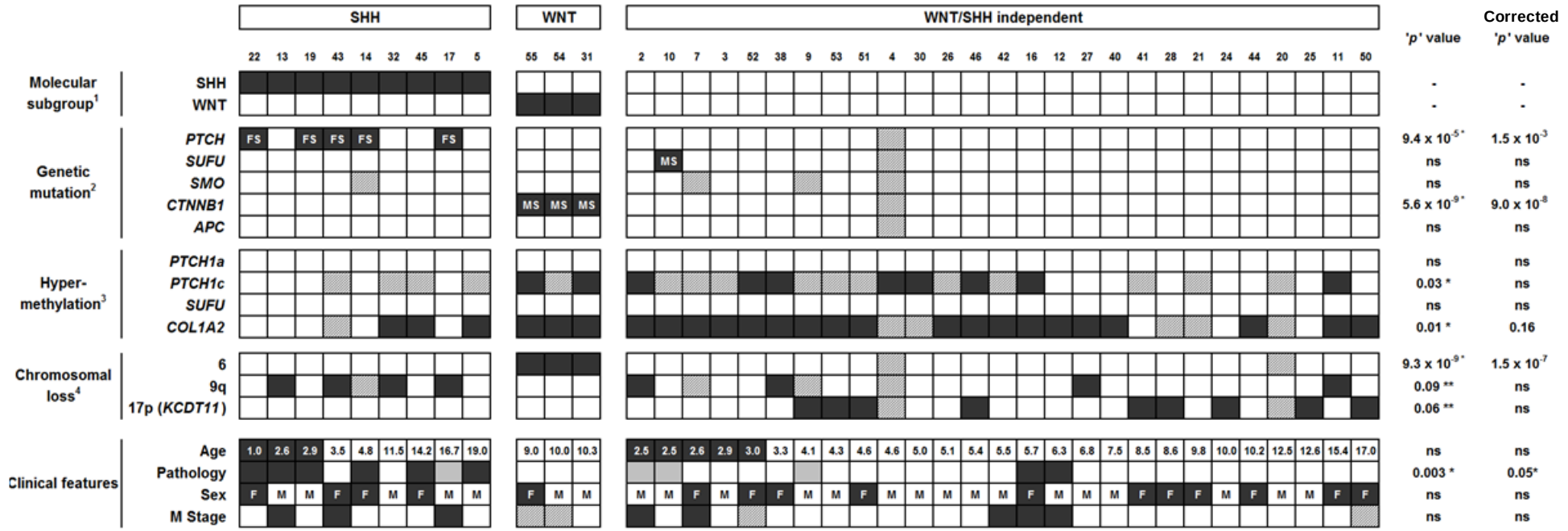


Table 3.8**. Associations between molecular subgroups and medulloblastoma genomic, epigenomic and clinical disease features**. NOTE. Cases are shown arranged by signature status (SHH, WNT, WNT / SHH independent) as determined in Figure 3.3, and then by ascending age, [a]molecular subgroup; [b]mutation status; [c]hypermethylation status (black, methylated; white, unmethylated); and [d]chromosomal loss (black, allelic; white, no loss). Age is shown in years and categorized into infants (≤3 years, black) and non-infants (>3 years, white). Pathology variant is indicated by a white square (classic), black square (DN) or grey square (LCA). Gender is indicated by black squares (F, female) and white squares (M, male). M- disease is indicated by white squares and M+ disease by black squares. Raw 'p' and 'p' values corrected for multiple hypothesis testing are shown for relationships between molecular / clinical correlates and subgroup membership (chi-squared tests, Bonferroni correction). Significant (*'P' < 0.05) and marginally significant (**'P' = 0.051–0.10) associations are marked. Diagonally hatched grey boxes indicate unavailable data. Abbreviations: FS, frameshift; MS, missense mutation; ns, not significant.

3.4.7 Genomic biomarkers of SHH and WNT pathway activation

An investigation of selected medulloblastoma chromosomal abnormalities (chromosome 6, 9q and 17p loss) and epigenomic defects (*COL1A2* status) of biological and / or prognostic significance (Pizer and Clifford, 2009; Anderton et al., 2008) was undertaken, for their associations with the SHH and WNT disease subgroups, and each other, to assess any utility as biomarkers of pathway activation.

Chromosome 6 loss, *CTNNB1* mutation and the absence of chromosome 9 and 17 abnormalities, were observed in all WNT cases in the primary cohort (Table 3.8), consistent with previous findings (Fattet et al., 2009; Kool et al., 2008; Clifford et al., 2006; Thompson et al., 2006). Across the combined cohort, evidence of loss of an entire copy of chromosome 6 was associated with 88% (14 / 16) of WNT cases with available data ('p' < 3 x 10^{-16}, Fisher's Exact test), however this relationship was not exclusive; chromosome 6 loss was also detected in occasional non-WNT cases (2 / 145 (1.4%)). 8 of 35 tumours tested (23%) showed evidence of genetic loss at the 9q22.3 region surrounding the *PTCH1* locus in our primary cohort (Table 3.8). 4 of 8 were in the SHH subgroup and 2 / 4 tumours with *PTCH1* mutations showed LOH of 9q22.3, however neither association reached significance ('*p'* = 0.06 and 0.22 respectively, Fisher's Exact test). A significant inverse association between 17p loss and membership of the SHH and WNT subgroups was observed; 17p losses were exclusively observed in WNT / SHH independent cases (17p LOH in 0 / 12 SHH or WNT cases *vs* 10 / 26 WNT / SHH independent cases ('*p'* = 0.02, Fisher's Exact test)). In addition, *COL1A2* hypermethylation was detected in 76% (25 / 33) of cases; an absence of *COL1A2* methylation was significantly associated with the SHH subgroup ('*p'* = 0.01, Chi-squared test), but this relationship was not maintained when a correction for multiple hypothesis testing was applied (Table 3.8).

3.4.8 Distinct clinical features of medulloblastoma molecular subgroups

Analysis of the medulloblastoma molecular disease subgroups, defined by the gene expression signatures, revealed striking differences in their clinical disease features.

3.4.8.1 Signature positive cases display different ages of incidence

WNT / SHH independent tumours comprised the majority of cases and had their peak incidence in the 3-to-6 year age group, but were extremely rare in the first 2 years of life (Figure 3.7). In contrast, SHH subgroup tumours had a major peak in infancy (21 / 42 (50%) of SHH cases were diagnosed ≤ 3 years of age). SHH subgroup tumours represented the majority of the infant clinical group (21 / 34 (62%) of cases ≤ 3 years of age), but were less common in non-infant children (16 / 127 (15%) cases > 3-15 years of age), and were the majority of adult cases (5 / 11 (45%) cases ≥ 16 years of age, overall 'p' = 5.8 x 10^{-9}, chi-squared test). Below 2 years of age, almost all cases (11 / 12 (92%)) were SHH positive. WNT subgroup cases were not observed in infants (minimum age observed was 5 years) and had a bi-modal age distribution with major and minor peaks at 10 and 20 years, respectively.

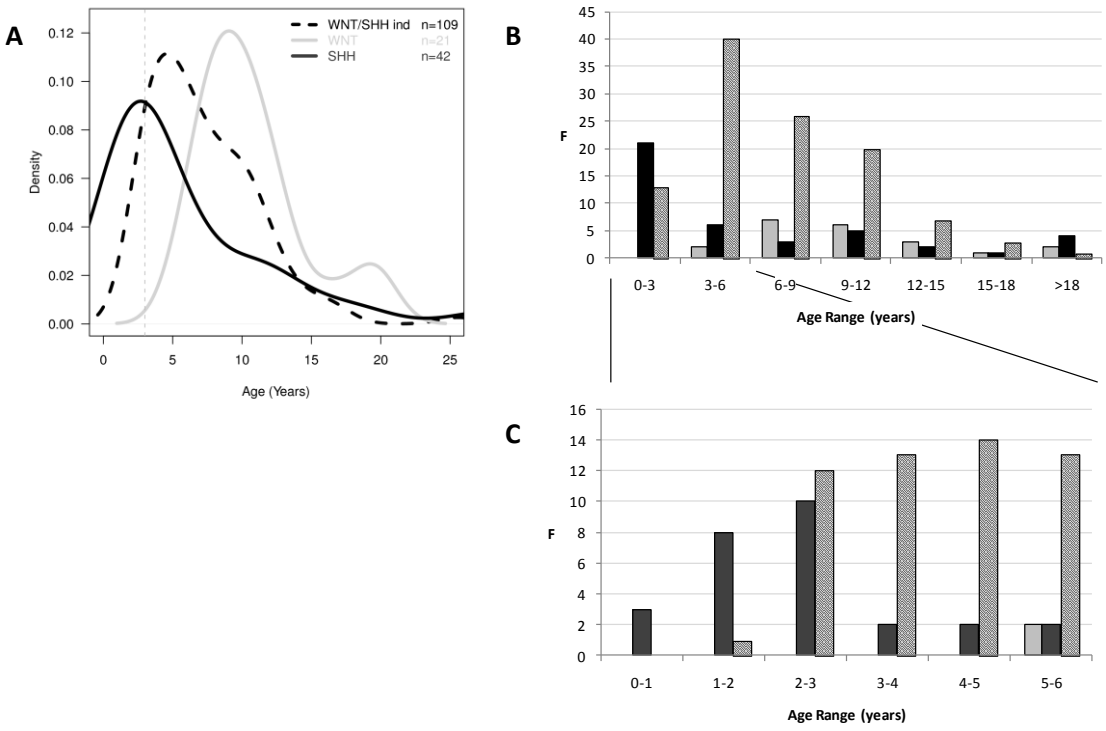


Figure 3.7. Medulloblastoma molecular subgroups show distinct age of incidence distributions. Data for the WNT (grey), SHH (black), and WNT / SHH independent (hatched) subgroups are shown, based on a combined cohort of 173 medulloblastomas. A, density plots show subgroup dependent ages of incidence. Case density represents the smoothed frequency of incidence within each of the 3 subgroups. Grey dotted line is plotted at 3 years of age. B, bar plots show age distribution of data set. C, bar plot shows age distribution of cases aged ≤6 years at diagnosis. F, frequency.

3.4.8.2 Subgroups show no differences in gender ratios

For cases with available gender information, WNT subgroup cases suggested gender parity (9 M: 8 F, ratio 1.1), in contrast to SHH (19 M: 14 F, ratio 1.4) and WNT / SHH independent cases (50 M: 33 F, ratio 1.5), which showed a male enrichment, although this was not significant ('p' = 0.85, chi-squared test) (Figure 3.8).

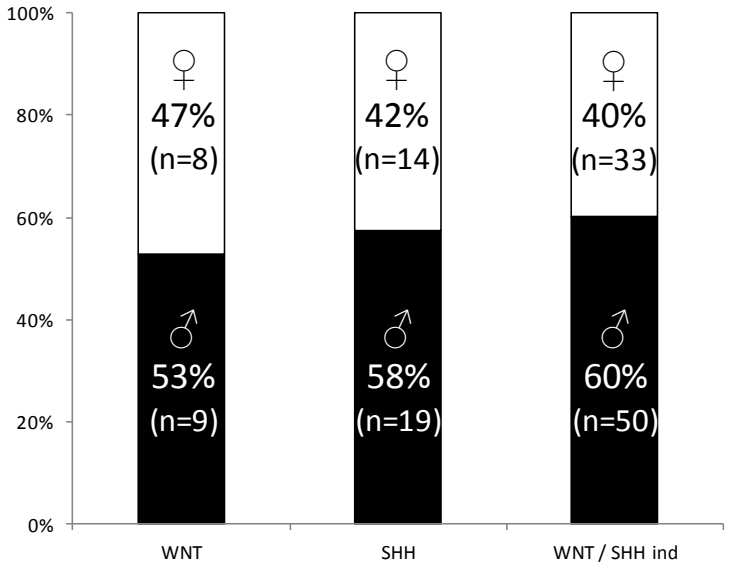


Figure 3.8. WNT subgroup suggests gender parity. SHH and WNT / SHH independent subgroup cases show enrichment for male cases (black) over female cases (white), in contrast to WNT subgroup cases, which had approximately equal gender ratio.

3.4.8.3 Signature positive cases show associations with specific histopathological subtypes

Significant differences were also observed in the distribution of medulloblastoma histological subtypes between the molecular subgroups ('p' = 9 X 10^{-12}, Chi-squared test; Figure 3.9, Figure 3.10). WNT subgroup cases exclusively displayed classic histology ('p' = 0.0003, Fisher's Exact test), and WNT / SHH independent tumours were also predominantly of the classic subtype, but DN and LCA cases were also observed. Consistent with previous studies (Kool et al., 2008; Thompson et al., 2006; Raffel et al., 1997; Wolter et al., 1997), SHH cases were overall strongly associated with DN histology ('p' = 2.1 X 10^{-8}, Fisher's exact test). However, this relationship was not absolute and LCA and classic cases were also observed in the SHH group. Most notably, examination of this large cohort revealed the relationship between SHH activation and DN pathology to be age dependent (Figure 3.9, Figure 3.10); DN cases made up the majority of infant (≤ 3 years old) SHH subgroup cases; all DN cases in this infant group displayed SHH activation. DN pathology may therefore serve as a surrogate marker of SHH activation in the infant group. In contrast, there were almost equal proportions of DN, LCA and classic cases in SHH-expressing non-infant cases, and the majority of non-infant DN tumours were not SHH activated ('p' < 0.0001, Fisher's Exact test). No significant evidence for differences in metastatic stage (WNT 7% (1 / 16) M+ disease, SHH 16% (5 / 32) and WNT / SHH-independent 24% (20 / 82)) were observed between the different expression subgroups ('p' = 0.20, Chi-squared).

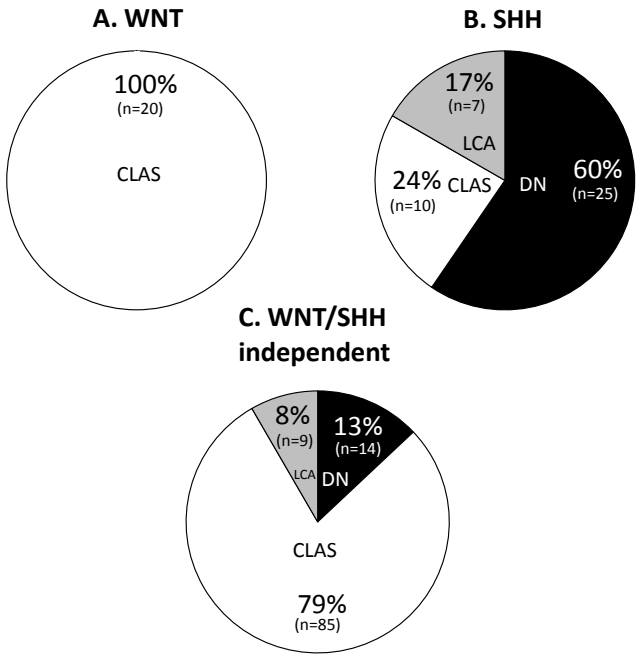


Figure 3.9. Molecular subgroups show relationships to medulloblastoma histological variants. Subgroup and histological information was available for 170 of 173 cases (Table 3.5). A, WNT subgroup (*n* = 20); B, SHH subgroup (*n* = 42); C, WNT / SHH independent cases (*n* = 108; 'p' = 3.1 x 10^{-11}, chi-squared test). White, classic (CLAS); grey, LCA; black DN histological variants.

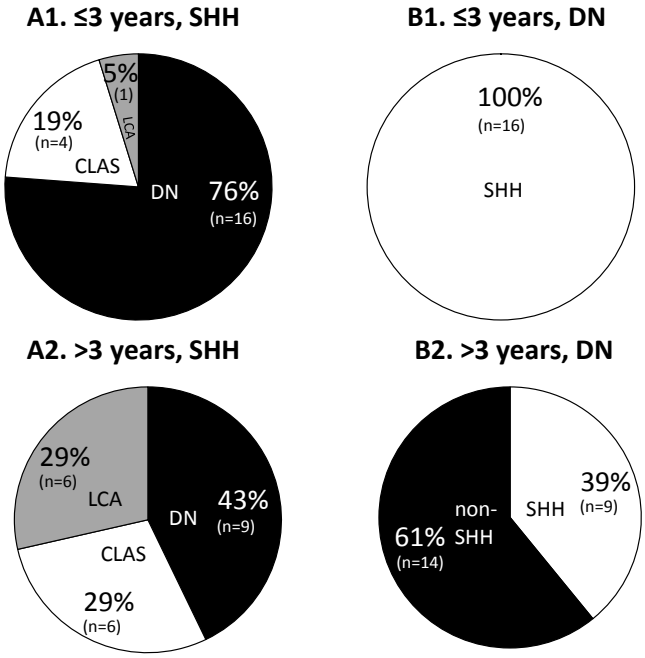


Figure 3.10. Associations between SHH subgroup medulloblastomas and DN histology are age-dependent. A, histological variants (white, classic (CLAS); grey, LCA; black DN) show significantly different distributions ('p' = 0.05; chi-squared test) in SHH subgroup cases arising in infants (≤3 years at diagnosis (*n* = 21); A1) and non-infants (>3 years (*n* = 21); A2). B, infant (*n* = 16; B1) and non-infant (*n* = 23; B2) DN medulloblastomas show significantly different relationships to the SHH subgroup ('p' = 8.6 x 10^{-5}; Fisher's exact test; SHH subgroup, white; non-SHH, black).

3.5 Discussion

3.5.1 Expression signatures of Wnt and SHH pathway activation

The identification of distinct medulloblastoma subgroups (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006) offers significant potential to improve the understanding of disease biology and clinical management. This chapter reported the development and validation of minimal diagnostic gene expression signatures, which can routinely be applied to identify the SHH, WNT and WNT / SHH independent medulloblastoma disease subgroups. These gene expression signatures are robust and informative for subgroup identification in RNA extracted from snap-frozen tumour material, using different gene expression assays. In particular, the GeXP assay reported here offers a number of advantages over microarray methodologies for the routine assignment of subgroup membership. The assay is easy to perform, takes less than a day to complete, is cost effective (less than one tenth the costs of performing microarray) and, importantly, can be performed on small amounts of total RNA (150 ng, compared to 500 ng to 5 µg for a typical expression microarray analysis). The removal of the significant disadvantages associated with microarray analysis (time consuming, complex and the need for expensive array analysis platforms) for subgroup assignment provides a strong basis for their clinical application; the GeXP method presented here is feasible for investigation in real time across multiple treatment centres during clinical treatment and in future clinical trials.

We have shown that the disease subgroups recognised by these signatures are equivalent and consistently identified in four independent medulloblastoma cohorts, allowing their assembly into a large combined cohort. Coupled with an extensive analysis of our novel primary cohort, this has allowed significant insights into the underlying molecular mechanisms, associated biomarkers, and clinical characteristics of these molecular disease subgroups.

3.5.2 Genetic and epigenetic subgroup defects

Our systematic investigation of specific medulloblastoma genetic and epigenetic defects in this study has informed their roles as determinants or correlates of the molecular subgroups identified.

3.5.2.1 Correlates of Wnt pathway activation

Consistent with previous studies (Kool et al., 2008; Clifford et al., 2006; Thompson et al., 2006), *CTNNB1* mutations were identified as the primary pathway activating event present in almost all WNT subgroup tumours, with chromosome 6 losses also affecting the majority of these cases.

3.5.2.2 Correlates of SHH pathway activation

PTCH1 mutation was the major mechanistic correlate of SHH activation, identified in ~34% of SHH cases. SHH-associated *PTCH1* mutations were detected both in conjunction with chromosome 9q loss, and in the heterozygous state, indicating disruption of a single *PTCH1* allele can be sufficient to cause SHH pathway disruption in medulloblastoma. An absence of *COL1A2* hypermethylation is also significantly associated with SHH subgroup medulloblastomas, most strongly in infant cases. Notably, a number of the previously reported determinants of SHH activation that we examined (*PTCH1* exon 1c methylation (Diede et al., 2010), *SUFU* missense mutation (Taylor et al., 2002) and 17p (*REN* (*KCTD11*) (Di Marcotullio et al., 2004) allelic loss) were not specifically associated with the SHH subgroup, indicating any role they may play in medulloblastoma is SHH-independent. Additionally, other SHH pathway defects examined (*PTCH1* exon 1a hypermethylation), including events previously reported in medulloblastoma (*SMO* mutations (Reifenberger et al., 1998) or *SUFU* truncating mutations (Taylor et al., 2002)) were not observed at all, suggesting their roles are either less common than previously thought, or are restricted to limited tumour subsets less well represented in our mutation screening cohort. This is likely the case for *SUFU* mutations, which appear to be associated with germline inheritance and have their peak incidence in infants (Brugieres et al., 2010; Scott et al., 2006; Taylor et al., 2002). Further mechanisms of pathway activation remain to be identified in the majority of SHH cases.

3.5.2.3 Correlates of the WNT / SHH independent subgroup

Chromosome 17 defects were the only events significantly correlated with the most common WNT / SHH-independent subgroup, suggesting a role for chromosome 17 genes in these cases. This disease subgroup however remains the least well

characterised at the molecular level. Sub-division of this group has been proposed on the basis of its transcriptomic and genomic patterns however, unlike the SHH and WNT groups, inconsistent results have been reported from different studies (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006), and the identification of specific genes and pathways associated with its pathogenesis will be critical to future advances in our understanding of its molecular basis and any underlying heterogeneity.

The significant associations observed between medulloblastoma molecular subgroups and specific gene, pathway and chromosomal defects (i) strongly support the existence of molecularly distinct SHH and WNT subgroups, (ii) inform the contributory mechanisms involved in their pathogenesis, and (iii) provide potential biomarkers for subgroup identification. When assessed for suitability as primary biomarkers, only *CTNNB1* mutations, which were specifically observed in all but one WNT subgroup case, have sufficient sensitivity and specificity to have utility. Nuclear localisation of β-catenin has also been widely reported as a positive marker of WNT pathway activation(Clifford et al., 2006; Ellison et al., 2005; Eberhart et al., 2000), although its relationship to our WNT expression signature and *CTNNB1* mutations could not be investigated in the present study due to tissue limitations. Likewise, *COL1A2* status may have utility in the identification of SHH subgroup infant desmoplastic medulloblastomas (this study and (Anderton et al., 2008)), particularly in cases where biopsy limitations do not allow assessment of the DN pathological variant. The remainder of gene and chromosomal defects investigated were not suitable as primary biomarkers for positive subgroup discrimination, as a result of either their (i) non-exclusivity, (ii) limitation to subsets of subgroup cases, or (iii) inverse correlation with pathway activation. In comparison, gene expression signatures positively identified all subgroup cases and provide an accurate diagnostic test for subgroup membership. The genomic markers examined may therefore provide useful secondary or confirmatory markers, when used in conjunction with these signatures.

Consensus for the precise number of transcriptomic disease subgroups has yet to be reached (section 1.3.12.1), with estimates of 4, 5 and 6 disease subgroups reported (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006). The

GeXP assay reported in this chapter could be extended to other WNT / SHH independent subgroups once they are better defined.

3.5.3 Clinical relevance of reported expression signatures

The combination of molecular and clinico-pathological data from four independent cohorts for meta-analysis, totalling 173 cases, has facilitated clear and significant insights to the clinical features of the medulloblastoma molecular subgroups, which have either not been apparent or not shown statistical significance in individual analyses of the smaller component cohorts reported to date (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006). The SHH (24% of cases), WNT (12%) and WNT / SHH-independent (64%) groups show different age distributions and relationships to disease histopathology. SHH subgroup tumours peak in infancy and are intimately correlated with DN pathology in this group, to the extent that DN pathology may be considered as a surrogate marker for SHH activation in medulloblastomas arising in infants <3 years old at diagnosis, although classic and LCA cases also constitute a minority of SHH subgroup cases in this age group. This relationship breaks down in non-infants (≥3 years at diagnosis), where SHH tumours are less common, and show equivalent proportions of DN, classic and LCA disease; SHH-independent DN cases are also commonly observed in this age group. These data strongly indicate that (i) SHH subgroup and (ii) DN tumours, arising in the infant and non-infant age groups, have different biological and clinical characteristics, and that SHH-positive DN tumours of infancy represent a unique disease subgroup associated with a favourable clinical behaviour (Garre et al., 2009; McManamy et al., 2007; Rutkowski et al., 2005), and a characteristic molecular pathogenesis (*COL1A2* unmethylated (Anderton et al., 2008)) and mutational spectrum (*SUFU* (Brugieres et al.; Taylor et al., 2002)). Conversely, WNT tumours display classic pathology and occur in non-infants. Notably, both the SHH and WNT subgroups show at least two different incidence peaks in their age distribution (both have second peaks in adults), suggesting additional clinical and molecular heterogeneity within these groups.

The lack of association between M stage and molecular subgroups (‘*p’* = 0.20) is in disagreement with the previous study by Kool *et al*., (2008), who reported metastatic tumours being less common in WNT and SHH pathway activated medulloblastomas.

This could be due to the different measurement criteria for metastasis between the studies (the present study compared M0 / 1 versus M2 / 3, while Kool *et al.* compared M0 versus M1 / M2 / M3). Alternatively, the increased numbers in this study (130 *versus* 58 with M stage data in the Kool *et al.* study) may have enabled a more accurate characterisation of the relationship between signalling pathway activation and metastasis, and future large clinically controlled studies should be informative in this regard.

3.5.4 Summary

The identification of medulloblastoma molecular subgroups has significant prognostic and predictive potential to improve therapeutic delivery and disease outlook in the clinical setting, and could represent a first step in the molecular diagnostic triage of medulloblastomas, to guide treatment decisions. In addition to distinct clinical features, molecular subgroups also appear to have characteristic clinical behaviours; the favourable prognosis of WNT subtype medulloblastomas is now established in multiple clinical cohorts (Korshunov et al., 2010; Fattet et al., 2009; Gajjar et al., 2006; Ellison et al., 2005), and will form the basis of treatment reductions in the forthcoming international molecularly-driven PNET 5 / 6 clinical trials (Pizer and Clifford, 2009). Combined data from this and other studies indicate SHH-positive DN tumours arising in infants represent a similarly favourable prognosis subgroup with a distinct molecular basis (Garre et al., 2009; Anderton et al., 2008; McManamy et al., 2007; Rutkowski et al., 2005). Recently, it has been recognised that paediatric and adult SHH medulloblastomas are molecularly and clinically distinct (Northcott et al., 2011), with the survival advantage of desmoplasia in infants not prognostic in adult SHH cases, raising the concept of disease prognostication within subgroups. Molecularly targeted SHH inhibitors are also currently under pre-clinical and clinical development, and have shown early evidence of activity in medulloblastoma (Rudin et al., 2009; Romer et al., 2004). Assessment of the prognostic impact of the remaining molecular subgroups will now be essential to determine their utility to direct the selection of adjuvant therapy.

The ability to accurately diagnose the SHH molecular subgroup will thus be important for the targeted delivery of these novel agents, and our findings have identified SHH-

positive subgroups of medulloblastomas which would be predicted to benefit most from SHH inhibition strategies. However, the SHH pathway plays a key role in normal, including cerebellar, development, and its transient inhibition in young mice causes permanent defects in growth plate formation and bone structure (Kimura et al., 2008). In view of such potential toxicities, a cautious approach to their application should be undertaken, particularly in the infant age group where SHH subgroup tumours predominate.

This chapter reported an assay that can robustly identify WNT, SHH and WNT / SHH independent medulloblastoma subgroups and investigated their clinical correlations in a wider meta-analysis, although the precise number of subgroups remains unclear (section 1.3.12.1). Based on these observations, the following chapter reports an investigation into whether epigenetic patterns of DNA methylation have additional utility for disease subclassification.

Chapter 4. DNA methylomics identifies clinically significant subgroups of medulloblastoma

4.1 Introduction

The study of epigenetic changes in medulloblastoma offers potential for the identification of novel biomarkers and disease targets. Until recently, the study of DNA methylation in medulloblastoma was limited to single candidate gene approaches. These have identified CpG island hypermethylation of *RASSF1A, CASP8* and *HIC1*, associated with transcriptional silencing in significant proportions of primary tumours (Lindsey et al., 2005). More recently, genome-wide experiments comparing expression in MB cell lines before and after treatment with the demethylating agent 5'-Aza-2'-deoxycytidine (Anderton et al., 2008; Kongkham et al., 2008) or the histone deacetylase inhibitor trichostatin A (Vibhakar et al., 2007) have identified methylation dependent silencing of genes (*COL1A2, SPINT2, DKK1*, respectively) that is recapitulated in primary tumour tissue. A novel approach from Diede et al. used a genome-wide screen that exploited the increased melting temperature of methylated DNA to identify that the epigenetic silencing of the *PTCH1-C* promoter may be a contributing factor towards tumourigenesis (Diede et al., 2010).

Previous work has demonstrated that *COL1A2* CpG island hypermethylation identifies infant DN cases, with an associated favourable prognosis, demonstrating how methylation of specific CpG islands can be useful for disease subgrouping, as well as predicting disease behaviour and patient outcome (Anderton et al., 2008). Until now, the utility of DNA methylation to subclassify medulloblastoma had not previously been investigated.

This part of the study made use of a newly-available technique for the characterisation of the DNA methylome. The Golden Gate Cancer Panel I methylation microarray (Bibikova et al., 2006) (see section 2.7) measures the methylation status of 1505 probes corresponding to 807 genes. Genome-wide transcriptomic approaches have previously been applied to identify subgroups of medulloblastoma (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006), identifying between 4 and 6 subgroups (see section 1.3.12.1). The application of a novel method for the characterisation of the medulloblastoma DNA methylome, independent of transcriptomic data, represents an opportunity to assess, firstly, whether methylomic data is amenable for the classification of medulloblastoma, and, if so, whether there is

any relationship to the previously posited transcriptomic disease subgroups. This approach could also identify specific molecular events in medulloblastoma development.

The merits of subgrouping the disease are clear; to date, it has enabled the characterisation of molecular processes driving the disease, which, in turn, has led to the introduction of drug treatments for SHH cases (Rudin et al., 2009) (section 1.3.11), and the recognition that medulloblastomas with an activated WNT pathway are associated with a good prognosis (Fattet et al., 2009; Clifford et al., 2006; Ellison et al., 2005). However, the underlying mechanisms and prognostic potential for any remaining, pathway independent subgroups remain poorly understood.

Once consensus has been reached for the number of disease subgroups, efforts can begin to identify the unique features of disease subgroups, initially for the identification of subgroup specific biomarkers to assign subgroup and for the identification of correlates of clinical behaviours. Following functional work, the identification of putative novel drug targets may be possible. The prospect of prognostication within disease subgroups also becomes possible; by identifying intra-subgroup variant disease features, it may be possible to identify prognostic markers that may also have functional relevance.

This chapter reports an investigation into the characterisation of methylomic subgroups of the disease, using a training tumour cohort for class discovery, and a test cohort for class validation. The robustness of the assigned classes is characterised. Additionally, the relationship between subgroup membership and clinico-pathological and molecular correlates is assessed, along with WNT and SHH signalling pathway activation, using the GeXP assay described in chapter 3.

Previous attempts to subclassify the disease have all been based on single cohorts (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006). While attempts have been made to validate findings in other publically available transcriptomic datasets, these findings have not been robustly validated (Cho et al., 2011; Northcott et al., 2010) (see section 1.3.12.1). The methylomic subclassification of the disease reported here is the first time that a combined test and training cohort have been applied in the same study for validation of identified subgroups using

identical methods. Moreover, this is the first high-throughput study to test a clinical trials-based cohort for the prognostic potential of any identified subgroups. This, thus far, unique opportunity to identify subgroups, validate them in a test cohort and test for survival differences in a trials cohort form the basis of investigations for both chapter 4 and 5.

4.2 Aims

Genome-wide assessment of the medulloblastoma DNA methylome had not previously been investigated. Using a novel DNA methylation array approach, this part of the study aimed to investigate:

- Whether medulloblastoma samples display differential patterns of DNA methylation and any ability of such patterns for disease subclassification and if subgroups can be identified, to investigate:
    - Whether subgroups can be validated in an independent medulloblastoma cohort.
    - Whether there is a relationship between subgroups and previously posited transcriptomic disease subgroups
    - The identification of subgroup specific methylation biomarkers
    - The clinico-pathological and molecular correlates of any identified methylomic subgroups
    - Whether there are survival differences between subgroups

4.3 Materials and Methods

4.3.1 Cohort description

An initial training cohort of 108 medulloblastoma samples, comprising 101 DNA samples extracted from frozen tumour tissue and 7 from FFPE tissue, was selected for array analysis (section 2.1). This cohort was chosen to include all known medulloblastoma histopathological subtypes, as well as samples with known specific medulloblastoma-associated mutations (*CTNNB1* (β-catenin) mutation (Ellison et al., 2005) or *PTCH1* mutation (Zurawel et al., 2000)). Histopathological subtype was confirmed on review using WHO criteria (Louis et al., 2007) by Professor David Ellison. Newcastle and North Tyneside Research Ethics Committee approval was obtained for the collection, storage, and biological study of all material.

The test cohort comprised 143 primary medulloblastoma samples from patients enrolled in the PNET3 clinical trial (Taylor et al., 2003), for which full survival and clinical information (age, histopathological subtype, gender, M stage, *MYCC / MYCN* amplification status (see section 2.1)) was available . This trial recruited childhood medulloblastomas (for this trial, defined as aged 3-16 years old at diagnosis) from across Europe (see section 2.1). DNA from this cohort was exclusively derived from FFPE materials.

4.3.2 Nucleic acid extraction

DNA was extracted from FFPE samples using a Qiagen DNeasy kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions. DNA and RNA were extracted from frozen tumour samples using Trizol (Invitrogen, Carlsbad, CA, USA) according to manufacturer's protocols, as detailed in section 2.2.1. Aliquots of 1 µg of DNA at 100 ng / µl, measured by Nanodrop (section 2.3.1), were sent to the array facility for processing.

4.3.3 Methylation Microarray assay

Microarray methylation analysis was performed on the Illumina GoldenGate Cancer Panel I methylation array at the Wellcome Trust Centre for Human Genetics, Oxford, UK according to manufacturer's protocols (Illumina, San Diego, CA, USA). Arrays were

imaged using a BeadArray Reader scanner. Raw experimental data were imported into BeadStudio v3.2 (Illumina, San Diego, CA), an analysis suite for bead array format microarrays (section 2.8).

4.3.4 Quality control, array normalisation and validation

Quality control (QC) was performed using the inbuilt QC metrics within methylation module v3.2 of BeadStudio, as well as the extra QC measures contained within the R (R Development Core Team, 2008) package, *beadarray* 1.6.0 (Dunning et al., 2007), to assess spatial distribution of anomalous signal intensities and to identify poorly performing samples by scatter plots of intensities (section 2.8.1). Sample replicates were run to assess intra- (n = 4 samples run in duplicate on the same array) and inter- (n = 4 sample duplicates run on different arrays) array reproducibility was assessed by running sample replicates. Data were initially normalised in BeadStudio as previously described (Ladd-Acosta C., 2007) and subjected to the quality control measures contained within that program. Subsequently, β values were re-derived using the BASH algorithm (section 2.8.1) to mask out regions with anomalous intensities. Subsequently, the β value was calculated using the equation shown in section 2.8.1. The software reports a β value for each locus. The β value can range from 0 (fully unmethylated) to 1 (fully methylated). Quality control plots were examined to identify poorly performing samples (section 2.8.1).

4.3.4.1 Bisulfite sequencing to confirm observed methylation patterns

It has been reported (Ladd-Acosta C., 2007; Bibikova et al., 2006) that the GoldenGate methylation array provides high quality reproducible data that correlates well with validation data derived from bisulfite sequencing. In order to confirm this correlation, a panel of 7 discriminatory probes (ASCL2_P360_F, HFE_E273_R, NOS2A_E117_R, COL1A2_E299_F, CCKAR_P270_F, SPDEF_P6_R and MSH2_P1008_F), corresponding to the genes *ASCL2, HFE, NOS2A, COL1A2, CCKAR, SPDEF* and *MSH2* were selected for bisulfite sequencing in 18 medulloblastoma samples selected from the methylation array cohort (section 2.6). PCR primers flanking the probes of interest were designed using MethPrimer (Li and Dahiya, 2002) and used to amplify the region of interest from bisulfite treated DNA. Sequencing reactions were performed and read using a

Beckman-Coulter CEQ 8800 (Beckman, Fullerton CA, USA). The methylation status at specific nucleotides was estimated by dividing the relative peak height of the G residue (representing a methylated cytosine) by the sum of the peak heights of the G and T residues (representing an unmethylated cytosine converted to a uracil residue by the bisulfite treatment). The design and execution of the validation work was undertaken in conjunction with Dr. Janet Lindsey. The primer sequences for this analysis are listed in Table 4.1. The differences between array and bisulfite sequencing derived estimates was visualised by plotting Bland-Altman plots (Bland and Altman, 1986), and density plots.

Gene	Probe	F primer (5′ – 3′)	R primer (5′ – 3′)
ASCL2	ASCL2_P360_F	GGGAATTTGA ATTTTTTATTT	AAACTAAATTCC TACTAAACCCC
CCKAR	CCKAR_P270_F	ATTGTTTTTTTATAA GGAGGTAGAATATA	CTAAATACAAACA ACCTAACTACCC
COL1A2	COL1A2_E299_F	AGGTATTTTAGGG TTAGGGAAATTTT	ATTACTACAAACA ACAACAAAATCC
HFE	HFE_E273_R	GGTAATAGTTGTA GGGTGATTTTTG	CAAATCCTCCAA AATTAACAAACTC
MSH2	MSH2_P1008_F	GGTAGAAGATTT TTTGGGTTTAAA	CACCATCCTAAAC AACATAATAAAAC
NOS2A	NOS2A_E117_R	AAAAATAATTTTT TGGATGGTATGG	TTACAACTAACTA CACTACCTCCCC
SPDEF	SPDEF_P6_R	TTGTTTGTGGTTT GAGGTAAGTAAG	CCCTCAAAAAAT AACCCTCTAAAAT

Table 4.1. Primer sequences used for bisulfite validation of methylation levels observed with GoldenGate methylation array. Sequences are listed 5′ – 3′.

4.3.5 Unsupervised cluster analysis

An unsupervised cluster analysis was performed on the methylation array data derived from the training cohort, as described in section 2.9.4. Briefly, an optimal combination of metagenes and clusters was identified by an iterative, NMF- and *k* means-based, consensus clustering approach, testing every combination of 2:6 metagenes and 2:6 clusters. After selecting the optimum metagene and cluster numbers, samples that were not consistently classified to the same cluster (i.e. assigned to the same cluster in less than 80% of repeat analyses) were assigned to the NC class. Using the same consensus clustering approach with the optimal combination of metagenes and clusters, the same cutoffs were applied to the test cohort to identify additional NC samples.

H matrices were extracted from the training dataset and used to project an H matrix for the test cohort (section 2.9.3). The training cohort H matrix was used to construct a classifier based on the class assignments from consensus clustering.

4.3.6 Construction of a classifier

Sample members from the test dataset were assigned to clusters by constructing a SVM classifier from the H matrix derived from the training dataset and applying it to the test dataset (section 2.12).

4.3.6.1 Support vector machine for classification of medulloblastoma subgroup

Using the group assignments derived from *k* means clustering of the training dataset, a classifier was constructed using the H matrix derived from NMF of the training dataset. SVM (Chang, 2001) was used to classify the data. A linear-kernel multi-class (MC)-SVM was tuned using LOOCV (section 2.12.3). Using the optimal cost parameter of 2, the MC-SVM was constructed. The SVM was used to predict class membership in the training cohort.

4.3.6.2 Validation of SVM classifier on an independent methylation array medulloblastoma data set

The projected H values for the test dataset (section 2.9.3) were used as the basis for prediction of class membership using the SVM classifier. The methylation probes that correlated or anti-correlated most closely with the training set H values were selected as being representative markers for each class and were visually compared for patterns of similarity in the test dataset.

A dataset containing the joined H values and class designations from both the test and training cohort was constructed and used as the basis for subsequent analyses. To assess cluster assignment, silhouette plots (section 2.10.1) (Rousseeuw, 1987) were used to assess class membership and identify samples that were incorrectly classified.

4.3.7 Principal Component Analysis to visually identify subgroup differences

Principal component analysis (sections 2.9.2) was applied to the datasets to visually compare group assignments. Principal components were generated from the H values of the training dataset and used to project test cohort principal components from the test cohort H values. The projected principal components for the test dataset were appended to the training dataset principal components and plotted as a biplot.

4.3.8 GeXP assay to identify WNT and SHH pathway activation

In order to examine the correlation between signalling pathway activation and the methylomic subgroups identified, the GeXP assay (Schwalbe et al., 2011) described in section 3.3.6 was used to assign expression status (WNT, SHH or WNT / SHH independent) in 88 / 100 cases in the training cohort for which mRNA was available. The experimental procedure is detailed in section 3.3.7. Briefly, experiments were run in triplicate on 25 ng mRNA per reaction to measure the expression of eight SHH signature genes and five WNT signature genes. Expression values were normalised against expression of *28S rRNA*. Class membership was assigned as described in section 3.4.4, using stacked barplots and bi-plots. The test cohort was derived from FFPE materials, so was unsuitable for this analysis. However, antibody assignment of WNT

and SHH status had previously been reported for the PNET3 cohort (Ellison et al., 2011a).

4.3.9 Antibody assignment of WNT and SHH status in FFPE samples

The antibody assignment of WNT and SHH status in FFPE was carried out as described (Ellison et al., 2011a) in section 2.1. Data was available for 5 / 7 members of the training cohort that were derived from DNA extracted from FFPE tissue and from 115 / 143 (80%) members of the test cohort, which was derived exclusively from DNA extracted from FFPE tissue. This data was integrated into downstream analyses of clinico-pathological correlates of the identified subgroups.

4.3.10 Global differences in patterns of methylation across subgroups

The identified subgroups were tested to identify large-scale changes in the proportion of methylated and unmethylated probes. For the purposes of this analysis, unmethylated probes were defined as having a β score < 0.333, and methylated probes as having a β score > 0.667. Analyses were conducted on the whole dataset after removal of QC failures (n = 230) (having excluded X chromosome probes), and on subsets of the dataset, comprising probes situated within and outside of CpG islands. ANOVA tests were used to identify significant differences. Where a significant difference was reported, Tukey's Honest Significance Difference test was applied to identify the subgroup pairs for which a significant difference was detected. This test compares all possible pairs of means, and is conceptually similar to a t test, except that there is a correction for multiple testing that reduces the risk of type I errors (false positives). Boxplots were constructed to visualise these comparisons.

4.3.11 Analysis and integration of clinico-pathological and molecular correlates with methylomic data

The training and test cohorts were joined for subsequent investigations of the relationship of subgroup membership to clinico-pathological and molecular correlates and survival.

Clinico-pathological correlate data was obtained from a variety of sources (section 2.1). Clinico-pathological data was available for gender, M stage, histological subtype, age at diagnosis, and survival. Molecular data for chromosome 17 LOH, chromosome 6 LOH, *CTNNB1* mutation and the amplification of the *MYCC* and *MYCN* oncogenes was also available.

Chi-squared tests of association were used to assess the relationships between subgroup membership and clinico-pathological and molecular correlates. To identify the specific subgroups driving significant differences in relationship to the correlate, chi-squared residuals (subgroups driving the significant chi-squared test will have a larger chi squared residual, which is a measure of the difference between the observed and expected values) were plotted on a heatmap.

4.3.12 Relationship of methylomic subgroups to survival

In order to determine the relationship of subgroups to survival, it was first necessary to verify that previously identified risk markers were identifiable, to test whether the survival cohort was representative of previously described medulloblastoma cohorts. Kaplan-Meier plots were constructed for LCA, M stage (M- *vs* M+), chromosome 17 LOH, age (under 3 *vs* others) and *MYCC / MYCN* amplification. Next, a plot of methylation subgroup membership was constructed for the whole cohort and for a subset of the cohort containing cases aged 3-16, matching the age profile for recruitment into the PNET3 trial.

4.4 Results

4.4.1 A genome-wide screen of methylation patterns in medulloblastoma

This chapter aimed to test the hypothesis that differential patterns of methylation exist within medulloblastomas and to test the ability of any such patterns to classify the disease. A total of 251 arrays were sent off for analysis on the Golden Gate Cancer Panel I methylation microarray.

4.4.2 Quality control and validation

For the training and test datasets, intra- and inter-array reproducibility was assessed by running sample replicates. Linear regression of training set replicates showed good correlation. Comparisons were made both between samples run twice on separate arrays (inter-array) and samples run twice on the same array (intra-array). Inter-array (n = 4) Pearson's r^2 values ranged from 0.97 to 0.99. Intra-array (n = 4) r^2 values ranged from 0.96 to 0.98. Similarly, inter-array replicates (n = 2) were run for the test cohort, with both r^2 values of 0.94.This was judged to be satisfactory, both for frozen DNA (Figure 4.1) and FFPE DNA (Figure 4.2).

Sample QC, as described in section 2.8.1, was implemented to identify poorly performing samples that were subsequently removed from further analyses. QC failures, classified by failure in greater than one test, are summarised in Table 4.2. The failure rate was 7.4% (8 / 108 samples) for the training cohort, which was predominantly comprised of frozen DNA. This was close to the failure rate of the test cohort (9.1% (13 / 143 samples), which was comprised exclusively of FFPE DNA.

After removal of 8 QC failure samples from the training cohort and 13 QC failure samples from the test (PNET3) cohort, final cohort sizes were 100 cases for the training cohort and 130 cases for the test cohort. The demographics of the final cohorts are shown in Table 4.3. Exploratory PCA analysis demonstrated no clustering by sample type (DNA extracted from fresh frozen / FFPE tissue) or from contributing centre (data not shown).

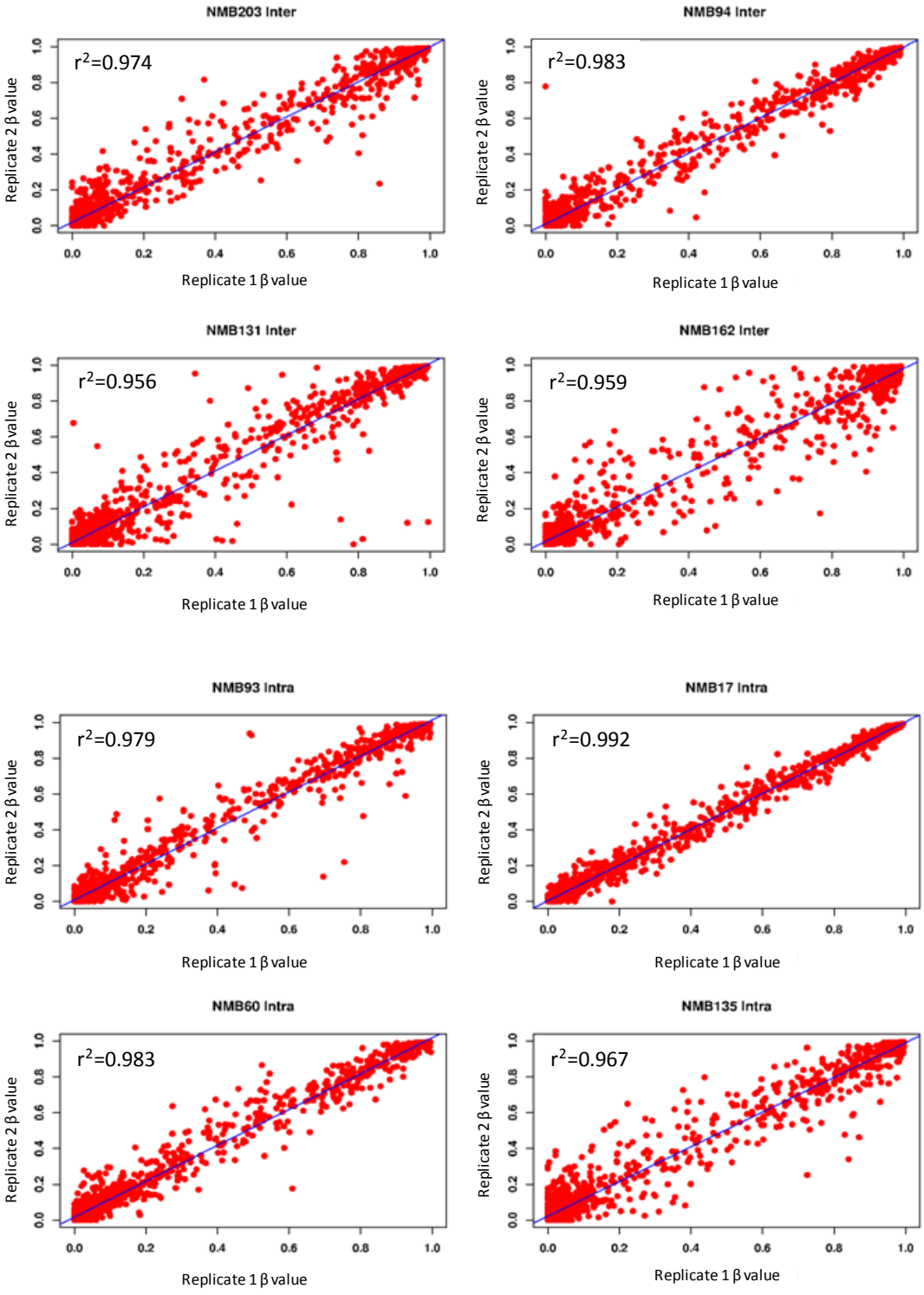


Figure 4.1. Inter- and intra-array replicates demonstrate reproducibility of Golden Gate methylation array. Scatter plots show matching β scores across two replicates. For each plot, sample ID and nature of comparison (intra-, inter-array) is shown. Pearson's r^2 score is shown.

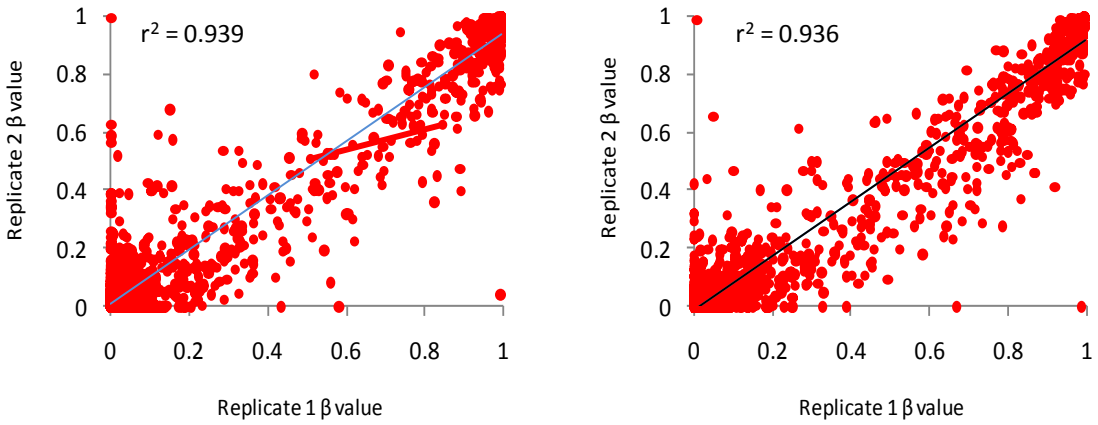


Figure 4.2. Inter-array replicates on FFPE array demonstrate good reproducibility for FFPE-derived samples. Scatter plots show matching β scores across two replicates. For each plot, sample ID and nature of comparison (intra-, inter-array) is shown. Pearson's r^2 score is shown.

	Allele Specific Extension	Bisulfite Conversion	Contamination	Extension Gap	Gender Control	First Hybridisation	Density Plot	Remove
Training Cohort Samples	NMB108	NMB108		NMB108	NMB108	NMB108	NMB108	Y
	NMB113	NMB113		NMB113	NMB113	NMB113	NMB113	Y
	NMB153							N
	NMB251	NMB251		NMB251	NMB251	NMB251	NMB251	Y
	NMB69B	NMB69B		NMB69B	NMB69B	NMB69B	NMB69B	Y
	PNET3209	PNET3209		PNET3209		PNET3209	PNET3209	Y
	RJG114					RJG114	RJG114	Y
	RJG115	RJG115		RJG115	RJG115	RJG115	RJG115	Y
	RJG118	RJG118		RJG118	RJG118	RJG118	RJG118	Y
						NMB70		N
						NMB127		N
Test cohort (PNET3) FFPE Samples	31							N
	32						32	Y
	41						41	Y
	43	43			43	43	43	Y
	178	178	178		178	178	178	Y
		205					205	Y
						209	209	Y
	50001	50001	50001		50001	50001	50001	Y
	50060	50060	50060			50060	50060	Y
		50086			50086	50086	50086	Y
	50128					50128	50128	Y
		50147						N
	50197	50197				50197	50197	Y
	50204					50204	50204	Y
	50250						50250	Y

Table 4.2. List of quality control measures and samples that failed these measures. Samples removed from further analyses are marked. Allele specific extension, bisulfite conversion, contamination, extension gap, gender control and first hybridisation are Bead Studio quality controls. Density plot was calculated using beadarray (Dunning et al., 2007). Samples failing more than one test were deemed QC failures and removed from subsequent analyses.

Demographic	Cohort	
	Training	**Test (PNET3)**
Number (frozen / FFPE)	100 (94 / 6) (94% / 6%)	130 (0 / 130) (0% / 100%)
Gender		
Male	62 (62%)	78 (60%)
Female (M:F ratio)	38 (38%) (1.6:1)	52 (40%) (1.5:1)
Age in years: median (range)	7.4 (0.1 – 43.0)	8.4 (3.1 – 15.6)
Age: Infant cases		
≤ 3 years of age	15 (15%)	0 (0%)
> 3 years of age	85 (85%)	130 (100%)
Histological subtype		
Classic	72 (72%)	110 (85%)
Desmoplastic / nodular	18 (18%)	9 (7%)
Large cell / anaplastic	10 (10%)	11 (8%)
M Stage		
M-	72 (72%)	105 (81%)
M+	17 (17%)	25 (19%)
NA	11 (11%)	0 (0%)
***CTNNB1* mutation**		
0 – no - negative	87 (87%)	118 (91%)
1 – yes - positive	10 (10%)	9 (7%)
NA	3 (3%)	3 (2%)
Chromosome 6 LOH		
0 – no - negative	89 (89%)	121 (93%)
1 – yes - positive	11 (11%)	8 (6%)
NA	0 (0%)	1 (1%)
Chromosome 17 LOH		
0 – no - negative	24 (24%)	102 (78%)
1 – yes - positive	9 (9%)	27 (21%)
NA	67 (67%)	1 (1%)
***MYCC* amplification**		
0 – no - negative	74 (74%)	128 (98%)
1 – yes - positive	3 (3%)	2 (2%)
NA	23 (23%)	0 (0%)
***MYCN* amplification**		
0 – no - negative	73 (73%)	125 (96%)
1 – yes - positive	4 (4%)	5 (4%)
NA	23 (23%)	0 (0%)
Signalling pathway mRNA signature		
WNT	6 (6%)	0 (0%)
SHH	19 (19%)	0 (0%)
WNT / SHH independent	63 (63%)	0 (0%)
NA	12 (12%)	130 (100%)
Antibody status		
WNT	5 (5%)	16 (12%)
SHH	0 (0%)	23 (18%)
WNT / SHH independent	0 (0%)	76 (58%)
NA	95 (95%)	15 (12%)

Table 4.3. Demographics, clinico-pathological and molecular correlates of 100 member medulloblastoma training cohort and the 130 member test medulloblastoma cohort. For each cohort, the number and percentage in parentheses for each measured correlate (excluding age range) are shown. NA indicates missing data. For the training cohort, the single MBEN case was linked to the DN cases. Data for gender, histological subtype, M stage, *CTNNB1* mutation, chromosome 6 LOH, chromosome 17 LOH, *MYCC* amplification, *MYCN* amplification, signalling pathway expression signature and signalling pathway antibody staining is shown.

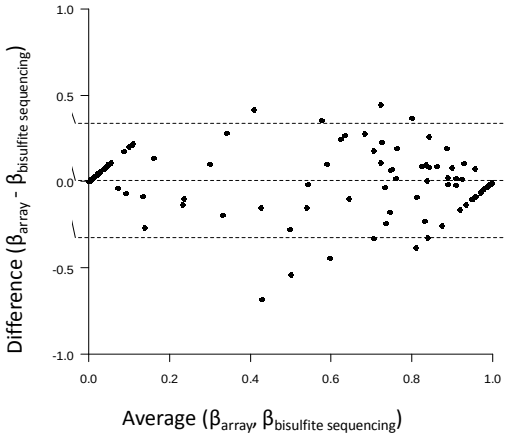
4.4.2.1 Verification of methylation values reported by bisulfite sequencing

Although previous work had reported that the methylation scores reported by the Golden Gate Cancer Panel I array accurately reflect the true methylation score (Bibikova et al., 2006), it was necessary to validate this for the datasets used in these experiments. Work carried out in conjunction with Dr. Janet Lindsey verified a panel of 7 probes that showed differential methylation across 18 medulloblastomas.

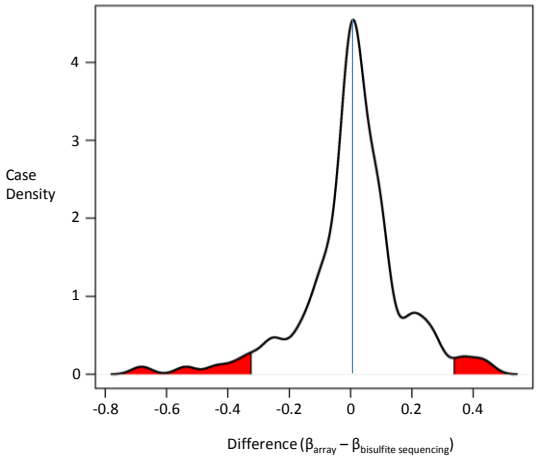
A Bland-Altman plot (Bland and Altman, 1986) of β values estimated by array and bisulfite sequencing is shown in Figure 4.3. It shows how at extremes of methylation (β close to 0 or 1), there is strong agreement between estimates. This is weakened at intermediate methylation scores (β close to 0.5), where the difference between estimates is increased. The density plot in panel B of Figure 4.3 shows the average deviation between the estimates of β made for individual data points using the two alternative methods. The mode of the distribution is 0.02, with a mean β score difference of 0.006. Coincidentally, the standard deviation of this distribution is 0.167, close to the reported level of sensitivity (0.17) for the Golden Gate assay (Bibikova et al., 2006). No one gene or sample was responsible for the few large-scale disagreements in β value observed.

Finally, the measurement of β values by bisulfite sequencing is by its nature imprecise, since it relies on the measurement of peak heights generated from direct DNA sequence analysis traces. A grouped comparison was made comparing methylation values when classified into three groups (unmethylated – $\beta < 0.333$, hemimethylated - $0.333 < \beta < 0.666$, methylated – $\beta > 0.667$). Under this classification scheme, there was agreement for 104 / 126 measurements (83%). For cases with disagreement, this was generally from one category to its adjacent category, rather than, for example, methylated in array and unmethylated in bisulfite estimates of β, which occurred in only 2 / 126 (1.5%) measurements.

A



B



C

Array	Bisulfite Sequencing: Unmethylated	Bisulfite Sequencing: Hemi-methylated	Bisulfite Sequencing: Methylated
Unmethylated	53	1	2
Hemi-methylated	3	5	6
Methylated	0	10	46

Figure 4.3. Validation of array β values by bisulfite sequencing.

A. Bland-Altman plot showing direct comparison between bisulfite sequencing estimation of methylation and Golden Gate array-estimated methylation. Data from 18 samples at 7 loci (*ASCL2, CCKAR, COL1A2, HFE, MSH2, NOS2A, SPDEF*) are shown. The x axis shows the average score from the two estimations of β value and the y axis shows the difference between bisulfite sequencing and array estimates of methylation. Horizontal dotted lines are plotted at 2 standard deviations of the difference and at the mean difference.

B. Density plot showing distribution of deviation between bisulfite sequencing and array estimates of methylation. The case density (y axis) is a measure of frequency. A blue line indicates the modal value for deviation between estimates. Measurements more than two standard deviations from the mean deviation are shaded in red.

C. Combined table / heatplot showing correlation between bisulfite and array estimates after transforming the data (Unmethylated – β score < 0.333; hemi-methylated - 0.334 < β score < 0.666; methylated - 0.667 < β score). The number in each cell is proportional to the colour it is filled with (the background of larger numbers will tend towards black).

4.4.3 Global patterns of DNA methylation

For the training set, 1058 / 1505 probes showed evidence of methylation (β score > 0.333) in one or more samples, of which 494 showed substantial variation between samples (defined as having standard deviation greater than 0.17). For the test set, 1166 / 1505 probes showed evidence of methylation (β score > 0.333) in one or more samples. 554 probes showed substantial variation between samples.

The distribution of average methylation status of each probe within and between training / test cohorts is summarised in Table 4.4. In particular, there was a marked difference in the methylation status of probes located within and outside of CpG islands. In the joined training and test datasets, 869 / 1044 probes (83%) located within CpG islands were unmethylated (β score < 0.333). This contrasted with non-CpG island probes, where the majority of probes (288 / 461 (62%)) were methylated (β score > 0.666). This difference was highly significant ('p' < 2×10^{-16}). Chi-squared tests for probe distribution differences between the training and test datasets were not significant.

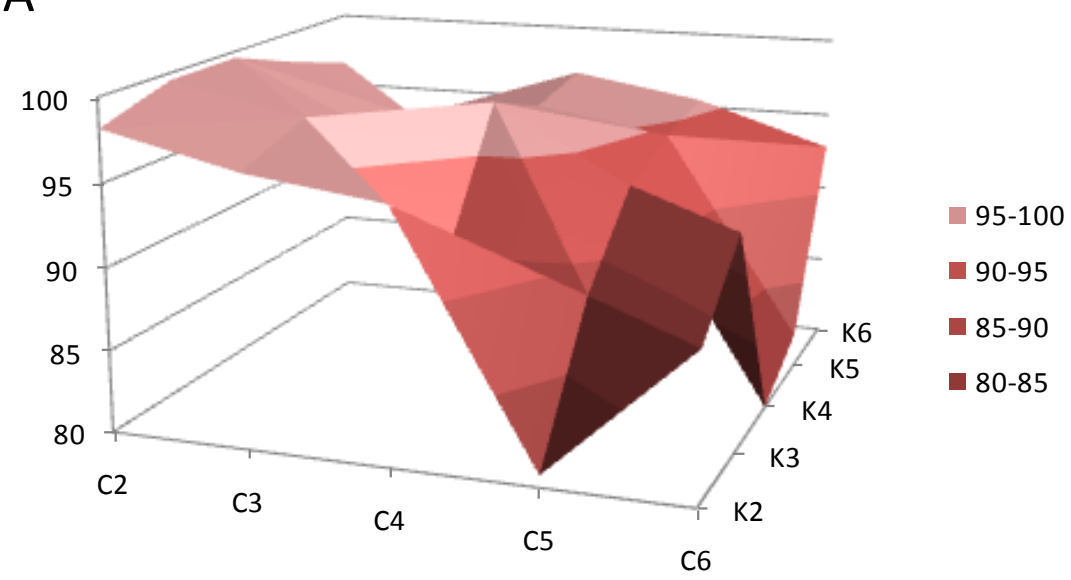
	Probe classification	Training Cohort (*n* = 100)	Test Cohort (*n* = 130)	Total Cohort (*n* = 230)	'P' value
All probes	Unmethylated	976 (65%)	947 (63%)	960 (64%)	0.129
	Hemimethylated	165 (11%)	148 (10%)	155 (10%)	
	Methylated	364 (24%)	410 (27%)	390 (26%)	
CpG Islands	Unmethylated	877 (84%)	861 (82%)	869 (83%)	0.211
	Hemimethylated	78 (7%)	71 (7%)	73 (7%)	
	Methylated	89 (9%)	112 (11%)	102 (10%)	
Non-CpG Islands	Unmethylated	99 (21%)	86 (19%)	91 (20%)	0.294
	Hemimethylated	87 (19%)	77 (17%)	82 (18%)	
	Methylated	275 (60%)	298 (65%)	288 (62%)	

Table 4.4 Global methylation distribution among training and test cohorts. Average methylation scores were calculated for all probes across the training and test cohort. For three probe categories (all probes, probes located within CpG islands and probes located outside of CpG islands), the number of unmethylated (average $\beta < 0.333$), methylated (average $\beta > 0.667$) and hemi-methylated probes ($0.333 <$ average $\beta \leq 0.666$) are shown for the training cohort, test cohort and total cohort. Chi-squared tests between the training and test cohort do not show any significant differences in the probe distributions; for the total cohort, there were significant differences between methylation classifications within and outside of CpG islands ('p' $< 2 \times 10^{-16}$, chi squared test).

4.4.4 Unsupervised cluster analysis identifies 4 distinct DNA methylation subgroups

An NMF / consensus clustering approach was initially applied to subsets of the training dataset (*n* = 100) to identify the optimal number of metagenes and clusters (see section 2.9.4). The average percentage of times within the 100 iterations that each sample was classified into the same cluster was calculated for each combination of metagenes and clusters. The overall average sample classification percentage was calculated by averaging across all sample averages, giving a measure of the reproducibility of each combination of metagenes and clusters (Figure 4.4). Although the highest reproducibility of sample clustering was observed for a two cluster, four metagene solution (99.59%) , this was discounted since it ran counter to what is known about the disease, where consensus exists for the existence of at least three subgroups (Cho et al., 2011; Schwalbe et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006). In addition, the cophenetic correlation coefficient (section 2.9.3.1) was also considered in choosing an optimal number of metagenes and clusters. An increase in cophenetic correlation coefficient from 0.997 to 0.998 was observed with a 4 metagene solution compared to a 3 metagene solution, indicating a better clustered 4 metagene solution, so for this reason, a 4 cluster 4 metagene solution (with an average modal sample assignment of 97.80%) was chosen in preference to a 3 cluster, 3 metagene solution (sample assignment stability 97.84%).

A



B

		Clusters				
		C2	C3	C4	C5	C6
Metagenes	K2	98.19	96.29	95.23	80.72	88.90
	K3	99.58	97.84	89.64	88.58	93.06
	K4	99.59	95.71	97.80	93.21	74.16
	K5	97.96	94.79	96.03	94.66	82.10
	K6	96.56	88.29	96.85	95.46	92.80

Figure 4.4. Determination of optimal combination of metagenes (K – 2 to 6 metagenes) and clusters (C – 2 to 6 clusters) within the training dataset. A. The average percentage assignment of samples to the same cluster over 100 iterations is shown as a 3D surface plot for each tested combination of metagenes and clusters. Data peaks are the optimal combinations of metagenes and clusters. B. The data shown in the surface plot is tabulated. The chosen optimal number of 4 metagenes, 4 clusters is highlighted yellow.

Having chosen a 4 metagene, 4 cluster approach, NC samples were identified. At a stability cutoff of 80%, three samples were not consistently classified: NMB63 (44%), NMB90 (64%) and NMB182 (63%) (percentages in parentheses indicate the classification percentage frequency of the modal class). These were assigned to the NC group at this time and removed from the training dataset. To pre-emptively identify NC samples within the test dataset, the same consensus clustering approach was undertaken with a 4 metagene, 4 cluster solution. At the same 80% cutoff, seven samples were identified as outliers: PNET30141 (52%), PNET30148 (63%), PNET30150 (72%), PNET350012 (78%), PNET350041 (55%), PNET350120 (60%) and PNET350291 (79%). These were assigned to the NC group of samples and removed from the test dataset.

NMF was iteratively applied to the refined, 97 member training cohort 200 times, and the best performing solution was selected for a 4 metagene factorisation. The H matrix was extracted and *k* means was used to assign samples to one of four subgroups (group 1 – 21 samples; group 2 – 23 samples; group 3 – 10 samples; group 4 – 43 samples).

Next, these assignments were used to train a SVM classifier on the training set H matrix. Initially, the classifier was tested against the training cohort. A confusion matrix (see section 2.12.4) between the cluster assignment and the SVM prediction of class assignment (Table 4.5) demonstrates that the classifier perfectly recapitulates the cluster assignments derived from *k* means analysis of the training set H matrix (section 2.9.5).

	Class	Cluster Assignment			
		1	2	3	4
SVM prediction	1	21	0	0	0
	2	0	23	0	0
	3	0	0	10	0
	4	0	0	0	43

Table 4.5. Confusion matrix shows that SVM classifier perfectly classifies the training dataset into the four classes assigned by the *k* means algorithm.

4.4.5 Clusters are reproducible and can be validated in a test dataset

After verifying the satisfactory performance of the SVM classifier on the training dataset, it was applied to the test dataset. H values for the test set were projected from the training set H values as described in section 2.9.3. The projected H values were applied to the SVM classifier. The class designations predicted from the SVM are compared against the class designations for the training set in Table 4.6.

Subgroup	Training *n*, (%)	Test *n*, (%)
1 (SHH)	21 (22%)	29 (24%)
2 (Grp II)	23 (24%)	25 (20%)
3 (WNT)	10 (10%)	18 (15%)
4 (Grp I)	43 (44%)	51 (41%)

Table 4.6. Comparison between sample numbers for each class for training and test datasets. Number is given with percentage in parentheses.

A silhouette plot (section 2.10.1) of the combined training and test cohort (*n* = 220) classified cases identified 4 samples (NMB128, NMB133, NMB250 and NMB137) as having a negative silhouette score, indicating an incorrect classification and these were re-assigned to the not-classifiable category. A silhouette plot (Figure 4.5) demonstrates that each of the remaining 216 classified samples have a positive silhouette score, indicating a correct clustering. A biplot (section 2.9.2.1) along principal components 2 and 3 of the joined dataset demonstrates clear separation between the classes and shows how subgroups are reproducible across the training and test datasets.

The synthesis of the classification of methylomic subgroups of medulloblastoma in training and test cohorts and the clinico-pathological and molecular correlates, metagene expression and correlative methylation probe data is shown in Figure 4.5. It illustrates the similar metagene profiles and methylation values for probes that correlate and anti-correlate with metagene expression between training and test

datasets. It is notable that there is evidence for intermediate expression of the group I associated metagene V3 in the majority of cases classified as group II, although the reciprocal relationship is not apparent for group I cases. Non-classified samples show a mixed metagene expression profile, with most expressing the V4 metagene (whose expression is associated with group II membership) in addition to one or more additional metagenes.

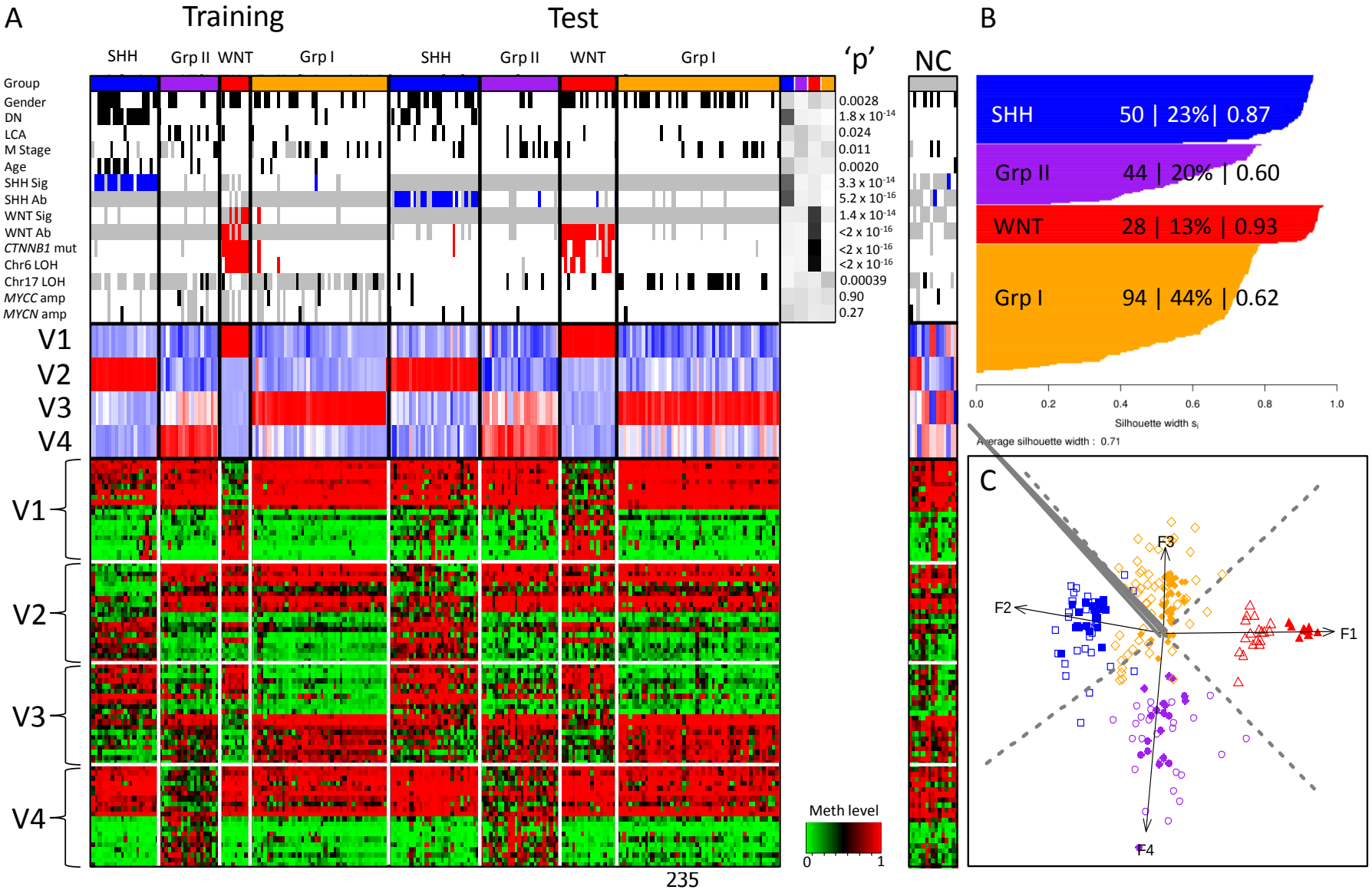

A
Training
Test
B
SHH
Grp II
WNT
Grp I
SHH
Grp II
WNT
Grp I
'p'
NC
Group
Gender
DN
LCA
M Stage
Age
SHH Sig
SHH Ab
WNT Sig
WNT Ab
CTNNB1 mut
Chr6 LOH
Chr17 LOH
MYCC amp
MYCN amp
0.0028
1.8 x 10-14
0.024
0.011
0.0020
3.3 x 10-14
5.2 x 10-16
1.4 x 10-14
<2 x 10-16
<2 x 10-16
<2 x 10-16
0.00039
0.90
0.27
V1
V2
V3
V4
V1
V2
V3
V4
Meth level
0
1
235
SHH 50 | 23% | 0.87
Grp II 44 | 20% | 0.60
WNT 28 | 13% | 0.93
Grp I 94 | 44% | 0.62
0.0
0.2
0.4
0.6
0.8
1.0
Silhouette width si
Average silhouette width : 0.71
C
F1
F2
F3
F4

Figure 4.5. DNA methylomics identifies 4 classes of medulloblastoma with significant relationships to clinico-pathological and molecular markers.

A: NMF-based consensus clustering of the training dataset identifies 4 subgroups that are validated in the test dataset. First panel shows subgroup membership, with clinico-pathological and molecular correlates of subgroups for both datasets. Subgroup membership is indicated by colour: SHH – blue; Group II – purple; WNT – red; Group I – orange. Female gender, desmoplastic /nodular histology (DN), large cell / anaplastic histology (LCA), M+ disease (M Stage) and age <3 years (Age) is shown in black. SHH gene signature positivity (SHH Sig) and SHH antibody positivity (SHH Ab) are labelled blue. Wnt expression signature positivity (WNT Sig), Wnt antibody positivity (WNT Ab), *CTNNB1* mutation and chromosome 6 loss of heterozygosity (Chr6 LOH) are labelled red. Chromosome 17 loss of heterozygosity (Chr17 LOH), *MYCC* amplification and *MYCN* amplification are labelled black. Missing data is labelled grey. Panel to the right of clinico-pathological and molecular correlates displays chi-squared test residuals that indicate any over-representation of each correlate across subgroups. Under-represented groups are white, whilst over-represented groups are black. 'P' values, derived from chi-squared tests of association between subgroup-assigned samples, are also shown. To the right of the training and test sets, the non-classified (NC) (n = 14) cases are also shown. Second row displays magnitudes of 4 metagenes (V1 to V4). Highly expressed metagenes are red, lowly expressed are blue. Third panel displays the top 10 most highly correlated and anti-correlated methylation probes for each metagene. Methylated probes are red, unmethylated probes are green, and hemi-methylated probes are black, as shown in the methylation key.

B.Silhouette plots (top right figure) of assigned subgroups demonstrates correctly clustered (silhouette score > 0) samples for 216 / 216 classified samples of the joined training and test cohort. A silhouette width (s_i) close to one indicates a very well clustered sample, whereas samples with close to zero width lie between two clusters; samples with negative silhouette widths are potentially placed into the wrong cluster. For each cluster, the number of cluster members, the percentage of cluster members and average silhouette width are given.

C: Bi-plot of training and test datasets (bottom right figure) demonstrates reproducibility of clusters across datasets. Arrows show projections of 4 metagenes along second and third principal components, labelled with their metagene number. For all clusters, training set samples are shown as filled shapes (WNT – red triangles, SHH – blue squares, Group I – orange diamonds, Group II – purple circles), with test samples as empty shapes.

4.4.5.1 Two training dataset methylation subgroups are associated with activation of developmental signalling pathways

The GeXP to assign WNT and SHH pathway activation status (Schwalbe et al., 2011) was applied to 88 / 100 training dataset cases with available mRNA. Sample pathway activation was assigned in the same way as described in section 3.3.9. WNT pathway activation was assigned to 6 (7%) cases, SHH pathway activation to 18 (20%) cases and WNT / SHH independent to 64 (73%) cases. A biplot showing the class assignments is shown in Figure 4.6. Striking associations between methylomic subgroup membership and signalling pathway activation were observed. Subgroup 1 was highly significantly associated with activation of the SHH pathway (17 / 19 cases assessed (89%); 'p'=3.3 x 10^{-14}, chi-squared test), whilst subgroup 3 was associated with activation of the WNT pathway (5 / 5 cases assessed (100%); 'p'=1.4 x 10^{-14}, chi-squared test). The relationship was not entirely exclusive, however, with one case each activated for WNT and SHH pathway being observed in subgroup 4.

4.4.5.2 Two test dataset methylation subgroups are associated with staining for antibody markers associated with activation of developmental signalling pathways

The expression signature used to assign WNT and SHH signalling pathway activation in the training cohort was unsuitable for RNA extracted from FFPE tissues. However, data reporting WNT and SHH pathway activation was available through an immunohistochemical technique for the majority of test set samples (115 / 130 (88%)) (see section 4.3.9). This work was carried out by Ellison and colleagues, who were working on the same PNET3 trials cohort that comprised the test dataset (Ellison et al., 2011a).

The same associations observed between methylation subgroups, pathway activation and gene expression signature in the training cohort were recapitulated in the test cohort with antibody status. (The following association tests also include training cohort members derived from FFPE materials where antibody status was available (*n* = 5)). Subgroup 1 was highly significantly associated with the SHH antibody GAB1 positivity ('p' = 5.2 x 10^{-16}, chi-squared test). Subgroup 3 was highly significantly associated with nuclear accumulation of β-catenin ('p' < 2 x 10^{-16}, chi-squared test).

Again, mirroring what was observed in the training cohort with expression signatures, the relationship was not exclusive. There was one SHH antibody positive member in both subgroups 2 and 3, and one WNT antibody positive member in subgroup 1.

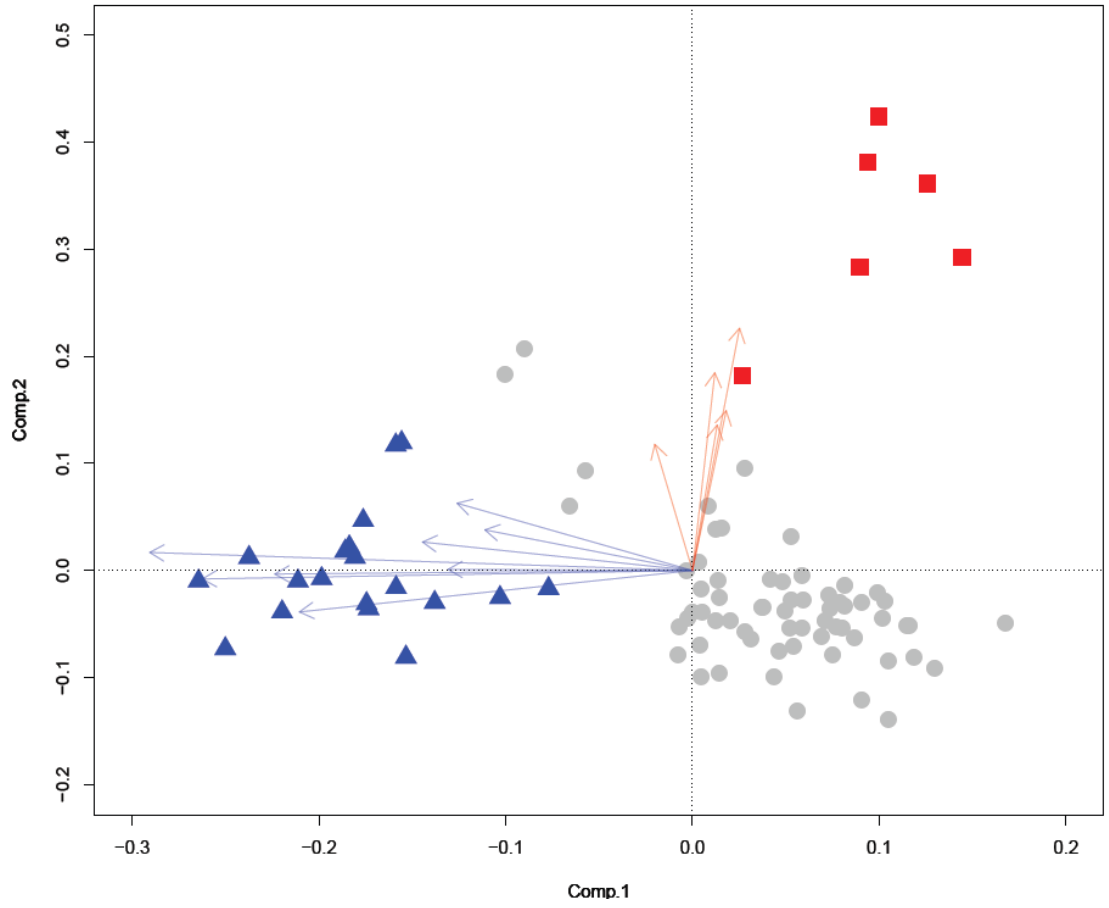


Figure 4.6. Bi-plot of WNT / SHH mRNA expression signature for 88 medulloblastomas, assessed by GeXP analysis (chapter 3). Axes show principal component loadings along component 1 (x axis) and component 2 (y axis). Arrows show projections of expression for each gene (SHH signature genes, blue; WNT signature genes, red). Cases are coloured according to assignment. SHH positive, blue triangle; WNT positive, red square; WNT / SHH independent, grey circle).

4.4.6 Global patterns of methylation within subgroups

The 4 subgroups identified were assessed for any global differences in methylation patterns. Comparisons were made between the absolute number of hypo- and hyper-methylated probes across subgroups, for the whole dataset (minus X chromosome probes), and for probes situated within and outside of CpG islands (Figure 4.7).

No evidence for CIMP (see section 1.3.7.3.2) was detected in any group. While there were no between-group differences in methylation patterns across the whole dataset, when the dataset was divided into probes within and outside of CpG islands, significant differences became apparent. There were highly significant differences between CpG island hypo-methylated probes ('p' = 4.6×10^{-8}, ANOVA) and non-CpG island hyper-methylated probes ('p' = 8.7×10^{-5}, ANOVA).

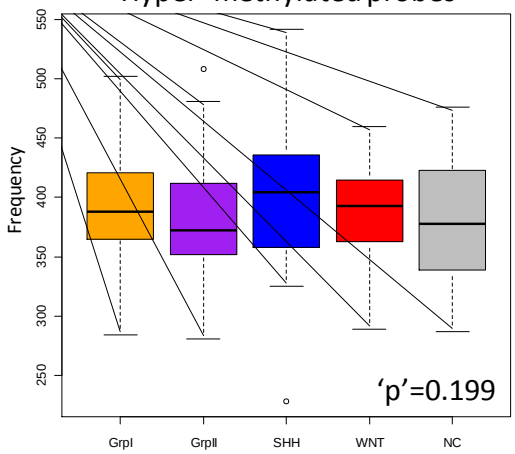
Tukey HSD tests were applied to identify the pairs of classes that were significantly different. For CpG island hypo-methylated probes, there were significant increases in the number of hypo-methylated group I probes compared to group II ('p' = 6.0×10^{-7}) and SHH ('p' = 7.2×10^{-7}) group members. For the non-CpG island hyper-methylated probes, there were significant increases in the number of hyper-methylated group I probes compared to group II ('p' = 1.8×10^{-5}). In summary, group I cases had an excess of hypo-methylated CpG island probes and an increase in hyper-methylated non-CpG island probes relative to group II.

4.4.7 Non-classified samples share few molecular features and are not qualitatively different from classified samples

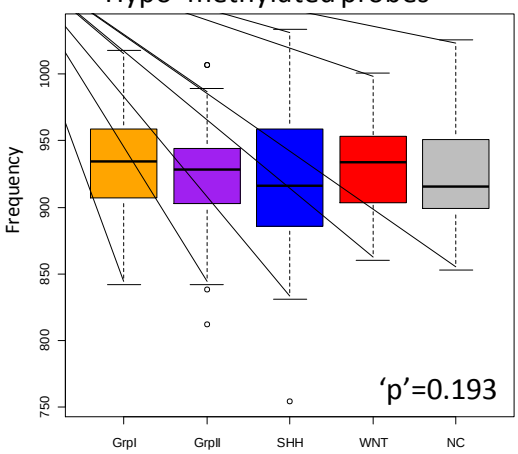
NC tumours (*n* = 14) comprised 6% of cases within the combined cohort. The NC samples were tested against the classified samples (*n* = 216) for differences in clinico-pathological and molecular correlates. No significant differences were detected (Table 4.7). Their clinico-pathological correlates are shown in Table 4.8. Two NC samples were SHH positive (one SHH GeXP positive, one SHH antibody positive). All other cases were negative for both WNT and SHH signature / antibody status. All NC cases were classic type histology.

All non-X chromosome probes

Hyper- methylated probes

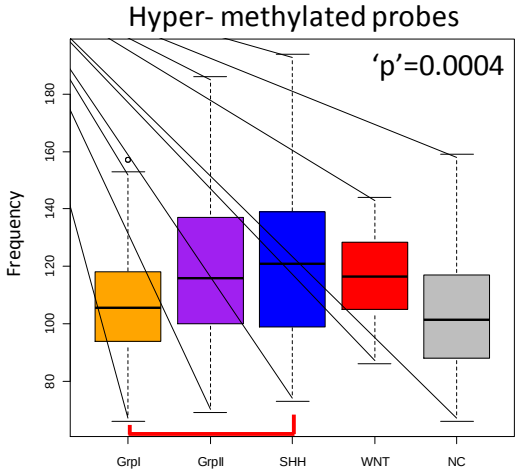


Hypo- methylated probes

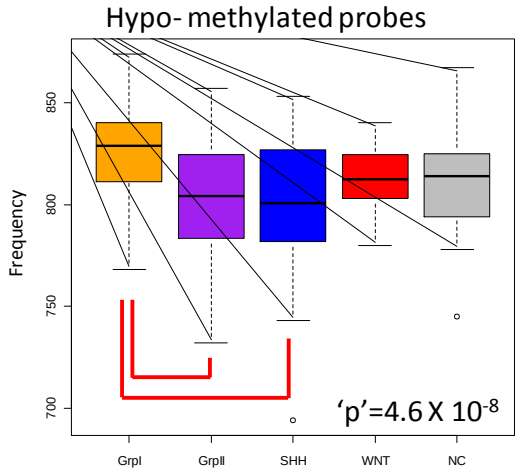


All non-X chromosome CpG island probes

Hyper- methylated probes

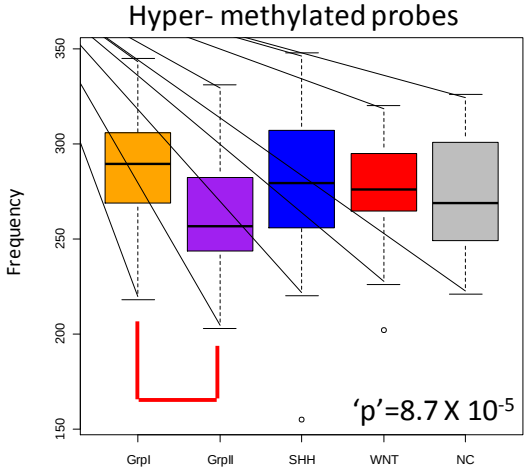


Hypo- methylated probes



All non-X chromosome non-CpG island probes

Hyper- methylated probes



Hypo- methylated probes

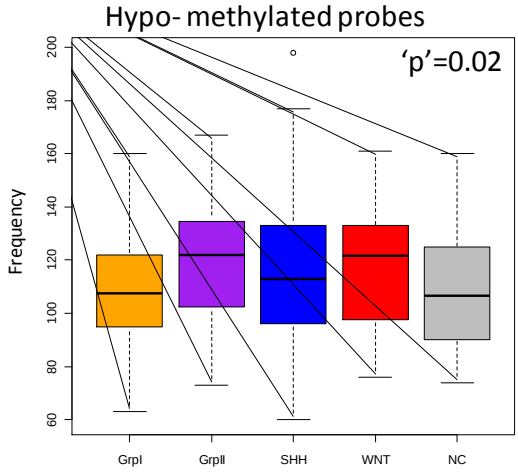


Figure 4.7. Comparison of global methylation patterns across identified subgroups. Boxplots show absolute numbers of hyper- or hypo-methylated probes across whole dataset after removal of X chromosome probes (row 1); across probes situated within a CpG island (row 2); across probes situated outside of CpG islands (row 3). Data on CpG island membership was supplied by Illumina in the annotation file for the Golden Gate array. Number of probes is shown on the y axis. 'P' values given are from ANOVA tests. Where this 'p' value was less than 0.05, Tukey Honest Significant Different tests were applied to identify group pairs where the mean significantly differed. Significant differences with 'p' < 0.0001 are indicated with a red connecting line.

Correlate	*'p'* value
Sex	1
DN histology	0.382
LCA histology	0.623
M stage	1
Age – under 3 *vs* over 3	1
Age (years)	0.440
SHH Sig	1
SHH Ab	1
WNT Sig	1
WNT Ab	0.580
Chr 6 loss	0.613
Chr 17 loss	0.685
MYCC amp	0.279
MYCN amp	0.445

Table 4.7. Comparison of clinico-pathological and molecular correlates between classified (*n* = 216) and non-classified samples (*n* = 14). Fisher's exact test was used for all comparisons except age in years, for which a t test was used. No correction for multiple testing was undertaken since all results were non-significant before any correction.

Sample ID	NMB128	NMB133	NMB250	X148	X150	X141	NMB137	X50120	X50012	X50041	X50291	NMB63	NMB182	NMB90
Sex	M	F	M	M	F	M	F	M	F	M	M	M	M	F
DN	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LCA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M Stage	NA	0	0	0	0	0	1	0	1	0	0	0	0	0
Infant (age < 3)	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Age (years)	4.5	6.9	4.8	4.2	8.9	6.4	15.1	13.5	4.2	10.8	6.7	11.5	9.6	3
SHH Sig	0	NA	0	NA	NA	NA	0	NA	NA	NA	NA	1	0	0
SHH Ab	NA	NA	NA	NA	0	0	NA	1	0	0	NA	NA	NA	NA
WNT Sig	0	NA	0	NA	NA	NA	0	NA	NA	NA	NA	0	0	0
WNT Ab	NA	NA	NA	NA	0	0	NA	0	0	0	NA	NA	NA	NA
CTNNB1 mutation	0	0	0	0	0	0	NA	0	0	0	0	0	0	0
Chr 6 LOH	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Chr 17 LOH	NA	NA	NA	0	1	0	NA	0	0	0	0	0	NA	0
MYCC amp	0	0	NA	0	0	0	0	0	1	0	0	0	0	0
MYCN amp	1	0	NA	0	0	0	0	0	0	0	0	0	0	0

Table 4.8. Clinico-pathological and molecular correlates of non-classifiable cohort. Gender (M, male; F, female). Histology (DN – desmoplastic / nodular; LCA – large cell / anaplastic). M stage (M-, 0; M+, 1). Age (>3 years, 0; ≤3 years, 1). SHH Sig – assessment of SHH expression signature positivity (0, signature negative; 1, signature positive). SHH Ab – assessment of positivity for SHH antibody staining (0 / 1, negative / positive for GAB1 antibody staining). WNT Sig – assessment of WNT expression signature positivity (0, signature negative; 1, signature positive). WNT Ab – assessment of positivity for nuclear accumulation of β-catenin (0 / 1, negative / positive for CTNNB1 antibody staining). *CTNNB1* mutation status (0, no mutation detected; 1 – mutation detected). Chromosome 6 and 17 loss, assessed by LOH analysis (0, no loss detected; 1, loss detected). *MYCC* and *MYCN* amplification, assessed by real time PCR (0, no amplification; 1 – amplification). NA – not available.

4.4.8 Assignment of names to identified subgroups

On the basis of the reported associations between subgroups and activation of the WNT and SHH signalling pathways described above, group names were assigned. The two SHH and WNT pathway-associated groups were named SHH and WNT. The remaining two groups, characterised predominantly by an absence of signalling pathway activation, were named group I and group II, with the largest group being assigned to group I. This designation is shown in Table 4.9.

K means cluster	Name	Training, *n* (%)	Test, *n* (%)	Joined, *n* (%)
1	SHH	21 (22%)	29 (24%)	50 (23%)
2	Group II	23 (24%)	25 (20%)	48 (22%)
3	WNT	10 (10%)	18 (15%)	28 (13%)
4	Group I	43 (44%)	51 (41%)	94 (43%)

Table 4.9. Named methylomic subgroups of medulloblastoma. The numbers and percentage comprising each group in the training, test and joined cohorts are shown.

4.4.9 Identification of probes that correlate with metagenes for the identification of novel biomarkers

The methylation probes that defined each metagene were identified by carrying out a Pearson's correlation of the probe methylation levels to the expression values of the metagenes within the training cohort. Since there was an exact correlation between metagene expression and subgroup membership, these probes also represented biomarkers for subgroup membership. The top 10 most correlative and anti-correlative probes are shown in Table 4.10 and shown graphically in Figure 4.5. It was not possible to carry out Gene Set Enrichment Analysis (Subramanian et al., 2005), firstly due to the low resolution of the array, but secondly because the array is highly enriched for cancer specific genes, so that even a random selection of genes present on the array will identify pathways and gene sets relevant to cancer and report them as significantly enriched against a genomic background.

The 80 significant probes listed in Table 4.10 were selected and used to assess subgroup assignment by PCA in the combined classified training and test cohort. Sample classification was assessed by plotting covariate spheroids along the 95% confidence intervals (Figure 4.8).

	Probe	Gene	Pearson's Correlation	Average β Value (group members)	Average β Value (non-group members)	Difference in β value
WNT Metagene	TRIM29_P135_F	*TRIM29*	-0.88	0.49	0.95	-0.46
	HDAC7A_P344_F	*HDAC7A*	-0.82	0.27	0.94	-0.66
	IAPP_E280_F	*IAPP*	-0.7	0.57	0.96	-0.4
	GP1BB_E23_F	*GP1BB*	-0.69	0.38	0.89	-0.51
	SPP1_P647_F	*SPP1*	-0.68	0.56	0.91	-0.35
	RARA_P1076_R	*RARA*	-0.67	0.49	0.92	-0.43
	SERPINB5_P19_R	*SERPINB5*	-0.67	0.79	0.97	-0.18
	DDR2_E331_F	*DDR2*	-0.62	0.27	0.87	-0.6
	RAB32_P493_R	*RAB32*	-0.62	0.22	0.84	-0.62
	FER_P581_F	*FER*	-0.6	0.35	0.84	-0.49
	HOXA9_P303_F	*HOXA9*	0.7	0.57	0.1	0.47
	PTHR1_P258_F	*PTHR1*	0.72	0.87	0.34	0.53
	ASCL2_P609_R	*ASCL2*	0.75	0.95	0.24	0.71
	POMC_P400_R	*POMC*	0.76	0.73	0.1	0.62
	TNFRSF10C_E109_F	*TNFRSF10C*	0.79	0.5	0.12	0.38
	ASCL2_P360_F	*ASCL2*	0.81	0.95	0.12	0.84
	MT1A_E13_R	*MT1A*	0.82	0.8	0.15	0.65
	TNFRSF10C_P7_F	*TNFRSF10C*	0.82	0.68	0.17	0.51
	ASCL2_E76_R	*ASCL2*	0.83	0.91	0.08	0.83
	MT1A_P49_R	*MT1A*	0.85	0.77	0.08	0.69
SHH Metagene	VAV1_E9_F	*VAV1*	-0.89	0.31	0.95	-0.63
	VAV1_P317_F	*VAV1*	-0.84	0.24	0.86	-0.62
	AATK_E63_R	*AATK*	-0.79	0.68	0.96	-0.28
	AATK_P709_R	*AATK*	-0.78	0.23	0.71	-0.48
	LCN2_P86_R	*LCN2*	-0.77	0.47	0.82	-0.36
	AATK_P519_R	*AATK*	-0.75	0.41	0.84	-0.43
	SPDEF_P6_R	*SPDEF*	-0.74	0.27	0.69	-0.42
	TGFB1_P833_R	*TGFB1*	-0.7	0.46	0.95	-0.49
	SPDEF_E116_R	*SPDEF*	-0.69	0.13	0.47	-0.34
	GFAP_P1214_F	*GFAP*	-0.65	0.38	0.75	-0.37
	PLG_E406_F	*PLG*	0.64	0.87	0.41	0.46
	IL16_P93_R	*IL16*	0.64	0.8	0.29	0.51
	MMP7_P613_F	*MMP7*	0.65	0.85	0.59	0.26
	CCKAR_E79_F	*CCKAR*	0.67	0.82	0.23	0.59
	DSC2_E90_F	*DSC2*	0.72	0.61	0.07	0.55
	CCKAR_P270_F	*CCKAR*	0.73	0.85	0.24	0.61
	KIAA1804_P689_R	*KIAA1804*	0.74	0.62	0.15	0.47
	MBD2_P233_F	*MBD2*	0.78	0.62	0.19	0.43
	MSH2_P1008_F	*MSH2*	0.88	0.87	0.14	0.73
	BAX_E281_R	*BAX*	0.89	0.49	0.15	0.34

	Probe	Gene	Pearson's Correlation	Average β Value (group members)	Average β Value (non-group members)	Difference in β value
Group I Metagene	CCKAR_E79_F	*CCKAR*	-0.83	0.08	0.58	-0.5
	PIK3R1_P307_F	*PIK3R1*	-0.79	0.08	0.6	-0.52
	PLG_E406_F	*PLG*	-0.76	0.29	0.68	-0.39
	IFNGR2_P377_R	*IFNGR2*	-0.75	0.09	0.52	-0.43
	CCKAR_P270_F	*CCKAR*	-0.75	0.1	0.59	-0.49
	RAN_P581_R	*RAN*	-0.75	0.21	0.64	-0.43
	ASCL2_P609_R	*ASCL2*	-0.65	0.12	0.49	-0.37
	RHOH_P121_F	*RHOH*	-0.65	0.41	0.82	-0.41
	ASCL2_P360_F	*ASCL2*	-0.64	0.04	0.37	-0.33
	ZNF264_P397_F	*ZNF264*	-0.63	0.09	0.46	-0.36
	ACTG2_P346_F	*ACTG2*	0.7	0.86	0.64	0.22
	IL8_P83_F	*IL8*	0.7	0.95	0.68	0.27
	SPP1_P647_F	*SPP1*	0.7	0.98	0.78	0.2
	SPP1_E140_R	*SPP1*	0.72	0.75	0.37	0.38
	HLA-DOB_E432_R	*HLA-DOB*	0.73	0.87	0.59	0.28
	MMP10_E136_R	*MMP10*	0.73	0.82	0.42	0.4
	HLA-DPA1_P28_R	*HLA-DPA1*	0.76	0.85	0.5	0.35
	WNT10B_P993_F	*WNT10B*	0.76	0.76	0.33	0.42
	LEFTY2_P719_F	*LEFTY2*	0.77	0.86	0.49	0.37
	TRIM29_E189_F	*TRIM29*	0.82	0.74	0.33	0.41
Group II Metagene	ZNFN1A1_E102_F	*ZNFN1A1*	-0.8	0.41	0.9	-0.49
	SERPINA5_E69_F	*SERPINA5*	-0.77	0.54	0.88	-0.34
	AIM2_E208_F	*AIM2*	-0.75	0.76	0.96	-0.21
	KRT1_P798_R	*KRT1*	-0.71	0.54	0.93	-0.39
	B3GALT5_P330_F	*B3GALT5*	-0.68	0.87	0.98	-0.11
	IL12B_E25_F	*IL12B*	-0.67	0.78	0.97	-0.19
	KRT5_E196_R	*KRT5*	-0.67	0.78	0.98	-0.2
	BLK_P668_R	*BLK*	-0.65	0.26	0.82	-0.56
	PLA2G2A_E268_F	*PLA2G2A*	-0.65	0.47	0.88	-0.41
	CEACAM1_P44_R	*CEACAM1*	-0.65	0.66	0.93	-0.27
	TGFB2_E226_R	*TGFB2*	0.72	0.5	0.09	0.41
	LOX_P313_R	*LOX*	0.72	0.3	0.03	0.27
	FES_P223_R	*FES*	0.73	0.6	0.13	0.47
	FRZB_E186_R	*FRZB*	0.74	0.52	0.08	0.44
	RARRES1_P426_R	*RARRES1*	0.75	0.63	0.21	0.41
	BCR_P346_F	*BCR*	0.75	0.3	0.06	0.24
	IL1RN_P93_R	*IL1RN*	0.76	0.58	0.13	0.45
	FES_E34_R	*FES*	0.8	0.58	0.1	0.48
	MMP14_P13_F	*MMP14*	0.83	0.53	0.09	0.45
	WRN_P969_F	*WRN*	0.84	0.65	0.17	0.49

Table 4.10. The top 10 most anti-correlative and correlative probes that define each metagene represent novel biomarkers for the methylomic subgroups of medulloblastoma. Anti-correlative probes are shown with a white background. Correlative probes are shown with a grey background. Probe name, gene name, Pearson correlation, β scores of group and non-group members, as well as the difference in β values between group and non-group members are given.

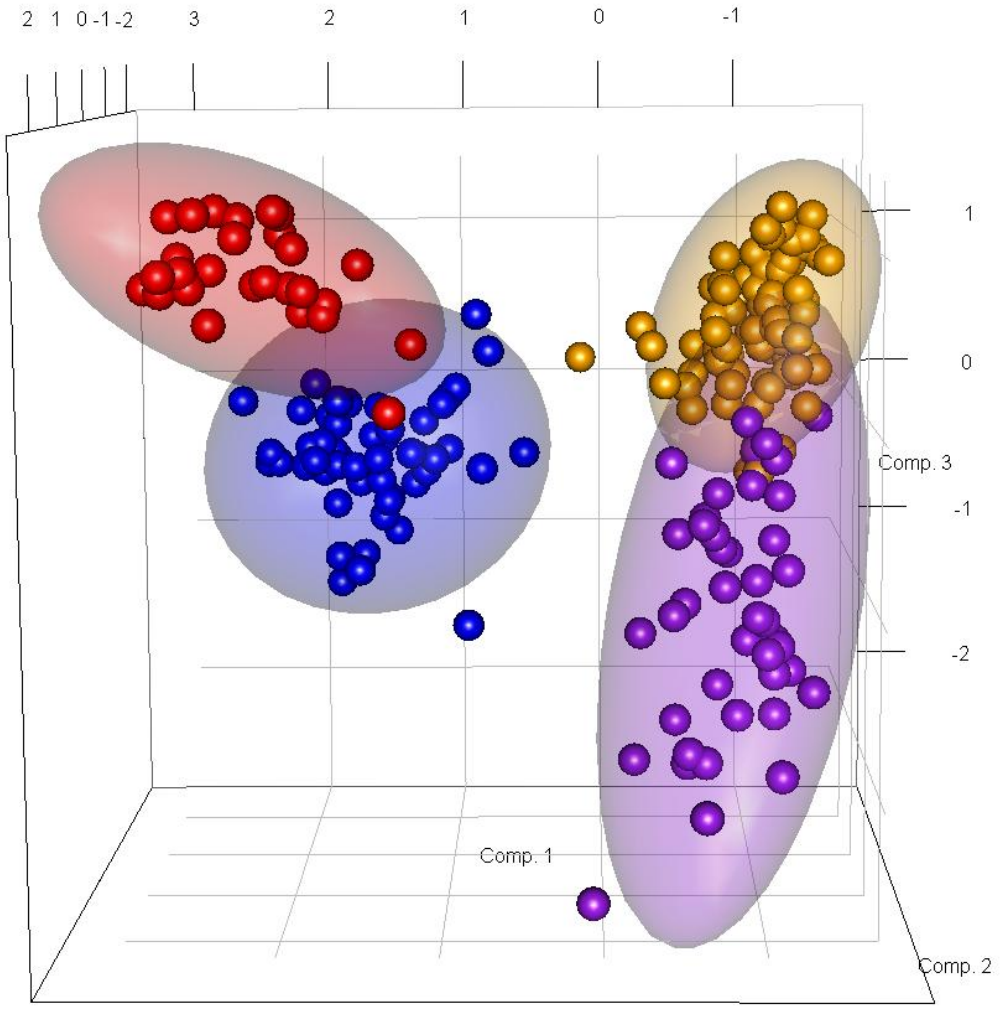


Figure 4.8. Selection of the top 10 most correlative and anti-correlative probes for each metagene recapitulates the assigned classes. Principal component analysis loadings plot, derived from 80 methylation probes showing highest correlation and anti-correlation to metagenes is shown. Class designations for WNT (red), SHH (blue), Group I (orange) and Group II (purple) are shown for 216 classifiable cases in the training and test cohorts. Covariance spheroids with a 95% confidence interval have been plotted to demonstrate group memberships.

4.4.10 DNA methylomic subgroups show different clinical disease features and behaviours

The classification of medulloblastomas into discrete subgroups enables differences in clinico-pathological and molecular correlates between subgroups to be assessed (Figure 4.9). Significant differences were assessed using chi-squared tests, with the contributing subgroup being identified by examining the chi-squared residuals for each subgroup (Figure 4.5). The WNT and SHH pathway activated subgroups were strongly associated with previously published markers for these subgroups (Ellison et al., 2011a; Schwalbe et al., 2011; Clifford et al., 2006) (see section 4.4.5.1, 4.4.5.2). The WNT methylomic subgroup was characterised by *CTNNB1* mutation ('p' < 2×10^{-16}), chromosome 6 loss ('p' < 2×10^{-16}), WNT expression signature (training cohort, 'p'=1.4×10^{-14}) and nuclear accumulation of β-catenin (test cohort, 'p' < 2×10^{-16}). The SHH methylomic subgroups was associated with DN histology ('p'=1.8×10^{-14}), SHH expression signature (training cohort, 'p' = 3.3×10^{-14}), GAB1 staining ('p'=5.2×10^{-16}), indicating SHH subgroup (test cohort) and infant status (training cohort only, 'p'=3.8×10^{-6}). There were few instances of cases positive for these markers being present in other subgroups (in the training cohort, one case with SHH pathway activation and one case with WNT pathway activation were classified into group I; in the test cohort, one case with WNT antibody positivity and *CTNNB1* mutation was classified into the SHH group, and two SHH antibody positive cases were not classified into the SHH methylomic subgroup).

Significant differences in incidence for gender ('p' = 0.0028), LCA histology (enriched in group II cases, 'p' = 0.024), M stage (enriched in group I and II cases, 'p' = 0.011), and chromosome 17 LOH (enriched in group I cases, 'p' = 0.00039) were also observed, although the significant differences observed for LCA phenotype and M+ disease become non-significant after correction for multiple testing (Figure 4.9).

The classification of medulloblastomas into four subgroups enables direct comparison between the non-signalling pathway activated medulloblastomas group I and group II. When this direct comparison is made, there is a significant enrichment for LCA cases in group II ('p' = 0.0063, chi-squared test) and a marginally significant enrichment for chromosome 17 LOH in group I cases ('p' = 0.033, Fisher's exact test). The age

distribution graph shown in the last panel of Figure 4.9 demonstrates that group II cases occur predominantly in younger cases compared to group I cases. As has previously been reported (Schwalbe et al., 2011), WNT positive cases were not observed in infants. Although there is a significant enrichment for infant cases within the SHH methylomic subgroup ('p' = 3.8 x 10^{-6}, chi squared), overall, there was no significant difference in ages between subgroups ('p' = 0.60, ANOVA). Whilst SHH and WNT cases showed balanced male and female cases and a paucity of metastatic cases, conversely group I and group II cases showed an excess of male cases ('p' = 0.0028, chi squared). This indicates that the previously reported gender imbalance (1.7:1 M:F cases (section 1.3) is being driven by methylomic subgroups I and II, which also showed an increased incidence of metastatic cases (Figure 4.9).

The relationship between subgroup membership and survival is shown in Figure 4.10. The WNT subgroup is clearly associated with an improved EFS compared to the remaining methylomic subgroups, which do not show any differences in their survival. The NC samples show a trend for improved survival, although only 12 / 14 NC samples had available survival data. Significant survival differences were observed for *MYCC / MYCN* gene amplification status, large cell anaplastic histology, metastasis and infant status, confirming that previously reported prognostic markers were reflective of the results of previous studies. No survival difference was seen for chromosome 17 LOH status; additionally, a relationship was observed between female gender and improved survival (Ellison et al., 2011b; Pfister et al., 2010; Curran et al., 2009). These relationships are shown in Figure 4.11.

Cox models were constructed to test for prognostic utility of metagene expression and for validation of previously reported survival correlates (Table 4.11). In univariate analyses, expression of the WNT- ('p' = 0.0013) and group II- ('p' = 0.05) associated metagenes were significant, although only the WNT metagene was significant after correction for multiple testing. In a multivariate analysis including the verified prognostic factors metastatic status, LCA histology and infant status (Pizer and Clifford, 2009), in addition to the expression of the metagene, WNT ('p' = 0.002) metagene expression remained prognostic, with expression of the group I metagene of marginal significance ('p' = 0.05), although again the group I metagene expression is not

significant after correction for multiple hypothesis testing.

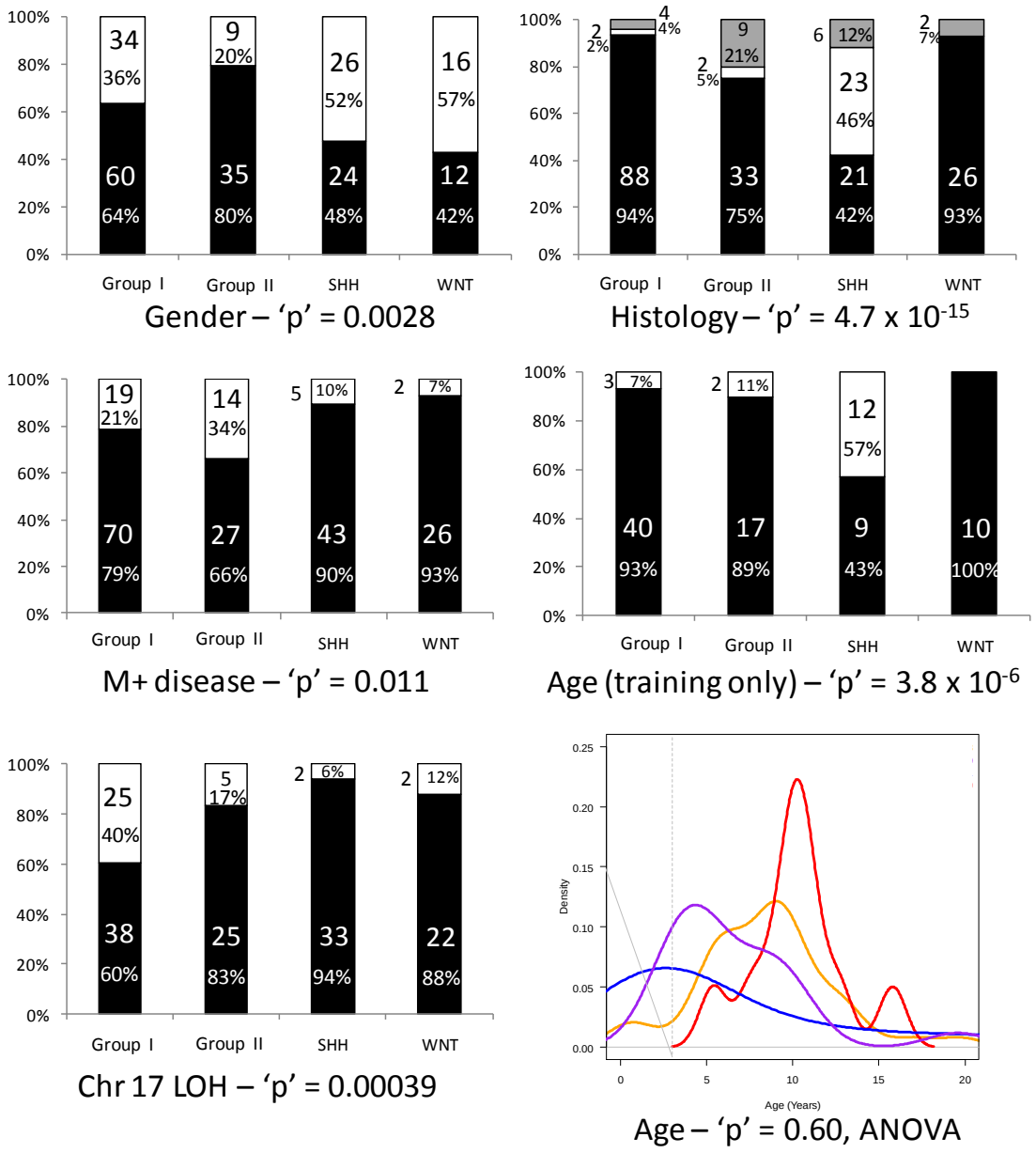


Figure 4.9. Clinical and molecular correlates of methylomic subgroups of medulloblastoma.

Number and percentage incidence are shown for selected clinical and molecular correlates of the disease across methylomic subgroups. Gender – male, black; female – white. Histology – classic, black; large cell / anaplastic, grey; desmoplastic / nodular – white. Metastasis – M- black; M+, white. Age (training cohort only, since this included infant cases) – black – over 3 years of age at diagnosis; white – under 3 years of age at diagnosis. Chr 17 LOH – chromosome 17p loss of heterozygosity assessed by HOMOD method. No loss detected – black; loss detected – white. 'P' values from chi squared tests are shown. The final figure (bottom right) shows the age distributions across the subgroups. WNT cases (red trace) are non-infant and peak in incidence at 10 years of age. SHH cases (blue trace) represent the largest proportion of cases observed in infancy, whilst group I cases (orange trace) have a peak of incidence at 9 years, in contrast to group II cases (purple trace) which have a peak of incidence at 4.5 years. The 'p' value from an ANOVA test to compare group means is shown.

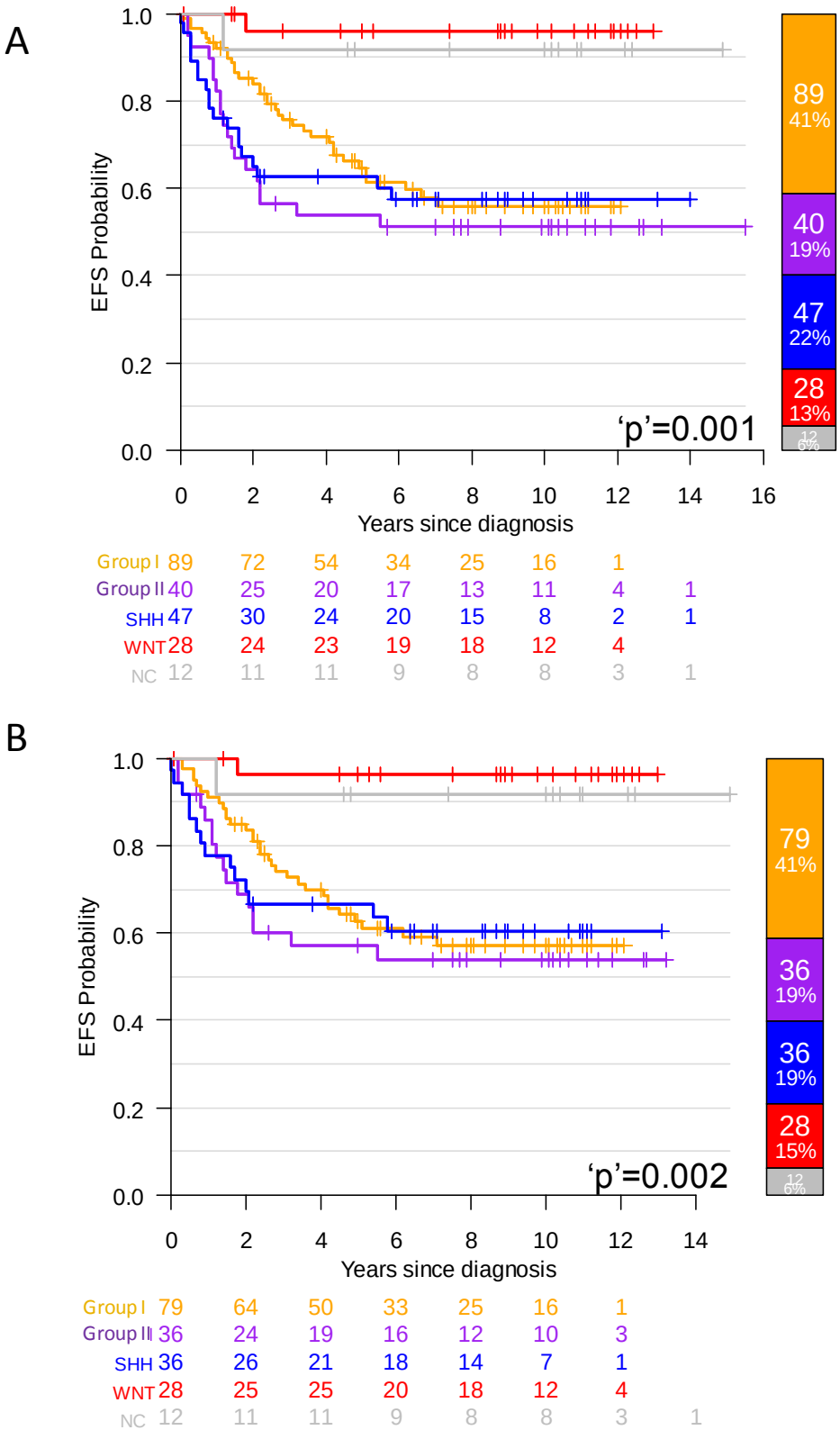


Figure 4.10. Differential survival of methylomic subgroups of medulloblastoma. A. Kaplan-Meier plot shows EFS for joined training and test cohorts with all available survival data (*n* = 216). B. Kaplan-Meier plot shows EFS data for PNET3 clinical trials cohort plus age-matched (aged 3-16 years) non-trials cases (*n* = 191). Subgroups are coloured as shown in at-risk table below plot. Stacked bar on right hand side of plot indicates number and percentage composition of each subgroup. 'P' values from log-rank tests are given.

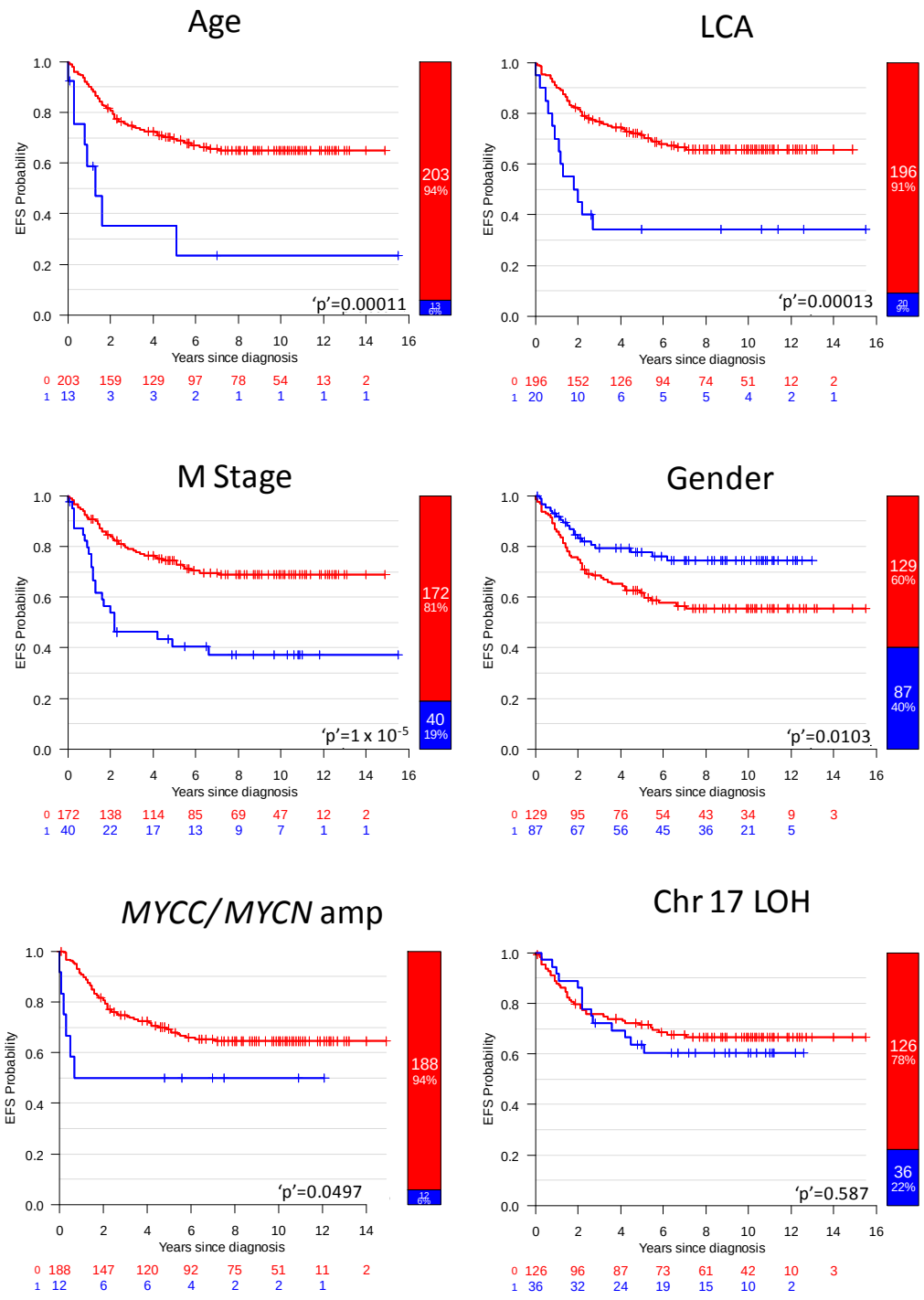


Figure 4.11. Investigation of relationship between previously reported molecular and clinico-pathological correlates with survival in the classified cohort with available survival data (*n* = 216). Each plot shows Kaplan-Meier plot, bar plot showing group membership and at-risk table, as explained in Figure 2.20. 'P' values derived from log-rank tests are shown. Age – 1– under 3; 0 - over 3 at diagnosis. LCA – 1 – LCA histological subtype; 0 – LCA negative. M stage – M-, 0; M+, 1; Gender – 0 – male; 1 – female. *MYCC / MYCN* amp – 0 – no amplification of *MYCC / MYCN*; 1 – amplification of *MYCC / MYCN*. Chr 17 LOH – 0 – no evidence for chromosome 17 LOH; 1 – evidence for chromosome 17 LOH.

	Univariate		Multivariate	
Correlate	'P' value	HR (95% CI)	'P' value	HR (95% CI)
V1 (WNT)	0.013	0.19 (0.069 – 0.52)	0.0020	0.18 (0.061 – 0.53)
V2 (SHH)	0.31	1.40 (0.74 – 2.65)	0.12	1.69 (0.87 – 3.28)
V3 (Group I)	0.14	1.67 (0.85 – 3.30)	0.05	2.11 (1.00 – 4.45)
V4 (Group II)	0.05	2.14 (1.01 – 4.59)	0.40	1.44 (0.62 – 3.33)
Large cell / anaplastic	0.0006	2.89 (1.58 – 5.27)	NA	NA
M+	9.6×10^{-5}	2.68 (1.64 – 4.42)	NA	NA
Age (under 3)	0.0005	3.72 (1.78 – 7.81)	NA	NA
Chr17 LOH	0.70	1.13 (0.61 – 2.07)	0.62	1.17 (0.63 – 2.20)
MYCC / MYCN amplification	0.03	2.55 (1.10 – 5.91)	0.002	4.07 (1.70-9.75)
Female Gender	0.007	0.48 (0.29 – 0.82)	0.03	0.55 (0.32 – 0.93)

Table 4.11. Cox model investigation of metagene expression and other previously postulated prognostic markers in classified cohort with survival data available (*n* = 204) identifies potentially prognostic metagenes. Univariate analysis considers expression of each metagene or correlate positivity in isolation Correlates considered include large cell / anaplastic histology, M+ disease, infant status, chromosome 17 LOH, amplification of *MYCC / MYCN*, and gender. 'P' value and HR (hazard ratio) (with 95% confidence intervals) are shown. The multivariate analysis included large cell / anaplastic histology, M stage and infant status as prognostic factors in the base model. Subsequently, each metagene or correlate that was not already in the model was added. The 'p' value for metagene and correlate expression in this multi-factorial model is reported, as well as hazard ratio and 95% confidence intervals. NA – not applicable, since large cell /anaplastic histology, M+ disease and infant status were covariates included in the base model for multivariate analysis.

4.4.11 Summary of the described subgroups and their characteristics

A consensus clustering approach was applied to a training cohort of 100 samples. 4 clusters were identified and validated in a test cohort of 130 samples. 94% (216 / 230) of samples could be confidently classified. Subgroup members showed highly significant differences in their clinico-pathological and molecular correlates (Figure 4.5, Figure 4.9). The WNT subgroup was associated with an improved survival; there were no survival differences between the remaining subgroups (Figure 4.10). The characteristic features of the subgroups are summarised in Table 4.12.

	Group I	Group II	SHH	WNT
Molecular abnormalities	Chr 17p LOH			Chr 6 LOH
Expression Profile / Antibody profile			SHH Signature SHH antibody GAB1 immunopositivity	WNT Signature WNT antibody CTNNB1 nuclear accumulation
Disease Features	Fewer female cases	LCA histology Fewer female cases	DN histology Enriched for infants	Classic histology M- disease No infant cases Improved survival

Table 4.12. Summary of molecular and clinical characteristics of methylomic subgroups. Chr 6 / 17p LOH – chromosome 6 or 17p LOH.

4.5 Discussion

4.5.1 Golden Gate Cancer Panel I array is robust, reproducible and reports accurate estimates of DNA methylation

This chapter employed a novel methylome profiling technology for the identification of disease subgroups of medulloblastoma cohorts. The failure rate for both training and test datasets (section 4.4.1) was consistent (7% and 9% respectively) and sample reproducibility was also satisfactory (Figure 4.1, Figure 4.2). One intriguing aspect of the nature of the QC failures within the test cohort was the enrichment for *MYCN* amplification, with 6 / 13 (46%) of QC failures being *MYCN* amplified. This contrasts to the QC-passing samples, where the frequency of *MYCN* amplification was 6 / 124 (5%).

The intriguing enrichment for *MYCN* amplified cases in the test cohort which failed initial QC is difficult to explain, as it is exceedingly unlikely to occur by chance ('p' = 0.0001, Fisher's Exact test) and the same relationship was not observed in the training dataset QC failures (6 / 6 failures with available data were not *MYCN* amplified). It is possible that the process of DNA extraction in *MYCN* amplified medulloblastomas from FFPE tissue is unreliable, but it could also hint that there is something fundamentally different about the stability of DNA in *MYCN* amplified medulloblastomas. Ideally, a paired comparison between DNA extracted from fresh frozen tissue and from FFPE tissue on the same sample for *MYCN* amplified cases would be performed on the methylation array to determine whether the observed effect is a methodological consequence of the more fragmented DNA from FFPE tissues, although this class of samples would be scarce, due to the low incidence of *MYCN* amplified cases and the difficulty in obtaining paired DNA samples.

The ability to generate methylomic data from DNA extracted from FFPE tissues offers potential for further investigatory or confirmatory work in archival cohorts, for which nucleic acid derivatives from fresh frozen tissues are not generally available, greatly increasing the potential sample pool for future studies, with size always an important limitation in generating informative medulloblastoma cohorts. Moreover, this technology does not seem subject to the batch effects commonly encountered in transcriptomic experiments, where sample preparation and handling can be a very important determinant of array performance. This may be due to the inherent stability

of 5-methylcytsoine residues within DNA compared to the transitory half-life and increased fragility of mRNA transcripts.

Bisulfite validation of a panel of 7 discriminatory probes in 18 medulloblastomas confirmed that the reported β values from the array were a close match to values estimated from bisulfite sequencing, with a mean difference in β score of 0.006. The non-random nature of the minority of comparisons with a large deviation suggests that this is not a locus-specific phenomenon but rather a stochastic process, and is therefore not a cause for concern, particularly in the context of an experiment measuring methylation at 1505 loci. Any stochastic errors in methylation estimation at individual loci would not, by definition, be able to contribute towards any identified subgroups within the dataset.

The investigations into sample reproducibility and the fidelity of the reported methylation values proved satisfactory and enabled analysis of methylation patterns of medulloblastoma to proceed.

4.5.2 Medulloblastoma comprises 4 methylomic subgroups

The classification of medulloblastoma using methylomic arrays in a 100 member training cohort identified four disease subgroups (group I, group II, SHH and WNT), which were validated in a 130 member test cohort. The validation of the identified subgroups in an independent second cohort demonstrates subgroup reproducibility, avoiding any confounding through over-fitting. The identification of subgroups is important in two ways: firstly, it represents the first time that methylomic approaches have been used to classify the disease, validating this approach and secondly, this provides the first classification of the disease independent of transcriptomic measures.

As discussed in section 1.3.12.1, estimates for the number of transcriptomic disease subgroups ranges from 4 to 6. Although the number varies, all studies report two discrete subgroups characterised by activation of the WNT or SHH signalling pathways. Like all previous transcriptomic studies, methylomic classification identified two clear WNT and SHH subgroups. In this case, two additional pathway independent subgroups were identified. The high congruence between WNT and SHH pathway activation (Figure 4.5), assessed by GeXP in the training cohort and by antibody positivity in the

test cohort clearly identifies the subgroups and represents a positive three-way validation of each technique for the assessment of WNT and SHH pathway status.

The estimated number of pathway-independent methylomic subgroups matches the number of pathway-independent transcriptomic subgroups identified by Northcott et al., although it is, without further investigation, impossible to ascertain whether there is any correlation between the two independently-identified pathway-independent disease subgroups. The investigation of this question is discussed in section 4.5.6.

The intermediate levels of expression of the V3 metagene by the majority of group II medulloblastomas (in addition to the high expression of the group II-defining metagene, V4), could indicate that group I and group II tumours arise from a similar cell of origin, with group II tumours acquiring additional defects.

4.5.2.1 Non-classifiable samples are distinct and share few clinico-pathological and molecular correlates

There were no common clinico-pathological, molecular or methylomic correlates that defined the 14 / 230 samples assigned as NC (Table 4.8). One possible explanation for the lack of common features is that the reason for the NC status is sample independent and is actually due to a cryptic technical error that was not identified with the QC measures available. If true, this technical error could be procedural or could lie with the quality of the original DNA sample. Another possible explanation is that these samples are atypical medulloblastomas and are simply not amenable to classification when referenced against typical medulloblastomas. It is also possible that the NC group do indeed represent a separate class of medulloblastomas and that the resolution of the array was not sufficient to include common features absent in classified medulloblastomas, although caution would have to be applied when using a higher resolution approach, since this, of course, would be prone to over-fitting (section 2.11).

The improved survival observed in the 12 / 14 NC cases compared to other non-WNT cases for which survival data was available is intriguing, although any interpretation of the result would necessarily have to be tempered with the consideration of the small sample size involved. Indeed, if a log-rank test is performed between NC and classified

samples in a binary comparison, the improved survival observed for NC samples is not significant ('p' = 0.06). No NC case was positive for any WNT subgroup correlates (WNT expression signature, WNT antibody positivity, chromosome 6 loss or β-catenin mutation, meaning that the improved survival observed is not being driven by its association to WNT pathway activation (Clifford et al., 2006; Ellison et al., 2005). This represents an interesting observation that requires further study in additional cohorts, to establish whether or not the NC class is consistent across studies and by running additional high throughput genomic and transcriptomic arrays, whether there are any copy number abnormalities or gene expression changes that are common to this subset of medulloblastomas.

4.5.2.2 Global patterns of DNA methylation

Array structure was investigated to identify global patterns of medulloblastoma methylation. The majority of CpG island probes (83%) were unmethylated, in contrast to non-CpG island probes with a majority of methylated probes (62%). This is consistent with what has been previously reported, with the majority of CpG islands being unmethylated and promoter regions of genes without CpG islands being predominantly methylated (Deaton and Bird, 2011).

When the analysis was extended to individual subgroups, differences became apparent. Group I samples were associated with an increase of hypo-methylated probes compared to group II within CpG islands and an increase of hyper-methylated probes outside of CpG islands relative to group II probes.

It is possible that group I tumours are displaying patterns of methylation that more closely match the underlying methylation patterns within normal tissue (CpG islands predominantly unmethylated, promoter regions without a CpG island predominantly methylated), and that the group II tumours, whilst similar to group I, have acquired additional defects that further diverge from a more normal, cerebellum-like state. This would be in concordance with the hypothesis that the group I and II tumours share a common cell of origin, as discussed in section 4.5.2. This could easily be tested by examining the methylation status of normal cerebellum samples.

4.5.3 Identification of novel biomarkers for disease identification

The biomarkers listed in Table 4.10 greatly increase the number of testable biomarkers for the identification of medulloblastoma subgroups. The PCA plot in Figure 4.8 demonstrates that these probes accurately identify sample groupings across the combined classifiable training and test cohort (*n* = 216), emphasising the potential for subgrouping using a reduced number of methylation biomarkers. There is clear separation of the WNT and SHH subgroups, with some overlap on the boundary between group I and group II cases, showing that the boundary between them is likely to be fuzzy and indicates that any future methylomic assays for the assignment of subgroup status would likely include a measure of uncertainty, as was implemented in this study (where to be assigned to a group, a sample had to have a greater than 80% modal assignment score (see section 2.9.4)). They may also represent key subgroup-specific genes and pathways in medulloblastoma that may become more apparent when DNA methylation is assessed with a higher resolution platform.

The identification of these biomarkers raises several possibilities for future investigations. It remains an open question whether these biomarkers are functionally relevant or serve only as markers of disease type. By running transcriptomic arrays in conjunction with methylation arrays on paired samples, anti-correlative pairs of methylation and gene expression probes can be identified and verified, using sodium bisulfite and RT-PCR based techniques respectively, which in turn could identify subgroup specific transcriptomic differences with a putative functional relevance.

Future approaches, discussed in detail in section 4.5.6, could be directed towards the generation of a testable assay for the assessment of subgroup determinant methylation probes.

4.5.4 Methylomic subgroups show significant differences in clinico-pathological and molecular correlates

The identified methylomic subgroups show clear differences in their clinical and molecular features. While the WNT and SHH pathway-activated subgroups are characterised by known and previously described features, other important novel relationships were observed.

The SHH subgroup cases had a major peak of incidence in infancy (in the training cohort, 9 / 21 SHH cases were ≤ 3 years of age; the test cohort contained no infant cases). Interestingly, the majority of adult cases (>16 years of age – 4 / 7 cases total) were SHH cases, in agreement with previous studies (Northcott et al., 2011; Remke et al., 2011b).

The WNT subgroup displayed previously reported characteristics of this group of tumours (Schwalbe et al., 2011); there were highly significant associations with chromosome 6 loss and activating mutations of β-catenin. There were no WNT subgroups cases from infants, and this subgroup displayed a tight age range, peaking at 10 years of age. Both WNT and SHH subgroups were characterised by a small excess of females (SHH – 26 / 50 (52%); WNT – 16 / 28 (57%)) and a paucity of M+ cases (SHH – 5 / 48 (10%); WNT – 2 / 28 (7%)).

This gender ratio was in contrast to the non-signalling pathway activated subgroups, which had a large excess of male cases, in agreement with previous group-wide transcriptomic studies, which found approximate gender parity in the WNT and SHH subgroups, with an excess of male cases in the WNT / SHH pathway independent subgroups (Cho et al., 2011; Northcott et al., 2010). This was especially pronounced for group II members (group II – 35 / 43 (80%); group I – 60 / 94 (64%)). This and other differences between group I and group II represent evidence to support the hypothesis that these represent discrete disease subgroups. Group I was almost entirely composed of classic type medulloblastomas (88 / 94; 94%), with group II cases showing the highest subgroup incidence (9 / 43; 21%) of large cell / anaplastic cases. One third of group II cases (14 / 41; 34%) were M+ disease, compared to 19 / 89 (21%) group I cases ('p' = 0.13, Fisher's Exact test).

There are two pieces of evidence that show a potential agreement between the 4 methylomic subgroups described here and the 4 transcriptomic subgroups described by Northcott et al. Firstly, the age distribution profile for groups I and II (Figure 4.9) shows similar patterns of incidence in the two WNT / SHH independent transcriptomic subgroups C and D. Group II and group C (4 years, methylomic, 3-5 years, transcriptomic) show an earlier peak incidence than group I and group D (9 years methylomic, 11-15 years transcriptomic). Secondly, there is a significant enrichment

for chromosome 17 LOH in group I, and a significant enrichment for isochromosome 17q is observed in transcriptomic subgroup D. The processes by which any putative overlap between the methylomic and transcriptomic subgroups could be assessed is discussed in section 4.5.6.

The incidence of *MYCC* and *MYCN* amplification was not sufficient to identify any significant relationships, since there was no subgroup exclusivity for amplification. *MYCC* and *MYCN* amplification was observed in SHH, group I and group II cases. The WNT subgroup did not contain any *MYCC* or *MYCN* amplified cases, although this subgroup is also the smallest and the absence could be due to sampling. *MYCN* amplification is more common (9 *MYCN*: 5 *MYCC*) than *MYCC* amplification.

4.5.5 The WNT subgroup is associated with an improved survival

Survival investigations of the defined methylomic subgroups were undertaken to identify prognostic differences (Figure 4.11). Initially, previously reported survival markers were tested to ascertain whether this was a representative medulloblastoma cohort. M+ status ('p' = 5 x 10^{-5}) (Rutkowski et al., 2010), LCA histological subtype ('p' = 0.0003) (Pizer and Clifford, 2009), infant status ('p' = 0.0002) (Rutkowski et al., 2005), *MYCC* or *MYCN* amplification ('p' = 0.02) (Ellison et al., 2011b) and male gender ('p' = 0.0055) (Curran et al., 2009) were all associated with a worse prognosis, recapitulating previous studies. Conversely, no relationship between survival and chromosome 17 LOH was observed ('p' = 0.70), contradicting previous studies (Pan et al., 2005; Di Marcotullio et al., 2004), but in agreement with others (Ellison et al., 2011b; McCabe et al., 2011; Jung et al., 2004; Emadian et al., 1996), which found no relationship between LOH at chromosome 17 and survival. The recapitulation of the established clinical markers (M+ disease, LCA histology and infant status) demonstrates that this medulloblastoma cohort is representative of the wider disease and provides hope that any identification of additional, novel, prognostic markers may not be the result of an unbalanced cohort.

The prognostic potential of metagene expression was assessed in univariate and multivariate Cox proportional hazards models (Table 4.11). Similar to the Kaplan-Meier plots shown in Figure 4.10, the WNT metagene V1 was the only metagene related to survival. The improved survival observed for the WNT methylomic subgroup (Figure

4.10) has previously been reported using other measures of WNT pathway activation (Ellison et al., 2011b; Schwalbe et al., 2011; Clifford et al., 2006; Ellison et al., 2005). There were no significant differences in survival for the remaining subgroups in either the full cohort or within the PNET3 clinical trials cohort with age-matched non trials cases. This is perhaps surprising for the group II cases, which show an enrichment for LCA histology and M+ staging. This could mean that the effects of these poorly understood markers is pleiotropic, emphasising both the need for the identification of additional prognostic markers for the assignment of patient risk in non-WNT medulloblastomas and also the need to analyse survival correlates within sample subgroups.

Previous work has reported differences in survival for non-WNT transcriptomic subgroups, although this work is limited; until now, array-based classification of subgroup status in medulloblastoma has not been tested in a clinical trials cohort of the disease. Cho et al. reported that membership of the c1 subgroup, characterised by expression of *MYCC* and related translational / ribosomal signatures (Cho et al., 2011), was associated with a poor prognosis. Northcott et al. defined 4 transcriptomic subgroups and reported that these subgroups could be classified using single antibody subgroup identifiers (Northcott et al., 2010). The antibody-based classification was subsequently applied to a non-overlapping independent trials cohort to assess survival differences. Subgroup C was identified as being associated with a poor survival, although it is notable that there were significant differences in age distribution between subgroup C and subgroup D in the transcriptomic-classified and antibody-classified cohorts, raising the possibility that the transcriptomic subgroups were not perfectly recapitulated in the survival cohort, which was classified solely by antibody staining and emphasising the need for direct cross-validation in a trials-based cohort classified using the same technique used to define subgroups.

Until the number of transcriptomic clusters reaches consensus, and the means for assigning subgroup membership become more robust, it will remain problematical to ascribe differences in survival between subgroups, not least because of the difficulties in reproducing classifications across studies. This study represents the first stage of the process, defining molecular disease subgroups and testing survival in a trials-based cohort whose subgroup status had been molecularly defined (Figure 4.10).

Once again, DNA methylation may provide a more reliable means of assessing subgroup membership. DNA methylation is inherently less labile than measuring gene expression and was robust across DNA from both frozen and FFPE samples. It appears that the antibody assignment of WNT and SHH activation (Ellison et al., 2011a) is also robust, since the correlation between the SHH methylomic subgroup and SHH antibody positivity in the test dataset was high. However, until the reproducibility of antibody assignment of WNT / SHH independent medulloblastoma subgroups (Northcott et al., 2010) is validated, an alternative, non-immunohistochemical approach might be preferable. The use of methylation markers might represent an alternative approach, for example using an MS-MLPA (methylation sensitive, multiplex ligation-dependent probe amplification) (Nygren et al., 2005) assay, which could be used routinely in a diagnostic setting, since the technique it is based upon, MLPA (Schouten et al., 2002), is widely used in diagnostic labs.

4.5.6 Further work

While this chapter has identified four methylomic subgroups of the disease, this represents only the beginning of the investigations into methylomic classification of the disease.

At the inception of the project, two transcriptomic studies (Kool et al., 2008; Thompson et al., 2006), which reported 5 transcriptomic subgroups of the disease, were used as the basis for designing a GeXP assay to assign signalling pathway activation, and this assay was used to confirm the high congruence between WNT and SHH pathway activation and membership of the WNT and SHH methylomic subgroups (chapter 3).

Subsequent reports (Cho et al., 2011; Northcott et al., 2010) have identified four and six subgroups, respectively. Work being undertaken in 2011 by Dr. Dan Williamson (Brain Tumour Research Group, Newcastle University) has taken all published medulloblastoma transcriptomic datasets (Cho et al., 2011; Northcott et al., 2010; Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006) and undertaken a meta-analysis to identify the current best estimate of subgroup number. This has identified four transcriptomic subgroups (WNT, SHH and two non-signalling pathway subgroups). It will be very important to know whether the methylomic subgroups described in this

chapter are recapitulating the transcriptomic subgroups already described, or whether they represent a novel means of disease classification.

At the time of writing, a second GeXP assay to classify samples based on these four transcriptomic subgroups is being designed by Dr. Dan Williamson and Dr. Matthew Partington (Brain Tumour Research Group, Newcastle University), and the resultant assay will be applied to the training dataset cohort, in order to assess whether there is any correlation between the non-signalling pathway activated transcriptomic subgroups and methylomic subgroups identified in this study. Subsequently, it may be possible to identify and validate protein markers for the routine classification of medulloblastoma transcriptomic subgroup by immunohistochemistry in the routine setting of hospital pathology labs.

One intriguing hypothesis for future work is that, having defined four disease subgroups, there will be additional variability within subgroups that could help identify intra-cohort prognostic markers and also explain intra-subgroup disease aetiology and progression. This remains unexplored at the moment, due to limitations of sample size; once subgroups are defined, sample size within each subgroup is reduced, and to remedy this in future, there will be a need for, in the context of medulloblastoma, extremely large (>500) sample cohorts to provide sufficiently large subgroup numbers for these sorts of intra-group analyses.

Technological advances in the assaying of DNA methylation will also be instrumental in further elucidating the nature of the medulloblastoma methylome. Since the inception of the project, the Golden Gate Cancer Panel I array has been superseded by higher resolution platforms. The Illumina Infinium methylation array (Baker, 2010; Thirlwell et al., 2010), which provides information at 27,000 CpG dinucleotides, with probes present within the promoter regions of 14,475 genes became available in 2009. Early in 2011, the next generation Illumina HD 450k methylation arrays became publicly available (Sandoval et al., 2011). This array measures CpG methylation at > 450,000 CpG dinucleotides, representing CpGs within promoter regions, gene bodies, 3’ UTR and inter-genic regions. Within promoter regions, CpGs located within CpG islands, but also in proximal regions (CpG island shore and shelves) are assayed (Irizarry et al., 2009), enabling a truly comprehensive characterisation of the methylome.

This enables questions to be asked that were not possible with the Golden Gate methylation array; patterns of differential methylation could be subjected to gene ontology and gene set enrichment analysis, to identify co-repressed networks of genes, silenced by methylation. Analysis of medulloblastoma by these 450k arrays can provide a rich source of potential disease biomarkers and, in conjunction with transcriptomic analysis, could for the first time identify relevant genes whose expression is mediated by promoter DNA methylation patterns in medulloblastoma. It will also become possible to identify any large, megabase scale gene silencing through hyper-methylation, leading to a functional LOH, as has previously been described in colorectal cancer cell lines (Frigola et al., 2006). Pilot studies of medulloblastomas on the 450k array have been undertaken in spring 2011, and appear to demonstrate satisfactory reproducibility on DNA extracted from both frozen and FFPE tissues.

In a post-array era, it is likely that whole-genome shotgun bisulfite sequencing (WGSBS) (Laird, 2010) will become the technique of choice for characterisation of DNA methylomes. This technique will provide a truly whole-genome summary of DNA methylation, although there are still some limitations to overcome, due to the reduction in sequence complexity following bisulfite conversion. Nevertheless, a recent study described WGSBS of human peripheral blood mononuclear cells, covering 92% of the genome, indicating that this approach is possible and will likely be more widely adopted in the future (Li et al., 2010).

The widespread availability and adoption of nanopore based sequencing techniques, which are able to directly read 5-methylcytosine (Clarke et al., 2009), while still several years away, will eventually provide a single molecule readout of DNA methylation within a cell, and will provide an unprecedented insight into the heterogeneity of DNA methylation and sequence within tumours. However, from a classification perspective, insights into disease heterogeneity are less important, and the global, genome-wide measure of DNA methylation of a tumour sample provided by the 450k array will be well suited to this task.

The recent description of 5-hydroxymethylcytosine within the primary sequence of DNA derived from brain tissue (Kriaucionis and Heintz, 2009) and the subsequent report that the TET1-mediated hydroxylation of 5-methylcytosine to 5-

hydroxymethylcytosine contributes to the age-related depletion of 5-methylcytosine in the adult brain (Guo et al., 2011), illustrates that the mechanisms and functional consequence of methylation in the brain are still poorly understood, with the tantalising possibility that 5-hydroxymethylcytosine represents an intermediate step in the demethylation of DNA. The simultaneous assay of both 5-methylcytosine and 5-hydroxymethylcytosine in medulloblastoma might further enhance our understanding of the role of DNA methylation in the pathogenesis of brain tumours.

4.5.7 Summary

This chapter has identified and validated the existence of 4 methylomic subgroups of medulloblastoma. While the identified subgroups have striking differences in their clinico-pathological and molecular correlates, it is noticeable that, WNT subgroup apart, membership of which is associated with an improved survival, there is no difference in survival between the remaining subgroups. Since there is no discernible prognostic value to membership of the SHH, group I or group II subgroups, the next chapter investigates the utility of methylomic biomarkers to augment survival models of medulloblastoma, for the generation of novel survival models with an improved performance over the current paradigm for clinical classification.

Chapter 5. Construction of a novel model for disease risk stratification using DNA methylomic biomarkers

5.1 Introduction

Although there has been significant progress in the evaluation of patient risk in medulloblastoma, and a 5 year PFS rate of 70-85% in non-infant cases is now achieved (Tamayo et al., 2011; Packer et al., 2006; Gilbertson, 2004), there remains a large proportion of high risk (~30%) and standard risk patients (~15%) for whom current treatments are ineffective. For infant cases, the survival rate is approximately 30-40% at 5 years, which may be related to the reluctance to administer radiation therapy because of the devastating side effects that this engenders.

Although the survival rates outlined above are improving, medulloblastoma survivors are commonly burdened with deleterious neuro-cognitive and neuro-endocrine *sequelae* (Garre et al., 1994). Efforts are being made to identify patients likely to have a good prognosis, so that their treatment intensity can be reduced, reducing their risk of deleterious *sequelae* whilst still maintaining a cure. Conversely, patients with a poor prognosis can continue to be aggressively treated, with the primary goal being to achieve a cure.

The current risk assignment scheme, which stratifies patients into two risk groups, does not adequately reflect the true heterogeneity of disease risk. Currently, patients positive for M+ disease, residual disease following surgical excision of the tumour or aged under 3 at diagnosis are classified as high risk. Remaining patients are classified as standard risk (Gilbertson, 2004). Large cell / anaplastic histology is becoming more widely accepted as a marker of poor prognosis (Brown et al., 2000) and is becoming incorporated into state of the art classification schemes (Ellison et al., 2011b; Tamayo et al., 2011).

The recognition that WNT subgroup medulloblastomas are associated with an improved prognosis (Clifford et al., 2006; Ellison et al., 2005) has rapidly been incorporated into classification schemes that will be assessed in future clinical trials and demonstrates how molecular classification of the disease can enhance classification (Ellison et al., 2011b; Pizer and Clifford, 2009). The PNET 5 / 6 trials, due to commence in early 2012, are investigating a refined classification scheme, where WNT subgroup membership, assigned by testing for the nuclear accumulation of β-catenin, confers assignment to a low risk group (Pizer and Clifford, 2009). Poor risk

cases are defined by LCA histology, M+ disease, *MYCC / MYCN* amplification or the presence of residual disease. Remaining cases are defined as standard risk (Pizer and Clifford, 2009).

More recently, several novel classification schemes have been proposed that incorporate additional molecular markers to existing, prognostic, clinico-pathological correlates (Ellison et al., 2011b; Tamayo et al., 2011; Pfister et al., 2009; de Haas et al., 2008), summarised in Table 5.1. However, these are limited, either by the complexity of the proposed scheme (Tamayo et al., 2011), or because they were not tested in clinical trials cohorts (Tamayo et al., 2011; Pfister et al., 2009; de Haas et al., 2008).

In addition, differential survival between transcriptomic disease subgroups has been investigated (Cho et al., 2011; Northcott et al., 2010) (summarised in Table 5.1). The major limitations to currently reported survival analyses of medulloblastomas profiled using high-throughput techniques is that they are non-clinical trials based (Cho et al., 2011; Tamayo et al., 2011; Northcott et al., 2010), that subgroup membership was not directly assessed by transcriptomic methods (Northcott et al., 2010) (discussed in section 4.5.5) and that consensus for the number and determinants of subgroups has not yet been reached.

In this chapter, a non-infant trials-based cohort is used as the basis for testing whether there are pleiotropic effects of survival correlates across subgroups. The utility of methylomic subgroups for disease prognostication are assessed, and a novel disease risk stratification scheme is proposed, incorporating novel DNA methylation biomarkers, which out-performs other state of the art classification schemes.

Year	Study	Classifier	Cohort size (*n*)	Clinical trials cohort	Performance	Reference
2011	Predicting relapse in patients with medulloblastoma by integrating evidence from clinical and genomic features	No simple classifier offered. Nomogram for assigning patient risk considered 7 chromosomal abnormalities, six 6-gene signatures, transcriptomic subgroup, M+ disease status, histology, *MYC* activation	Training cohort - 96 Test cohort - 78	N	Optimal model had AUC of 0.87, 0.80 in training / test cohorts respectively	Tamayo et al., 2011
2010	Definition of disease-risk stratification groups in childhood medulloblastoma using combined clinical, pathologic, and molecular variables	Low-risk: b-catenin nucleopositive with absence of M+ disease and / or LCA histology and / or *MYC* amplification High-risk: M+ disease and / or LCA histology and / or *MYC* amplification Standard-risk: All other cases	207	Y - PNET3	5 year EFS: Low-risk - 92% Standard-risk - 77% High-risk - 43%	Ellison et al., 2011b
2009	Outcome prediction in paediatric medulloblastoma based on DNA copy number aberrations of chromosome 6q and 17q and the *MYC* and *MYCN loci*	5 risk categories (in order of increasing risk): 6q loss 6q / 17q balanced 17q gain 6q gain *MYCC / MYCN* amplified	Discovery cohort - 80 Test cohort - 260	N	5 year EFS: 6q loss - 88% 6q / 17q balanced - 73% 17q gain - 48% 6q gain - 19% *MYCC / MYCN* amp - 5%	Pfister et al., 2009
2008	Molecular risk stratification of medulloblastoma patients based on immunohistochemical analysis of MYC, LDHB , and CCNB1 expression	Immunohistochemical scoring: Low -risk: MYC - Standard-risk: MYC+, LDHB / CCNB1 - High-risk: MYC +, LDHB / CCNB1 +	109	N	5 year EFS: Low-risk - 91% Standard-risk - 66% High-risk - 18%	de Haas et al., 2008
2010	Medulloblastoma comprises four distinct molecular variants	Four transcriptomic disease subgroups: WNT, SHH, Group C, Group D	236	N	5 year OS: WNT - 95% SHH - 88% Group C - 35% Group D - 76%	Northcott et al., 2910
2010	Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome	Six transcriptomic disease subgroups: c1, c2, c3 (SHH), c4, c5, c6 (WNT)	115	N	5 year EFS: c1 - 41% c2 - 64% c3 (SHH) - 65% c4 - 77% c5 - 68% c6 (WNT) -82%	Cho et al., 2011

Table 5.1. Recently reported prognostication schemes for medulloblastoma. For each study, the year, title of study, description of disease classifier, cohort size, indication as to whether cohort is comprised of clinical trials cases and an estimate of performance is given. 5 year EFS and OS percentages are estimated from Kaplan-Meier curves if not explicitly stated in article text. The reference for each study is given. The latter two studies, separated by a grey bar, are from transcriptomic analysis of the disease, and only examined the relationships between transcriptomic disease subgroup membership and survival.

5.2 Aims

This chapter aimed to investigate the utility of DNA methylomics for disease subclassification and prognostication in medulloblastoma:

1. To determine whether previously identified survival correlates of medulloblastoma exhibit pleiotropic behaviours across methylomic subgroups of the disease
2. To investigate the utility of methylated loci as biomarkers for disease prognostication.
3. To investigate any identified prognostic methylated loci for their integration into an improved risk stratification scheme for medulloblastoma.

5.3 Materials and methods

5.3.1 PNET3 clinical trials cohort

The survival cohort comprised 136 PNET3 cases (Taylor et al., 2003) (section 2.1) which passed methylation array QC (section 2.8.1). Histopathological subtype was confirmed on review by Professor David Ellison (St Jude Children's Research Hospital, Memphis, TN, USA). Newcastle and North Tyneside Research Ethics Committee approval was obtained for the collection, storage, and biological study of all material. EFS was considered in preference to OS, for the reasons outlined in section 2.13.1.

5.3.1.1 Available clinico-pathological correlates

Data for M+ disease stage, chromosome 17 LOH, *MYCC* / *MYCN* amplification and gender were available, and were measured as described in section 2.1.

5.3.2 Age matched cohort

An additional, 55 member, age-matched (3 – 16 years at diagnosis) cohort with available survival data was appended to the PNET3 cohort described above. This additional cohort comprised cases from the training dataset described in section 2.1. A log-rank test (section 2.13.2) identified any survival difference between the PNET3 and age-matched cohorts. Tests of association (Fisher's exact, chi-squared and *t* tests) were carried out to identify differences in incidence of previously reported clinical correlates.

5.3.3 Validation of known correlates

Log-rank tests were applied to validate previously identified survival factors in medulloblastoma. Tests were applied to M+ disease stage (Chang et al., 1969), LCA histology (Ellison et al., 2011b; Brown et al., 2000), *MYCC / MYCN* amplification (Lamont et al., 2004), gender (Curran et al., 2009) and chromosome 17 LOH (Ellison et al., 2011b; Pan et al., 2005) in both the whole survival cohort and, where possible, among the methylomic subgroups group I, group II and SHH. For many correlates, the numbers of cases positive for the correlate within each subgroup is too low to draw any inference about any subgroup-specific behaviour. To identify subgroups where sample size might be sufficient to detect subgroup-specific correlate effects, power

calculations were performed using the software PS, version 3.0.43 (http://biostat.mc.vanderbilt.edu/PowerSampleSize). Since the WNT and NC groups of samples were associated with a low risk (section 4.5.5), these subgroups were excluded from subgroup-specific analyses of survival correlates because they would necessarily be non-informative, but were included in whole cohort tests. Since information regarding extent of tumour resection was incomplete (information unavailable for 59 / 191 (31%) cases), this was not considered as a prognostic factor in these investigations.

5.3.4 Identification of additional testable markers for assigning disease risk

Potential prognostic methylation markers were first filtered to ensure that only potentially useful markers were tested (section 2.15). Next, the bimodality index (Wang et al., 2009a) for the selected probes was calculated to identify potentially prognostic probes suitable for consideration. As discussed in section 2.16, an ideal methylation prognostic biomarker would have a bimodal distribution with two modes close to zero and one, easily separable in future assays to measure prognostic biomarkers.

5.3.5 Cox boost algorithm for integration of cross-validated high-dimensional data with existing clinical covariates

The Cox boost algorithm (section 2.17) was used to identify additionally prognostic markers on a base model consisting of M+ disease status, LCA histology and *MYCC / MYCN* amplification. After filtering for potentially useable prognostic probes (section 2.15), the 200 most bi-modal methylation probes (section 2.16) plus methylomic subgroup membership were considered as potential additionally prognostic markers on a cohort from which WNT cases had been removed.

The optimal algorithm parameters were calculated. First, the optimal number of boosting steps was calculated using LOOCV. The penalty score was also derived for the optimal number of boosting steps. After running the algorithm, optional covariates with additional prognostic power were identified and formally introduced into a Cox

model which included the base model covariates shown above. 'P' values for the optional covariates are subsequently calculated using LOOCV.

5.3.6 Cox proportional hazards model to identify additional prognostic covariates

A Cox proportional hazards model (section 2.14) was constructed for a model which included LCA histology, M+ disease and *MYCC / MYCN* amplification and the selected optional covariates identified using the Cox boost algorithm (section 5.3.5). The performance of the novel Cox model for predicting survival was compared against currently used survival models (current clinical model – LCA histology, M+ disease stage as markers of poor risk; PNET 5 / 6 clinical trials model – LCA histology, M+ disease stage, *MYCC / MYCN* amplification; N.B. WNT as a favourable prognostic marker was removed, since the cohort, by definition, did not contain WNT cases (described in section 5.3.3)) using ROC curves (section 2.18).

5.3.7 Characterisation of additional covariates

Cox proportional hazards models (section 2.14) for the selected additional covariates were constructed in a univariate and multivariate setting, with a base model constructed from LCA histology, M+ disease stage and *MYCC / MYCN* amplification.

5.3.8 Formulation of a novel medulloblastoma risk stratification scheme

The potential utility of identified prognostic probes was investigated. It is likely that, in a routine hospital laboratory setting, an assay to measure methylation would not report a score at the same level of precision as the methylation array. Rather, it is more likely that the assessment of methylation status will be determined based on a pre-determined cutoff. To test whether the selected methylation covariates were amenable to being classified by cutoff, the methylation data from the selected covariates was binarised at cutoffs (ranging from 0.9 to 0.1 in increments of 0.1) to determine whether binarised data behaved similarly to the continuous variable. Univariate Cox proportional hazards models were constructed at each cutoff and any probe that did not behave similarly to a univariate model using continuous methylation data, at appropriate cutoffs, was rejected.

Bisulfite sequencing-based estimates (section 2.6) of DNA methylation are able to distinguish unmethylated (β score ≤0.33) from hemi-methylated (0.33 < β score ≤ 0.67) from methylated DNA (β score > 0.67). Covariates amenable to classification were assigned to an appropriate cutoff (0.33 or 0.67) and integrated into the model, replacing the previously continuous methylation covariates. Cox proportional hazards models (section 2.14) for the binary-classified markers were constructed in univariate and multivariate analyses. ROC curves (section 2.18) were plotted to assess classifier performance between the Cox model utilising continuous methylation covariates and the Cox model with binary-classified covariates.

Finally, the Cox model comprising the base covariates plus additional methylation-based binary-classified biomarkers was adapted to form the basis of an easily-determined clinical algorithm. It was built upon the PNET 5 / 6 treatment strategy, which is currently being assessed in clinical trials (Pizer and Clifford, 2009). This scheme places cases positive for LCA histology and / or M+ disease and / or *MYCC* or *MYCN* into a poor prognosis, high risk group. Cases positive for nuclear accumulation of *CTNNB1* (equivalent to WNT pathway activation), but negative for any high risk factors are placed into a low risk group. Remaining cases are assigned as standard risk.

Unlike PNET 5 / 6, it was decided that, since the WNT subgroup behaves clinically so differently from other medulloblastomas (section 4.5.5), that membership of this subgroup would, regardless of the status of any other risk factor, confer membership of the low risk group.

A Cox proportional hazards model was constructed using the selected risk factors in the non-WNT survival cohort. A nomogram (section 2.19) was plotted to visualise the magnitude of the included hazard variables and to aid assignment of risk score boundaries.

The novel biomarkers were integrated with the other included covariates to form an additive risk stratification model. For comparison, a classifier derived from CART (classification and regression tree) analysis (section 2.20) was constructed from the non-WNT survival cohort, considering potential high risk variables M + disease, LCA histology, *MYCC / MYCN* amplification, and *MXI1* and *IL8* methylation. This was augmented with WNT subgroup membership to derive a CART-based classifier, which

was compared to the additive risk model by plotting Kaplan-Meier curves (section 2.13.2) and ROC curves (section 2.18). In the same way, the additive classifier was subsequently compared with current disease classification models and recently reported classification schemes.

5.4 Results

The methylomic subgroups identified in chapter 4 were used to investigate the prognostic potential for methylomic classification of the disease. Following cohort validation, any subgroup specific, pleiotropic effects of previously described risk markers of medulloblastoma were investigated. Subsequently, a novel classification scheme for medulloblastoma prognostication was investigated, adding methylation markers to augment the current clinical risk stratification schemes.

5.4.1 The age-matched cohort is not significantly different from the PNET3 cohort

The survival characteristics of the age matched cohort (*n* = 55) were compared against the survival of the PNET3 cohort samples (*n* = 136) in order to verify their suitability for adding to the PNET3 cohort. No difference in survival was found (Figure 5.1). Similarly, no difference in any survival correlate tested was observed (Table 5.2).

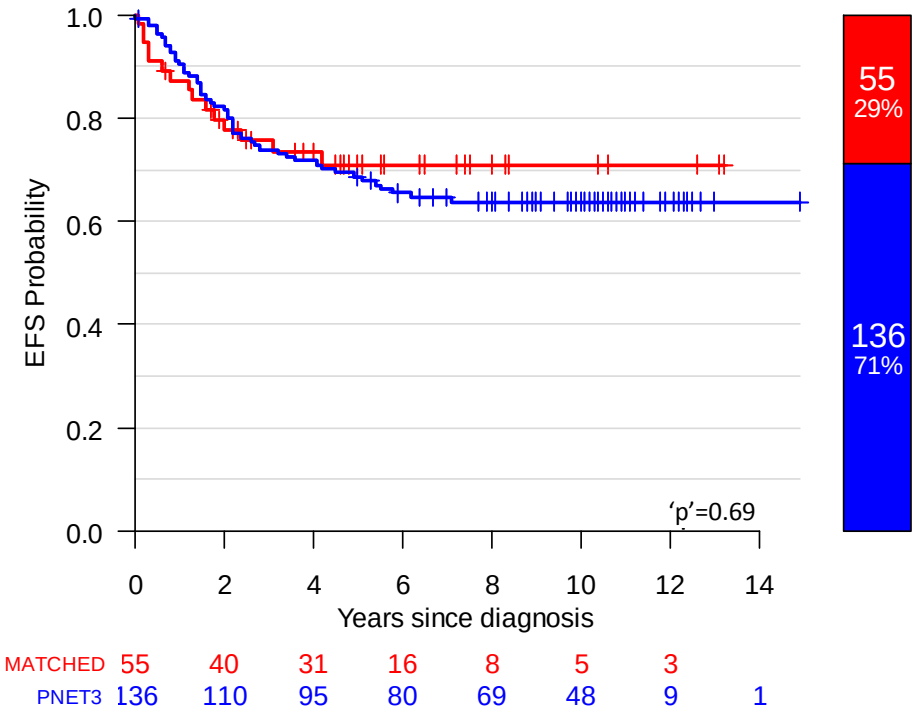


Figure 5.1. No survival differences observed between the PNET3 trials cohort (*n* = 136; PNET3; blue) and age-matched non-trials samples (*n* = 55; MATCHED; red). At-risk table is shown below plot, and log-rank test 'p' value is shown at bottom right of plot.

Demographic	Cohort		'p' value
	PNET3	Age Matched	
Gender			1
Male	80 (59%)	33 (60%)	
Female (M:F ratio)	56 (41%) (1.4:1)	22 (40%) (1.5:1)	
Age in years: median (range)	8.45 (3.1 – 15.6)	8.6 (3.2 – 15.1)	0.29
Histological subtype			0.33
Classic	115 (85%)	42 (76%)	
Desmoplastic / nodular	9 (7%)	7 (13%)	
Large cell / anaplastic	12 (9%)	6 (11%)	
M Stage			0.69
M-	111 (82%)	43 (78%)	
M+	25 (18%)	12 (22%)	
NA	0 (0%)	0 (%)	
***CTNNB1* mutation**			1
0 – no - negative	118 (87%)	48 (87%)	
1 – yes - positive	13 (10%)	5 (9%)	
NA	5 (4%)	2 (4%)	
Chromosome 17 LOH			0.41
0 – no - negative	106 (78%)	15 (27%)	
1 – yes - positive	28 (21%)	6 (11%)	
NA	2 (1%)	34 (62%)	
***MYCC* amplification**			0.33
0 – no - negative	134 (99%)	53 (96%)	
1 – yes - positive	2 (1%)	2 (4%)	
NA	0 (0%)	0 (%)	
***MYCN* amplification**			1
0 – no - negative	131 (96%)	53 (96%)	
1 – yes - positive	5 (4%)	2 (4%)	
NA	0 (0%)	0 (0%)	

Table 5.2. No difference between survival correlates in PNET3 and age matched cohort. For each cohort, the number and percentage in parentheses for each potential survival correlate (excluding age range) are shown. NA indicates missing data. Data for gender, histological subtype, age, M stage, *CTNNB1* mutation, chromosome 17 LOH, *MYCC* amplification, and *MYCN* amplification is shown. The 'p' value shows results from testing for differences in correlate occurrence using Fisher's Exact test (gender, M stage, *CTNNB1* mutation, Chromosome 17 LOH, *MYCC* amplification and *MYCN* amplification), chi-squared test (histological subtype) and *t* test (age).

5.4.2 Validation of previously reported clinico-pathological correlates

The behaviour of previously reported prognostic markers in the joined, 191 member cohort, consisting of cases from the PNET3 trials cohort (*n* = 136) and age-matched non-trials cohort cases (*n* = 55), was investigated. Log-rank tests identified significant relationships between survival and large cell / anaplastic histology ('p' = 2×10^{-5}), M+ disease stage ('p' = 2.8×10^{-6}), gender ('p' = 0.013) , *MYCC / MYCN* amplification ('p' = 0.013) and membership of the WNT methylomic subgroup (data was complete for this measure of WNT pathway activation, in contrast to *CTNNB1* mutation, missing in 7 cases, 'p' = 0.00142). No significant relationship between chromosome 17 LOH and survival was observed ('p' = 0.34). These relationships are shown in Figure 5.2. A multivariate Cox proportional hazards model of the significant variables demonstrates that these covariates are all independently prognostic (Table 5.3).

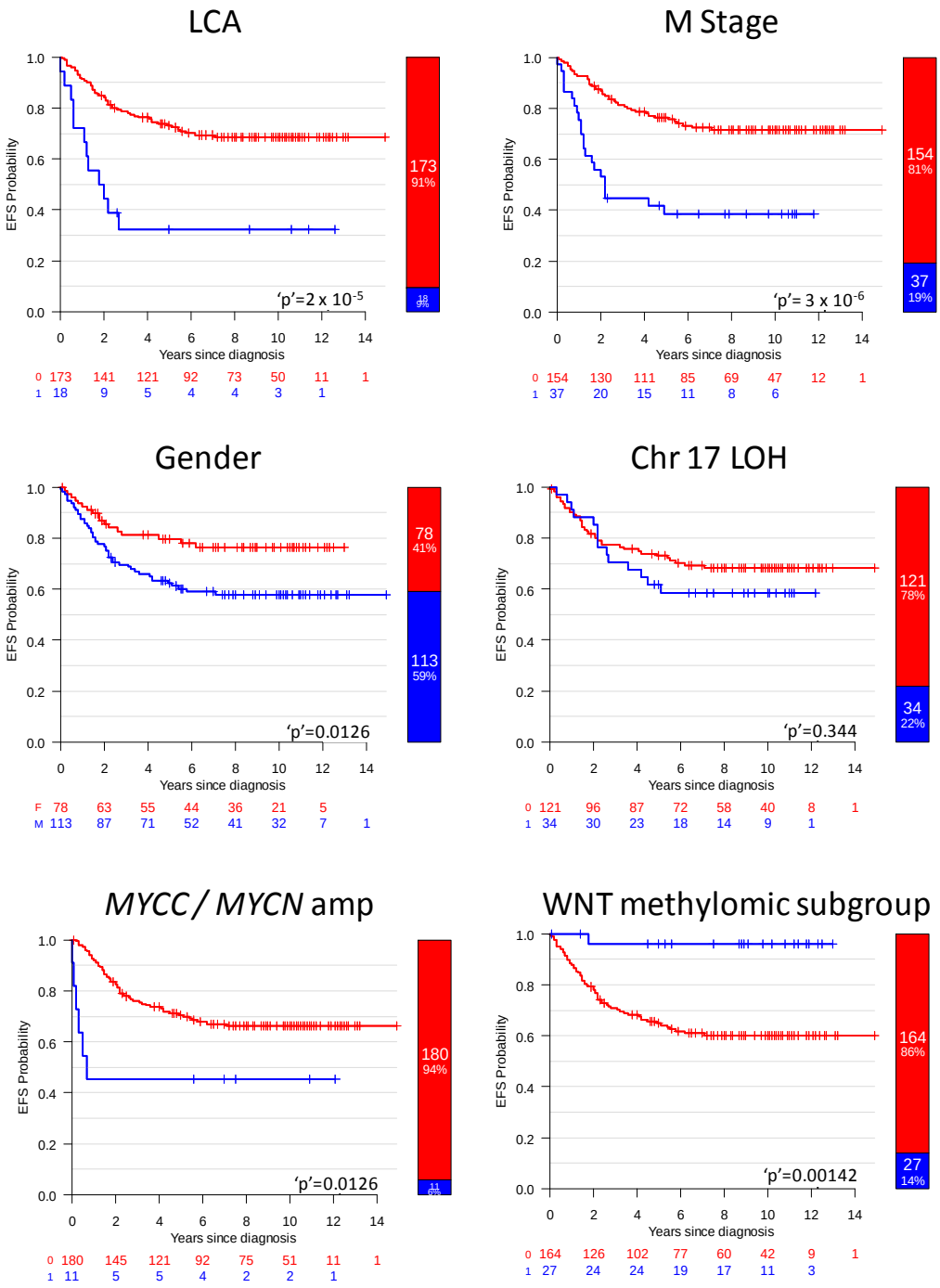


Figure 5.2. Significant relationships were observed for previously reported disease prognostic markers in the combined cohort. Kaplan-Meier plots are shown for large cell / anaplastic histology (LCA) (0 (red) , LCA negative; 1 (blue), LCA positive), M Stage (M+ disease stage) (0 (red), M-; 1 (blue) ,M+), gender (F (red) – female; M (blue) – male), chromosome 17 loss of heterozygosity (Chr 17 LOH) (0 (red), chr 17 LOH negative; 1 (blue), chr 17 LOH positive), *MYCC / MYCN* amplification (*MYCC / MYCN* amp) (0 (red) – no *MYCC / MYCN* amplification detected; 1 (blue) – *MYCC / MYCN* amplification detected) and WNT subgroup membership (0 (red) – not a member of WNT subgroup; 1 (blue) – member of WNT subgroup). At-risk tables and 'p' values from log-rank tests are shown.

Covariate	Hazard Ratio (95% CI)	'p' value
LCA histology	4.33 (2.24 – 8.34)	1.2×10^{-5}
M+ disease	2.73 (1.61 – 4.63)	0.00019
MYCC / MYCN amplification	3.34 (1.39 – 8.05)	0.0072
Male gender	2.14 (1.20 – 3.83)	0.010
WNT methylomic subgroup	0.11 (0.015 – 0.77)	0.026

Table 5.3. Previously reported prognostic markers are independently prognostic in the combined cohort. Table shows a multivariate Cox proportional hazards model that incorporated all previously reported survival correlates that were significant on logrank test. For each covariate, hazard ratio (95% confidence interval in parentheses) and 'p' value are shown.

5.4.3 Validation of current clinical model and PNET5 / 6 clinical trials model

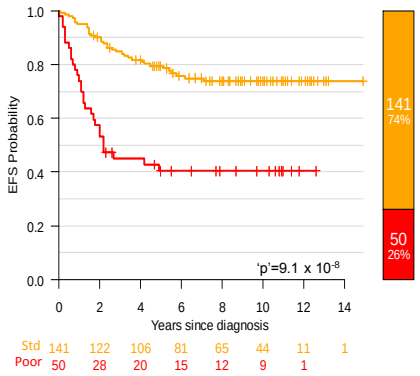
Before any investigation of novel risk stratification models, it was necessary to investigate the prognostic potential of existing classification schemes in the investigation survival cohort. The current clinical classification scheme, as applied to non-infant cases, classifies patients positive for LCA histology and / or M+ disease into a poor risk subgroup. In a Cox proportional hazards model, both LCA histology and M+ disease were strong predictors of adverse outcome (LCA histology, HR = 3.57, 95% CI 1.89 – 6.77, 'p' = 9.3×10^{-5}; M+ disease, HR = 3.25, 95% CI 1.92 - 5.48, 'p' = 1.1×10^{-5}) and identified a standard risk group with a 5 year EFS of 78% and a poor risk group with a 5 year EFS of 40% (Figure 5.3A and B).

The PNET 5 / 6 clinical trials model (Figure 5.3C and D) classifies patients positive for any of LCA histology, M+ disease or *MYCC / MYCN* amplification into a poor risk subgroup, while patients positive for WNT subgroup membership, without any poor risk correlates, are classified as low risk. Remaining cases are classed as standard risk. In the investigation cohort, Cox proportional hazards models showed that all three high risk disease correlates were strongly predictive of disease outcome (LCA histology, HR = 3.84, 95% CI 2.01 – 7.32, 'p' = 4.5×10^{-5}; M+ disease, HR = 2.90, 95% CI 1.72 – 4.90,

'p' = 7.0 x 10^{-5}; *MYCC / MYCN* amplification, HR = 2.44, 95% CI 1.04 – 5.71, 'p' = 0.0039) and that membership of the WNT subgroup conferred an improved survival (HR 0.09, 95% CI 0.01 – 0.68, 'p' = 0.019). Kaplan-Meier curves showed 5 year EFS of 95%, 76% and 42% for the low, standard and poor risk disease stratification groups, respectively.

The current clinical model and the PNET 5 / 6 clinical trials models have significant prognostic potential in the investigation cohort. The addition of WNT subgroup membership to the PNET 5 / 6 stratification model enables a new low risk category, defined by WNT status, to be identified. Previous investigations in chapter 4 revealed that the WNT methylomic subgroup of the disease is associated with an improved prognosis (Figure 4.10), with no survival differences apparent for the remaining subgroups. This emphasises the need for the identification of additional prognostic biomarkers for non-WNT medulloblastomas, and is the approach taken in the remainder of this chapter (section 5.4.5).

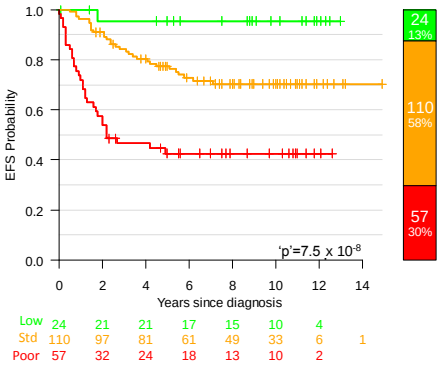
A



B

Covariate	Hazard Ratio (95% CI)	'p' value
LCA histology	3.57 (1.89 – 6.77)	9.3 x 10^{-5}
M+ disease	3.25 (1.92 – 5.48)	1.1 x 10^{-5}

C



D

Covariate	Hazard Ratio (95% CI)	'p' value
LCA histology	3.84 (2.01 – 7.32)	4.5 x 10^{-5}
M+ disease	2.90 (1.72 – 4.90)	7.0 x 10^{-5}
MYCC / MYCN amplification	2.44 (1.04 – 5.71)	0.039
WNT subgroup membership	0.09 (0.01 – 0.68)	0.019

Figure 5.3. **Both the current clinical model and the PNET 5 / 6 trials model for disease stratification are prognostic in the investigation cohort**. **A**. Kaplan-Meier plot of current clinical model applied to investigation cohort (Std – standard risk (orange) – negative for LCA histology and M+ disease; Poor – poor risk (red) – positive for LCA histology and / or M+ disease). At-risk table and 'p' values from log-rank tests are shown. **B**. Cox proportional hazards model of clinical covariates considered in current clinical risk stratification. For each covariate, hazard ratio (95% confidence interval in parentheses) and 'p' value are shown. **C**. Kaplan-Meier plot of PNET 5 / 6 clinical model applied to investigation cohort (Low – low risk (green) – WNT positive cases negative for LCA histology, M+ disease and *MYCC / MYCN* amplification; Std – standard risk (orange) – cases negative for WNT activation, LCA histology, M+ disease and *MYCC / MYCN* amplification; Poor – poor risk (red) – positive for LCA histology and / or M+ disease and / or *MYCC / MYCN* amplification). At-risk table and 'p' values from log-rank tests are shown. **D**. Cox proportional hazards model of clinical covariates considered in current clinical risk stratification. For each covariate, hazard ratio (95% confidence interval in parentheses) and 'p' value are shown.

5.4.4 Investigation into the feasibility of identifying pleiotropic effects of previously identified survival correlates between methylomic subgroups

The risk factors tested across the whole cohort in section 5.4.2 were next investigated for potential pleiotropic effects between disease subgroups. Power calculations, using the observed subgroup correlate incidence, showed that for all but one combination of subgroup and correlate (gender in group I subgroup cases), any investigation into any pleiotropic effect of correlates between subgroups would be under-powered (i.e. power less than 0.80) and would not be worth pursuing in this cohort (Table 5.4); moreover, since the one comparison that was adequately powered did not have an adequately powered comparator subgroup, any inference about the nature of a pleiotropic effect would not be possible.

Risk Factor	Methylomic Subgroup					
	SHH		Group I		Group II	
	Incidence	Power	Incidence	Power	Incidence	Power
LCA histology 0 1	31 (86%) 5 (14%)	0.28	75 (95%) 4 (5%)	0.25	29 (81%) 7 (19%)	0.35
M Stage 0 1	32 (89%) 4 (11%)	0.24	62 (78%) 17 (22%)	0.67	24 (67%) 12 (33%)	0.46
Chromosome 17 LOH 0 1	29 (94%) 2 (6%)	0.15	38 (61%) 24 (39%)	0.71	23 (85%) 4 (15%)	0.23
MYCC / MYCN amplification 0 1	31 (86%) 5 (14%)	0.28	77 (97%) 2 (3%)	0.15	33 (92%) 3 (8%)	0.19
Gender M F	17 (47%) 19 (53%)	0.53	49 (62%) 30 (38%)	0.81	28 (78%) 8 (22%)	0.38

Table 5.4. Investigation into the feasibility of testing for pleiotropic effects of prognostic markers across methylomic disease subgroups. Power to detect pleiotropic effects for the methylomic subgroups SHH, group I and group II were calculated from the observed incidence of each risk factor (0 – absence of risk factor; 1 – presence of risk factor). The incidence and power for each comparison is shown.

5.4.4.1 MYCC / MYCN amplification

While this study was under-powered for the identification of subgroup specific correlate effects (Table 5.4), the differential subgroup specific survival effects observed for *MYCC / MYCN* amplification, while anecdotal, bear further scrutiny. *MYCC / MYCN* amplification was relatively rare in the whole cohort (11 / 191 cases (6%)), but there were striking differences in outcome between the subgroups. *MYCC / MYCN* amplification positive cases in group I (2 / 79 cases (3%), 'p' = 0.30) and group II (3 / 36 cases (8%), 'p' = 0.81) did not show significant survival differences, whereas there was a dismal prognosis for *MYCC / MYCN* amplification in SHH cases (5 / 36 cases (%), 'p' = 1.3×10^{-11}, log-rank test)(Figure 5.4).

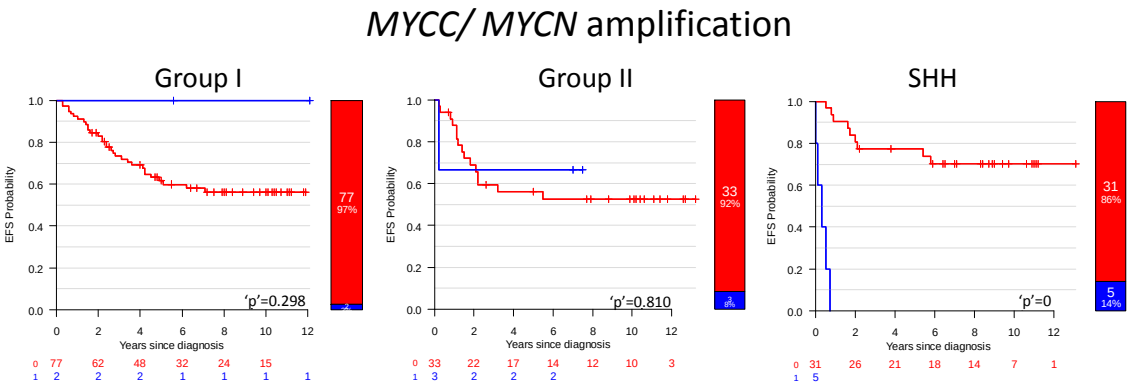


Figure 5.4. *MYCC / MYCN* amplified cases display striking differences in clinical behaviours across disease subgroups. Kaplan-Meier plots are shown for *MYCC / MYCN* amplification (*MYCC / MYCN* amp) (0 (red) – no *MYCC / MYCN* amplification detected; 1 (blue) – *MYCC / MYCN* amplification detected). At-risk tables and 'p' values from log-rank tests are shown.

5.4.5 Integration of methylomic correlates with previously reported survival markers

The Cox boost algorithm (Binder et al., 2009) was next used to identify significant methylation markers or subgroup membership with additional prognostic potential, in addition to the current clinical risk stratification strategy (section 2.17). This algorithm employs a boosting approach to avoid overfitting (section 2.11), generating survival models which include mandatory clinical covariates, plus a small number of prognostic high-dimensional covariates, limited by the addition of a penalty score for every additional term included in the model. Cross-validated 'p' values for high-dimensional covariates are subsequently calculated using LOOCV.

WNT cases were removed from the cohort prior to running the algorithm, since there were so few WNT cases that relapsed (1 / 28 WNT cases). The first stages of the model identified optimal parameter sizes. The cross-validated, optimal number of boosting steps was 17. The optimal penalty parameter was 13000. The algorithm identified three additionally prognostic optional covariates (cross-validated 'p' values are shown in parentheses): IL8_P83_F ('p' = 0.027), MXI1_P1269_F ('p' = 0.00065), and POMC_P400_R ('p' = 0.021). Methylomic subgroup membership was not prognostic.

5.4.5.1 Assessment of suitability of identified methylation probes

The suitability of the three selected methylation probes was assessed. First, their prognostic potential in univariate and multivariate Cox models was examined (Table 5.5). All three putative probes were significant ('p' < 0.05) in both univariate and multivariate analyses. Next, the methylation probes were considered as binary variables at a range of cutoffs (section 5.3.8). Since no cutoff was significant for the POMC_P400_R probe, and it was only marginally significant in a multivariate analysis ('p' = 0.047), it was removed from further consideration (Table 5.6).

5.4.5.2 Selected methylation probes show subgroup specific patterns of methylation

The distribution of β scores for the selected probes, MXI1_P1269_F and IL8_P83_F, were assessed by plotting combined boxplots and strip-plots (Figure 5.6). The WNT and SHH subgroups show a methylated MXI1_P1269_F β score, with the remaining

subgroups (plus the NC cases) showing a mixed methylation profile. IL8_P83_F also had a mixed methylation profile, with group I and group II cases being predominantly methylated, and SHH and WNT cases showing a mixed methylation profile.

5.4.6 An improved survival model for medulloblastoma incorporating continuous methylomic markers

A finalised Cox proportional hazards model for non-WNT medulloblastomas is shown in Table 5.7. It includes existing covariates (LCA histology, M+ disease, *MYCC / MYCN* amplification, section 5.4.3) and two methylation probes (MXI1_P1269_F, IL8_P83_F), selected on the basis of (i) their prognostic potential, assessed by the Cox boost algorithm (section 2.17) and (ii) their suitability for assay using alternative techniques (see section 5.3.8, 5.4.7.1). Gender was not considered as a prognostic covariate, since the prognostic role of gender is unclear in the disease (Ellison et al., 2011b; Curran et al., 2009; Alston et al., 2003; Weil et al., 1998). Each covariate in the survival model is independently significant (Table 5.7). The performance of the extended model was assessed by plotting ROC curves (section 2.18), shown in Figure 5.5. This demonstrates the improved performance of this model at both 5 and 10 years relative to Cox proportional hazards models derived from the PNET 5 / 6 clinical trials model (omitting WNT membership, since this is not relevant to a non-WNT survival cohort), and current clinical model (section 5.4.3), with an improvement in AUC of 0.060 and 0.069 at 5 years and 0.085 and 0.090 at 10 years, respectively.

Gene	Probe	Univariate		Multivariate	
		Hazard Ratio (95% CI)	'P' value	Hazard Ratio (95% CI)	'P' value
MXI1	MXI1_P1269_F	3.29 (1.30 – 8.29)	0.012	4.92 (1.90 – 12.74)	0.0011
IL8	IL8_P83_F	8.22 (1.46 – 46.37)	0.017	6.44 (1.17 – 35.59)	0.033
POMC	POMC_P400_R	0.15 (0.02 – 0.86)	0.034	0.17 (0.03 – 0.97)	0.047

Table 5.5. Assessment of putative prognostic methylation markers as continuous variables in univariate and multivariate analyses of the non-WNT subgroup survival cohort. For each methylation covariate, hazard ratios (95% confidence intervals in parentheses) and 'p' value, uncorrected for multiple testing, are shown for univariate and multivariate analyses (adding to a base model that included LCA histology, M+ status and *MYCC / MYCN* amplification as covariates).

Gene	*MXI1*		*IL8*		*POMC*	
Probe	MXI1_P1269_F		IL8_P83_F		POMC_P400_R	
Cutoff	Hazard Ratio (95% CI)	'P' value	Hazard Ratio (95% CI)	'P' value	Hazard Ratio (95% CI)	'P' value
Continuous	3.29 (1.30 – 8.29)	0.017	8.22 (1.46 – 46.37)	0.017	0.15 (0.02 – 0.86)	0.034
0.9	1.18 (0.68 – 2.06)	0.559	3.60 (1.55 – 8.38)	0.0029	0 (0 – inf)	1
0.8	1.87 (1.14 – 3.09)	0.014	3.66 (1.33 – 10.01)	0.012	0 (0 – inf)	1
0.7	2.57 (1.47-4.50)	0.00092	3.22 (1.17 – 8.88)	0.024	0 (0 – inf)	1
0.6	2.65 (1.41 – 4.98)	0.00247	2.54 (0.92 – 6.99)	0.072	0 (0 – inf)	1
0.5	1.87 (1.00 – 3.51)	0.0516	2.61 (0.82 – 8.33)	0.105	0.15 (0.02 – 1.11)	0.064
0.4	1.90 (0.96 – 3.73)	0.0643	3.41 (0.83 – 13.97)	0.088	0.25 (0.07 – 1.04)	0.057
0.3	2.08 (0.95 – 4.57)	0.0677	1.95 (0.48 – 7.97)	0.35	0.74 (0.33 – 1.63)	0.45
0.2	2.04 (0.88 – 4.73)	0.0973	3.15 (0.44 – 22.72)	0.26	0.60 (0.30 – 1.22)	0.16
0.1	2.12 (0.77 – 5.84)	0.146	NA	NA	0.59 (0.31 – 1.11)	0.10

Table 5.6. Consideration of the prognostic potential for methylation probes as binary variables at variable cutoffs. For each methylation probe, the methylation score was converted into a binary variable using the cutoffs listed in the first column. A univariate Cox proportional hazards model was constructed for each binary variable in the non-WNT survival cohort. For comparison, univariate Cox proportional hazards models are shown for the methylation scores as a continuous variable. The hazard ratio (95% confidence intervals in parentheses) and 'p' value at each cutoff are shown. Where confidence intervals are reported as '0 – inf', this represents a situation in which it was not possible to assign confidence intervals.

	Univariate		Multivariate	
Covariate	Hazard ratio (95% CI)	'P' value	Hazard ratio (95% CI)	'P' value
Large cell / anaplastic	3.90 (2.06 – 7.38)	2.86×10^{-5}	3.35 (1.74 – 6.45)	0.00030
M+ disease	3.03 (1.80 – 5.12)	3.31×10^{-5}	3.57 (2.07 – 6.15)	5.0×10^{-6}
MYCC / MYCN amplification	2.33 (1.00 – 5.42)	0.049	3.15 (1.32 – 7.50)	0.0095
Continuous *MXI1* methylation	3.29 (1.30 – 8.29)	0.012	6.39 (2.35 – 17.4)	0.00028
Continuous *IL8* methylation	8.22 (1.46 – 46.37)	0.017	9.59 (1.67 – 55.09)	0.011

Table 5.7. An extended Cox proportional hazards survival model for medulloblastoma incorporates continuous methylation markers. For each selected covariate, hazard ratio (95% confidence intervals in parentheses) and 'p' value are shown for both univariate and multivariate analyses. All selected covariates are significant in univariate analyses and independently significant in multivariate analyses.

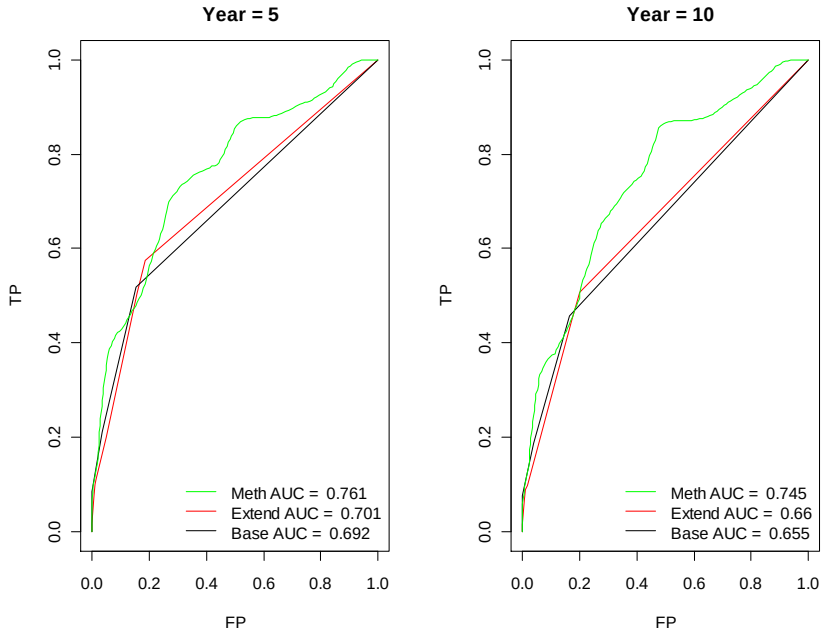


Figure 5.5. ROC curves demonstrate improved performance in a non-WNT survival cohort of novel Cox proportional hazards model that includes methylation markers as continuous variables at 5 and 10 years. ROC curves are shown for Cox proportional hazards models derived from base model covariates (Base - variables LCA histology and M+ disease, shown black), extended PNET 5 / 6 clinical model covariates (section 5.4.3) (Extend – PNET 5 / 6 clinical model - variables LCA histology, M+ disease and *MYCC / MYCN* amplification, shown red) and methylation augmented model (Meth – variables LCA histology, M+ disease, *MYCC / MYCN* amplification, *MXI1* methylation and *IL8* methylation, shown green). The AUC is shown for each model. FP – false positive rate, x axis; TP – true positive rate, y axis.

5.4.7 An improved model for prediction of survival in medulloblastoma, using cutoffs to assign DNA methylation biomarker status

The Cox model described in section 5.4.6 demonstrates that the addition of methylation markers to existing prognostic markers can add additional predictive value to survival models in medulloblastoma. As discussed in section 5.3.8, for a clinically useful test, a methylation score defined by a cutoff, comparable to the assessment of methylation status by bisulfite sequencing, is necessary. Realistically, bisulfite sequencing is able to assign a CpG locus to methylated, hemi-methylated and unmethylated classes (in terms of β score, this is equivalent to unmethylated – $\beta \leq 0.33$; hemi-methylated – $0.33 < \beta \leq 0.67$; methylated – $\beta > 0.67$).

5.4.7.1 Selection of a cutoff for binary classification of methylation

Table 5.6 shows that a cutoff of 0.67 would appear to be satisfactory for both MXI1_P1269_F and IL8_P83_F probes, since MXI1_P1269_F was significant at cutoffs of 0.6 ('p' = 0.0025) and 0.7 ('p' = 0.00092), and IL8_P83_F was significant at a cutoff of 0.7 ('p' = 0.024), and marginally significant at a cutoff of 0.6 ('p' = 0.072). The distribution of MXI1_P1269_F and IL8_P83_F methylation in the non-WNT survival cohort is shown in Figure 5.6. Log-rank tests applied to this binary classified methylation data at a cutoff of 0.67 (Figure 5.7), demonstrated significant differences in survival. 66 / 163 (40%) cases were unmethylated for *MXI1*, associated with an improved survival ('p' = 0.00098, log-rank test). 26 / 163 (16%) cases were unmethylated for *IL8*, also associated with an improved survival ('p' = 0.017, log-rank test). On this basis, subsequent investigations were carried out using binary classified methylation scores at a cutoff of 0.67.

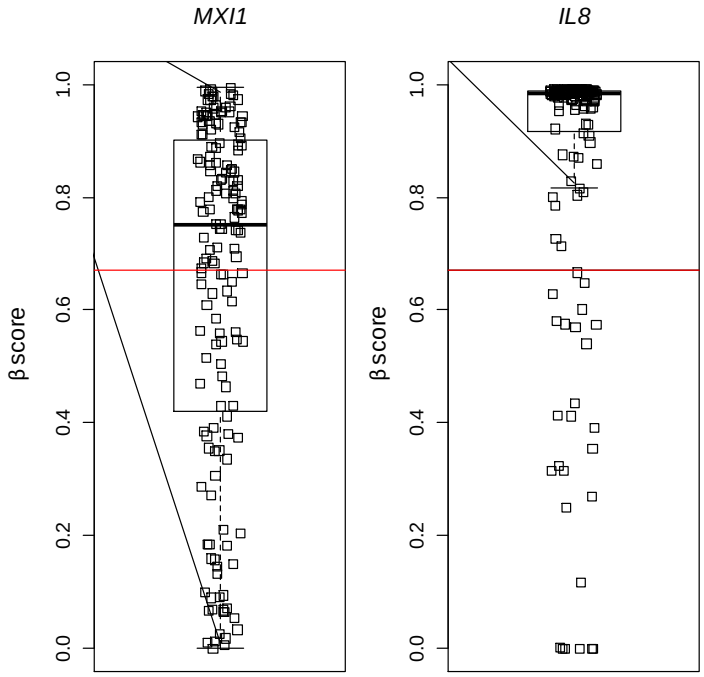


Figure 5.6. Distribution of β scores for selected prognostic methylation probes. Strip-plots show distribution of β scores across non-WNT samples. Boxplots show median score (thick black line) and inter-quartile ranges (extent of box). The whisker (dotted line) shows the lowest and highest data-points that lie within 1.5 inter-quartile ranges of the lower and upper quartile, respectively. The selected β score cutoff of 0.67 is shown for both selected loci as a red line.

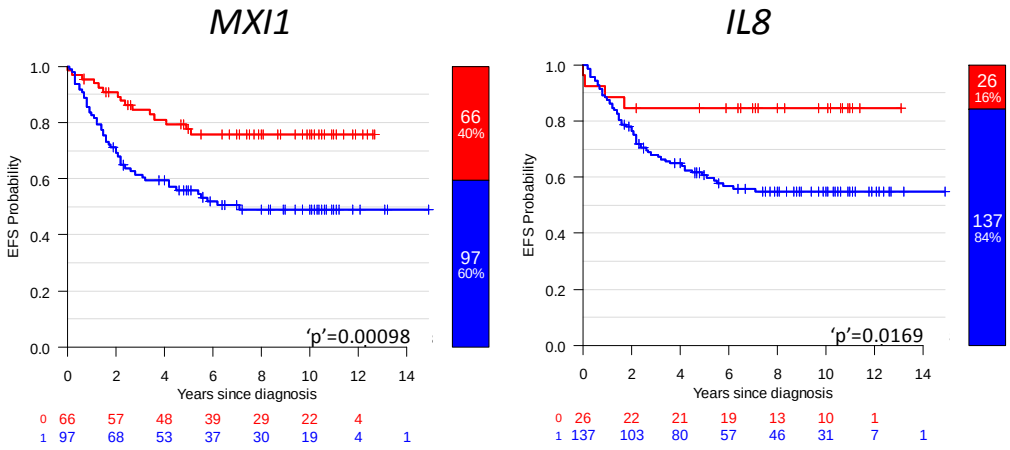


Figure 5.7. Binary classification of methylation markers recapitulates significant survival differences. The significant relationships observed here reiterate the significant relationships between methylation and survival when considered as continuous variables (Table 5.5). Kaplan-Meier plots are shown for *MXI1* and *IL8* methylation at a cutoff of 0.67 (0 (red) – methylation < 0.67; 1 (blue) – methylation ≥ 0.67). At-risk tables and 'p' values from log-rank tests are shown.

5.4.7.2 Cox model for medulloblastoma survival using binary-classified methylated probes is equivalent to model based on continuous methylation variables

The replacement of the β score for *MXI1* and *IL8* methylation with a binary classification at a 0.67 cutoff in the original Cox model shown in Table 5.7 is demonstrated in Table 5.8. Each covariate remains significant with this binary replacement, demonstrating that the integrity of the Cox model remains intact. Interestingly, the hazard ratios for all 5 covariates were similar in magnitude. The Cox model with binary methylated variables (binary) performs equivalently to the counterpart Cox model with continuous methylated variables (continuous) (Figure 5.8). At 5 years, there is very little difference between the AUC from continuous (AUC = 0.761) and binary (AUC = 0.767) Cox models. At 10 years, the binary model (AUC = 0.763) slightly outperforms the continuous model (AUC = 0.745).

	Univariate		Multivariate	
Covariate	Hazard ratio (95% CI)	'P' value	Hazard ratio (95% CI)	'P' value
Large cell / anaplastic	3.90 (2.06 – 7.38)	2.86×10^{-5}	4.09 (2.11 – 7.93)	2.9×10^{-5}
M+ disease	3.03 (1.80 – 5.12)	3.31×10^{-5}	3.39 (1.98 – 5.81)	8.2×10^{-6}
MYCC / MYCN amplification	2.33 (1.00 – 5.42)	0.049	2.79 (1.18 – 6.55)	0.019
Binary *MXI1* methylation	2.57 (1.44 – 4.60)	0.0015	3.48 (1.92 – 6.31)	4.1×10^{-5}
Binary *IL8* methylation	3.22 (1.17 – 8.88)	0.024	3.35 (1.20 – 9.37)	0.021

Table 5.8. Binary classification of methylation probes does not substantively change the previously identified Cox model. For each selected covariate, hazard ratio (95% confidence intervals in parentheses) and 'p' value are shown in both univariate and multivariate analyses.

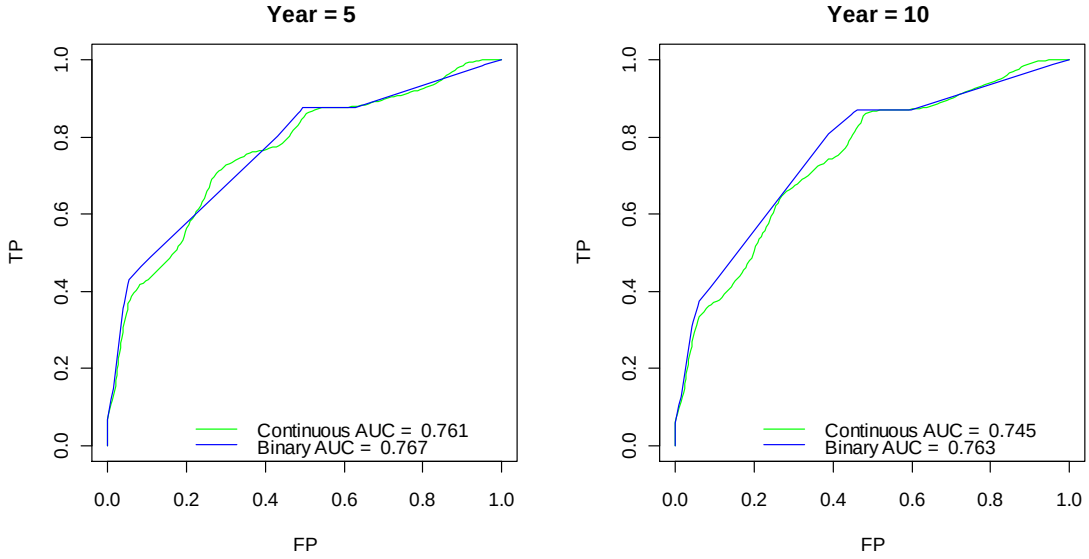


Figure 5.8. Binary classification of *MXI1* and *IL8* methylation does not affect classifier performance at 5 or 10 years. ROC curves demonstrate equivalent performance of Cox proportional hazards model applied to the non-WNT survival cohort that includes continuous or binary-classified methylation markers. ROC curves are shown for binary model (Binary - variables LCA histology, M+ disease, *MYCC / MYCN* amplification, binary *MXI1* and *IL8* methylation, shown blue) and continuous methylation model (Continuous – variables LCA histology, M+ disease, *MYCC / MYCN* amplification, *MXI1* and *IL8* methylation, shown green). The area under the curve (AUC) is shown for each model. FP – false positive rate, x axis; TP – true positive rate, y axis.

5.4.7.3 Formulation of a testable survival model for assigning patient risk in medulloblastoma

An extended classification scheme for the assignment of disease risk in medulloblastoma was formulated. It built upon the PNET 5 / 6 treatment strategy, which is about to enter trials (Pizer and Clifford, 2009). This scheme places cases positive for large cell / anaplastic histology and / or M+ disease and / or *MYCC* or *MYCN* amplification into a poor prognosis, high risk group. Cases positive for nuclear accumulation of *CTNNB1* (equivalent to WNT pathway activation), but without any high risk features, are placed into a low risk group. Remaining cases are assigned as standard risk.

Unlike PNET 5 / 6, it was decided that, since the WNT methylomic subgroup behaves clinically so differently from other medulloblastomas (section 4.5.5), that membership of this subgroup would, regardless of the status of any other risk factor, confer membership of the low risk group. The Cox model shown in section 5.4.7.2 was used as the basis for the remainder of the classifier, comprising the high risk features, LCA histology, M+ disease, *MYCC* or *MYCN* amplification, *MXI1* and *IL8* methylation status.

The Cox proportional hazards model described in Table 5.8 demonstrated that the hazard ratios for all covariates were broadly similar, recapitulated by a nomogram (section 2.19) of hazards in the non-WNT survival cohort (Figure 5.9), which demonstrates approximately equal points scores for each risk factor considered. Risk stratification has been defined according to the total points score from combinations of risk factor positivity, which is, in turn, consistent with an additive disease risk model, in which each risk factor is considered with an equal weighting.

In stratification models, it was noted that survival decreased in proportion to the total number of independently significant high risk features present (Figure 5.10). In order to simplify a classification scheme based on the number of risk factors, non-WNT cases with 0 or 1 risk factors were, like WNT cases, assigned to the low risk group. Non-WNT cases with 2 risk factors were assigned to standard risk group. Finally, non-WNT cases with greater than 2 risk factors were assigned to the poor risk group. This classification scheme is summarised and compared with previous disease risk stratification schemes in Figure 5.12. Most importantly, this scheme would classify an additional 64 patients

as low risk compared to the PNET 5 / 6 clinical trials classification scheme, representing an increase of 267% in low risk patients, without changing event-free survival (PNET 5 / 6 classification – low risk 95% EFS, 24 patients; novel methylation scheme – low risk 90% EFS, 88 patients). While there is a concomitant reduction in EFS for standard risk and poor risk disease under the novel methylation classification scheme, this is due to the large numbers of good prognosis samples being assigned to the low risk disease group.

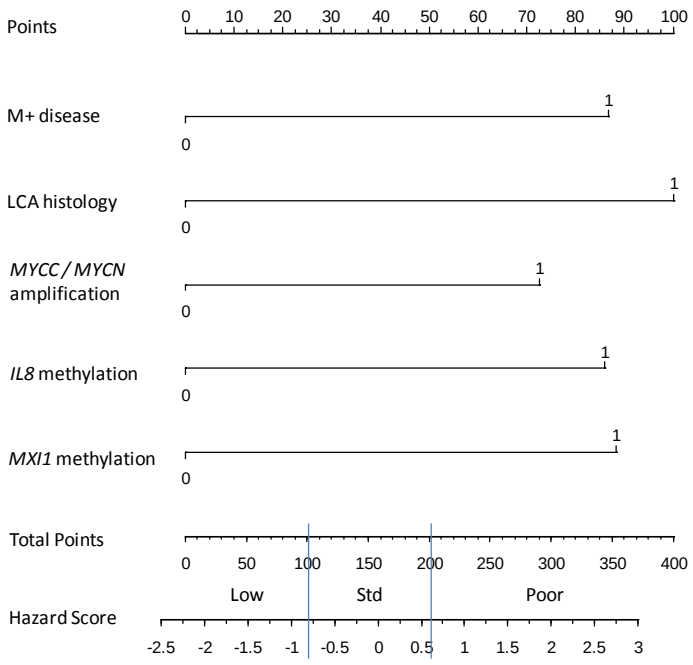


Figure 5.9. Nomogram of risk factors in Cox proportional hazards model derived from non-WNT survival cohort (*n* = 163, Table 5.8), demonstrates similar magnitudes of hazards. Risk boundaries are shown (Low – low risk; Std – standard risk; Poor – poor risk), defined by the total number of points conferred by risk factor positivity and delineated by blue lines. In the illustrated stratification scheme, the absence of any risk factor or a single risk factor would confer membership of the low risk group. Any combination of two risk factors would confer membership of the standard risk group and cases with three or more risk factors would be classified as poor risk. This validates the utility of the risk factors as equally predictive risk markers in an additive model of risk stratification, shown in Figure 5.10.

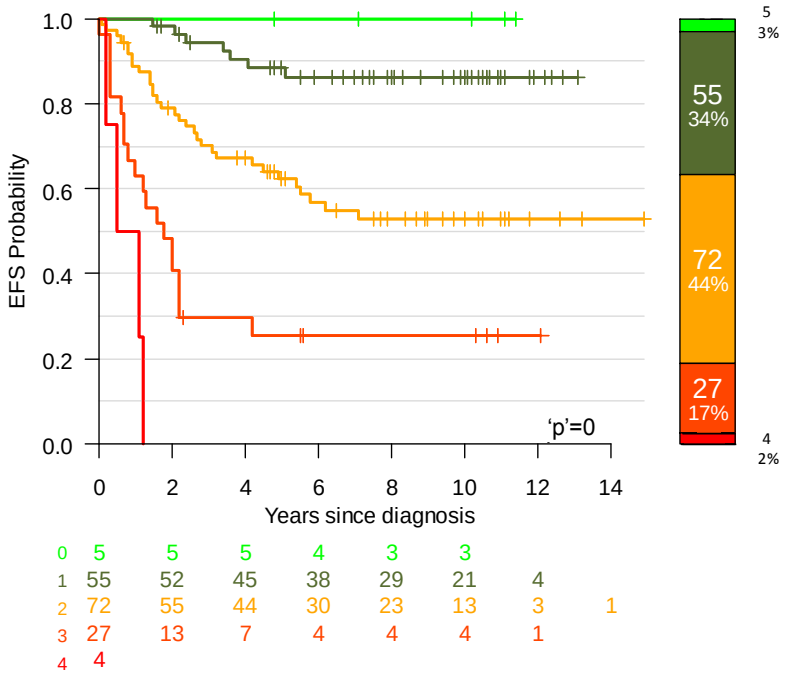
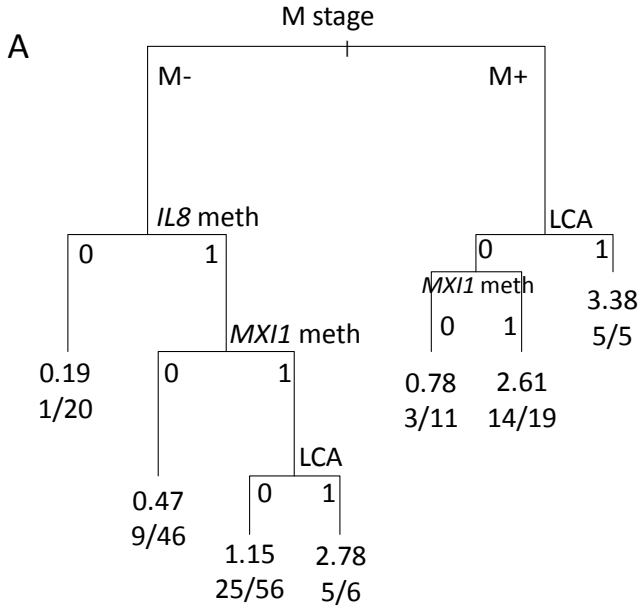


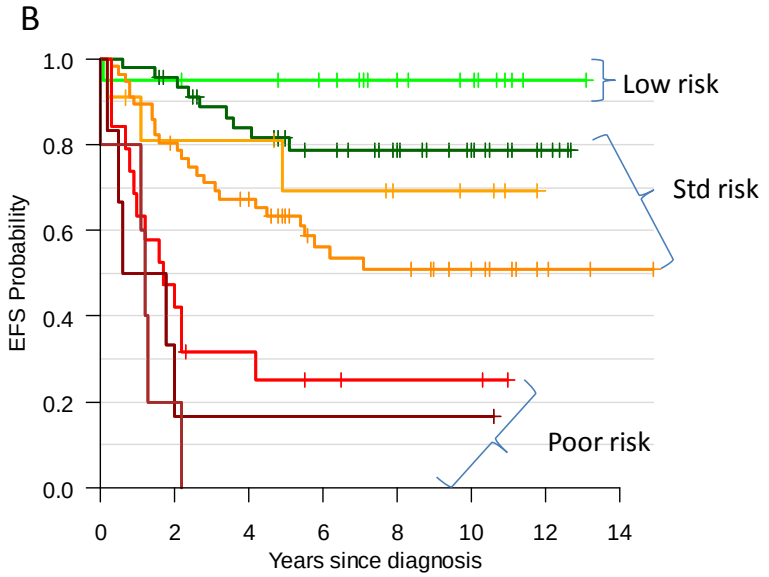
Figure 5.10. The number of risk factors in the non-WNT survival cohort (*n* = 163) determines survival. Kaplan-Meier curves are shown for each occurrence of risk factor frequency (0 – green; 1 – dark green; 2 – orange; 3 – dark orange; 4 – red). Where frequency and percentage on bar chart are too small to discern, they have been plotted to the right of the bar chart. In the final proposed model, cases with 0 and 1 risk factors are merged with all WNT cases to form a single low risk category. Non-WNT cases with 2 risk factors are categorised as standard risk. Non-WNT cases with 3 or 4 risk factors are assigned to a poor risk category.

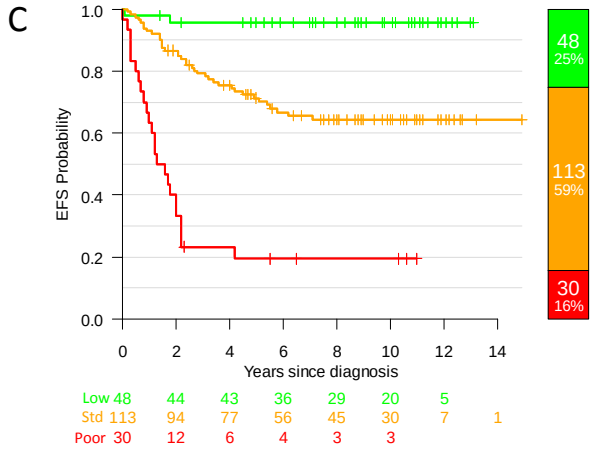
5.4.7.4 Derivation of a survival model by classification and regression tree (CART)

The proposed model described in section 5.4.7.3 was compared against a model derived using a different method. The classification and regression tree method, CART (section 2.20), was used to identify variables that most effectively stratify patients into different disease groups. In the non-WNT survival cohort, the poor risk factors M+ disease, LCA histology, *MYCC / MYCN* amplification, *MXI1* methylation and *IL8* methylation were considered, as described in section 5.3.8. A total of seven different outcome categories were identified (Figure 5.11A). In contrast to the additive survival model proposed in section 5.4.7.3, the decision tree did not utilise *MYCC / MYCN* amplification as a prognostic variable, although both *MXI1* and *IL8* methylation were important prognostic factors.

The Kaplan-Meier curve plotted from this categorisation (Figure 5.11B shows that the groups lend themselves to further consolidation in classification, analogous to the categorical simplification proposed in section 5.4.7.3). After automatically assigning WNT cases to a low risk group, this consolidation enables the identification of three groups (low risk (*n* = 48), 5 year EFS 95%; standard risk (*n* = 113), 5 year EFS 71%; poor risk (*n* = 30), 5 year EFS 19%, Figure 5.11C). Finally, the stratification based on CART analysis was compared against the stratification from an additive model of risk assignment described in section 5.4.7.3. ROC curves (Figure 5.11D) show that the additive model marginally outperforms the CART-derived model (additive model – AUC 0.799 at 5 years, 0.796 at 10 years; CART-derived model – AUC 0.769 at 5 years, 0.761 at 10 years). Notwithstanding this slight improvement in classifier performance, the major advantages to the additive model of risk assignment compared to the CART-derived model are (i) its simplicity relative to the relatively complex scheme shown in Figure 5.11A, and (ii) the large increase in the number of patients assigned to low risk (including WNT cases, 88 low risk patients for additive model, 48 low risk patients for CART-derived model, an increase of 83%).


A
M stage
M-
M+
IL8 meth
0
1
MXI1 meth
0.19
1/20
0.47
9/46
LCA
1.15
25/56
2.78
5/6
LCA
MXI1 meth
0.78
3/11
2.61
14/19
3.38
5/5


B
EFS Probability
Years since diagnosis
Low risk
Std risk
Poor risk


C
EFS Probability
Years since diagnosis
48
25%
113
59%
30
16%
Low 48 44 43 36 29 20 5
Std 113 94 77 56 45 30 7 1
Poor 30 12 6 4 3 3

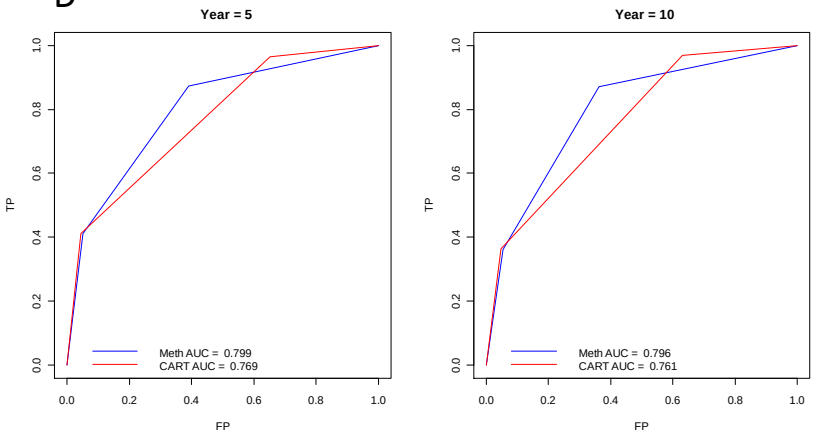

D
Year = 5
Year = 10
TP
FP
Meth AUC = 0.799
CART AUC = 0.769
Meth AUC = 0.796
CART AUC = 0.761

Figure 5.11. Classification and regression tree approach for formulation of risk stratification scheme. **A**. Decision tree derived from consideration of non-WNT survival cohort with 5 potentially prognostic variables (LCA histology (LCA), M+ disease (M stage), *MYCC / MYCN* amplification, *MXI1* methylation (*MXI1* meth) and *IL8* methylation (*IL8* meth)). At each terminal node, the resulting hazard ratio is shown. Underneath, the number of relapsing cases divided by the total number of cases that fall within that particular category, are shown. **B**. Kaplan-Meier plots are shown for the seven terminal nodes of the non-WNT survival cohort described in part A. To simplify the classification scheme, the seven curves have been simplified into three categories as shown by curly brackets. **C**. Simplified classification scheme based on CART classification of risk categories. Kaplan-Meier plots are shown for low risk (Low – green – cases identified as low risk shown in **B** plus WNT cases defined by methylomic subgroup); standard risk (Std – orange) and poor risk (Poor - red) categories. Barplot shows number and percentage of each category. At-risk tables and 'p' values from log-rank tests are shown. **D**. A comparison of risk stratification performance using the additive classification scheme (section 5.4.7.3) and the scheme implemented from the CART classification. ROC curves are shown for additive model (Meth - poor risk variables LCA histology, M+ disease, *MYCC / MYCN* amplification, binary *MXI1* and *IL8* methylation; low risk variable WNT status, shown blue) and CART model (shown red). The area under the curve (AUC) is shown for each model. FP – false positive rate, x axis; TP – true positive rate, y axis.

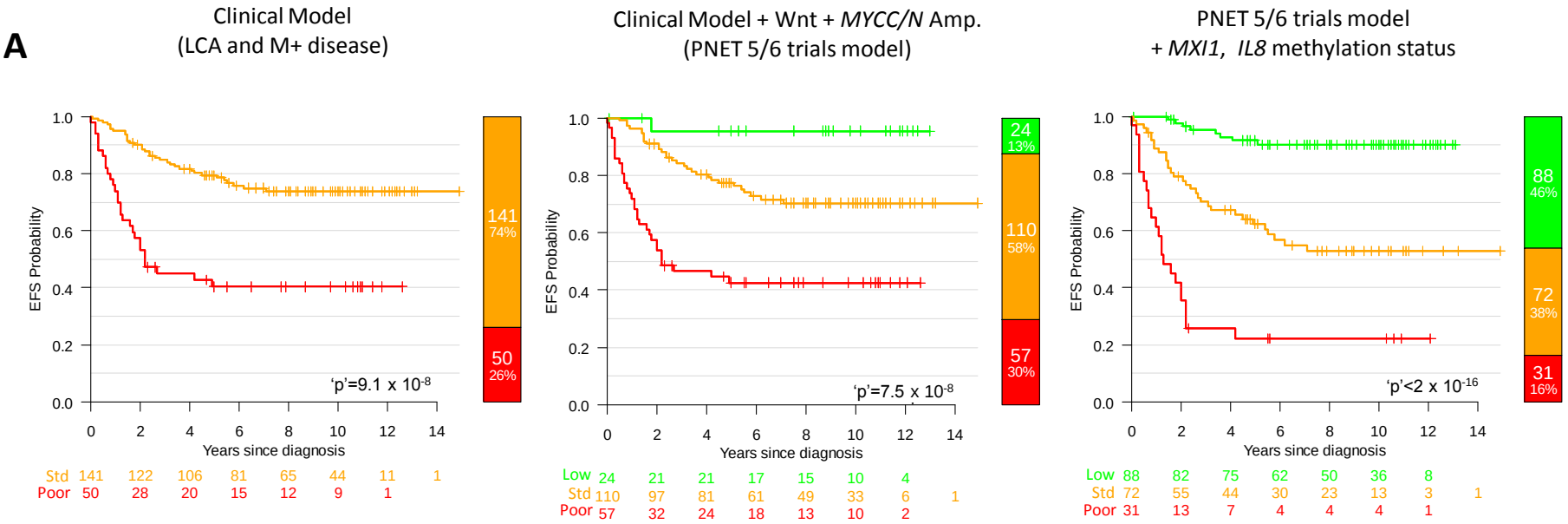
5.4.7.5 Methylation augmented survival model outperforms current disease classification schemes

Not only does the novel methylation classification scheme identify a substantially enlarged low risk group, it outperformed previous classification schemes, as shown by ROC curves in Figure 5.13. The methylation scheme performs better than the current classification scheme (AUC increases of 0.122 at 5 years, 0.153 at 10 years) and the PNET 5 / 6 trials model (AUC increases of 0.081 at 5 years, 0.107 at 10 years)

5.4.7.6 Patterns of risk-factor co-occurrence in non WNT survival cohort

The frequency of risk factor co-occurrence is shown in Table 5.9. This shows how, for cases with combinations of two risk factors (identified as standard risk in the proposed model), this always involves at least one methylation marker (72 / 72 cases (100%)). Most cases with two or more previously identified risk factors (any combination of LCA histology, M+ disease or *MYCC / MYCN* amplification) are also associated with a methylated *MXI1* and / or *IL8* biomarker (11 / 12 cases (92%)). Finally, where only one risk factor was observed, this was a methylation biomarker in 53 / 55 cases (96%).

A


Clinical Model
(LCA and M+ disease)
Clinical Model + Wnt + *MYCC/N* Amp.
(PNET 5/6 trials model)
PNET 5/6 trials model
+ *MXI1*, *IL8* methylation status
EFS Probability
Years since diagnosis
'p'=9.1 x 10-8
'p'=7.5 x 10-8
'p'<2 x 10-16
141 74%
50 26%
24 13%
110 58%
57 30%
88 46%
72 38%
31 16%
Std 141 122 106 81 65 44 11 1
Poor 50 28 20 15 12 9 1
Low 24 21 21 17 15 10 4
Std 110 97 81 61 49 33 6 1
Poor 57 32 24 18 13 10 2
Low 88 82 75 62 50 36 8
Std 72 55 44 30 23 13 3 1
Poor 31 13 7 4 4 4 1

B

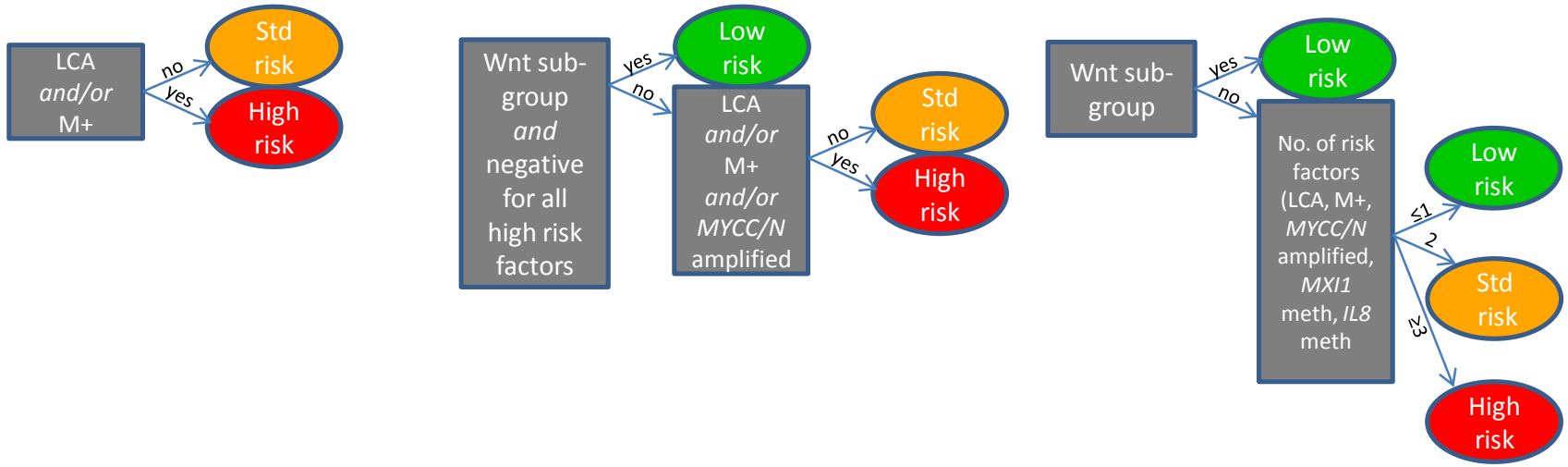

LCA *and/or* M+
no
yes
Std risk
High risk
Wnt sub-group *and* negative for all high risk factors
yes
no
Low risk
LCA *and/or* M+ *and/or* *MYCC/N* amplified
no
yes
Std risk
High risk
Wnt sub-group
yes
no
Low risk
No. of risk factors (LCA, M+, *MYCC/N* amplified, *MXI1* meth, *IL8* meth
≤1
2
≥3
Low risk
Std risk
High risk

Figure 5.12. DNA methylomics offers potential for an improved risk classification of medulloblastoma. A. Refinement of clinical risk stratification applied to non-infant PNET3 clinical trials cohort (*n* = 190), illustrated using Kaplan-Meier plots of EFS. First panel shows current clinical staging, second panel shows modified risk stratification, used for assigning treatment in current PNET 5 / 6 clinical trials. Third panel displays a novel cumulative risk stratification scheme that utilises existing clinical markers plus two cross-validated prognostic DNA biomarkers (*MXI1*, *IL8*), identified in our methylomic screen. Low risk is coloured green, standard risk amber and high risk red. **B.** Flowcharts for assignment of patient risk in current clinical model (first panel), PNET 5 / 6 clinical model (second panel) and PNET 5 / 6 clinical model augmented by methylation biomarkers (third panel), used to derive Kaplan-Meier plots shown in part A. LCA: large cell / anaplastic; M+: M stage greater than 1; *MYCC / N* amplified: amplification of *MYCC or MYCN* oncogenes; *MXI1* methylation: methylation score of ≥ 0.67; *IL8* methylation: methylation score of ≥ 0.67.

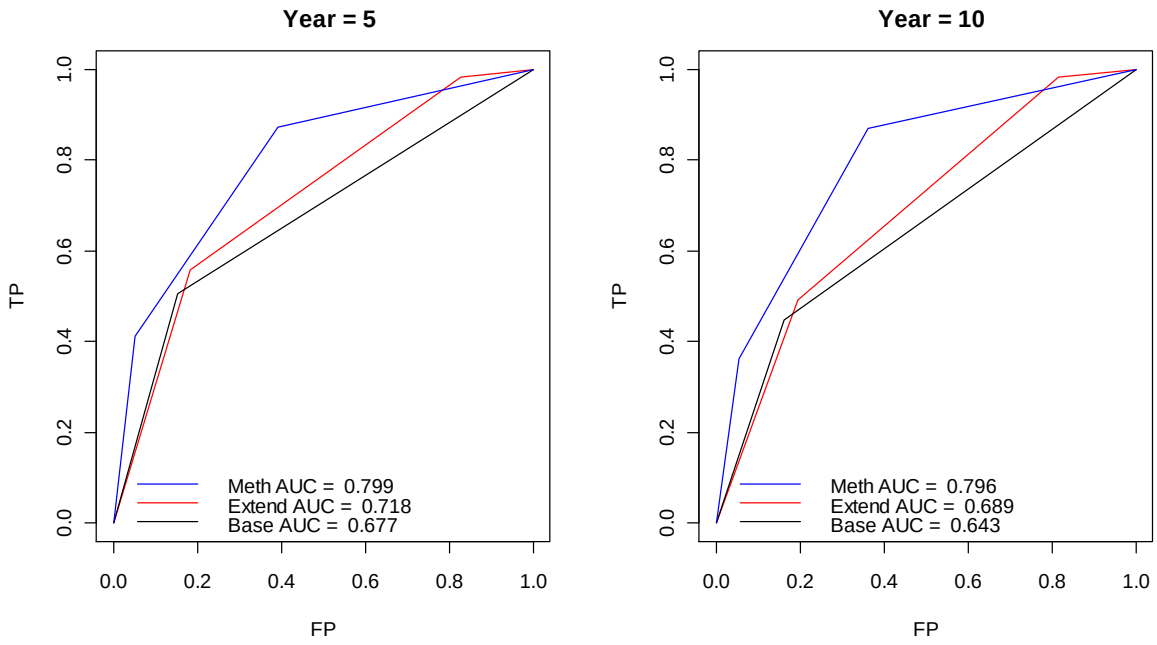


Figure 5.13. Novel methylation-based classification scheme out-performs current risk stratification models at 5 and 10 years. ROC curves are shown for current clinical model (Base – high risk variables: LCA histology and M+ disease, shown black), extended PNET 5 / 6 trial clinical model (Extend – high risk variables: LCA histology, M+ disease, *MYCC / MYCN* amplification. Low risk variable: WNT pathway activation and absence of high risk factors, shown red), methylation augmented clinical model, outlined in part B of Figure 5.12 (Meth - high risk variables: LCA histology, M+ disease, *MYCC / MYCN* amplification, binary *MXI1* and *IL8* methylation. Low risk variable: WNT status, shown blue). The area under the curve (AUC) is shown for each model. FP – false positive rate, x axis; TP – true positive rate, y axis.

Frequency	LCA histology	M+ disease	*MYCC / MYCN* amplification	*MXI1* methylation	*IL8* methylation	Risk
5	0	0	0	0	0	Low
40	0	0	0	0	1	Low
13	0	0	0	1	0	Low
1	0	0	1	0	0	Low
1	0	1	0	0	0	Low
0	1	0	0	0	0	Low
52	0	0	0	1	1	Std
1	0	0	1	0	1	Std
9	0	1	0	0	1	Std
5	1	0	0	0	1	Std
1	0	0	1	1	0	Std
4	0	1	0	1	0	Std
0	1	0	0	1	0	Std
0	0	1	1	0	0	Std
0	1	0	1	0	0	Std
0	1	1	0	0	0	Std
4	0	0	1	1	1	High
15	0	1	0	1	1	High
4	1	0	0	1	1	High
1	0	1	1	0	1	High
2	1	1	0	0	1	High
0	1	0	1	0	1	High
0	0	1	1	1	0	High
0	1	0	1	1	0	High
0	1	1	0	1	0	High
1	1	1	1	0	0	High
0	0	1	1	1	1	High
2	1	0	1	1	1	High
2	1	1	0	1	1	High
0	1	1	1	0	1	High
0	1	1	1	1	0	High
0	1	1	1	1	1	High

Table 5.9. Co-occurrence of risk factors in non-WNT survival cohort. Table shows occurrence of risk factors and their frequency among the non-WNT survival cohort (*n* = 163). For each risk factor (LCA histology, M+ disease, *MYCC / MYCN* amplification, *MXI1* methylation and *IL8* methylation, 1 denotes the presence of a risk factor, and 0 the absence of a risk factor. The resultant risk, determined by the total number of risk factors is shown.

5.5 Discussion

Investigations in chapter 4 demonstrated that, WNT and NC medulloblastoma cases apart, there was no discernible difference in survival between the group I, group II and SHH methylomic subgroups. In the age-matched, trials-based cohort, this finding was repeated in multivariate analyses which showed no additional prognostic value for the non-WNT subgroups, in contrast to previous studies which showed that transcriptomic subgroups C (Northcott et al., 2010) and c1 (Cho et al., 2011) were associated with a poor prognosis. This chapter investigated the prognostic utility of methylomic biomarkers and identified a novel risk stratification scheme that outperforms previously published schemes in medulloblastoma.

5.5.1 Survival cohort recapitulates previously reported survival markers and classification schemes

After showing that the age-matched cohort (*n* = 55) showed equivalent survival to the PNET3 clinical-trials cohort (*n* = 136) (Figure 5.1), the two datasets were joined to form a unified survival cohort (*n* = 191). The verification that previously reported markers of survival (LCA histology, M+ disease and *MYCC / MYCN* amplification (Ellison et al., 2011b; von Hoff et al., 2009; Brown et al., 2000)) were associated with significant differences in survival (section 5.4.2, Figure 5.2) demonstrated that this was a representative cohort of medulloblastomas. Next, the verification that both current clinical risk stratification and the proposed PNET 5 / 6 classification schemes (Figure 5.3) perform well with the survival cohort enabled comparisons to be made between novel and existing stratification schemes.

5.5.2 Investigation of the pleiotropic effects of prognostic markers across methylomic subgroups

The determination that the cohort investigated here was representative of previously reported medulloblastoma cohorts verified that investigations into any pleiotropic effects of previously identified clinical correlates would be possible. However, the power calculations shown in Table 5.4 demonstrate that such investigations would not be adequately powered, highlighting the inevitable problem of sample sizes decreasing as medulloblastomas are placed into discrete subgroups, and the need for

investigations in larger cohorts, in a post-transcriptomic / methylomic disease classification setting, as discussed in section 4.5.1, since the numbers per subgroup will be much reduced. This need could be met by methylome profiling of historical FFPE trials-based cohorts, since clinical trials cohorts with archived fresh frozen materials are scarce.

Currently, the biological behaviour of discrete disease subgroups remains poorly understood and the effects on survival of clinico-pathological correlates need to be investigated.

5.5.2.1 MYCC / MYCN amplified SHH cases have a dismal prognosis

The effects of *MYCC* / *MYCN* amplification on survival for SHH subgroup cases were striking, where 5 / 36 (14%) cases with amplification (4 *MYCN* / 1 *MYCC*) all relapsed within 1 year (Figure 5.4). It is unfortunate that too few cases of either oncogene were observed individually to make any inference about the separate behaviours of *MYCC* and *MYCN*. Nevertheless, this dismal survival was not observed in group I or group II cases with *MYCC / MYCN* amplification, where survival was consistent with non-amplified cases. The role of *MYCN* in medulloblastoma cases and in SHH medulloblastoma in particular, is becoming more widely recognised, although its precise contribution to survival is unclear. Poor outcome of *MYCC / MYCN* amplified cases has previously been reported (Pfister et al., 2009), although in a more recent study, a poor outcome for *MYCC* amplified cases only was observed (Ellison et al., 2011b). The frequent association of *MYCN* amplification with *TP53* mutation (Pfaff et al., 2010) may be a causative factor for the poor survival of *MYCN* cases observed within the SHH subgroup, while it has also been observed that paediatric SHH cases are enriched for *MYCN* amplification relative to adult SHH cases (Northcott et al., 2011). The poorer survival of SHH *MYCN* amplified cases represents an intriguing finding that should be investigated in other cohorts, since the observations reported above remain anecdotal, and there are too few cases to determine whether this is an important event in disease prognostication. With a reported incidence of *MYCC / MYCN* amplification of 4% and 7% respectively (Pfister et al., 2009), and a SHH subgroup incidence of 23% (Table 4.9), a cohort size of 1240 would be required to identify

approximately 20 SHH, *MYCN* amplified cases, emphasising the need for global cooperation in sample ascertainment for the next phase of elucidating the effects of survival markers in a post-subgrouped disease

5.5.3 Female gender is associated with an improved survival

A significant survival advantage was observed across the whole cohort for female cases ('p' = 0.013, log-rank test). The prognostic role of gender in medulloblastomas is unclear. The largest study addressing this question in medulloblastoma (Curran et al., 2009) of 1049 subjects (668 males, 381 females), reported that in children over 3, females had significantly greater survival than males, a finding consistent with the increase in female survival observed in this study, although this has been contradicted elsewhere (Ellison et al., 2011b; Curran et al., 2009; Alston et al., 2003; Weil et al., 1998). Investigations in chapter 4 identified that there was approximately gender parity in the WNT and SHH methylomic subgroups, with a male excess in group I and group II cases. Gender has not been widely reported as a prognostic factor in medulloblastoma until recently and requires further investigation in larger cohorts to more fully understand the relationship between gender and subgroup before considering its introduction into future disease risk stratification schemes. The gender imbalance observed in medulloblastoma has been widely reported, with a male : female disease ratio estimated from the central brain tumour registry of the United States of 1.6 : 1 (Crawford et al., 2007), and this investigation has demonstrated how this male excess is being driven by the group I and group II methylomic subgroup medulloblastoma cases. Recent work has demonstrated a potentially protective role for oestrogen in mouse models of medulloblastoma (Mancuso et al., 2010), demonstrating for the first time, how female gender could mediate improved survival by the interaction of oestrogen and insulin growth factor 1 (IGF-1) mediated pathways. Further investigation of the potential prognostic impact of gender in medulloblastoma is needed, both in infant and childhood cohorts before its effect is considered in survival models and risk stratification schemes.

5.5.4 Addition of MXI1 and IL8 methylation status to current medulloblastoma risk stratification models

The identification of a novel prognostic Cox model, augmented with methylation biomarkers, which outperforms the current clinical risk stratification schemes, represents a first precedent for the use of methylomic data for prognostication in medulloblastoma. While this is a single cohort study and is therefore liable to over-fitting (section 2.11), the Cox boost algorithm used took steps to minimise this problem (section 2.17). Nevertheless, this model would require additional validation in an independent cohort, discussed below in section 5.5.7.

The similar hazard ratios for the included poor risk markers in the prognostic Cox model (Figure 5.9, Figure 5.10), enabled the creation of a novel additive risk classifier (section 5.4.7.3), based on the total number of risk factors (Figure 5.12). This classifier is an intuitive and attractive method for the assignment of patient risk in a clinical setting, which compares favourably with both alternative methods for deriving classifiers (CART, Figure 5.11), and other reported classification schemes, discussed below.

Although the CART-derived classification scheme did not include *MYCC / MYCN* amplification as prognostic factors, both *MXI1* and *IL8* methylation were selected as prognostic risk factors in the classification tree, validating their prognostic utility by using an independent method to stratify risk. *MYCC / MYCN* amplification was selected as a prognostic risk factor in the selected additive risk model, since it was prognostic both in univariate and multivariate analyses (Figure 5.2, Table 5.3) and has previously been reported to be a significant marker of poor prognosis (Ellison et al., 2011b; Pfister et al., 2009).

5.5.5 Comparison with state-of-the-art risk stratification schemes

A report from Tamayo and colleagues (Tamayo et al., 2011) presented a selection of 4 novel classification schemes (Table 5.1), with AUC ranging from 0.58 to 0.84 at 30 months in a training cohort (n = 96). In an independent test cohort (n = 78), the performance of the schemes range from an AUC of 0.73 to 0.80 (Tamayo et al., 2011). The cohort used for the derivation of the classification scheme presented in this

chapter (*n* = 191) was double the size of the training cohort used by Tamayo and colleagues (Tamayo et al., 2011).

For the risk stratification scheme presented by Pfister and colleagues, their novel classifier (Table 5.1) had an AUC of approximately 0.67 at 30 months (Pfister et al., 2009). The scheme presented in this chapter had an AUC of 0.86 at 30 months, slightly outperforming the optimally performing scheme in the training dataset of Tamayo and colleagues (Tamayo et al., 2011) and comfortably outperforming the scheme proposed by Pfister and colleagues (Pfister et al., 2009). Moreover, the scheme presented here considers only six covariates (WNT status, LCA histology, M+ disease, *MYCC / MYCN* amplification, *MXI1* and *IL8* methylation), in contrast to the scheme chosen by Tamayo and colleagues, which proposed a nomogram to assign risk that considered the effects of 17 covariates, 6 of which are 6-member expression signatures (a total of 47 variables) (Tamayo et al., 2011). While this might be feasible in the context of a discovery paper, this approach would not be practicable in a routine setting, in contrast to the scheme presented here, which adds only two additional variables to the PNET 5 / 6 clinical trials model, easily testable in a routine diagnostic lab setting (discussed in section 5.5.7). Furthermore, under the proposed classification scheme, WNT status, *MXI1* and *IL8* methylation are assigned simultaneously using DNA methylomics, leaving only *MYCC* and *MYCN* amplification status to be measured using different means.

As previously discussed (section 4.5.5), methylation markers are less labile than transcriptomic measurements, are more suitable for measurement and are also suitable for use in archival cohorts from which nucleic acid derivatives extracted from FFPE tissue are available. The methylation augmented classification scheme is comparable in complexity to the scheme proposed by Pfister and colleagues, which considered copy number aberrations of chromosome 6q, 17q and the *MYCC* and *MYCN* loci (Pfister et al., 2009), although the scheme presented here comfortably outperforms their classifier.

5.5.6 Investigation into the gene function of selected survival methylation biomarkers

As discussed in section 1.1.7, the widely-accepted understanding is that methylation of gene promoter regions is associated with a condensed chromatin structure (Figure 1.7). Therefore it is possible that for both *MXI1* and *IL8*, the increased risk described above associated with a methylated promoter region CpG dinucleotide is representative of the CpG island as a whole (since both *MXI1* and *IL8* have CpG islands spanning their promoter regions), which is in turn plays a role in gene silencing. This would need to be validated through bisulfite sequencing (section 2.6.1) of the CpG islands.

5.5.6.1 MXI1 is a negative regulator of MYC activity

The *MXI1* (MAX-interacting protein 1) gene is a six / seven exon gene (see below) located on chromosome 10q25.2 and encodes a protein, MXI1 that binds to MAX. Since MXI1 competitively competes with the potent oncoprotein MYC, it can act as a negative regulator of MYC activity by preventing MYC from binding to MAX, suppressing cell proliferation and is therefore a putative tumour suppressor (Eagle et al., 1995). The allelic loss of *MXI1* in prostate cancer has subsequently been shown to be common in prostate cancer, where a study identified allelic loss of *MXI1* in 22 / 40 (53%) tumours (Prochownik et al., 1998). This is potentially important, since dysregulation of the MYC pathway is fundamental to the tumour biology of 10-15% of medulloblastomas (Eberhart et al., 2004). Mutations in *MXI1* have previously been described in medulloblastoma (Scott et al., 2006), suggesting that allelic loss through either mutation or inactivation through DNA-methylation induced silencing may contribute towards disease pathogenesis by freeing the MYC / MAD / MAX pathway from inhibition (section 1.3.7.3.1).

The relationship between *MXI1* methylation and any silencing of gene expression warrants further investigation, and is discussed below in section 5.5.7. However, indirect evidence is available from the transcriptomic studies of medulloblastoma which assayed the expression of the *MXI1* gene (Cho et al., 2011; Northcott et al., 2010; Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006), shown in Figure 5.14. Every study demonstrated expression of the gene, with a large average dynamic range of expression across studies (average range 3.3 $\log_2$ units, representing a 10 fold

change). The absence of evidence for gene silencing might mean other mechanisms other than DNA methylation of the *MXI1* promoter are responsible for the regulation of transcription of this gene. Nevertheless, if there is a methylation-induced down-regulation of *MXI1* expression, this could be contributing to the release from inhibition of the MYC / MAD / MAX network. The patterns of methylation observed, whereby hypomethylation of the *MXI1* locus was associated with an improved prognosis, are not entirely consistent with this hypothesis, since the majority of non-WNT medulloblastomas assayed in this chapter (97 / 163 non WNT cases (60%)) had elevated levels (β score ≥ 0.67) of *MXI1* methylation. Alternatively, the differential methylation observed for *MXI1* could potentially be determining the isoform of *MXI1* expressed. Investigations in glioblastoma have identified an alternative *MXI1* transcript, *MXI1-0,* that is over-expressed in glioblastomas relative to the wild-type transcript. *MXI1-0* is transcribed from an alternative upstream exon (exon 0), located 18kb upstream of the wildtype first exon (Engstrom et al., 2004). Investigations of the MXI1-0 protein demonstrated that it failed to repress *MYCC* dependent transcription. Future investigations should aim to characterise the methylation status of both the conventional promoter region as well as methylation surrounding exon 0, in tandem with investigations of the expression of transcript starting at exon 0 relative to transcript starting at exon 1.

5.5.6.2 Expression of IL8 is associated with PTEN loss

IL8 is a four exon gene located at chromosome 4q13.3. It encodes a chemokine protein *IL8* (also known as *CXCL4*) whose function is to attract haematopoietic progenitor cells (Pruijt et al., 2002) for the initiation of angiogenesis. The expression of *IL8* has previously been shown to be relevant to glioblastoma (de la Iglesia et al., 2008). The loss and / or inactivation of phosphatase and tensin homologue (*PTEN)* have been shown to be an important factor in the pathogenesis of glioblastomas. Endogenous signal transducer and activation of transcription 3 (STAT3) signalling is inhibited in PTEN-deficient glioblastoma cells. Activated STAT3 directly occupies the *IL8* promoter and inhibits its transcription. Consistent with this chain of events, *IL8* was up-regulated in *PTEN*-deficient glioblastoma cells (de la Iglesia et al., 2008).

In a whole exome screen of 22 medulloblastoma cases, a single mutation in *PTEN* was identified. In a validation cohort of 88 cases, 3 / 88 (3.5%) were positive for mutations in *PTEN* (Parsons et al., 2011). In addition, the transcriptomic identification of six subgroups by Cho and colleagues demonstrated that 14 / 18 members of the c5 subgroup had focal *PTEN* and / or 10q loss, demonstrating that the inactivation of *PTEN* through loss or mutation is common in medulloblastoma (Cho et al., 2011). While it is clear that the expression of *IL8* bears further investigation in medulloblastoma, the mechanism by which *IL8* methylation can confer a poorer survival is unclear, since the methylation-dependent silencing of *IL8* is antithetical to the activation of *IL8* described above in PTEN-deficient glioblastoma cells.

As for *MXI1,* there is a need to investigate the relationship between differential methylation of *IL8* and its gene expression. Across the available transcriptomic datasets, there is evidence for variable *IL8* expression (Figure 5.14), with a large average dynamic range of expression across studies (average range 6.1 $\log_2$ units, representing a 70 fold change); within 4 studies (Cho et al., 2011; Northcott et al., 2010; Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006), a large proportion of cases showed very low levels of expression of *IL8* (Affymetrix expression array data reports a value of expression which is $\log_2$ transformed after adding a constant to prevent negative levels of expression. An absent or low level of expression is approximately 2-3 units on this scale). The higher expression levels observed by Northcott and colleagues (Northcott et al., 2010) are likely due to the differential sample processing since this experiment utilised Affymetrix exon arrays which are not directly comparable to the 3' *in vitro* transcription expression arrays used in the other four studies. In summary, the expression of *IL8* in transcriptomic studies is consistent with evidence of gene silencing is a minority of cases. However, this cannot be directly associated with methylation levels of the *IL8* promoter observed in this study, since the majority of medulloblastomas assayed in this chapter (137 / 163 non WNT cases (84%)) had elevated levels (β score ≥ 0.67) of *IL8* methylation.

Finally, transcriptomic investigations of medulloblastoma have suggested that group C medulloblastomas (Northcott et al., 2010) are enriched for genes associated with elevated levels of *MYC*, suggesting that any effect of *MXI1* methylation on the nature of *MXI1* expression may be modulated by methylomic subgroup membership.

However, for both *MXI1* and *IL8* methylation, after binary cutoffs at 0.67 were applied, there was no discernible survival difference for methylation status in group I and group II medulloblastomas, where in every case, an absence of methylation was associated with an improved survival (*MXI1* - group I, 'p' = 0.01, log rank test; group II, 'p' = 0.01, log rank test. *IL8* – group 1, 'p' =0.12, log rank test; group II, 'p' = 0.10, log rank test).

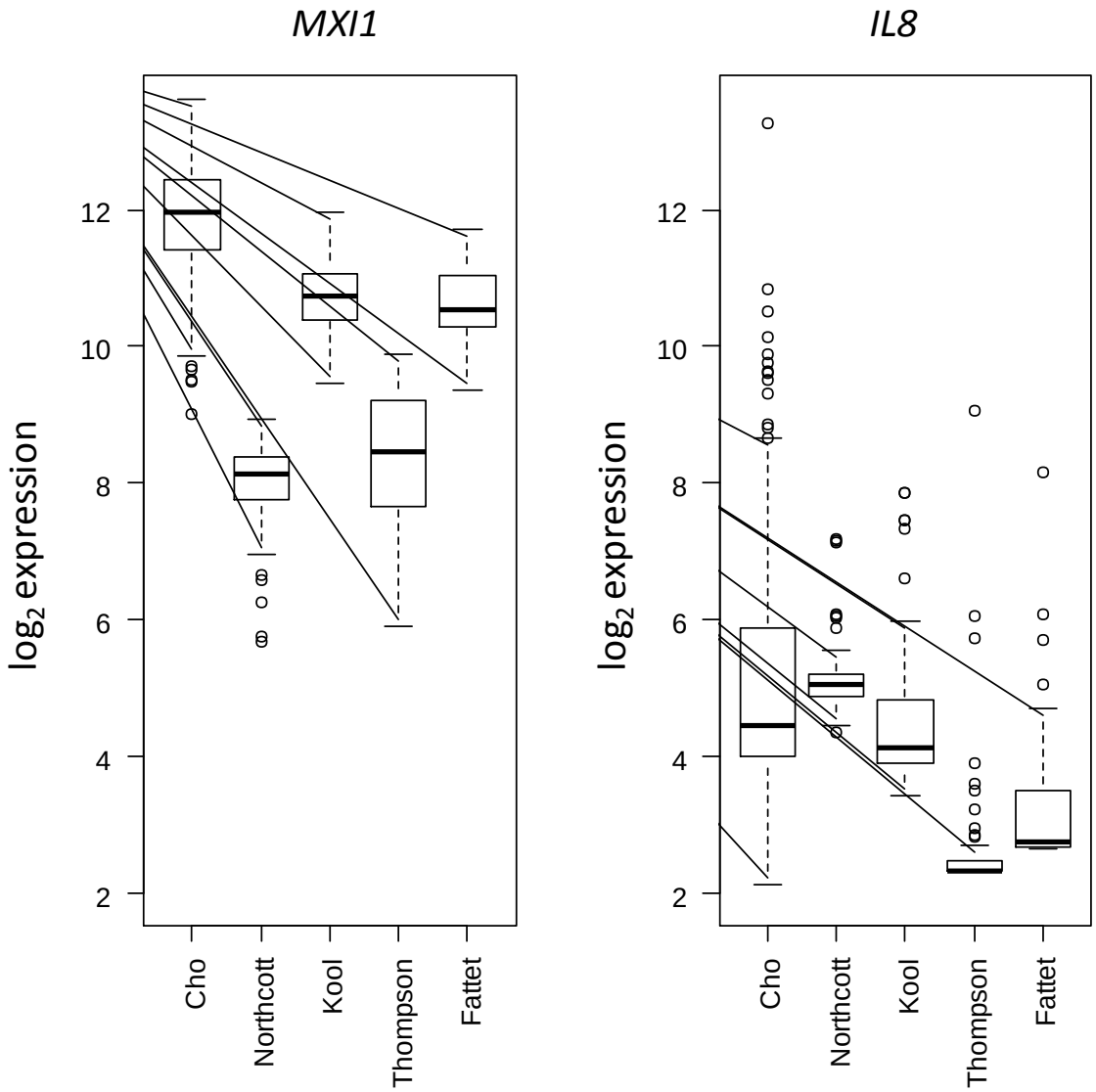


Figure 5.14. Box-plots show log_2 expression of *MXI1* and *IL8* from 5 studies which undertook transcriptomic analysis of medulloblastoma (Cho et al., 2011; Northcott et al., 2010; Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006). Each transcriptomic dataset is labelled by its first author, with studies ordered by decreasing number of samples (Cho et al. – *n* = 194; Northcott et al. – *n* = 103; Kool et al. – *n* = 62; Thompson et al. – *n* = 46; Fattet et al. – *n* = 40).

5.5.7 Further work

The aims described in the introduction to this chapter have been fulfilled, but more questions remain. In particular, the investigation of the pleiotropic survival effects of clinico-pathological and molecular correlates were shown to be inadequately powered, although a potential effect for *MYCC / MYCN* amplification was identified, and emphasised the need to expand the investigation into larger cohorts. There is now a

clear need to fully understand the effect of these well-validated survival correlates in the context of subgroup membership. This could provide a platform for the derivation of novel risk classification schemes. It is possible that incorporation of the interaction between subgroups and risk factors could greatly increase the predictive power of risk stratification, without the need to include additional molecular markers.

The proof of concept that methylation biomarkers can be useful as prognostic biomarkers sets a precedent for future investigations on higher resolution platforms in larger cohorts. In these future investigations, to identify methylation biomarkers that not only are prognostic, but also have a putative functional relevance, parallel investigations of the relationship between (i) methylation and survival and (ii) expression and survival could be undertaken on genome-wide array platforms, and any loci with significant relationships to survival for both gene expression and methylation could be investigated further. This need not necessarily come from paired methylation and expression array samples, and could be bootstrapped onto existing transcriptomic datasets. This would ensure that any potentially prognostic methylation probes are associated with a change in expression that may be functionally relevant.

Before any further insights can be gained from examining the prognostic potential of methylation markers, *MXI1* and *IL8* methylation status would need to be confirmed by bisulfite sequencing. Subject to successful validation, the relationship between methylation and gene expression could be investigated. This would aim to identify whether the prognostic methylation markers are simply biomarkers or whether they engender changes in gene expression that may have a putative functional impact.

The relationship between promoter methylation status and gene expression could be investigated by treating methylated medulloblastoma cell lines with 5-aza-cytidine, a DNA methyltransferase inhibitor (Lindsey et al., 2004), and measuring gene expression before and after treatment. If warranted, the expression of primary medulloblastoma samples could then be directly compared against the methylation state of the promoter regions.

This chapter has confirmed that methylomic biomarkers have potential for assigning patient risk in medulloblastoma. The identification of the two methylation biomarkers *IL8* and *MXI1* and their integration into a clinical model produces a classification

scheme that outperforms currently used risk stratifications. It is clear that since this is a single cohort study, and before any consideration of the utility of this scheme is made, it would be imperative to validate the prognostic nature of these biomarkers in an independent cohort.

Subject to successful validation of the fidelity of the reported methylation markers and validation of the classifier in an independent cohort, the development of an assay for the measurement of *MXI1* and *IL8* methylation in a routine diagnostic setting would be necessary. It would be relatively simple to integrate prognostic markers into a future MS-MLPA (Nygren et al., 2005) assay (see section 4.5.6) or Veracode assay (Illumina, 2011; Pedersen et al., 2011) that assigns medulloblastoma subgroup based on methylation biomarkers.

Finally, while this chapter has confined itself to examining paediatric medulloblastomas, exactly the same questions could and should be asked in further infant and adult medulloblastoma and joined (infant, paediatric and adult) cohorts.

5.5.8 Summary

This chapter has investigated the possibility of identifying differential behaviours of clinico-pathological and molecular correlates in medulloblastoma disease subgroups and identified the need to identify any differential behaviours in larger, independent disease cohorts.

The chapter has also demonstrated that methylation biomarkers have utility as prognostic biomarkers, and identified specific biomarkers (*MXI1* and *IL8* methylation) with prognostic utility. Subsequently, a novel risk stratification scheme for medulloblastoma was derived, which out-performed both currently-used and state-of-the-art risk stratification schemes and provides a platform for future further investigations for refining risk stratification of medulloblastoma.

Chapter 6. Summary and Discussion

6.1 Introduction

Medulloblastoma is the most common malignant brain tumour of childhood, accounting for 20% of all paediatric brain malignancies, with an incidence of 0.5 / 100,000 children / year (Crawford et al., 2007), corresponding to approximately 90 cases a year in the UK (Pizer and Clifford, 2009). Clinico-pathological correlates have proved useful for disease stratification; 80% of standard risk patients are cured of their disease (Packer et al., 2006). High risk patients, defined by positivity for any of metastatic disease, residual disease following surgical excision and aged under 3 at diagnosis, fare worse, with 5 year PFS rates of 50% (Pizer and Clifford, 2009).

Although the majority of newly diagnosed cases of medulloblastoma will be cured of their disease, survivors treated under current protocols are burdened with unacceptably severe neuro-cognitive and neuro-endocrine late effects. Typically, survivors will have a reduced IQ and lifelong social impairment, alongside defects in growth (section 1.3.5.2).

The recent discoveries that medulloblastoma is a molecularly heterogeneous disease (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006; Ellison et al., 2005; Raffel et al., 1997), comprising a series of molecular subgroups, offer hope that additional prognostic molecular markers may be identified that enable more appropriate treatment strategies, through improved disease-risk stratification and the targeted use of novel molecularly-targeted therapeutics, with the twin aims of (i) curing every patient and (ii) reducing treatment intensities in patients with a favourable prognosis, for an improvement in their quality of life as a survivor of medulloblastoma. This approach has started to bear fruit – the upcoming PNET 5 / 6 clinical trials (Pizer and Clifford, 2009) include activation of the WNT pathway as a marker of favourable prognosis and amplification of *MYCC / MYCN* oncogenes as markers of poor prognosis, in addition to established clinico-pathological disease correlates.

While the precise number of disease subgroups remains unclear (see section 1.3.12.1), there is hope that including subgroup information in survival models may also enable a more refined disease risk stratification. Moreover, the pleiotropic effects of established and novel risk covariates across disease subgroups remain to be

investigated in clinical trials cohorts. This, coupled with the increasing availability and coverage of methods to assess genomic, epigenomic and transcriptomic differences, should enable the completion of a comprehensive molecular characterisation of the disease over the next decade. This should, in turn, enable the derivation of novel stratification schemes, alongside improvements in pathway specific inhibitors (section 1.3.11), which show promise for inhibiting tumour growth whilst reducing radiotherapeutic intensities, to help fulfil the twin aims of curing every patient, whilst reducing the late effects of treatment as far as possible.

This project set out to investigate the nature and utility of medulloblastoma subgroups for disease prognostication. In chapter 3, transcriptomic signatures of WNT and SHH pathway activation were defined that accurately classified tumours across multiple studies (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006). A comprehensive analysis of the genetic and epigenetic determinants of pathway activation was undertaken in the primary investigation cohort. Subsequently, a meta-analysis of the clinico-pathological and molecular features of the WNT, SHH, and WNT / SHH independent subgroups of medulloblastoma was undertaken, incorporating data from three other published studies, assigning pathway activation status using the defined expression signatures.

In chapter 4, a novel DNA methylation microarray approach was used to determine whether medulloblastomas showed any differences in their patterns of methylation, and, if so, whether these patterns had any utility for disease classification. On a training dataset, a consensus-clustering, NMF-based approach (section 2.9.4) identified four subgroups, and defined the methylation probes that were determining metagene expression. The H values from the four-group clustering were projected onto the test dataset and a classifier used to assign group membership of the test dataset. Again, the clinico-pathological and molecular features of the identified subgroups were investigated.

Finally, a novel risk stratification scheme was identified in chapter 5, which used the Cox boost algorithm (Binder et al., 2009) to augment existing clinical variables with methylation markers. From the resultant Cox model (section 2.14), an additive risk

stratification scheme was defined, and compared against current, imminent (PNET 5 / 6) and state of-the-art molecular classification schemes.

6.2 Gene signatures of signalling pathway activation have utility for disease subclassification

The assessment of WNT and SHH pathway activation in medulloblastoma is fundamental to the assignment of treatment strategies: The improved prognosis of WNT-activated tumours is well recognised (Cho et al., 2011; Northcott et al., 2010; Fattet et al., 2009; Clifford et al., 2006; Ellison et al., 2005), and it is hoped that the reduced treatment intensities applied to WNT-positive cases in the upcoming PNET 5 clinical trial will ensure a cure and achieve a significant reduction in late effects of treatment. Moreover, the possibility of biological inhibition of the SHH pathway in SHH medulloblastomas (section 1.3.11) holds rich promise for a reduction in treatment side effects through targeted therapies that obviate or diminish the need for radiotherapy (Rudin et al., 2009). If the concerns about the premature fusion of growth plates in bones of young children treated with SHH inhibitors were to be dispelled (Kimura et al., 2008), this could also be a promising treatment option for the 62% of infants who show an activated SHH pathway (Schwalbe et al., 2011), a group of patients in whom radiotherapy is either not used, or used only sparingly, due to the unacceptable side effects of treatment in very young children.

Clinically, in order to assign subgroup status, an assay must be rapid, robust, reproducible and applicable in a standard hospital pathology lab. The described GeXP assay for the assignment of WNT and SHH pathway activation fulfils all of these criteria. Its applicability across *in vitro* and *in silico* datasets demonstrated its robustness and enabled a meta-analysis across multiple datasets to accurately identify the defining molecular and clinico-pathological correlates of the WNT, SHH and WNT / SHH independent subgroups.

Using three further transcriptomic datasets of medulloblastoma (Fattet et al., 2009; Kool et al., 2008; Thompson et al., 2006) available at the time of analysis, in addition to the primary investigation cohort, a total cohort of 173 cases was available for meta-analysis. Mutations in *CTNNB1* and *PTCH1* were exclusively detected in WNT and SHH pathway-positive cases, respectively. While *CTNNB1* mutation was reported in 95% of WNT cases, 34% of SHH cases investigated harboured a *PTCH1* mutation, indicating that other mechanisms of pathway activation must be responsible. Loss of

chromosome 9q was observed in SHH cases both with and without *PTCH1* mutation, indicating that disruption of a single *PTCH1* allele is sufficient to cause SHH pathway activation in medulloblastoma. A number of the previously reported determinants of SHH pathway activation (*SUFU* missense mutation (Taylor et al., 2002), *REN* allelic loss (Di Marcotullio et al., 2004), *PTCH 1c* methylation (Diede et al., 2010)) were not specifically associated with the SHH subgroup, indicating that additional activation mechanisms remain to be identified. Loss of chromosome 6 was also closely correlated with WNT pathway activation, whilst chromosome 17p loss was observed only in the WNT / SHH independent subgroup, indicating that there may be a role for chromosome 17p genes in these cases.

WNT / SHH independent tumours comprised the majority of cases (63%), and had a peak of incidence at between 3-6 years of age. SHH tumours were most common in infancy (50% of SHH cases were observed in infants), with a second peak in adult cases (45% of adult cases). Almost all cases under 2 years of age at diagnosis were SHH-positive. WNT positive cases peaked in incidence at 10 and 20 years and were not observed in infant cases.

There were significant differences in histopathological subtypes between the subgroups. WNT cases were exclusively classic type, whereas there was a strong association between DN histology and the SHH subgroup, although this relationship was not exclusive. DN cases made up the majority of infant SHH subgroup cases and, interestingly, all DN infants were of the SHH subgroup. In contrast, there were equal proportions of DN, LCA and classic cases in SHH-expressing non-infant cases, and the majority of non-infant DN tumours were not SHH activated. WNT / SHH independent tumours were primarily classic type, although both LCA and DN histologies were also observed. No significant differences in metastatic state or gender were identified, although the WNT cases showed gender parity, in contrast to SHH and WNT / SHH independent cases, which had an excess of males cases.

The significant associations observed between medulloblastoma subgroups and specific gene, pathway and chromosomal defects strongly supports the existence of discrete molecular WNT and SHH subgroups. In contrast to the investigated genetic and chromosomal correlates of the assigned subgroups, the expression signatures

positively identified all subgroup cases and provided an accurate diagnostic test for subgroup membership. Subsequently, the meta-analysis demonstrated how SHH subgroup and DN tumours, arising in the infant and non-infant age groups, have different biological and clinical characteristics, and that SHH-positive DN tumours of infancy may represent a unique disease subgroup associated with a favourable clinical behaviour (Garre et al., 2009; McManamy et al., 2007; Rutkowski et al., 2005). The bimodal age distributions of the WNT and SHH cases also suggests that there exists additional clinical and molecular heterogeneity within these groups.

In summary, an assay that robustly identified WNT, SHH and WNT / SHH independent medulloblastoma subgroups was reported and the clinico-pathological and molecular correlates were investigated in a wider meta-analysis. The precise number of disease subgroups remains unclear (section 1.3.12.1), and requires further investigation. In chapter 4, a novel method that assessed epigenetic patterns of DNA methylation was applied to primary medulloblastomas, to investigate whether they exhibited differences in their methylation patterns, and if so, whether the differences had any utility for further disease subclassification.

6.3 Patterns of methylation have utility for disease subclassification

The utility of DNA methylation for disease subclassification has not previously been assessed in medulloblastoma. The Golden Gate DNA methylation microarray (section 2.7) was used to test whether patterns of DNA methylation had any utility for subclassification of medulloblastoma. The observed subgroups were validated in a second independent trials cohort and their relationship to the previously discussed transcriptomic subgroups was assessed by running the expression signature assay of WNT and SHH pathway activation reported in chapter 3. The clinico-pathological and molecular correlates of the assigned subgroups were investigated, and subgroup methylation biomarkers identified. Finally, the prognostic implications of subgroup membership were investigated.

Consensus clustering of the training dataset (*n* = 100) identified four clusters that were well validated in the test dataset (*n* = 130), in addition to a small group of cases that were not classifiable (*n* = 14; 6%). Integration with transcriptomic subgroup assignments revealed two subgroups characterised by WNT and SHH pathway

activation, respectively, and two subgroups independent of the WNT and SHH pathways. The WNT subgroup (*n* = 28; 13%) was associated with a WNT expression signature (Chapter 3 and Schwalbe et al., 2011), *CTNNB1* mutation, nuclear accumulation of β-catenin protein (Eberhart et al., 2000) and chromosome 6 loss (Clifford et al., 2006). WNT cases were not observed in infants, and had a bi-modal age of incidence, with peaks at 10 and 17 years, and had a predominantly classic histology. In contrast, SHH cases (*n* = 50; 23%) were enriched in infant cases, with peaks of age of incidence at 3 years and in adult cases (>16 years). SHH cases were associated with previously defined markers of the SHH group (a SHH expression signature (Chapter 3 and Schwalbe et al., 2011), GAB1 immunopositivity (Ellison et al., 2011a) and were enriched for DN histology. Both SHH and WNT cases showed gender parity.

In contrast, the remaining subgroups, group I (*n* = 94; 44%) and group II (*n* = 44; 20%), were enriched for male cases, with the enrichment being particularly strong in group II cases. Group II was also enriched for LCA histology and metastatic disease, and peaked in incidence at 4 years. Group I, the largest subgroup, was similarly enriched for metastatic disease, but was uniquely enriched for cases that displayed loss of chromosome 17p and showed a peak of incidence at 9 years. An investigation of the prognostic potential for subgroup membership revealed that the WNT subgroup was associated with a favourable prognosis. No difference was apparent for the remaining subgroups, SHH, group I and group II.

Chapters 3 and 4 both investigated the subgroups of medulloblastoma using two different methodologies, and defined clinico-pathological and molecular features associated with each group. The findings from both chapters are summarised in Table 6.1. The equivalent groupings from both studies share clinical and molecular features, and have striking similarities in their incidence.

	Subgroup	WNT	SHH	WNT / SHH independent	
Transcriptomic Signatures (chapter 3)	Clinico-pathological features	· Non-infant cases, bi-modal age distribution, peaks at 10 and 20 years · Classic histology · Gender parity	· 62% of infant cases · 45% of adult cases · DN enrichment – all infant DN cases were SHH; majority of non-infant DN cases were not SHH · Male excess	· Peak of incidence at 3-6 years old · Predominantly classic histology · Male excess	
	Molecular features	· *CTNNB1* mutation (95%) · Chromosome 6 loss (88%)	· *PTCH1* mutation (34%) · Absence of *COL1A2* methylation	· Chromosome 17p loss	
	Incidence (%)	12	24	63	
DNA methylomic classification (chapter 4)	**Subgroup**	**WNT**	**SHH**	**Group I**	**Group II**
	Clinico-pathological features	· Good prognosis · Non-infant cases, bi-modal age distribution, peaks at 10 and 17 years · Classic histology · Gender parity	· Infant cases · DN enrichment · Gender parity	· Peak of incidence at 9 years · Metastatic disease · Male excess	· Peak of incidence at 4 years · Metastatic disease · Enrichment for LCA histology · Male excess
	Molecular features	· WNT expression signature · *CTNNB1* mutation · Chromosome 6 loss · Nuclear accumulation of β-catenin	· SHH expression signature · GAB1 immunopositivity	· Enrichment for chromosome 17p loss	
	Incidence (%)	13	23	44	20

Table 6.1. Summary of clinico-pathological and molecular features from transcriptomic classification and methylomic classification of medulloblastoma.

Previous transcriptomic investigations had reported from 4 to 6 disease subgroups (Cho et al., 2011; Northcott et al., 2010; Kool et al., 2008; Thompson et al., 2006); here, using a methylomic approach, 4 subgroups were identified. Two (WNT and SHH) were similar in nature to the previously described WNT and SHH transcriptomic subgroups, as shown in Table 6.1. The two WNT / SHH independent subgroups, group I and group II showed overlapping methylation patterns and appeared to be closely related; nevertheless, significant differences between these subgroups were identified. These differences (enrichment for loss of chromosome 17p in group I, LCA enrichment in group II) matched the enrichments of loss of chromosome 17p in Northcott transcriptomic group C (Northcott et al., 2010) and LCA histology in Northcott transcriptomic subgroup D (Northcott et al., 2010) respectively, suggesting that there is a degree of overlap between the transcriptomic subgroups reported by Northcott et al. and the methylomic subgroups reported in this project, and is therefore supportive of there being four subgroups of medulloblastoma. By running paired methylomic and transcriptomic arrays, the precise relationship between methylomically- and transcriptomically-determined subgroups could be assessed, and, by clustering the samples using the transcriptomic determinants of the six subgroups described by Cho and colleagues (Cho et al., 2011) to subgroup the disease, the relationship between the methylomic subgroups and their 6 transcriptomic subgroups could also be assessed.

The demonstration that DNA methylomics is well suited to the identification of medulloblastoma subgroups enables the wider assignment of medulloblastoma subgroups in historical cohorts where fresh frozen tumour material is unavailable, since the Golden Gate array works well with DNA derived from FFPE tissues. The Golden Gate array has proven to be a useful platform for investigation of genome-wide methylation patterns, showing reproducible results from low amounts of starting material. In addition, the stability of DNA methylation offers an advantage over transcriptomic assays, which typically perform poorly on RNA derived from FFPE tissue, due to the fragility of mRNA.

The work described in chapter 4 convincingly identified four disease subgroups, and identified significant relationships with clinico-pathological and molecular disease

features, validating the use of DNA methylation microarrays for disease subclassification.

Although the WNT subgroup was associated with a favourable prognosis, there was no prognostic value to membership of the SHH, group I or group II subgroups; the work reported in chapter 5 investigated the utility of methylomic biomarkers to augment survival models of medulloblastoma, for the generation of novel survival models.

6.4 Novel disease classification models

Chapter 5 reported an investigation into the prognostic potential of methylomic markers, in addition to a consideration of the feasibility of investigating established clinico-pathological risk factors for any evidence of distinct subgroup behaviours. Unfortunately, the survival cohort used in this chapter, despite being one of the largest medulloblastoma cohorts studied to date, was not adequately powered to detect such differences, so these investigations must wait for larger cohorts, demonstrating how subsequent studies need to be carefully designed to ensure adequate power for the detection of subgroup specific differences in the prognostic potential of established clinico-pathological correlates.

Since WNT cases were associated with a particularly favourable prognosis, they were removed from subsequent investigations into establishing prognostic methylation biomarkers. After verification that established clinical covariates were significantly prognostic in the test cohort, which consisted of PNET3 trials cases, plus additional, age-matched, non-trial, primary medulloblastomas, methylation probes amenable to development as biomarkers were selected, by calculating their bimodality (Wang et al., 2009a). The 200 most bimodal probes were applied to the Cox Boost algorithm (section 2.17), along with the mandatory clinico-pathological covariates LCA histology, M+ disease and *MYCC / MYCN* amplification.

The algorithm identified three cross-validated, potentially prognostic methylated loci, two of which, *MXI1* and *IL8,* were amenable to binary classification of their methylation status, analogous to the classification that could be made through bisulfite sequencing. After confirming that binary classification of methylation status did not affect multivariate Cox models, nomograms were constructed that demonstrated how each covariate selected by the Cox boost algorithm had similar hazard ratios, supportive of a simple additive model of disease risk. In the described model, non-WNT patients were assigned to one of three risk groups (low, standard and poor), depending on their cumulative number of risk factors. Patients with 0 or 1 of the risk factors LCA histology, M+ disease, *MYCC / MYCN* amplification, *MXI1* methylation or *IL8* methylation were assigned to the low risk group; patients with 2 risk factors were assigned to the standard risk group; patients with 3 or more risk factors were

assigned to the high risk group. WNT cases were assigned to the low risk group, regardless of the status of any other risk factor.

The proposed classification scheme performed well, comfortably out-performing the current risk stratification schemes, as well as the upcoming PNET 5 / 6 clinical trials model and other state-of-the-art classification schemes. Most importantly, 46% of patients were classified as low risk, compared to only 15% of patients using the PNET 5 / 6 classification(Pizer and Clifford, 2009). This would, of course, require validation in additional, independent cohorts, but nevertheless, it represents a validation that methylation markers can be useful for disease prognostication, particularly if care is taken to pick bimodal loci, which are, in turn, suitable for development in subsequent clinical assays.

6.5 Future work and progress towards improved treatments for medulloblastoma

The molecular heterogeneity of medulloblastoma is becoming clear. Assays for the routine assignment of subgroup identity are important, since they can identify groups of patients with a good prognosis (WNT) or patients for whom SHH inhibitors might be efficacious. The assay described in chapter 3 is a robust, rapid and low cost means of assigning subgroup status and could be useful in a routine hospital pathology laboratory. An integrative analysis of all the publicly available transcriptomic datasets, discussed in section 4.5.6, could help consensus for the number of transcriptomic disease subgroups to be reached.

A robust definition of disease subgroups, identified through comprehensive integrated transcriptomic, copy number, sequencing and methylation-based analyses in large, clinical trials cohorts is needed; this would enable the inter-relationship between gene copy number, mutational status, methylation status and gene expression to be assessed, and their contribution to determination of subgroup membership to be quantified. The integrative techniques that will enable these analyses include the application of NMF techniques (section 2.9.3) across multiple studies, Genomic Identification of Significant Targets in Cancer (GISTIC) analysis, for identification of copy number aberration enrichment (Beroukhim et al., 2007), geneset enrichment analysis (GSEA) (Subramanian et al., 2005), and tests for anti-correlation to identify

hyper / hypomethylation events associated with a reduced / increased gene expression. Subsequently, expression signatures that identify the definitively determined medulloblastoma subgroups could be developed, enabling assignment of disease subgroups in a single assay, which could be useful in a diagnostic laboratory setting and, in future, could have additional prognostic implications that are currently unknown.

In chapter 4, a first generation DNA methylation microarray (Bibikova et al., 2006) was used to identify four disease subgroups. Subsequently, the Infinium (27,000 probes) (Bibikova et al., 2009) and Infinium HD (450,000 probes) (Sandoval et al., 2011) methylation arrays have been released that achieve unprecedented resolution to enable a truly genome-wide characterisation of DNA methylation, a coverage which was lacking for the Golden Gate methylation arrays employed in chapter 4. By applying these arrays to larger cohorts, it might be possible to identify additional subgroups-within-subgroups. Alternatively, as WGSBS and the methods for its analysis become more mature (Li et al., 2010; Lister et al., 2009), the use of this approach could provide a record of CpG methylation at single nucleotide resolution. There is evidence from transcriptomic studies that paediatric and adult SHH medulloblastomas are clinically and molecularly distinct (Northcott et al., 2011; Remke et al., 2011b), and this increasing refinement in the molecular classification of medulloblastomas may, in turn, be useful for further refinements of future risk stratification schemes. The use of the subgroup-associated methylation biomarkers in assays that determine subgroup status is attractive, since DNA methylation is a more stable indicator than gene expression, and, through using an MS-MLPA assay (Nygren et al., 2005) to assign patient subgroup and, potentially, prognosis, it may be possible to automate the classification of subgroup assignment in a routine diagnostic setting; an automated classification is something which is difficult to achieve by using immunohistochemistry, which is subjective in nature. The 450k arrays will also enable questions to be asked that were not possible with the Golden Gate methylation array; patterns of differential methylation could be subjected to gene ontology and gene set enrichment analysis (GSEA), to identify co-repressed networks of genes, silenced by methylation. Analysis of medulloblastoma by these 450k arrays can provide a rich source of potential disease biomarkers and, in conjunction with transcriptomic analysis, could for the first time

identify relevant genes whose expression is mediated by promoter DNA methylation patterns in medulloblastoma.

As outlined in the introduction (section 1.3.16), the molecular heterogeneity of medulloblastoma necessitates a re-evaluation of disease prognostication in the context of the disease subgroups. The relative paucity of validated molecular markers of prognosis (WNT / *MYC* gene amplification excepted) may be because the nature of the subgroup can influence the clinical relevance of the marker, and this requires investigation in larger cohorts.

The improved survival model described in chapter 5 represented a proof of concept that methylation biomarkers can be useful as prognostic biomarkers, and sets a precedent for future investigations of the prognostic potential of DNA methylation biomarkers on higher resolution platforms, in larger cohorts. For the purposes of this study, it is important that the methylation status of the genes *MXI1* and *IL8* be verified using bisulfite sequencing to confirm that the array is accurately reporting the methylation status of the promoter region. Before any consideration of the potential incorporation of *MXI1* and *IL8* methylation status into current classification schemes, the prognostic value of these genes would need to be subjected to validation in independent, preferably trials-based, disease cohorts.

In these future investigations, to identify methylation biomarkers that not only are prognostic, but also have a putative functional relevance, parallel investigations of the relationship between (i) methylation and survival, and (ii) expression and survival, could be undertaken on genome-wide array platforms, and any loci with significant relationships to survival for both gene expression and methylation could be investigated further, ensuring that any selected, putatively prognostic methylation probes are associated with a change in expression that may be functionally relevant.

This project explored themes that will become increasingly important over the coming years – although tumour genome sequencing may eventually enable truly individualised patient treatment options to be considered, at least for the next decade, broad risk- based classification will remain clinically relevant – the increased sensitivity afforded by the inclusion of additional molecular (genomic, transcriptomic and methylomic) markers have the potential to improve patient outcome, both in terms of

patient survival and for the reduction of late effects, but will have to be done carefully, to ensure that only robust and validated molecular makers are incorporated.

Chapter 7. References

ABTA. (2010) 'A Primer of Brain Tumours'. Available at: http://www.abta.org/siteFiles/pdflibrary/ABTA_BrainTumorPrimer.pdf. (Accessed: October 1st, 2011).

Aldosari, N., Bigner, S. H., Burger, P. C., Becker, L., Kepner, J. L., Friedman, H. S. and McLendon, R. E. (2002) 'MYCC and MYCN Oncogene Amplification in Medulloblastoma', *Arch Pathol Lab Med,* 126, (5), pp. 540-544.

Alston, R. D., Newton, R., Kelsey, A., Newbould, M. J., Birch, J. M., Lawson, B. and McNally, R. J. (2003) 'Childhood medulloblastoma in northwest England 1954 to 1997: incidence and survival', *Dev Med Child Neurol,* 45, (5), pp. 308-314.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool', *J Mol Biol,* 215, (3), pp. 403-410.

Anderton, J. A., Lindsey, J. C., Lusher, M. E., Gilbertson, R. J., Bailey, S., Ellison, D. W. and Clifford, S. C. (2008) 'Global analysis of the medulloblastoma epigenome identifies disease-subgroup-specific inactivation of COL1A2', *Neuro Oncol,* 10, (6), pp. 981-994.

Avet-Loiseau, H., Venuat, A. M., Terrier-Lacombe, M. J., Lellouch-Tubiana, A., Zerah, M. and Vassal, G. (1999) 'Comparative genomic hybridization detects many recurrent imbalances in central nervous system primitive neuroectodermal tumours in children', *Br J Cancer,* 79, (11-12), pp. 1843-1847.

Baeza, N., Masuoka, J., Kleihues, P. and Ohgaki, H. (2003) 'AXIN1 mutations but not deletions in cerebellar medulloblastomas', *Oncogene,* 22, (4), pp. 632-636.

Bailey, C. C., Gnekow, A., Wellek, S., Jones, M., Round, C., Brown, J., Phillips, A. and Neidhardt, M. K. (1995) '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', *Medical and Pediatric Oncology,* 25, (3), pp. 166-178.

Bailey, P. and Cushing, H. (1925) 'Medulloblastoma Cerebelli: a Common Type of Midcerebellar Glioma of Childhood', *Arch Neurol Psychiatry,* 2, p. 192.

Bair, E. and Tibshirani, R. (2004) 'Semi-supervised methods to predict patient survival from gene expression data', *PLoS Biol,* 2, (4), pp. 511-522.

Baker, M. (2010) 'Epigenome: mapping in motion', *Nat Meth,* 7, (3), pp. 181-186.

Barker, N. and Clevers, H. (2006) 'Mining the Wnt pathway for cancer therapeutics', *Nat Rev Drug Discov,* 5, (12), pp. 997-1014.

Barnfield, P. C., Zhang, X., Thanabalasingham, V., Yoshida, M. and Hui, C. C. (2005) 'Negative regulation of Gli1 and Gli2 activator function by Suppressor of fused through multiple mechanisms', *Differentiation,* 73, (8), pp. 397-405.

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I. F., Soboleva, A., Tomashevsky, M. and Edgar, R. (2007) 'NCBI GEO: mining tens of millions of expression profiles--database and tools update', *Nucleic Acids Res,* 35, (Database issue), pp. D760-D765.

Baryawno, N., Sveinbjornsson, B., Eksborg, S., Chen, C. S., Kogner, P. and Johnsen, J. I. (2010) 'Small-molecule inhibitors of phosphatidylinositol 3-kinase/Akt signaling inhibit Wnt/beta-catenin pathway cross-talk and suppress medulloblastoma growth', *Cancer Res,* 70, (1), pp. 266-276.

Baylin, S. B. and Jones, P. A. (2011) 'A decade of exploring the cancer epigenome — biological and translational implications', *Nat Rev Cancer,* 11, (10), pp. 726-734.

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J. C., Huang, J. H., Alexander, S., Du, J., Kau, T., Thomas, R. K., Shah, K., Soto, H., Perner, S., Prensner, J., Debiasi, R. M., Demichelis, F., Hatton, C., Rubin, M.

A., Garraway, L. A., Nelson, S. F., Liau, L., Mischel, P. S., Cloughesy, T. F., Meyerson, M., Golub, T. A., Lander, E. S., Mellinghoff, I. K. and Sellers, W. R. (2007) 'Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma', *Proc Natl Acad Sci U S A,* 104, (50), pp. 20007-20012.

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R. and Gunderson, K. L. (2009) 'Genome-wide DNA methylation profiling using Infinium® assay', *Epigenomics,* 1, (1), pp. 177-200.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E. W., Wu, B., Doucet, D., Thomas, N. J., Wang, Y., Vollmer, E., Goldmann, T., Seifart, C., Jiang, W., Barker, D. L., Chee, M. S., Floros, J. and Fan, J. B. (2006) 'High-throughput DNA methylation profiling using universal bead arrays', *Genome Res,* 16, (3), pp. 383-393.

Bigner, S. H., Friedman, H. S., Vogelstein, B., Oakes, W. J. and Bigner, D. D. (1990) 'Amplification of the c-myc gene in human medulloblastoma cell lines and xenografts', *Cancer Res,* 50, (8), pp. 2347-2350.

Bigner, S. H., Mark, J., Friedman, H. S., Biegel, J. A. and Bigner, D. D. (1988) 'Structural chromosomal abnormalities in human medulloblastoma', *Cancer Genetics and Cytogenetics,* 30, (1), pp. 91-101.

Bigner, S. H., McLendon, R. E., Fuchs, M., McKeever, P. E. and Friedman, H. S. (1997) 'Chromosomal characteristics of childhood brain tumors', *Cancer Genetics and Cytogenetics,* 97, (2), pp. 125-134.

Binder, H., Allignol, A., Schumacher, M. and Beyersmann, J. (2009) 'Boosting for high-dimensional time-to-event data with competing risks', *Bioinformatics,* 25, (7), pp. 890-896.

Bird, A. (2002) 'DNA methylation patterns and epigenetic memory', *Genes Dev,* 16, (1), pp. 6-21.

Bird, A. P. (1980) 'DNA methylation and the frequency of CpG in animal DNA', *Nucleic Acids Res,* 8, (7), pp. 1499-1504.

Blackburn, E. H. (1991) 'Structure and function of telomeres', *Nature,* 350, (6319), pp. 569-573.

Bland, J. M. and Altman, D. G. (2004) 'The logrank test', *BMJ,* 328, (7447), p. 1073.

Bland, M. J. and Altman, D. G. (1986) 'Statistical methods for assessing agreement between two methods of clinical measurement', *The Lancet,* 327, (8476), pp. 307-310.

Bradburn, M. J., Clark, T. G., Love, S. B. and Altman, D. G. (2003) 'Survival analysis part II: multivariate data analysis--an introduction to concepts and methods', *Br J Cancer,* 89, (3), pp. 431-436.

Brown, H. G., Kepner, J. L., Perlman, E. J., Friedman, H. S., Strother, D. R., Duffner, P. K., Kun, L. E., Goldthwaite, P. T. and Burger, P. C. (2000) '"Large cell/anaplastic" medulloblastomas: a Pediatric Oncology Group Study', *J Neuropathol Exp Neurol,* 59, (10), pp. 857-865.

Brugieres, L., Pierron, G., Chompret, A., Paillerets, B. B., Di Rocco, F., Varlet, P., Pierre-Kahn, A., Caron, O., Grill, J. and Delattre, O. (2010) 'Incomplete penetrance of the predisposition to medulloblastoma associated with germ-line SUFU mutations', *J Med Genet,* 47, (2), pp. 142-144.

Brunet, J.-P., Tamayo, P., Golub, T. R. and Mesirov, J. P. (2004) 'Metagenes and molecular pattern discovery using matrix factorization', *Proc Natl Acad Sci U S A,* 101, (12), pp. 4164-4169.

Brunner, A. L., Johnson, D. S., Kim, S. W., Valouev, A., Reddy, T. E., Neff, N. F., Anton, E., Medina, C., Nguyen, L., Chiao, E., Oyolu, C. B., Schroth, G. P., Absher, D. M.,

Baker, J. C. and Myers, R. M. (2009) 'Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver', *Genome Res,* 19, (6), pp. 1044-1056.

Cairns, J. M., Dunning, M. J., Ritchie, M. E., Russell, R. and Lynch, A. G. (2008) 'BASH: a tool for managing BeadArray spatial artefacts', *Bioinformatics,* 24, (24), pp. 2921-2922.

Carlotti Jr, C. G., Smith, C. and Rutka, J. T. (2008) 'The molecular genetics of medulloblastoma: an assessment of new therapeutic targets', *Neurosurg Rev,* 31, (4), pp. 359-368.

Cavenee, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphree, A. L., Strong, L. C. and White, R. L. (1983) 'Expression of recessive alleles by chromosomal mechanisms in retinoblastoma', *Nature,* 305, (5937), pp. 779-784.

CBTRUS (2011) *Selected primary (malignant and non-malignant) brain and central nervous system tumor age-specific incidence rates by age at diagnosis, CBTRUS, 1998-2002*. Available at: http://cbtrus.org/2005-2006/tables/2006.table12.pdf (Accessed: September 15th, 2011).

CCRG (2010) *Survival from Childhood Cancer, Great Britain, 1971-2005*. Available at: http://www.ccrg.ox.ac.uk/datasets/survivalrates.htm (Accessed: September 20th, 2011).

Chang, C. C., Lin, C.J. (2001) *LIBSVM : a library for support vector machines* (3.1). Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm/ (Accessed: August 14th, 2011).

Chang, C. H., Housepian, E. M. and Herbert, C., Jr. (1969) 'An operative staging system and a megavoltage radiotherapeutic technic for cerebellar medulloblastomas', *Radiology,* 93, (6), pp. 1351-1359.

Charles, J. S. (2004) 'Principles of Tumor Suppression', *Cell,* 116, (2), pp. 235-246.

Chen, W. Y., Zeng, X., Carter, M. G., Morrell, C. N., Chiu Yen, R. W., Esteller, M., Watkins, D. N., Herman, J. G., Mankowski, J. L. and Baylin, S. B. (2003) 'Heterozygous disruption of Hic1 predisposes mice to a gender-dependent spectrum of malignant tumors', *Nat Genet,* 33, (2), pp. 197-202.

Chi, A. S. and Bernstein, B. E. (2009) 'Developmental biology. Pluripotent chromatin state', *Science,* 323, (5911), pp. 220-221.

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. (2011) 'Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome', *J Clin Oncol,* 29, (11), pp. 1424-1430.

Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S. and Bayley, H. (2009) 'Continuous base identification for single-molecule nanopore DNA sequencing', *Nat Nanotechnol,* 4, (4), pp. 265-270.

Clifford, S. C., Lusher, M. E., Lindsey, J. C., Langdon, J. A., Gilbertson, R. J., Straughton, D. and Ellison, D. W. (2006) 'Wnt/Wingless pathway activation and chromosome 6 loss characterize a distinct molecular sub-group of medulloblastomas associated with a favorable prognosis', *Cell Cycle,* 5, (22), pp. 2666-2670.

Costello, J. F., Fruhwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., Gao, X., Wright, F. A., Feramisco, J. D., Peltomaki, P., Lang, J. C., Schuller, D. E., Yu, L., Bloomfield, C. D., Caligiuri, M. A., Yates, A., Nishikawa, R., Su Huang, H., Petrelli, N. J., Zhang, X., O'Dorisio, M. S., Held, W. A., Cavenee, W. K. and Plass, C. (2000) 'Aberrant CpG-island methylation has non-random and tumour-type-specific patterns', *Nat Genet,* 24, (2), pp. 132-138.

Cox, D. R. (1972) 'Regression Models and Life-Tables', *Journal of the Royal Statistical Society. Series B (Methodological),* 34, (2), pp. 187-220.

Crawford, J. R., MacDonald, T. J. and Packer, R. J. (2007) 'Medulloblastoma in childhood: new biological advances', *The Lancet Neurology,* 6, (12), pp. 1073-1085.

CRUK (2011) *CancerStats Incidence 2008 - UK*. Available at: http://info.cancerresearchuk.org/prod_consump/groups/cr_common/@nre/@sta/documents/generalcontent/cr_072111.pdf (Accessed: September 19, 2011).

Curran, E. K., Sainani, K. L., Le, G. M., Propp, J. M. and Fisher, P. G. (2009) 'Gender affects survival for medulloblastoma only in older children and adults: a study from the Surveillance Epidemiology and End Results Registry', *Pediatr Blood Cancer,* 52, (1), pp. 60-64.

Dahmane, N. and Ruiz i Altaba, A. (1999) 'Sonic hedgehog regulates the growth and patterning of the cerebellum', *Development,* 126, (14), pp. 3089-3100.

Dahmen, R. P., Koch, A., Denkhaus, D., Tonn, J. C., Sorensen, N., Berthold, F., Behrens, J., Birchmeier, W., Wiestler, O. D. and Pietsch, T. (2001) 'Deletions of AXIN1, a component of the WNT/wingless pathway, in sporadic medulloblastomas', *Cancer Res,* 61, (19), pp. 7039-7043.

Dammann, R., Li, C., Yoon, J. H., Chin, P. L., Bates, S. and Pfeifer, G. P. (2000) 'Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3', *Nat Genet,* 25, (3), pp. 315-319.

de Bustros, A., Nelkin, B. D., Silverman, A., Ehrlich, G., Poiesz, B. and Baylin, S. B. (1988) 'The short arm of chromosome 11 is a "hot spot" for hypermethylation in human neoplasia', *Proc Natl Acad Sci U S A,* 85, (15), pp. 5693-5697.

de Haas, T., Hasselt, N., Troost, D., Caron, H., Popovic, M., Zadravec-Zaletel, L., Grajkowska, W., Perek, M., Osterheld, M. C., Ellison, D., Baas, F., Versteeg, R. and Kool, M. (2008) 'Molecular risk stratification of medulloblastoma patients based on immunohistochemical analysis of MYC, LDHB, and CCNB1 expression', *Clin Cancer Res,* 14, (13), pp. 4154-4160.

de la Iglesia, N., Konopka, G., Lim, K. L., Nutt, C. L., Bromberg, J. F., Frank, D. A., Mischel, P. S., Louis, D. N. and Bonni, A. (2008) 'Deregulation of a STAT3-interleukin 8 signaling pathway promotes human glioblastoma cell proliferation and invasiveness', *J Neurosci,* 28, (23), pp. 5870-5878.

De Smaele, E., Di Marcotullio, L., Ferretti, E., Screpanti, I., Alesse, E. and Gulino, A. (2004) 'Chromosome 17p deletion in human medulloblastoma: a missing checkpoint in the Hedgehog pathway', *Cell Cycle,* 3, (10), pp. 1263-1266.

Deaton, A. M. and Bird, A. (2011) 'CpG islands and the regulation of transcription', *Genes Dev,* 25, (10), pp. 1010-1022.

Di Marcotullio, L., Ferretti, E., De Smaele, E., Argenti, B., Mincione, C., Zazzeroni, F., Gallo, R., Masuelli, L., Napolitano, M., Maroder, M., Modesti, A., Giangaspero, F., Screpanti, I., Alesse, E. and Gulino, A. (2004) 'REN(KCTD11) is a suppressor of Hedgehog signaling and is deleted in human medulloblastoma', *Proc Natl Acad Sci U S A,* 101, (29), pp. 10833-10838.

Diede, S. J., Guenthoer, J., Geng, L. N., Mahoney, S. E., Marotta, M., Olson, J. M., Tanaka, H. and Tapscott, S. J. (2010) 'DNA methylation of developmental genes in pediatric medulloblastomas identified by denaturation analysis of methylation differences', *Proc Natl Acad Sci U S A,* 107, (1), pp. 234-239.

Dijkgraaf, G. J., Alicke, B., Weinmann, L., Januario, T., West, K., Modrusan, Z., Burdick, D., Goldsmith, R., Robarge, K., Sutherlin, D., Scales, S. J., Gould, S. E., Yauch, R. L. and de Sauvage, F. J. (2011) 'Small molecule inhibition of GDC-0449 refractory Smoothened mutants and downstream mechanisms of drug resistance', *Cancer*

Res, 71, (2), pp. 435-444.

Dobrovic, A. and Simpfendorfer, D. (1997) 'Methylation of the BRCA1 gene in sporadic breast cancer', *Cancer Res,* 57, (16), pp. 3347-3350.

Down, T. A., Rakyan, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E., Graf, S., Johnson, N., Herrero, J., Tomazou, E. M., Thorne, N. P., Backdahl, L., Herberth, M., Howe, K. L., Jackson, D. K., Miretti, M. M., Marioni, J. C., Birney, E., Hubbard, T. J., Durbin, R., Tavare, S. and Beck, S. (2008) 'A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis', *Nat Biotechnol,* 26, (7), pp. 779-785.

Druker, B. J., Guilhot, F., O'Brien, S. G., Gathmann, I., Kantarjian, H., Gattermann, N., Deininger, M. W. N., Silver, R. T., Goldman, J. M., Stone, R. M., Cervantes, F., Hochhaus, A., Powell, B. L., Gabrilove, J. L., Rousselot, P., Reiffers, J., Cornelissen, J. J., Hughes, T., Agis, H., Fischer, T., Verhoef, G., Shepherd, J., Saglio, G., Gratwohl, A., Nielsen, J. L., Radich, J. P., Simonsson, B., Taylor, K., Baccarani, M., So, C., Letvak, L. and Larson, R. A. (2006) 'Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia', *N Engl J Med,* 355, (23), pp. 2408-2417.

Dunning, M. J., Smith, M. L., Ritchie, M. E. and Tavare, S. (2007) 'beadarray: R classes and methods for Illumina bead-based data', *Bioinformatics,* 23, (16), pp. 2183-2184.

Eagle, L. R., Yin, X., Brothman, A. R., Williams, B. J., Atkin, N. B. and Prochownik, E. V. (1995) 'Mutation of the MXI1 gene in prostate cancer', *Nat Genet,* 9, (3), pp. 249-255.

Eberhart, C. G. (2011) 'Molecular diagnostics in embryonal brain tumors', *Brain Pathol,* 21, (1), pp. 96-104.

Eberhart, C. G. and Burger, P. C. (2003) 'Anaplasia and grading in medulloblastomas',

Brain Pathol, 13, (3), pp. 376-385.

Eberhart, C. G., Kepner, J. L., Goldthwaite, P. T., Kun, L. E., Duffner, P. K., Friedman, H. S., Strother, D. R. and Burger, P. C. (2002) 'Histopathologic grading of medulloblastomas: a Pediatric Oncology Group study', *Cancer,* 94, (2), pp. 552-560.

Eberhart, C. G., Kratz, J., Wang, Y., Summers, K., Stearns, D., Cohen, K., Dang, C. V. and Burger, P. C. (2004) 'Histopathological and molecular prognostic markers in medulloblastoma: c-myc, N-myc, TrkC, and anaplasia', *J Neuropathol Exp Neurol,* 63, (5), pp. 441-449.

Eberhart, C. G., Tihan, T. and Burger, P. C. (2000) 'Nuclear localization and mutation of beta-catenin in medulloblastomas', *J Neuropathol Exp Neurol,* 59, (4), pp. 333-337.

Ellison, D. (2002) 'Classifying the medulloblastoma: insights from morphology and molecular genetics', *Neuropathology and Applied Neurobiology,* 28, (4), pp. 257-282.

Ellison, D. W., Clifford, S. C., Gajjar, A. and Gilbertson, R. J. (2003) 'What's new in neuro-oncology? Recent advances in medulloblastoma', *European Journal of Paediatric Neurology,* 7, (2), pp. 53-66.

Ellison, D. W., Clifford, S.C., Giangaspero, F. (2006) 'Medulloblastoma', in McLendon, R. E., Rosenblum, M.K., Bigner, D.D. (ed), *Russell & Rubenstein's Pathology of Tumours of the Nervous System.* Seventh ed. London: Hodder Arnold, pp. 247-264.

Ellison, D. W., Dalton, J., Kocak, M., Nicholson, S. L., Fraga, C., Neale, G., Kenney, A. M., Brat, D. J., Perry, A., Yong, W. H., Taylor, R. E., Bailey, S., Clifford, S. C. and Gilbertson, R. J. (2011a) 'Medulloblastoma: clinicopathological correlates of SHH, WNT, and non-SHH/WNT molecular subgroups', *Acta Neuropathol,* 121,

(3), pp. 381-396.

Ellison, D. W., Kocak, M., Dalton, J., Megahed, H., Lusher, M. E., Ryan, S. L., Zhao, W., Nicholson, S. L., Taylor, R. E., Bailey, S. and Clifford, S. C. (2011b) 'Definition of disease-risk stratification groups in childhood medulloblastoma using combined clinical, pathologic, and molecular variables', *J Clin Oncol,* 29, (11), pp. 1400-1407.

Ellison, D. W., Onilude, O. E., Lindsey, J. C., Lusher, M. E., Weston, C. L., Taylor, R. E., Pearson, A. D. and Clifford, S. C. (2005) 'beta-Catenin status predicts a favorable outcome in childhood medulloblastoma: the United Kingdom Children's Cancer Study Group Brain Tumour Committee', *J Clin Oncol,* 23, (31), pp. 7951-7957.

Emadian, S. M., McDonald, J. D., Gerken, S. C. and Fults, D. (1996) 'Correlation of chromosome 17p loss with clinical outcome in medulloblastoma', *Clin Cancer Res,* 2, (9), pp. 1559-1564.

Engstrom, L. D., Youkilis, A. S., Gorelick, J. L., Zheng, D., Ackley, V., Petroff, C. A., Benson, L. Q., Coon, M. R., Zhu, X., Hanash, S. M. and Wechsler, D. S. (2004) 'Mxi1-0, an alternatively transcribed Mxi1 isoform, is overexpressed in glioblastomas', *Neoplasia,* 6, (5), pp. 660-673.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A. and Bickmore, W. A. (2010) 'Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination', *Mol Cell,* 38, (3), pp. 452-464.

Esteller, M. (2006) 'The necessity of a human epigenome project', *Carcinogenesis,* 27, (6), pp. 1121-1125.

Esteller, M., Corn, P. G., Baylin, S. B. and Herman, J. G. (2001) 'A gene hypermethylation profile of human cancer', *Cancer Res,* 61, (8), pp. 3225-3229.

Esteller, M., Garcia-Foncillas, J., Andion, E., Goodman, S. N., Hidalgo, O. F., Vanaclocha, V., Baylin, S. B. and Herman, J. G. (2000) 'Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents', *N Engl J Med,* 343, (19), pp. 1350-1354.

Esteller, M., Hamilton, S. R., Burger, P. C., Baylin, S. B. and Herman, J. G. (1999) 'Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia', *Cancer Res,* 59, (4), pp. 793-797.

Fattet, S., Haberler, C., Legoix, P., Varlet, P., Lellouch-Tubiana, A., Lair, S., Manie, E., Raquin, M. A., Bours, D., Carpentier, S., Barillot, E., Grill, J., Doz, F., Puget, S., Janoueix-Lerosey, I. and Delattre, O. (2009) 'Beta-catenin status in paediatric medulloblastomas: correlation of immunohistochemical expression with mutational status, genetic profiles, and clinical characteristics', *J Pathol,* 218, (1), pp. 86-94.

Fearon, E. R. and Vogelstein, B. (1990) 'A genetic model for colorectal tumorigenesis', *Cell,* 61, (5), pp. 759-767.

Ferguson, S. and Lesniak, M. S. (2005) 'Percival Bailey and the classification of brain tumors', *Neurosurg Focus,* 18, (4), pp. e7. 1-6.

Fernandez, A. F., Assenov, Y., Martin-Subero, J. I., Balint, B., Siebert, R., Taniguchi, H., Yamamoto, H., Hidalgo, M., Tan, A. C., Galm, O., Ferrer, I., Sanchez-Cespedes, M., Villanueva, A., Carmona, J., Sanchez-Mut, J. V., Berdasco, M., Moreno, V., Capella, G., Monk, D., Ballestar, E., Ropero, S., Martinez, R., Sanchez-Carbayo, M., Prosper, F., Agirre, X., Fraga, M. F., Grana, O., Perez-Jurado, L., Mora, J., Puig, S., Prat, J., Badimon, L., Puca, A. A., Meltzer, S. J., Lengauer, T., Bridgewater, J., Bock, C. and Esteller, M. (2011) 'A DNA methylation fingerprint of 1628 human samples', *Genome Res* (in press).

Ferretti, E., De Smaele, E., Di Marcotullio, L., Screpanti, I. and Gulino, A. (2005)

'Hedgehog checkpoints in medulloblastoma: the chromosome 17p deletion paradigm', *Trends Mol Med,* 11, (12), pp. 537-545.

Fleck, B. J., Pandya, A., Vanner, L., Kerkering, K. and Bodurtha, J. (2001) 'Coffin-Siris syndrome: review and presentation of new cases from a questionnaire study', *Am J Med Genet,* 99, (1), pp. 1-7.

Fossati, P., Ricardi, U. and Orecchia, R. (2009) 'Pediatric medulloblastoma: toxicity of current treatment and potential role of protontherapy', *Cancer Treat Rev,* 35, (1), pp. 79-96.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R. A., Peinado, M. A. and Clark, S. J. (2006) 'Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band', *Nat Genet,* 38, (5), pp. 540-549.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L. and Paul, C. L. (1992) 'A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands', *Proc Natl Acad Sci U S A,* 89, (5), pp. 1827-1831.

Fruhwald, M. C., O'Dorisio, M. S., Dai, Z., Tanner, S. M., Balster, D. A., Gao, X., Wright, F. A. and Plass, C. (2001) 'Aberrant promoter methylation of previously unidentified target genes is a common abnormality in medulloblastomas--implications for tumor biology and potential clinical utility', *Oncogene,* 20, (36), pp. 5033-5042.

Fults, D., Pedone, C., Dai, C. and Holland, E. C. (2002) 'MYC expression promotes the proliferation of neural progenitor cells in culture and in vivo', *Neoplasia,* 4, (1), pp. 32-39.

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M. and Haussler, D. (2000) 'Support vector machine classification and validation of cancer tissue

samples using microarray expression data', *Bioinformatics,* 16, (10), pp. 906-914.

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. (2006) '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), pp. 813-820.

Gajjar, A. and Clifford, S. C. (2010) *'Embryonal Tumors'* in Estlin, E.J, Gilbertson, R.J., Wynn, R.F. (ed), *Pediatric Hematology and Oncology.* First ed. Oxford: Blackwell, p. 36.

Gao, Y. and Church, G. (2011) 'Improving molecular cancer class discovery through sparse non-negative matrix factorization', *Bioinformatics,* 21, (21), pp. 3970-3975.

Garber, K. (2009) 'Drugging the Wnt Pathway: Problems And Progress', *J Natl Cancer Inst,* 101, (8), pp. 548-550.

Gardiner-Garden, M. and Frommer, M. (1987) 'CpG Islands in vertebrate genomes', *Journal of Molecular Biology,* 196, (2), pp. 261-282.

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. (2009) 'Medulloblastoma variants: age-dependent occurrence and relation to Gorlin syndrome--a new clinical perspective', *Clin Cancer Res,* 15, (7), pp. 2463-2471.

Garre, M. L., Gandus, S., Cesana, B., Haupt, R., De Bernardi, B., Comelli, A., Ferrando, A.,

Stella, G., Vitali, M. L., Picco, P. and et al. (1994) 'Health status of long-term survivors after cancer in childhood. Results of an uniinstitutional study in Italy', *Am J Pediatr Hematol Oncol,* 16, (2), pp. 143-152.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. and Zhang, J. (2004) 'Bioconductor: open software development for computational biology and bioinformatics', *Genome Biol,* 5, (10), pp. R80.

Giangaspero, F., Wellek, S., Masuoka, J., Gessi, M., Kleihues, P. and Ohgaki, H. (2006) 'Stratification of medulloblastoma on the basis of histopathological grading', *Acta Neuropathol,* 112, (1), pp. 5-12.

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. (2010) 'Subtypes of medulloblastoma have distinct developmental origins', *Nature,* 468, (7327), pp. 1095-1099.

Gilbertson, R., Wickramasinghe, C., Hernan, R., Balaji, V., Hunt, D., Jones-Wallace, D., Crolla, J., Perry, R., Lunec, J., Pearson, A. and Ellison, D. (2001) 'Clinical and molecular stratification of disease risk in medulloblastoma', *Br J Cancer,* 85, (5), pp. 705-712.

Gilbertson, R. J. (2004) 'Medulloblastoma: signalling a change in treatment', *Lancet Oncol,* 5, (4), pp. 209-218.

Gilbertson, R. J. (2011) 'Finding the Perfect Partner for Medulloblastoma Prognostication', *J Clin Oncol,* 29, (29), pp. 3841-3842.

Gilbertson, R. J. and Ellison, D. W. (2008) 'The Origins of Medulloblastoma Subtypes', *Annual Review of Pathology: Mechanisms of Disease,* 3, (1), pp. 341-365.

Glockner, S. C., Dhir, M., Yi, J. M., McGarvey, K. E., Van Neste, L., Louwagie, J., Chan, T. A., Kleeberger, W., de Bruine, A. P., Smits, K. M., Khalid-de Bakker, C. A., Jonkers, D. M., Stockbrugger, R. W., Meijer, G. A., Oort, F. A., Iacobuzio-Donahue, C., Bierau, K., Herman, J. G., Baylin, S. B., Van Engeland, M., Schuebel, K. E. and Ahuja, N. (2009) 'Methylation of TFPI2 in stool DNA: a potential novel biomarker for the detection of colorectal cancer', *Cancer Res,* 69, (11), pp. 4691-4699.

Goelz, S. E., Vogelstein, B., Hamilton, S. R. and Feinberg, A. P. (1985) 'Hypomethylation of DNA from benign and malignant human colon neoplasms', *Science,* 228, (4696), pp. 187-190.

Goldberg, E. K., Glendening, J. M., Karanjawala, Z., Sridhar, A., Walker, G. J., Hayward, N. K., Rice, A. J., Kurera, D., Tebha, Y. and Fountain, J. W. (2000) 'Localization of Multiple Melanoma Tumor-Suppressor Genes on Chromosome 11 by Use of Homozygosity Mapping-of-Deletions Analysis', *The American Journal of Human Genetics,* 67, (2), pp. 417-431.

Goodrich, L. V., Milenkovic, L., Higgins, K. M. and Scott, M. P. (1997) 'Altered neural cell fates and medulloblastoma in mouse patched mutants', *Science,* 277, (5329), pp. 1109-1113.

Gorlin, R. J., Vickers, R. A., Kellen, E. and Williamson, J. J. (1965) 'Multiple Basal-Cell Nevi Syndrome. An Analysis of a Syndrome Consisting of Multiple Nevoid Basal-Cell Carcinoma, Jaw Cysts, Skeletal Anomalies, Medulloblastoma, and Hyporesponsiveness to Parathormone', *Cancer,* 18, pp. 89-104.

Grandori, C., Cowley, S. M., James, L. P. and Eisenman, R. N. (2000) 'The MYC/MAX/MAD Network and the Transcriptional Control of Cell Behavior',

Annual Review of Cell and Developmental Biology, 16, (1), pp. 653-699.

Greger, V., Debus, N., Lohmann, D., Hopping, W., Passarge, E. and Horsthemke, B. (1994) 'Frequency and parental origin of hypermethylated RB1 alleles in retinoblastoma', *Hum Genet,* 94, (5), pp. 491-496.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M. and et al. (1991) 'Identification and characterization of the familial adenomatous polyposis coli gene', *Cell,* 66, (3), pp. 589-600.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., Eussen, B. H., de Klein, A., Wessels, L., de Laat, W. and van Steensel, B. (2008) 'Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions', *Nature,* 453, (7197), pp. 948-951.

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. and Young, R. A. (2007) 'A chromatin landmark and transcription initiation at most promoters in human cells', *Cell,* 130, (1), pp. 77-88.

Gunderson, K. L., Kruglyak, S., Graige, M. S., Garcia, F., Kermani, B. G., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J., Doucet, D., Milewski, M., Yang, R., Siegmund, C., Haas, J., Zhou, L., Oliphant, A., Fan, J. B., Barnard, S. and Chee, M. S. (2004) 'Decoding randomly ordered DNA arrays', *Genome Res,* 14, (5), pp. 870-877.

Guo, J. U., Su, Y., Zhong, C., Ming, G. L. and Song, H. (2011) 'Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain', *Cell,* 145, (3), pp. 423-434.

Gupta, S., Takebe, N. and Lorusso, P. (2010) 'Targeting the Hedgehog pathway in cancer', *Ther Adv Med Oncol,* 2, (4), pp. 237-250.

Gurney, J. G., Smith, M. A., Bunin, G. R. (1999) 'CNS and miscellaneous intracranial and intraspinal neoplasms', in Ries, L. A. G., Smith,M.A.,Gurney,J.G.,Linet,M.,Tamra,T.,Young,J.L.,Bunin,G.R. (ed), *Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975-1995.* Bethseda, MD: NIH Pub. No. 99-4649.

Hahn, H., Wicking, C., Zaphiropoulous, P. G., Gailani, M. R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A. B., Gillies, S., Negus, K., Smyth, I., Pressman, C., Leffell, D. J., Gerrard, B., Goldstein, A. M., Dean, M., Toftgard, R., Chenevix-Trench, G., Wainwright, B. and Bale, A. E. (1996) 'Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome', *Cell,* 85, (6), pp. 841-851.

Halperin, E. C., Watson, D. M. and George, S. L. (2001) 'Duration of symptoms prior to diagnosis is related inversely to presenting disease stage in children with medulloblastoma', *Cancer,* 91, (8), pp. 1444-1450.

Hamilton, S. R., Liu, B., Parsons, R. E., Papadopoulos, N., Jen, J., Powell, S. M., Krush, A. J., Berk, T., Cohen, Z., Tetu, B. and et al. (1995) 'The molecular basis of Turcot's syndrome', *N Engl J Med,* 332, (13), pp. 839-847.

Hanahan, D. and Weinberg, R. A. (2000) 'The hallmarks of cancer', *Cell,* 100, (1), pp. 57-70.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell,* 144, (5), pp. 646-674.

Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Hatada, I., Shibata, H., Kawai, J., Hirose, K., Watanabe, S., Fushiki, S., Wada, S. and et al. (1993) 'Restriction landmark genomic scanning method and its various applications', *Electrophoresis,* 14, (4), pp. 251-258.

Heagerty, P. J., Lumley, T. and Pepe, M. S. (2000) 'Time-dependent ROC curves for

censored survival data and a diagnostic marker', *Biometrics,* 56, (2), pp. 337-344.

Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D. and Baylin, S. B. (1996) 'Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands', *Proc Natl Acad Sci U S A,* 93, (18), pp. 9821-9826.

Herman, J. G., Latif, F., Weng, Y., Lerman, M. I., Zbar, B., Liu, S., Samid, D., Duan, D. S., Gnarra, J. R., Linehan, W. M. and et al. (1994) 'Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma', *Proc Natl Acad Sci U S A,* 91, (21), pp. 9700-9704.

Herman, J. G., Merlo, A., Mao, L., Lapidus, R. G., Issa, J. P., Davidson, N. E., Sidransky, D. and Baylin, S. B. (1995) 'Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers', *Cancer Res,* 55, (20), pp. 4525-4530.

Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J. P., Markowitz, S., Willson, J. K., Hamilton, S. R., Kinzler, K. W., Kane, M. F., Kolodner, R. D., Vogelstein, B., Kunkel, T. A. and Baylin, S. B. (1998) 'Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma', *Proc Natl Acad Sci U S A,* 95, (12), pp. 6870-6875.

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T. S., Harte, R. A., Hsu, F., Hillman-Jackson, J., Kuhn, R. M., Pedersen, J. S., Pohl, A., Raney, B. J., Rosenbloom, K. R., Siepel, A., Smith, K. E., Sugnet, C. W., Sultan-Qurraie, A., Thomas, D. J., Trumbower, H., Weber, R. J., Weirauch, M., Zweig, A. S., Haussler, D. and Kent, W. J. (2006) 'The UCSC Genome Browser Database: update 2006', *Nucleic Acids Res,* 34, (Database issue), pp. D590-598.

Hirsch, B., Shimamura, A., Moreau, L., Baldinger, S., Hag-alshiekh, M., Bostrom, B., Sencer, S. and D'Andrea, A. D. (2004) 'Association of biallelic BRCA2/FANCD1

mutations with spontaneous chromosomal instability and solid tumors of childhood', *Blood,* 103, (7), pp. 2554-2559.

Hooper, J. E. and Scott, M. P. (2005) 'Communicating with Hedgehogs', *Nat Rev Mol Cell Biol,* 6, (4), pp. 306-317.

Huang, H., Mahler-Araujo, B. M., Sankila, A., Chimelli, L., Yonekawa, Y., Kleihues, P. and Ohgaki, H. (2000) 'APC mutations in sporadic medulloblastomas', *Am J Pathol,* 156, (2), pp. 433-437.

Huang, S.-M. A., Mishina, Y. M., Liu, S., Cheung, A., Stegmeier, F., Michaud, G. A., Charlat, O., Wiellette, E., Zhang, Y., Wiessner, S., Hild, M., Shi, X., Wilson, C. J., Mickanin, C., Myer, V., Fazal, A., Tomlinson, R., Serluca, F., Shao, W., Cheng, H., Shultz, M., Rau, C., Schirle, M., Schlegl, J., Ghidelli, S., Fawell, S., Lu, C., Curtis, D., Kirschner, M. W., Lengauer, C., Finan, P. M., Tallarico, J. A., Bouwmeester, T., Porter, J. A., Bauer, A. and Cong, F. (2009) 'Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling', *Nature,* 461, (7264), pp. 614-620.

Huang, T. H., Perry, M. R. and Laux, D. E. (1999) 'Methylation profiling of CpG islands in human breast cancer cells', *Hum Mol Genet,* 8, (3), pp. 459-470.

Illumina (2011) *VeraCode GoldenGate Methylation Assay*. Available at: http://www.illumina.com/technology/veracode_methylation_assay.ilmn (Accessed: September 15th, 2011).

Ingham, P. W. and McMahon, A. P. (2001) 'Hedgehog signaling in animal development: paradigms and principles', *Genes Dev,* 15, (23), pp. 3059-3087.

Irgon, J., Huang, C. C., Zhang, Y., Talantov, D., Bhanot, G. and Szalma, S. (2010) 'Robust multi-tissue gene panel for cancer detection', *BMC Cancer,* 10, pp. 319.

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B. and Speed, T. P. (2003) 'Summaries of Affymetrix GeneChip probe level data', *Nucleic Acids Res,* 31, (4),

pp. e15.

Irizarry, R. A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S. A., Jeddeloh, J. A., Wen, B. and Feinberg, A. P. (2008) 'Comprehensive high-throughput arrays for relative methylation (CHARM)', *Genome Res,* 18, (5), pp. 780-90.

Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J. B., Sabunciyan, S. and Feinberg, A. P. (2009) 'The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores', *Nat Genet,* 41, (2), pp. 178-186.

Issa, J. P., Ottaviano, Y. L., Celano, P., Hamilton, S. R., Davidson, N. E. and Baylin, S. B. (1994) 'Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon', *Nat Genet,* 7, (4), pp. 536-540.

Jacobs, I. J., Kohler, M. F., Wiseman, R. W., Marks, J. R., Whitaker, R., Kerns, B. A., Humphrey, P., Berchuck, A., Ponder, B. A. and Bast, R. C., Jr. (1992) 'Clonal origin of epithelial ovarian carcinoma: analysis by loss of heterozygosity, p53 mutation, and X-chromosome inactivation', *J Natl Cancer Inst,* 84, (23), pp. 1793-1798.

Jänne, P. A., Engelman, J. A. and Johnson, B. E. (2005) 'Epidermal Growth Factor Receptor Mutations in Non–Small-Cell Lung Cancer: Implications for Treatment and Tumor Biology', *J Clin Oncol,* 23, (14), pp. 3227-3234.

Jeong, S., Liang, G., Sharma, S., Lin, J. C., Choi, S. H., Han, H., Yoo, C. B., Egger, G., Yang, A. S. and Jones, P. A. (2009) 'Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA', *Mol Cell Biol,* 29, (19), pp. 5366-5376.

Johnson, R. L., Rothman, A. L., Xie, J., Goodrich, L. V., Bare, J. W., Bonifas, J. M., Quinn, A. G., Myers, R. M., Cox, D. R., Epstein, E. H., Jr. and Scott, M. P. (1996) 'Human

homolog of patched, a candidate gene for the basal cell nevus syndrome', *Science,* 272, (5268), pp. 1668-1671.

Jones, P. A. and Baylin, S. B. (2002) 'The fundamental role of epigenetic events in cancer', *Nat Rev Genet,* 3, (6), pp. 415-428.

Jones, P. A. and Laird, P. W. (1999) 'Cancer epigenetics comes of age', *Nat Genet,* 21, (2), pp. 163-167.

Jones, T. D., Carr, M. D., Eble, J. N., Wang, M., Lopez-Beltran, A. and Cheng, L. (2005) 'Clonal origin of lymph node metastases in bladder carcinoma', *Cancer,* 104, (9), pp. 1901-1910.

Jung, H. L., Wang, K. C., Kim, S. K., Sung, K. W., Koo, H. H., Shin, H. Y., Ahn, H. S., Shin, H. J. and Cho, B. K. (2004) 'Loss of heterozygosity analysis of chromosome 17p13.1-13.3 and its correlation with clinical outcome in medulloblastomas', *J Neurooncol,* 67, (1-2), pp. 41-46.

Kadouri, L., Hubert, A., Rotenberg, Y., Hamburger, T., Sagi, M., Nechushtan, C., Abeliovich, D. and Peretz, T. (2007) 'Cancer risks in carriers of the BRCA1/2 Ashkenazi founder mutations', *J Med Genet,* 44, (7), pp. 467-471.

Kaplan, E. L. and Meier, P. (1958) 'Nonparametric Estimation from Incomplete Observations', *Journal of the American Statistical Association,* 53, (282), pp. 457-481.

Katzenellenbogen, R. A., Baylin, S. B. and Herman, J. G. (1999) 'Hypermethylation of the DAP-kinase CpG island is a common alteration in B-cell malignancies', *Blood,* 93, (12), pp. 4347-4353.

Kaufman, L. and Rousseeuw, P. J. (2008) 'Chapter 1. Introduction', in *Finding Groups in Data.* John Wiley & Sons, Inc., pp. 1-67.

Kenney, A. M., Cole, M. D. and Rowitch, D. H. (2003) 'Nmyc upregulation by sonic hedgehog signaling promotes proliferation in developing cerebellar granule neuron precursors', *Development,* 130, (1), pp. 15-28.

Kent, W. J. (2002) 'BLAT--the BLAST-like alignment tool', *Genome Res,* 12, (4), pp. 656-664.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R. A., Niveleau, A., Cedar, H. and Simon, I. (2006) 'Evidence for an instructive mechanism of de novo methylation in cancer cells', *Nat Genet,* 38, (2), pp. 149-153.

Khosravi-Far, R. and Der, C. J. (1994) 'The Ras signal transduction pathway', *Cancer Metastasis Rev,* 13, (1), pp. 67-89.

Kim, J., Tang, J. Y., Gong, R., Kim, J., Lee, J. J., Clemons, K. V., Chong, C. R., Chang, K. S., Fereshteh, M., Gardner, D., Reya, T., Liu, J. O., Epstein, E. H., Stevens, D. A. and Beachy, P. A. (2010) 'Itraconazole, a Commonly Used Antifungal that Inhibits Hedgehog Pathway Activity and Cancer Growth', *Cancer Cell,* 17, (4), pp. 388-399.

Kim, J. Y., Nelson, A. L., Algon, S. A., Graves, O., Sturla, L. M., Goumnerova, L. C., Rowitch, D. H., Segal, R. A. and Pomeroy, S. L. (2003) 'Medulloblastoma tumorigenesis diverges from cerebellar granule cell differentiation in patched heterozygous mice', *Dev Biol,* 263, (1), pp. 50-66.

Kim, P. M. and Tidor, B. (2003) 'Subsystem Identification Through Dimensionality Reduction of Large-Scale Gene Expression Data', *Genome Res,* 13, (7), pp. 1706-1718.

Kimura, H., Ng, J. M. Y. and Curran, T. (2008) 'Transient Inhibition of the Hedgehog Pathway in Young Mice Causes Permanent Defects in Bone Structure', *Cancer Cell,* 13, (3), pp. 249-260.

Klaus, A. and Birchmeier, W. (2008) 'Wnt signalling and its impact on development and cancer', *Nat Rev Cancer,* 8, (5), pp. 387-398.

Kleihues, P., Louis, D. N., Scheithauer, B. W., Rorke, L. B., Reifenberger, G., Burger, P. C. and Cavenee, W. K. (2002) 'The WHO classification of tumors of the nervous system', *J Neuropathol Exp Neurol,* 61, (3), pp. 215-225; discussion 226-229.

Kleihues, P., Schauble, B., zur Hausen, A., Esteve, J. and Ohgaki, H. (1997) 'Tumors associated with p53 germline mutations: a synopsis of 91 families', *Am J Pathol,* 150, (1), pp. 1-13.

Knudson, A. G., Jr. (1971) 'Mutation and cancer: statistical study of retinoblastoma', *Proc Natl Acad Sci U S A,* 68, (4), pp. 820-823.

Kogerman, P., Krause, D., Rahnama, F., Kogerman, L., Unden, A. B., Zaphiropoulos, P. G. and Toftgard, R. (2002) 'Alternative first exons of PTCH1 are differentially regulated in vivo and may confer different functions to the PTCH1 protein', *Oncogene,* 21, (39), pp. 6007-6016.

Kongkham, P. N., Northcott, P. A., Croul, S. E., Smith, C. A., Taylor, M. D. and Rutka, J. T. (2010) 'The SFRP family of WNT inhibitors function as novel tumor suppressor genes epigenetically silenced in medulloblastoma', *Oncogene,* 29, (20), pp. 3017-3024.

Kongkham, P. N., Northcott, P. A., Ra, Y. S., Nakahara, Y., Mainprize, T. G., Croul, S. E., Smith, C. A., Taylor, M. D. and Rutka, J. T. (2008) 'An epigenetic genome-wide screen identifies SPINT2 as a novel tumor suppressor gene in pediatric medulloblastoma', *Cancer Res,* 68, (23), pp. 9945-9953.

Kool, M., Koster, J., Bunt, J., Hasselt, N. E., Lakeman, A., van Sluis, P., Troost, D., Meeteren, N. S., Caron, H. N., Cloos, J., Mrsic, A., Ylstra, B., Grajkowska, W., Hartmann, W., Pietsch, T., Ellison, D., Clifford, S. C. and Versteeg, R. (2008)

'Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features', *PLoS ONE,* 3, (8), pp. e3088.

Korshunov, A., Benner, A., Remke, M., Lichter, P., von Deimling, A. and Pfister, S. (2008) 'Accumulation of genomic aberrations during clinical progression of medulloblastoma', *Acta Neuropathol,* 116, (4), pp. 383-390.

Korshunov, A., Remke, M., Werft, W., Benner, A., Ryzhova, M., Witt, H., Sturm, D., Wittmann, A., Schottler, A., Felsberg, J., Reifenberger, G., Rutkowski, S., Scheurlen, W., Kulozik, A. E., von Deimling, A., Lichter, P. and Pfister, S. M. (2010) 'Adult and Pediatric Medulloblastomas Are Genetically Distinct and Require Different Algorithms for Molecular Risk Stratification', *J Clin Oncol,* 28, (18), pp. 3054-3060.

Kriaucionis, S. and Heintz, N. (2009) 'The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain', *Science,* 324, (5929), pp. 929-930.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., Adli, M., Kasif, S., Ptaszek, L. M., Cowan, C. A., Lander, E. S., Koseki, H. and Bernstein, B. E. (2008) 'Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains', *PLoS Genet,* 4, (10), pp. e1000242.

Kwabi-Addo, B., Giri, D., Schmidt, K., Podsypanina, K., Parsons, R., Greenberg, N. and Ittmann, M. (2001) 'Haploinsufficiency of the Pten tumor suppressor gene promotes prostate cancer progression', *Proc Natl Acad Sci U S A,* 98, (20), pp. 11563-11568.

Ladd-Acosta C., P. J., Sabunciyan S., Yolken R.H., Webster M.J., Dinkins T., Callinan P.A., Fan J.B., Potash J.B., Feinberg A.P. (2007) 'DNA methylation signatures within the human brain.', *Am J Hum Genet,* 81, (6), pp. 1304-1315.

Laird, P. W. (2010) 'Principles and challenges of genome-wide DNA methylation analysis', *Nat Rev Genet,* 11, pp. 191-203.

Lamont, J. M., McManamy, C. S., Pearson, A. D., Clifford, S. C. and Ellison, D. W. (2004) 'Combined histopathological and molecular cytogenetic stratification of medulloblastoma patients', *Clin Cancer Res,* 10, (16), pp. 5482-5493.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001) 'Initial sequencing and analysis of the human genome', *Nature,* 409, (6822), pp. 860-921.

Langdon, J. A., Lamont, J. M., Scott, D. K., Dyer, S., Prebble, E., Bown, N., Grundy, R. G., Ellison, D. W. and Clifford, S. C. (2006) 'Combined genome-wide allelotyping and copy number analysis identify frequent genetic losses without copy number reduction in medulloblastoma', *Genes Chromosomes Cancer,* 45, (1), pp. 47-60.

Lee, D. D. and Seung, H. S. (1999) 'Learning the parts of objects by non-negative matrix

factorization', *Nature,* 401, (6755), pp. 788-791.

Lee, W. H., Morton, R. A., Epstein, J. I., Brooks, J. D., Campbell, P. A., Bova, G. S., Hsieh, W. S., Isaacs, W. B. and Nelson, W. G. (1994) 'Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis', *Proc Natl Acad Sci U S A,* 91, (24), pp. 11733-11737.

Lee, Y. and Lee, C.-K. (2003) 'Classification of multiple cancer types by multicategory support vector machines using gene expression data', *Bioinformatics,* 19, (9), pp. 1132-1139.

Li, F. P. and Fraumeni, J. F., Jr. (1969) 'Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome', *J Natl Cancer Inst,* 43, (6), pp. 1365-1373.

Li, L. C. and Dahiya, R. (2002) 'MethPrimer: designing primers for methylation PCRs', *Bioinformatics,* 18, (11), pp. 1427-1431.

Li, Y., Zhu, J., Tian, G., Li, N., Li, Q., Ye, M., Zheng, H., Yu, J., Wu, H., Sun, J., Zhang, H., Chen, Q., Luo, R., Chen, M., He, Y., Jin, X., Zhang, Q., Yu, C., Zhou, G., Sun, J., Huang, Y., Zheng, H., Cao, H., Zhou, X., Guo, S., Hu, X., Li, X., Kristiansen, K., Bolund, L., Xu, J., Wang, W., Yang, H., Wang, J., Li, R., Beck, S., Wang, J. and Zhang, X. (2010) 'The DNA Methylome of Human Peripheral Blood Mononuclear Cells', *PLoS Biol,* 8, (11), pp. e1000533.

Lim, D. A., Huang, Y. C., Swigut, T., Mirick, A. L., Garcia-Verdugo, J. M., Wysocka, J., Ernst, P. and Alvarez-Buylla, A. (2009) 'Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells', *Nature,* 458, (7237), pp. 529-533.

Lin, J. C., Jeong, S., Liang, G., Takai, D., Fatemi, M., Tsai, Y. C., Egger, G., Gal-Yam, E. N. and Jones, P. A. (2007) 'Role of nucleosomal occupancy in the epigenetic

silencing of the MLH1 CpG island', *Cancer Cell,* 12, (5), pp. 432-444.

Lindsey, J. C., Anderton, J. A., Lusher, M. E. and Clifford, S. C. (2005) 'Epigenetic events in medulloblastoma development', *Neurosurg Focus,* 19, (5), pp. E10.

Lindsey, J. C., Hill, R. M., Megahed, H., Lusher, M. E., Schwalbe, E. C., Cole, M., Hogg, T. L., Gilbertson, R. J., Ellison, D. W., Bailey, S. and Clifford, S. C. (2011) 'TP53 mutations in favorable-risk Wnt/Wingless-subtype medulloblastomas', *J Clin Oncol,* 29, (12), pp. e344-6; author reply e347-348.

Lindsey, J. C., Lusher, M. E., Anderton, J. A., Bailey, S., Gilbertson, R. J., Pearson, A. D., Ellison, D. W. and Clifford, S. C. (2004) 'Identification of tumour-specific epigenetic events in medulloblastoma development by hypermethylation profiling', *Carcinogenesis,* 25, (5), pp. 661-668.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B. and Ecker, J. R. (2009) 'Human DNA methylomes at base resolution show widespread epigenomic differences', *Nature,* 462, (7271), pp. 315-322.

Lo, K. C., Rossi, M. R., Eberhart, C. G. and Cowell, J. K. (2007) 'Genome Wide Copy Number Abnormalities in Pediatric Medulloblastomas as Assessed by Array Comparative Genome Hybridization', *Brain Pathology,* 17, (3), pp. 282-296.

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W. and Kleihues, P. (2007) 'The 2007 WHO classification of tumours of the central nervous system', *Acta Neuropathol,* 114, (2), pp. 97-109.

Lu, D., Choi, M. Y., Yu, J., Castro, J. E., Kipps, T. J. and Carson, D. A. (2011) 'Salinomycin inhibits Wnt signaling and selectively induces apoptosis in chronic lymphocytic leukemia cells', *Proc Natl Acad Sci U S A,* 108, (32), pp. 13253-13257.

Lujambio, A., Calin, G. A., Villanueva, A., Ropero, S., Sanchez-Cespedes, M., Blanco, D., Montuenga, L. M., Rossi, S., Nicoloso, M. S., Faller, W. J., Gallagher, W. M., Eccles, S. A., Croce, C. M. and Esteller, M. (2008) 'A microRNA DNA methylation signature for human cancer metastasis', *Proc Natl Acad Sci U S A,* 105, (36), pp. 13556-13561.

Malkin, D. (2001) 'The role of p53 in human cancer', *J Neurooncol,* 51, (3), pp. 231-243.

Malkin, D., Li, F. P., Strong, L. C., Fraumeni, J. F., Jr., Nelson, C. E., Kim, D. H., Kassel, J., Gryka, M. A., Bischoff, F. Z., Tainsky, M. A. and et al. (1990) 'Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms', *Science,* 250, (4985), pp. 1233-1238.

Mancuso, M., Leonardi, S., Ceccarelli, M., Pasquali, E., De Stefano, I., Prisco, M. G., Rebessi, S., Tanori, M., Scambia, G., Di Majo, V., Pazzaglia, S., Saran, A. and Gallo, D. (2010) 'Protective role of 17 beta-estradiol on medulloblastoma development in Patched 1 heterozygous mice', *Int J Cancer,* 127, (12), pp. 2749-2757.

Mantel, N. (1966) 'Evaluation of survival data and two new rank order statistics arising in its consideration', *Cancer Chemother Rep,* 50, (3), pp. 163-170.

Marino, S. (2005) 'Medulloblastoma: developmental mechanisms out of control', *Trends in Molecular Medicine,* 11, (1), pp. 17-22.

Martin, G. S. (1970) 'Rous sarcoma virus: a function required for the maintenance of the transformed state', *Nature,* 227, (5262), pp. 1021-1023.

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M., Rowitch, D. H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S. J., Haussler, D., Marra, M. A., Hirst, M., Wang, T. and Costello, J. F. (2010) 'Conserved role of intragenic DNA methylation in

regulating alternative promoters', *Nature,* 466, (7303), pp. 253-257.

McCabe, M. G., Backlund, L. M., Leong, H. S., Ichimura, K. and Collins, V. P. (2011) 'Chromosome 17 alterations identify good-risk and poor-risk tumors independently of clinical factors in medulloblastoma', *Neuro Oncol,* 13, (4), pp. 376-383.

McCarthy, N. (2011) 'Medulloblastoma: Origins', *Nat Rev Cancer,* 11, (2), pp. 80-81.

McManamy, C. S., Lamont, J. M., Taylor, R. E., Cole, M., Pearson, A. D., Clifford, S. C. and Ellison, D. W. (2003) 'Morphophenotypic variation predicts clinical behavior in childhood non-desmoplastic medulloblastomas', *J Neuropathol Exp Neurol,* 62, (6), pp. 627-632.

McManamy, C. S., Pears, J., Weston, C. L., Hanzely, Z., Ironside, J. W., Taylor, R. E., Grundy, R. G., Clifford, S. C. and Ellison, D. W. (2007) 'Nodule formation and desmoplasia in medulloblastomas-defining the nodular/desmoplastic variant and its biological behavior', *Brain Pathol,* 17, (2), pp. 151-164.

Millenaar, F. F., Okyere, J., May, S. T., van Zanten, M., Voesenek, L. A. and Peeters, A. J. (2006) 'How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results', *BMC Bioinformatics,* 7, pp. 137.

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M. and Schubeler, D. (2008) 'Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors', *Mol Cell,* 30, (6), pp. 755-766.

Moinova, H. R., Chen, W. D., Shen, L., Smiraglia, D., Olechnowicz, J., Ravi, L., Kasturi, L., Myeroff, L., Plass, C., Parsons, R., Minna, J., Willson, J. K., Green, S. B., Issa, J. P. and Markowitz, S. D. (2002) 'HLTF gene silencing in human colon cancer', *Proc Natl Acad Sci U S A,* 99, (7), pp. 4562-4567.

Nagai, H., Kim, Y. S., Konishi, N., Baba, M., Kubota, T., Yoshimura, A. and Emi, M. (2002) 'Combined hypermethylation and chromosome loss associated with inactivation of SSI-1/SOCS-1/JAB gene in human hepatocellular carcinomas', *Cancer Lett,* 186, (1), pp. 59-65.

Newton, H. B. (2001) 'Review of the molecular genetics and chemotherapeutic treatment of adult and paediatric medulloblastoma', *Expert Opin Investig Drugs,* 10, (12), pp. 2089-2104.

Nguyen, C., Liang, G., Nguyen, T. T., Tsao-Wei, D., Groshen, S., Lubbert, M., Zhou, J. H., Benedict, W. F. and Jones, P. A. (2001) 'Susceptibility of nonpromoter CpG islands to de novo methylation in normal and neoplastic cells', *J Natl Cancer Inst,* 93, (19), pp. 1465-1472.

Nicholson, J., Wickramasinghe, C., Ross, F., Crolla, J. and Ellison, D. (2000) 'Imbalances of chromosome 17 in medulloblastomas determined by comparative genomic hybridisation and fluorescence in situ hybridisation', *Mol Pathol,* 53, (6), pp. 313-319.

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. (2011) 'Pediatric and adult sonic hedgehog medulloblastomas are clinically and molecularly distinct', *Acta Neuropathol,* 122, (2), pp. 231-240.

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. (2010) 'Medulloblastoma Comprises Four Distinct Molecular Variants', *J Clin Oncol,* 29, (11), pp. 1408-1414.

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. (2009) 'Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma', *Nat Genet,* 41, (4), pp. 465-472.

Noushmehr, H., Weisenberger, D. J., Diefes, K., Phillips, H. S., Pujara, K., Berman, B. P., Pan, F., Pelloski, C. E., Sulman, E. P., Bhat, K. P., Verhaak, R. G. W., Hoadley, K. A., Hayes, D. N., Perou, C. M., Schmidt, H. K., Ding, L., Wilson, R. K., Van Den Berg, D., Shen, H., Bengtsson, H., Neuvial, P., Cope, L. M., Buckley, J., Herman, J. G., Baylin, S. B., Laird, P. W. and Aldape, K. (2010) 'Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma', *Cancer Cell,* 17, (5), pp. 510-522.

Nusslein-Volhard, C. and Wieschaus, E. (1980) 'Mutations affecting segment number and polarity in Drosophila', *Nature,* 287, (5785), pp. 795-801.

Nygren, A. O., Ameziane, N., Duarte, H. M., Vijzelaar, R. N., Waisfisz, Q., Hess, C. J., Schouten, J. P. and Errami, A. (2005) 'Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences', *Nucleic Acids Res,* 33, (14), pp. e128.

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. (2003) 'Shared genetic susceptibility to breast cancer, brain tumors, and Fanconi anemia', *J Natl Cancer Inst,* 95, (20), pp. 1548-1551.

Ohm, J. E., McGarvey, K. M., Yu, X., Cheng, L., Schuebel, K. E., Cope, L., Mohammad, H. P., Chen, W., Daniel, V. C., Yu, W., Berman, D. M., Jenuwein, T., Pruitt, K., Sharkis, S. J., Watkins, D. N., Herman, J. G. and Baylin, S. B. (2007) 'A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing', *Nat Genet,* 39, (2), pp. 237-242.

Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X. and Bestor, T. H. (2007) 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature,* 448, (7154), pp. 714-717.

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. (2006) 'Phase III Study of Craniospinal Radiation Therapy Followed by Adjuvant Chemotherapy for Newly Diagnosed Average-Risk Medulloblastoma', *J Clin Oncol,* 24, (25), pp. 4202-4208.

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. (1999) '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), pp. 2127-2136.

Pan, E., Pellarin, M., Holmes, E., Smirnov, I., Misra, A., Eberhart, C. G., Burger, P. C., Biegel, J. A. and Feuerstein, B. G. (2005) 'Isochromosome 17q Is a Negative Prognostic Factor in Poor-Risk Childhood Medulloblastoma Patients', *Clinical Cancer Research,* 11, (13), pp. 4733-4740.

Paraf, F., Jothy, S. and Van Meir, E. G. (1997) 'Brain tumor-polyposis syndrome: two genetic diseases?', *J Clin Oncol,* 15, (7), pp. 2744-2758.

Parsons, B. L. (2008) 'Many different tumor types have polyclonal tumor origin: evidence and implications', *Mutat Res,* 659, (3), pp. 232-247.

Parsons, D. W., Jones, S., Zhang, X., Lin, J. C., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Siu, I. M., Gallia, G. L., Olivi, A., McLendon, R., Rasheed, B. A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D. A., Tekleab, H., Diaz, L. A., Jr., Hartigan, J., Smith, D. R., Strausberg, R. L., Marie, S. K., Shinjo, S. M., Yan, H., Riggins, G. J.,

Bigner, D. D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V. E. and Kinzler, K. W. (2008) 'An integrated genomic analysis of human glioblastoma multiforme', *Science,* 321, (5897), pp. 1807-1812.

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. (2011) 'The Genetic Landscape of the Childhood Cancer Medulloblastoma', *Science,* 331, (6016), pp. 435-439.

Peate, I. (2011) 'Men and cancer: the gender dimension', *Br J Nurs,* 20, (6), pp. 340, 342-3.

Pedersen, K. S., Bamlet, W. R., Oberg, A. L., de Andrade, M., Matsumoto, M. E., Tang, H., Thibodeau, S. N., Petersen, G. M. and Wang, L. (2011) 'Leukocyte DNA Methylation Signature Differentiates Pancreatic Cancer Patients from Healthy Controls', *PLoS ONE,* 6, (3), pp. e18223.

Pezzolo, A., Coco, S., Raso, A., Parodi, F., Pistorio, A., Valdora, F., Capra, V., Zollo, M., Aschero, S., Basso, E., Cama, A., Nozza, P., Gambini, C., Cinalli, G., Garre, M. L., Iolascon, A., Pistoia, V. and Tonini, G. P. (2011) 'Loss of 10q26.1-q26.3 in association with 7q34-q36.3 gain or 17q24.3-q25.3 gain predict poor outcome in pediatric medulloblastoma', *Cancer Lett,* 308, (2), pp. 215-24.

Pfaff, E., Remke, M., Sturm, D., Benner, A., Witt, H., Milde, T., von Bueren, A. O., Wittmann, A., Schöttler, A., Jorch, N., Graf, N., Kulozik, A. E., Witt, O., Scheurlen, W., von Deimling, A., Rutkowski, S., Taylor, M. D., Tabori, U., Lichter, P., Korshunov, A. and Pfister, S. M. (2010) 'TP53 Mutation Is Frequently Associated

With CTNNB1 Mutation or MYCN Amplification and Is Compatible With Long-Term Survival in Medulloblastoma', *J Clin Oncol,* 28, (35), pp. 5188-5196.

Pfister, S., Remke, M., Benner, A., Mendrzyk, F., Toedt, G., Felsberg, J., Wittmann, A., Devens, F., Gerber, N. U., Joos, S., Kulozik, A., Reifenberger, G., Rutkowski, S., Wiestler, O. D., Radlwimmer, B., Scheurlen, W., Lichter, P. and Korshunov, A. (2009) 'Outcome Prediction in Pediatric Medulloblastoma Based on DNA Copy-Number Aberrations of Chromosomes 6q and 17q and the MYC and MYCN Loci', *J Clin Oncol,* 27, (10), pp. 1627-1636.

Pfister, S., Schlaeger, C., Mendrzyk, F., Wittmann, A., Benner, A., Kulozik, A., Scheurlen, W., Radlwimmer, B. and Lichter, P. (2007) 'Array-based profiling of reference-independent methylation status (aPRIMES) identifies frequent promoter methylation and consecutive downregulation of ZIC2 in pediatric medulloblastoma', *Nucleic Acids Res,* 35, (7), pp. e51.

Pfister, S. M., Korshunov, A., Kool, M., Hasselblatt, M., Eberhart, C. and Taylor, M. D. (2010) 'Molecular diagnostics of CNS embryonal tumors', *Acta Neuropathol,* 120, (5), pp. 553-566.

Pietsch, T., Waha, A., Koch, A., Kraus, J., Albrecht, S., Tonn, J., Sorensen, N., Berthold, F., Henk, B., Schmandt, N., Wolf, H. K., von Deimling, A., Wainwright, B., Chenevix-Trench, G., Wiestler, O. D. and Wicking, C. (1997) 'Medulloblastomas of the desmoplastic variant carry mutations of the human homologue of Drosophila patched', *Cancer Res,* 57, (11), pp. 2085-2088.

Pizer, B. L. and Clifford, S. C. (2009) 'The potential impact of tumour biology on improved clinical practice for medulloblastoma: progress towards biologically driven clinical trials', *Br J Neurosurg,* 23, (4), pp. 364-375.

Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., Kato, R., Sasaki, H., Hirotsune, S., Okazaki, Y., Held, W. A., Hayashizaki, Y. and Chapman, V. M. (1996) 'Identification of Grf1 on mouse chromosome 9 as an imprinted

gene by RLGS-M', *Nat Genet,* 14, (1), pp. 106-109.

Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., Kim, J. Y. H., Goumnerova, L. C., Black, P. M., Lau, C., Allen, J. C., Zagzag, D., Olson, J. M., Curran, T., Wetmore, C., Biegel, J. A., Poggio, T., Mukherjee, S., Rifkin, R., Califano, A., Stolovitzky, G., Louis, D. N., Mesirov, J. P., Lander, E. S. and Golub, T. R. (2002) 'Prediction of central nervous system embryonal tumour outcome based on gene expression', *Nature,* 415, (6870), pp. 436-442.

Prochownik, E. V., Eagle Grove, L., Deubler, D., Zhu, X. L., Stephenson, R. A., Rohr, L. R., Yin, X. and Brothman, A. R. (1998) 'Commonly occurring loss and mutation of the MXI1 gene in prostate cancer', *Genes Chromosomes Cancer,* 22, (4), pp. 295-304.

Pruijt, J. F., Verzaal, P., van Os, R., de Kruijf, E. J., van Schie, M. L., Mantovani, A., Vecchi, A., Lindley, I. J., Willemze, R., Starckx, S., Opdenakker, G. and Fibbe, W. E. (2002) 'Neutrophils are indispensable for hematopoietic stem cell mobilization induced by interleukin-8 in mice', *Proc Natl Acad Sci U S A,* 99, (9), pp. 6228-6233.

R Development Core Team. (2008) *R: A language and environment for statistical computing.* Available at: http://cran.r-project.org/ (Accessed: July 20th, 2011).

Raffel, C., Jenkins, R. B., Frederick, L., Hebrink, D., Alderete, B., Fults, D. W. and James, C. D. (1997) 'Sporadic medulloblastomas contain PTCH mutations', *Cancer Res,* 57, (5), pp. 842-845.

Rai, A. J., Kamath, R. M., Gerald, W. and Fleisher, M. (2009) 'Analytical validation of the GeXP analyzer and design of a workflow for cancer-biomarker discovery using multiplexed gene-expression profiling', *Anal Bioanal Chem,* 393, (5), pp. 1505-1511.

Ransohoff, D. F. (2004) 'Rules of evidence for cancer molecular-marker discovery and

validation', *Nat Rev Cancer,* 4, (4), pp. 309-314.

Reid, S., Schindler, D., Hanenberg, H., Barker, K., Hanks, S., Kalb, R., Neveling, K., Kelly, P., Seal, S., Freund, M., Wurm, M., Batish, S. D., Lach, F. P., Yetgin, S., Neitzel, H., Ariffin, H., Tischkowitz, M., Mathew, C. G., Auerbach, A. D. and Rahman, N. (2007) 'Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer', *Nat Genet,* 39, (2), pp. 162-164.

Reifenberger, J., Wolter, M., Weber, R. G., Megahed, M., Ruzicka, T., Lichter, P. and Reifenberger, G. (1998) '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), pp. 1798-1803.

Remke, M., Hielscher, T., Korshunov, A., Northcott, P. A., Bender, S., Kool, M., Westermann, F., Benner, A., Cin, H., Ryzhova, M., Sturm, D., Witt, H., Haag, D., Toedt, G., Wittmann, A., Schottler, A., von Bueren, A. O., von Deimling, A., Rutkowski, S., Scheurlen, W., Kulozik, A. E., Taylor, M. D., Lichter, P. and Pfister, S. M. (2011a) 'FSTL5 Is a Marker of Poor Prognosis in Non-WNT/ Non-SHH Medulloblastoma', *J Clin Oncol,* 29, (29), pp. 3852-3861.

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. (2011b) 'Adult medulloblastoma comprises three major molecular variants', *J Clin Oncol,* 29, (19), pp. 2717-2723.

Rideout, W. M., 3rd, Coetzee, G. A., Olumi, A. F. and Jones, P. A. (1990) '5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes', *Science,* 249, (4974), pp. 1288-1290.

Ridola, V., Grill, J., Doz, F., Gentet, J. C., Frappaz, D., Raquin, M. A., Habrand, J. L., Sainte-Rose, C., Valteau-Couanet, D. and Kalifa, C. (2007) 'High-dose chemotherapy with autologous stem cell rescue followed by posterior fossa irradiation for local medulloblastoma recurrence or progression after

conventional chemotherapy', *Cancer,* 110, (1), pp. 156-163.

Rieber, J., Remke, M., Hartmann, C., Korshunov, A., Burkhardt, B., Sturm, D., Mechtersheimer, G., Wittmann, A., Greil, J., Blattmann, C., Witt, O., Behnisch, W., Halatsch, M. E., Orakcioglu, B., von Deimling, A., Lichter, P., Kulozik, A. and Pfister, S. (2009) 'Novel oncogene amplifications in tumors from a family with Li-Fraumeni syndrome', *Genes Chromosomes Cancer,* 48, (7), pp. 558-568.

Robertson, K. D. (2005) 'DNA methylation and human disease', *Nat Rev Genet,* 6, (8), pp. 597-610.

Rogers, L., Pattisapu, J., Smith, R. R. and Parker, P. (1988) 'Medulloblastoma in association with the Coffin-Siris syndrome', *Childs Nerv Syst,* 4, (1), pp. 41-44.

Roh, T.-Y., Cuddapah, S., Cui, K. and Zhao, K. (2006) 'The genomic landscape of histone modifications in human T cells', *Proc Natl Acad Sci U S A,* 103, (43), pp. 15782-15787.

Romer, J. T., Kimura, H., Magdaleno, S., Sasai, K., Fuller, C., Baines, H., Connelly, M., Stewart, C. F., Gould, S., Rubin, L. L. and Curran, T. (2004) 'Suppression of the Shh pathway using a small molecule inhibitor eliminates medulloblastoma in Ptc1(+/-)p53(-/-) mice', *Cancer Cell,* 6, (3), pp. 229-240.

Rood, B. R., Zhang, H., Weitman, D. M. and Cogen, P. H. (2002) 'Hypermethylation of HIC-1 and 17p allelic loss in medulloblastoma', *Cancer Res,* 62, (13), pp. 3794-3797.

Rousseeuw, P. J. (1987) 'Silhouettes: a graphical aid to the interpretation and validation of cluster analysis', *J. Comput. Appl. Math.,* 20, (1), pp. 53-65.

Rubin, L. L. and de Sauvage, F. J. (2006) 'Targeting the Hedgehog pathway in cancer', *Nat Rev Drug Discov,* 5, (12), pp. 1026-1033.

Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S. H., Masiarz, F. R., Munemitsu, S. and Polakis, P. (1993) 'Association of the APC gene product with beta-catenin', *Science,* 262, (5140), pp. 1731-1734.

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. (2009) 'Treatment of Medulloblastoma with Hedgehog Pathway Inhibitor GDC-0449', *N Engl J Med,* 361, (12), pp. 1173-1178.

Ruiz i Altaba, A. (1997) 'Catching a Gli-mpse of Hedgehog', *Cell,* 90, (2), pp. 193-196.

Ruiz i Altaba, A., Palma, V. and Dahmane, N. (2002) 'Hedgehog-GLI signaling and the growth of the brain', *Nat Rev Neurosci,* 3, (1), pp. 24-33.

Rutkowski, S., Bode, U., Deinlein, F., Ottensmeier, H., Warmuth-Metz, M., Soerensen, N., Graf, N., Emser, A., Pietsch, T., Wolff, J. E., Kortmann, R. D. and Kuehl, J. (2005) 'Treatment of early childhood medulloblastoma by postoperative chemotherapy alone', *N Engl J Med,* 352, (10), pp. 978-986.

Rutkowski, S., von Bueren, A., von Hoff, K., Hartmann, W., Shalaby, T., Deinlein, F., Warmuth-Metz, M., Soerensen, N., Emser, A., Bode, U., Mittler, U., Urban, C., Benesch, M., Kortmann, R. D., Schlegel, P. G., Kuehl, J., Pietsch, T. and Grotzer, M. (2007) 'Prognostic Relevance of Clinical and Biological Risk Factors in Childhood Medulloblastoma: Results of Patients Treated in the Prospective Multicenter Trial HIT'91', *Clin Cancer Res,* 13, (9), pp. 2651-2657.

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. (2010) 'Survival and Prognostic Factors of Early Childhood Medulloblastoma: An International Meta-Analysis', *J Clin Oncol,* 28, (33), pp. 4961-4968.

Sandoval, J., Heyn, H. A., Moran, S., Serra-Musach, J., Pujana, M. A., Bibikova, M. and

Esteller, M. (2011) 'Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome', *Epigenetics,* 6, (6), pp. 692-702.

Sarkar, C., Deb, P. and Sharma, M. C. (2005) 'Recent advances in embryonal tumours of the central nervous system', *Childs Nerv Syst,* 21, (4), pp. 272-293.

Sato, M., Mori, Y., Sakurada, A., Fujimura, S. and Horii, A. (1998) 'The H-cadherin (CDH13) gene is inactivated in human lung cancer', *Hum Genet,* 103, (1), pp. 96-101.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995) 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray', *Science,* 270, (5235), pp. 467-470.

Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwijnenburg, D., Diepvens, F. and Pals, G. (2002) 'Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification', *Nucleic Acids Res,* 30, (12), pp. e57.

Schuller, U., Heine, V. M., Mao, J., Kho, A. T., Dillon, A. K., Han, Y. G., Huillard, E., Sun, T., Ligon, A. H., Qian, Y., Ma, Q., Alvarez-Buylla, A., McMahon, A. P., Rowitch, D. H. and Ligon, K. L. (2008) 'Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma', *Cancer Cell,* 14, (2), pp. 123-134.

Schwab, M. (1990) 'Amplification of the MYCN oncogene and deletion of putative tumour suppressor gene in human neuroblastomas', *Brain Pathol,* 1, (1), pp. 41-46.

Schwalbe, E. C., Lindsey, J. C., Straughton, D., Hogg, T. L., Cole, M., Megahed, H., Ryan, S. L., Lusher, M. E., Taylor, M. D., Gilbertson, R. J., Ellison, D. W., Bailey, S. and Clifford, S. C. (2011) 'Rapid diagnosis of medulloblastoma molecular subgroups', *Clin Cancer Res,* 17, (7), pp. 1883-1894.

Scott, D. K., Straughton, D., Cole, M., Bailey, S., Ellison, D. W. and Clifford, S. C. (2006) 'Identification and analysis of tumor suppressor loci at chromosome 10q23.3-10q25.3 in medulloblastoma', *Cell Cycle,* 5, (20), pp. 2381-2389.

Scotting, P. J., Walker, D. A. and Perilongo, G. (2005) 'Childhood solid tumours: a developmental disorder', *Nat Rev Cancer,* 5, (6), pp. 481-488.

Secombe, J., Pierce, S. B. and Eisenman, R. N. (2004) 'Myc: a weapon of mass destruction', *Cell,* 117, (2), pp. 153-156.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K. (2001) 'dbSNP: the NCBI database of genetic variation', *Nucleic Acids Res,* 29, (1), pp. 308-311.

Silverman, L. R. and Mufti, G. J. (2005) 'Methylation inhibitor therapy in the treatment of myelodysplastic syndrome', *Nat Clin Pract Oncol,* 2 Suppl 1, pp. S12-23.

Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G. and Kasprzyk, A. (2009) 'BioMart--biological queries made easy', *BMC Genomics,* 10, pp. 22.

Smiraglia, D. J., Fruhwald, M. C., Costello, J. F., McCormick, S. P., Dai, Z., Peltomaki, P., O'Dorisio, M. S., Cavenee, W. K. and Plass, C. (1999) 'A new tool for the rapid cloning of amplified and hypermethylated human DNA sequences from restriction landmark genome scanning gels', *Genomics,* 58, (3), pp. 254-262.

Soffietti, R., Ruda, R. and Mutani, R. (2002) 'Management of brain metastases', *J Neurol,* 249, (10), pp. 1357-1369.

Staal, F. J. T., van der Burg, M., Wessels, L. F. A., Barendregt, B. H., Baert, M. R. M., van den Burg, C. M. M., Van Huffel, C., Langerak, A. W., van der Velden, V. H. J., Reinders, M. J. T. and van Dongen, J. J. M. (2003) 'DNA microarrays for comparison of gene expression profiles between diagnosis and relapse in

precursor-B acute lymphoblastic leukemia: choice of technique and purification influence the identification of potential diagnostic markers', *Leukemia,* 17, (7), pp. 1324-1332.

Statnikov, A., Aliferis, C. F., Tsamardinos, I., Hardin, D. and Levy, S. (2005) 'A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis', *Bioinformatics,* 21, (5), pp. 631-643.

Stearns, D., Chaudhry, A., Abel, T. W., Burger, P. C., Dang, C. V. and Eberhart, C. G. (2006) 'c-myc overexpression causes anaplasia in medulloblastoma', *Cancer Res,* 66, (2), pp. 673-681.

Stehelin, D., Varmus, H. E., Bishop, J. M. and Vogt, P. K. (1976) 'DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA', *Nature,* 260, (5547), pp. 170-173.

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G. and Pombo, A. (2007) 'Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells', *Nat Cell Biol,* 9, (12), pp. 1428-1435.

Strachan, T. and Read, A. P. (1999) 'Chapter 18: Cancer Genetics' in *Human Molecular Genetics, 2nd edition.* New York: Wiley-Liss.

Su, L. K., Vogelstein, B. and Kinzler, K. W. (1993) 'Association of the APC tumor suppressor protein with catenins', *Science,* 262, (5140), pp. 1734-1737.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. and Mesirov, J. P. (2005) 'Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles', *Proc Natl Acad Sci U S A,* 102, (43), pp. 15545-15550.

Suzuki, H., Gabrielson, E., Chen, W., Anbazhagan, R., van Engeland, M., Weijenberg, M. P., Herman, J. G. and Baylin, S. B. (2002) 'A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer', *Nat Genet,* 31, (2), pp. 141-149.

Suzuki, R. and Shimodaira, H. (2006) 'Pvclust: an R package for assessing the uncertainty in hierarchical clustering', *Bioinformatics,* 22, (12), pp. 1540-1542.

Swartling, F. J., Grimmer, M. R., Hackett, C. S., Northcott, P. A., Fan, Q. W., Goldenberg, D. D., Lau, J., Masic, S., Nguyen, K., Yakovenko, S., Zhe, X. N., Gilmer, H. C., Collins, R., Nagaoka, M., Phillips, J. J., Jenkins, R. B., Tihan, T., Vandenberg, S. R., James, C. D., Tanaka, K., Taylor, M. D., Weiss, W. A. and Chesler, L. (2010) 'Pleiotropic role for MYCN in medulloblastoma', *Genes Dev,* 24, (10), pp. 1059-1072.

Tabori, U., Baskin, B., Shago, M., Alon, N., Taylor, M. D., Ray, P. N., Bouffet, E., Malkin, D. and Hawkins, C. (2010) 'Universal poor survival in children with medulloblastoma harboring somatic TP53 mutations', *J Clin Oncol,* 28, (8), pp. 1345-1350.

Tamayo, P., Cho, Y.-J., Tsherniak, A., Greulich, H., Ambrogio, L., Schouten-van Meeteren, N., Zhou, T., Buxton, A., Kool, M., Meyerson, M., Pomeroy, S. L. and Mesirov, J. P. (2011) 'Predicting Relapse in Patients With Medulloblastoma by Integrating Evidence From Clinical and Genomic Features', *J Clin Oncol,* 29, (11), pp. 1415-1423.

Tamayo, P., Scanfeld, D., Ebert, B. L., Gillette, M. A., Roberts, C. W. M. and Mesirov, J. P. (2007) 'Metagene projection for cross-platform, cross-species characterization of global transcriptional states', *Proc Natl Acad Sci U S A,* 104, (14), pp. 5959-5964.

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. (2002) 'Mutations in SUFU predispose to medulloblastoma', *Nat Genet,* 31, (3), pp. 306-310.

Taylor, M. D., Mainprize, T. G., Rutka, J. T., Becker, L., Bayani, J. and Drake, J. M. (2001) 'Medulloblastoma in a child with Rubenstein-Taybi Syndrome: case report and review of the literature', *Pediatr Neurosurg,* 35, (5), pp. 235-238.

Taylor, R. E., Bailey, C. C., Robinson, K., Weston, C. L., Ellison, D., Ironside, J., Lucraft, H., Gilbertson, R., Tait, D. M., Walker, D. A., Pizer, B. L., Imeson, J. and Lashford, L. S. (2003) 'Results of a Randomized Study of Preradiation Chemotherapy Versus Radiotherapy Alone for Nonmetastatic Medulloblastoma: The International Society of Paediatric Oncology/United Kingdom Children's Cancer Study Group PNET-3 Study', *J Clin Oncol,* 21, (8), pp. 1581-1591.

Teodoridis, J. M., Hardie, C. and Brown, R. (2008) 'CpG island methylator phenotype (CIMP) in cancer: Causes and implications', *Cancer Lett,* 268, (2), pp. 177-186.

Teodoridis, J. M., Strathdee, G. and Brown, R. (2004) 'Epigenetic silencing mediated by CpG island methylation: potential as a therapeutic target and as a biomarker', *Drug Resist Updat,* 7, (4-5), pp. 267-278.

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Gayther, S. A., Apostolidou, S., Jones, A., Lechner, M., Beck, S., Jacobs, I. J. and Widschwendter, M. (2009) 'An epigenetic signature in peripheral blood predicts active ovarian cancer', *PLoS ONE,* 4, (12), pp. e8274.

Thirlwell, C., Eymard, M., Feber, A., Teschendorff, A., Pearce, K., Lechner, M., Widschwendter, M. and Beck, S. (2010) 'Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip', *Methods,* 52, (3), pp. 248-254.

Thomas, K. R. and Capecchi, M. R. (1990) 'Targeted disruption of the murine int-1

proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development', *Nature,* 346, (6287), pp. 847-850.

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. (2006) 'Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations', *J Clin Oncol,* 24, (12), pp. 1924-1931.

Turcot, J., Despres, J. P. and St Pierre, F. (1959) 'Malignant tumors of the central nervous system associated with familial polyposis of the colon: report of two cases', *Dis Colon Rectum,* 2, pp. 465-468.

Verma, S., Tavare, C. and Gilles, F. (2008) 'Histologic Features and Prognosis in Pediatric Medulloblastoma', *Pediatr Dev Pathol*, 11, (5), pp. 337-343.

Vibhakar, R., Foltz, G., Yoon, J. G., Field, L., Lee, H., Ryu, G. Y., Pierson, J., Davidson, B. and Madan, A. (2007) 'Dickkopf-1 is an epigenetically silenced candidate tumor suppressor gene in medulloblastoma', *Neuro Oncol,* 9, (2), pp. 135-144.

Vogelstein, B. and Kinzler, K. W. (2004) 'Cancer genes and the pathways they control', *Nat Med,* 10, (8), pp. 789-799.

von Hoff, K., Hartmann, W., von Bueren, A. O., Gerber, N. U., Grotzer, M. A., Pietsch, T. and Rutkowski, S. (2009) 'Large cell/anaplastic medulloblastoma: Outcome according to myc status, histopathological, and clinical risk factors', *Pediatr Blood Cancer,* 54, (3), pp. 369-376.

Wahlfors, J., Hiltunen, H., Heinonen, K., Hamalainen, E., Alhonen, L. and Janne, J. (1992) 'Genomic hypomethylation in human chronic lymphocytic leukemia', *Blood,* 80, (8), pp. 2074-2080.

Walsh, C. P., Chaillet, J. R. and Bestor, T. H. (1998) 'Transcription of IAP endogenous

retroviruses is constrained by cytosine methylation', *Nat Genet,* 20, (2), pp. 116-117.

Wang, J., Wen, S., Symmans, W. F., Pusztai, L. and Coombes, K. R. (2009a) 'The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data', *Cancer Inform,* 7, pp. 199-216.

Wang, X., Wang, M., MacLennan, G. T., Abdul-Karim, F. W., Eble, J. N., Jones, T. D., Olobatuyi, F., Eisenberg, R., Cummings, O. W., Zhang, S., Lopez-Beltran, A., Montironi, R., Zheng, S., Lin, H., Davidson, D. D. and Cheng, L. (2009b) 'Evidence for common clonal origin of multifocal lung cancers', *J Natl Cancer Inst,* 101, (8), pp. 560-570.

Wang, Y. and Leung, F. C. (2004) 'An evaluation of new criteria for CpG islands in the human genome as gene markers', *Bioinformatics,* 20, (7), pp. 1170-1177.

WCRF (2010) *World Cancer Statistics: Overall*. Available at: http://www.wcrf-uk.org/research/cancer_statistics/world_cancer_statistics_overall.php (Accessed: September 20th, 2011).

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L. and Schubeler, D. (2005) 'Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells', *Nat Genet,* 37, (8), pp. 853-862.

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Paabo, S., Rebhan, M. and Schubeler, D. (2007) 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nat Genet,* 39, (4), pp. 457-466.

Wechsler-Reya, R. J. and Scott, M. P. (1999) 'Control of Neuronal Precursor Proliferation in the Cerebellum by Sonic Hedgehog', *Neuron,* 22, (1), pp. 103-114.

Wei, S. H., Balch, C., Paik, H. H., Kim, Y.-S., Baldwin, R. L., Liyanarachchi, S., Li, L., Wang, Z., Wan, J. C., Davuluri, R. V., Karlan, B. Y., Gifford, G., Brown, R., Kim, S., Huang, T. H. M. and Nephew, K. P. (2006) 'Prognostic DNA Methylation Biomarkers in Ovarian Cancer', *Clin Cancer Res,* 12, (9), pp. 2788-2794.

Weil, M. D., Lamborn, K., Edwards, M. S. B. and Wara, W. M. (1998) 'Influence of a Child's Sex on Medulloblastoma Outcome', *JAMA: The Journal of the American Medical Association,* 279, (18), pp. 1474-1476.

West, R. W. and Barrett, J. C. (1993) 'Inactivation of a tumor suppressor function in immortal Syrian hamster cells by N-methyl-N'-nitro-N-nitrosoguanidine and by 5-aza-2'-deoxycytidine', *Carcinogenesis,* 14, (2), pp. 285-289.

Whang, J., Frei, E., 3rd, Tjio, J. H., Carbone, P. P. and Brecher, G. (1963) 'The Distribution of the Philadelphia Chromosome in Patients with Chronic Myelogenous Leukemia', *Blood,* 22, pp. 664-673.

Wheeler, J. M., Kim, H. C., Efstathiou, J. A., Ilyas, M., Mortensen, N. J. and Bodmer, W. F. (2001) 'Hypermethylation of the promoter region of the E-cadherin gene (CDH1) in sporadic and ulcerative colitis associated colorectal cancer', *Gut,* 48, (3), pp. 367-371.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D. J., Campan, M., Young, J., Jacobs, I. and Laird, P. W. (2007) 'Epigenetic stem cell signature in cancer', *Nat Genet,* 39, (2), pp. 157-158.

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A., Rappsilber, J. and Helin, K. (2011) 'TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity', *Nature,* 473, (7347), pp. 343-348.

Wilne, S., Collier, J., Kennedy, C., Koller, K., Grundy, R. and Walker, D. (2007) 'Presentation of childhood CNS tumours: a systematic review and meta-analysis', *Lancet Oncol,* 8, (8), pp. 685-695.

Wolter, M., Reifenberger, J., Sommer, C., Ruzicka, T. and Reifenberger, G. (1997) '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), pp. 2581-2585.

Wutz, A. and Gribnau, J. (2007) 'X inactivation Xplained', *Curr Opin Genet Dev,* 17, (5), pp. 387-393.

Xing, E. P., Nie, Y., Song, Y., Yang, G. Y., Cai, Y. C., Wang, L. D. and Yang, C. S. (1999) 'Mechanisms of inactivation of p14ARF, p15INK4b, and p16INK4a genes in human esophageal squamous cell carcinoma', *Clin Cancer Res,* 5, (10), pp. 2704-2713.

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. (2008) 'Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells', *Cancer Cell,* 14, (2), pp. 135-145.

Yauch, R. L., Dijkgraaf, G. J., Alicke, B., Januario, T., Ahn, C. P., Holcomb, T., Pujara, K., Stinson, J., Callahan, C. A., Tang, T., Bazan, J. F., Kan, Z., Seshagiri, S., Hann, C. L., Gould, S. E., Low, J. A., Rudin, C. M. and de Sauvage, F. J. (2009) 'Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma', *Science,* 326, (5952), pp. 572-574.

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. (1999) '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), pp. 832-845.

Zurawel, R. H., Allen, C., Chiappa, S., Cato, W., Biegel, J., Cogen, P., de Sauvage, F. and Raffel, C. (2000) 'Analysis of PTCH/SMO/SHH pathway genes in medulloblastoma', *Genes Chromosomes Cancer,* 27, (1), pp. 44-51.

Chapter 8. Publications